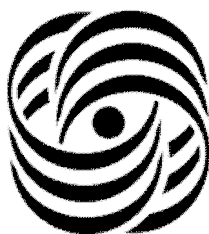


**Министерство образования и науки Российской Федерации
Московский государственный университет им. М.В. Ломоносова
Факультет Вычислительной математики и кибернетики**



**СОВЕТ МОЛОДЫХ УЧЁНЫХ
ВМК МГУ**



**Сборник тезисов
XVI Международной научной конференции студентов,
аспирантов и молодых учёных
«ЛОМОНОСОВ-2009»,
секция
«ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И КИБЕРНЕТИКА»**



**Москва
МГУ имени М. В. Ломоносова
13–18 апреля 2009 г.**

УДК 517.6+519.8
ББК 22
Л75

**Организационный комитет
Международной молодежной научной олимпиады «Ломоносов-2009»**

Сопредседатель
Организационного комитета
**Ректор МГУ им. М.В. Ломоносова,
академик В. А. Садовничий**

Сопредседатель
Организационного комитета
**Министр образования и науки
Российской Федерации А.А. Фурсенко**

Организационный комитет
XVI Международной научной конференции
студентов, аспирантов и молодых ученых
«Ломоносов-2009»
секции «Вычислительная математика и кибернетика»:

Е.И. Монсеев (*председатель*), **Шевцова И.Г.** (*ответственный секретарь*), **Дарьин А.Н.,**
Дьяконов А.Г., Ильин А.В., Коношин А.С., Майсурадзе А.И., Малышко В.В., Сальников А.Н.,
Столяров А.В., Фомичев В.В.

Л75

Ломоносов-2009: Материалы XVI Международной научной конференции студентов, аспирантов и молодых ученых: секция «Вычислительная математика и кибернетика»; 13-18 апреля; Москва, МГУ имени М.В. Ломоносова, факультет ВМК: Сборник тезисов/Сост. Ильин А.В., Шевцова И.Г. – М.: Издательский отдел факультета ВМК МГУ (лицензия ИД 05899 от 24.09.2001)), 2009. – 96с.
ISBN: 978-5-89407-328-6

В настоящий сборник вошли тезисы докладов участников Международной научной конференции «Ломоносов» по секции «Вычислительная математика и кибернетика». Тематика тезисов очень разнообразна и охватывает многие актуальные проблемы современной фундаментальной науки. Все представленные в сборнике материалы были рекомендованы к публикации Экспертным советом секции. Тезисы публикуются в том виде, в котором были представлены авторами, они не редактировались, а были лишь приведены к единому формату.

УДК 517.6+519.8
ББК 22

ISBN 978-5-89407-328-6

© Издательский отдел факультета вычислительной математики и кибернетики
МГУ им. М.В. Ломоносова, 2009
© Ильин А.В., Шевцова И.Г. составление, оформление, 2009

СОДЕРЖАНИЕ

Использование архитектуры Cell/B.E. для ускорения выполнения пакета молекулярного моделирования MOLKERN Алемасов Николай Александрович	10
Восстановление типов данных с использование информации о выполнении программы Антонов Вадим Юрьевич, Долгова Катерина Николаевна	11
О существовании и разрушении решения начально-краевой задачи для нелинейного уравнения соболевского типа с переменным коэффициентом Аристов Анатолий Игоревич	12
Вариационное усвоение данных наблюдений в задаче об адекватности приливной модели Ассовский Максим Владимирович, Ботвиновский Е.А.	13
Генерация признаков при наличии артефактов для задачи распознавания личности по форме ладони Бакина Ирина Геннадьевна	14
Об асимптотическом поведении хроматического индекса случайных гиперграфов Будников Юрий Александрович	15
Ситуации равновесия в обобщённом классе игр многих лиц с личными и общественными критериями Верхотурова Екатерина Евгеньевна	16
Применение генетических алгоритмов в задаче классификации сигналов (приложение в ВСІ) Власова Юлия Валерьевна	17
Автоматизация построения кода медиаторов для функционального тестирования web-приложений по технологии UniTESK Герлиц Евгений Анатольевич	18
Реакция программного интерфейса на действия пользователя Германов Константин Сергеевич	19
Алгоритм поиска дубликатов в базе видеопоследовательностей на основе сопоставления иерархии смен сцен Глазистов Иван Викторович, Паршин Александр Евгеньевич	20
Построение расписания процедуры тестирования автомобильной комбинации приборов Головенков Евгений Владимирович	21

Сравнение модификаций EM-алгоритма для декомпозиции волатильности финансовых временных рядов Горшенин Андрей Константинович	22
Управление внешней нормативно-справочной информацией в распределённых информационных системах Гудков Кирилл Сергеевич	23
О нижних оценках числа независимых множеств в некоторых классах графов Дайняк Александр Борисович	24
Восстановление управляющих конструкций языка высокого уровня из исходной программы на языке ассемблера Деревенец Егор Олегович, Долгова Катерина Николаевна	25
Исследование и разработка декомпилятора в язык Си Долгова Катерина Николаевна	26
Разрешимость и регулярность задач «нечёткой» разметки точечных конфигураций Дорофеев Николай Юрьевич	27
Оценка мимической динамики движения челюсти в процессе жевания по трехмерному видеоряду Дышкант Наталья Федоровна	28
Использование мультифрактальных показателей при оценивании параметров гауссовского и мультипликативного шума на изображениях Емец Юрий Владимирович	29
Эффективная работа с данными при решении задач коллаборативной фильтрации. Заборовский Никита Владимирович	31
Семейство разностных схем с дискретными прозрачными граничными условиями для уравнения Шрёдингера на полуоси Злотник Илья Александрович	32
Моделирование работы систем резервного копирования данных Казиков Виталий Гайясович	33
Критерий k -сингулярности системы точек и оптимальное разбиение на подсистемы Карпович Павел Алексеевич	34
Настройка над системой пакетной обработки задач пользователя MAUI с целью управления динамическими приоритетами. Князев Николай Александрович	35

О методах трёхмерного моделирования фарлей-бунемановской неустойчивости на современных многопроцессорных вычислительных комплексах Ковалёв Дмитрий Владимирович	36
Прогнозирование временных рядов при несимметричной функции потерь: прогнозирование плотности распределения и квантильная регрессия Коваль Артур Сергеевич	37
Компьютерная технология исследования потребительского спроса с помощью обобщенного непараметрического метода Кондраков Иван Александрович	38
Система генерации тестовых программ с использованием ограничений ТЕСЛА Корныхин Евгений Валерьевич	39
Определение параметров источника света по фотографиям Коробченко Дмитрий Александрович, Ильин Андрей Алексеевич	40
Внутренняя эллипсоидальная оценка пересечения двух концентрических эллипсов Косенкова Лидия Викторовна	41
Защита web-приложений Кочетков Вячеслав Владимирович	43
Система параллельной конвейерной обработки даннь Кривов Максим. Андреевич, Притула Михаил Николаевич	44
Об одном подходе к организации системы распознавания таблиц, содержащих статистические данные Кудинов Павел Юрьевич	45
Генетический подход к проблеме предсказания сроков сдачи ПО в эксплуатацию Кульдин Сергей Павлович	46
Непараметрическая модель динамики срочной структуры процентных ставок Лапшин Виктор Александрович	47
О задачах распределения ресурсов и проверки устойчивости для систем информационного мониторинга Лебедев Анатолий Анатольевич	48
Нелинейная аппроксимация функций отражательной способности материалов Лебедев Андрей Сергеевич, Ильин Андрей Алексеевич	49
Адаптация онтологий для повышения эффективности рассуждений в приложениях семантического web Левшин Дмитрий Владимирович	50

Численная оптимизация обеспечения для портфеля производных финансовых инструментов с учетом пакета активных заявок Лукина Анна Юрьевна, Долматов Андрей Сергеевич,	51
Нейросетевые и статистические методы обнаружения сигналов Ляликова Виктория Геннадиевна	52
Расчет времени деградации лазерной мишени Малинина Елена Александровна	53
Автоматизированный поиск повторяющихся элементов на ректифицированных изображениях фасадов городских зданий Мизин Иван Сергеевич, Якубенко Антон Анатольевич	54
Исследование многопроцессорного расписания Мирошниченко Роман Викторович	55
Оптимизация блочного умножения разреженных матриц на вектор для графических акселераторов Монаков Александр Владимирович, Кравец Алексей Сергеевич	56
Численная реализация метода внесённой границы для моделирования вязкой несжимаемой жидкости в областях сложной конфигурации Мортиков Евгений Валерьевич	57
Метод автоматического выделения терминалей дофаминергических нейронов на изображениях фронтальных срезов стриатума Мягков Артем Александрович, Яшина Вера Владимировна	58
Модификация многоуровневого метода Монте-Карло Нагапетян Тигран Артурович	59
Применение сеточных методов к задаче разделения смесей вероятностных распределений. Назаров Алексей Леонидович, Гапонова Маргарита Олеговна	60
Усиленный закон больших чисел для случайных процессов Наумов Алексей Александрович	61
Решение некоторой проблемы обработки сигналов ЯМР с использованием обобщенного спектрально-аналитического метода Новикова Дарья Андреевна	62
Программная поддержка языка лексико-синтаксических шаблонов Носков Алексей Анатольевич	63
Алгоритмы машинной графики в задачах геопространственного моделирования Овсянников Михаил Сергеевич	64

Информационные графы с автоматными функциями Пивоваров Александр Павлович	66
Программная реализация метода разделения области FETI-DP для массивно-параллельной вычислительной системы IBM Blue Gene/P Позднеев Александр Валерьевич	67
Оценка компактности задач распознавания образов Потепалов Дмитрий Николаевич	68
Кластеризация схем поведения пользователей веб-ресурсов Проскурнев Андрей Александрович	69
Поиск похожих пользователей в социальных сетях Пустовойтов Никита Юрьевич	70
Перестановочные критерии для одновыборочных задач проверки гипотез Рябенко Евгений Алексеевич	71
LSPL-шаблоны для решения задачи автоматизированного построения онтологий Седова Яна Анатольевна	72
Об одном приближенном способе решения задач оптимального подвижного управления для систем гиперболического типа Сейидова Улвия Вахаб	73
Задачи граничного слоя для уравнения колебаний сферического слоя Сергеев Сергей Андреевич	74
Матирование видео на основе оптического потока Синдеев Михаил Сергеевич, Конушин В.С.	75
Построение виртуальных сцен реального мира в системе реконструкции отражательных свойств материалов Синявский Виталий Александрович, Ильин Андрей Алексеевич	76
Экспериментальное исследование свойств методов индуктивного формирования знаний Смагин Сергей Владимирович	78
Применение технологии UniTESK для функционального тестирования инфраструктурного ПО Грид Смолов Сергей Александрович	79
О конструктивной характеристике пороговых функций Соколов Андрей Павлович	80

Модель вычислительной платформы для динамического анализа защищенного бинарного кода Соловьев Михаил Александрович	81
Семантическая сегментация дорожного покрытия Судаков Сергей Владимирович	82
Методы уменьшения количества временных узлов, создаваемых при исполнении запросов на языке XQuery Таранов Илья Сергеевич	83
Выделение лиц на изображениях в условиях искажающих факторов Тарасова Дарья Алексеевна, Шаров Андрей Николаевич	84
Построение композитно-иерархических скрытых марковских моделей для анализа поведения Темлянцев Александр Валерьевич	85
Метод байесовского оценивания параметров. Алгоритмы «Марковская цепь Монте Карло» Ушакова Анастасия Николаевна	86
Локальные методы прогнозирования временных рядов Федорова Валентина Павловна	88
Аппроксимация полигональной сетки головы человека с помощью параметризованной антропометрической модели Федюков Максим Александрович	89
Применение обобщенного спектрально-аналитического метода в задаче анализа биологических данных Филиппов Виталий Владимирович, Горчаков Михаил Александрович	90
Применение плоских кривых для программирования поверхностных эскизов Фомкина Мария Сергеевна	91
Применение базиса бинарных изображений в рентгеновских аппаратах сканирующего типа. Хасанов Ренат Ришатович, Хасанов Марат Ришатович	92
Оптимизация граничного управления смещением на одном конце струны с условием закрепления на другом конце в пространстве W_2^2 для задачи возбуждения Холомеева Анна Андреевна	93
Информационный критерий оценки качества цветных цифровых изображений Хорунжий Михаил Дмитриевич	94

Использование алгоритмов онлайн обучения при разработке системы семантической сегментации изображений Шапвалов Роман Викторович, Барина Ольга Вячеславовна.....	95
Кластеризация интересов пользователей социальной сети Шейнин Владимир Валерьевич.....	96
Оценка артефактов блочности JPEG-изображениях Шмаглит Лев Александрович	97

Использование архитектуры Cell/B.E. для ускорения выполнения пакета молекулярного моделирования MOLKERN¹

Алемасов Николай Александрович

Аспирант

Институт цитологии и генетики СО РАН, Новосибирск, Россия

E-mail: alemasov@bionet.nsc.ru

Программный комплекс MOLKERN [1] предназначен для моделирования структуры и динамики крупных комплексов биомолекул. Все алгоритмы программного комплекса имеют вычислительную сложность не выше, чем $O(N \log N)$. Пакет написан на языке C++ для платформы x86 с использованием библиотек STL, BOOST, BLAS, MPI и интерфейса OpenMP. Целью данной работы является перенос программного комплекса на платформу Cell/B.E, которая специально спроектирована для интенсивных расчётов.

Процессор Cell/B.E. [2] содержит 9 ядер (PPE + 8 SPE). PPE (PowerPC Processor Element) предназначен для взаимодействия с операционной системой. SPE (Synergistic Processor Element) используются как основные вычислительные устройства. Каждый SPE имеет локальную память LS (Local Storage) размером 256 Кб, 128 128-битных регистров и широкий набор SIMD-инструкций для работы с ними. Особенности Cell/B.E. позволяют получить существенный прирост скорости выполнения программ относительно архитектуры x86.

В результате работы сформирована архитектура программного комплекса MOLKERN в версии под Cell процессор. Большая часть кода MOLKERN выполняется на PPE. Наиболее вычислительно затратные функции расчета парных невалентных взаимодействий (4/5 от всего времени выполнения) выполняются на 8 SPE. Передача данных осуществляется блоками, размер которых выбирается с учётом LS. Блок данных включает всю необходимую информацию для расчета, чтобы исключить задержки обращения в память. Блоки независимы по данным, что позволяет выполнять их в различных потоках SPE. За сохранение результатов, полученных на разных SPE, отвечает PPE, тем самым избегаются блокировки данных. Написан небольшой (не более 300 строк) код для управления потоками и пересылки данных между PPE и SPE. В реализации использовалась библиотека libspe 2 IBM SDK [3].

В Cell/B.E версии программного комплекса MOLKERN время выполнения вычислительно затратных функций расчета невалентных взаимодействий уменьшилось от 3 (для дальних) до 5 раз (для ближних взаимодействий), что существенно, до 60% увеличило общую производительность комплекса.

Литература

1. Фомин Э.С., Алемасов Н.А., Чирцов А.С., Фомин А.Э. (2006) Библиотека программных компонент MOLKERN для построения программ молекулярного моделирования // Биофизика. – 2006. – Т.51, Вып.7, – С. 110-113.
2. Kahle J. A., Day M.N., Hofstee H.P., Johns C.R. (2005) Introduction to the Cell multiprocessor // IBM Journal of Research and Development. – 2005. – Vol. 49, № 4/5. – P.589-604.
3. www.ibm.com/developerworks/power/cell (IBM SDK for Multicore acceleration v3.0).

¹ Работа выполнена в рамках Междисциплинарного интеграционного проекта фундаментальных исследований СО РАН №26 «Математические модели, численные методы и параллельные алгоритмы для решения больших задач СО РАН и их реализация на многопроцессорных суперЭВМ» и Междисциплинарного интеграционного проекта фундаментальных исследований СО РАН № 113 «Разработка вычислительных методов, алгоритмов и аппаратурно-программного инструментария параллельного моделирования природных процессов». Автор благодарит компанию T-platforms за предоставление доступа к серверу с PowerXCell8i и своего научного руководителя к.ф.-м.н.Фомина Э.С. за помощь в работе.

Восстановление типов данных с использованием информации о выполнении программы

Антонов Вадим Юрьевич, Долгова Катерина Николаевна

Студент факультета ВМиК

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: avadim@gmail.com, katerina@ispras.ru

Обратная инженерия в индустрии программного обеспечения – процесс извлечения информации, неявно представленной в исполняемой программе. Результатом обратной инженерии является представление программы на более высоком уровне абстракции [1].

Одной из важных задач обратной инженерии является задача восстановления типов данных. Эта задача играет важную роль в восстановлении протоколов обмена информацией, декомпиляции, восстановлении интерфейсов приложений и анализе программ на наличие ошибок и угроз безопасности. Несмотря на то, что полное автоматическое восстановление типов данных в общем случае является алгоритмически неразрешимой задачей, существует ряд методов, позволяющих решать эту задачу для некоторых классов программ.

Можно выделить два подхода к восстановлению типов данных: статический и динамический. Статический подход использует только информацию, доступную в исполняемом файле программы, в то время как динамический подход учитывает информацию времени выполнения программы, в частности, для предлагаемого метода используется информация о значениях переменных.

В данной работе представляется метод сбора и анализа информации времени выполнения программы для восстановления типов данных по бинарному коду исполняемой программы. Цель анализа данных – восстановить знаковую целого типа и отличить целый тип от указательного, так как по ассемблерному коду восстановить это не всегда возможно. Ассемблерные команды, в мнемоники которых явно заложена знаковая, обычно двухадресные, и знаковая относится к одному из используемых операндов. Отличить, к какому именно операнду – при статическом анализе зачастую невозможно. Аналогичная ситуация возникает с распознаванием типа 32-х битных переменных (для 32-х битной архитектуры). Для работы с указателями и 32-х битными переменными используются одни и те же команды, для хранения – одни и те же регистры общего назначения. Если в программе для указательного типа не было разыменования, то при статическом анализе отличить его от целого типа не представляется возможным.

Предлагаемый метод сбора информации времени выполнения основан на том, что для каждой операции чтения и записи в память строится профиль загружаемых и считываемых значений. Разработан ряд эвристик, позволяющих по значениям переменной, которые она принимала во время выполнения программы, с некоторой вероятностью оценить является ли она знаковой или беззнаковой, а также является ли переменная указательного типа или это переменная целого типа.

Разработанный метод реализован в инструменте TDTrace, который позволяет анализировать адреса и значения, используемые во время выполнения программы, а также представляет полученную информацию в удобном для последующего использования виде. При реализации была использована среда Valgrind с открытыми исходными кодами для разработки инструментов анализа программ во время их выполнения[2].

Инструмент TDTrace был апробирован на нескольких программах с открытыми исходными кодами. Результаты этих экспериментов показали состоятельность подхода.

Литература

1. Chikofsky, E.J.; J.H. Cross II (January 1990). "Reverse Engineering and Design Recovery: A Taxonomy in IEEE Software". // IEEE Computer Society
2. Nicholas Nethercote and Julian Seward. (2007) Valgrind: A Framework for Heavyweight Dynamic Binary Instrumentation // Proceedings of ACM SIGPLAN 2007 Conference on Programming Language Design and Implementation (PLDI 2007), San Diego, California, USA, June 2007.

О существовании и разрушении решения начально-краевой задачи для нелинейного уравнения соболевского типа с переменным коэффициентом

Аристов Анатолий Игоревич

Выпускник 2008 г. ф-та Вычислительной математики и кибернетики Московского Государственного Университета имени М. В. Ломоносова Москва, Россия

e-mail: ai_aristov@mail.ru

Работа посвящена исследованию существования и разрушения обобщенного решения следующей задачи:

$$\begin{cases} \frac{\partial}{\partial t}(\Delta u - u) + \mu(t)u^3 = 0 \\ u(x, 0) = u_0(x) \\ u(x, t)|_{\partial\Omega} = 0 \end{cases}$$

и нахождению оценок времени разрушения (в случае разрушения за конечное время). Здесь $x \in \Omega$ - ограниченному подмножеству R^3 с границей $\partial\Omega \in C^{(2,\delta)}$, $\delta \in (0,1]$. Такое уравнение описывает нестационарные процессы в полупроводниках. Будем предполагать, что $u_0 \in H_0^1(\Omega)$. Коэффициент $\mu(t) \in C[0; \infty)$ - действительная

функция, причем $\mu > 0 \forall t \geq 0$ и $\int_0^\infty \mu(\tau) d\tau < \infty$. Обобщенным решением этой задачи будем

называть такое $u \in C^1[0, T; H_0^1(\Omega))$, что $\left\langle \frac{\partial}{\partial t}(\Delta u - u) + \mu u^3, w \right\rangle = 0 \forall w \in H_0^1(\Omega)$.

На основе принципа сжимающих отображений можно доказать следующее утверждение:

Теорема 1. $\forall u_0 \in H_0^1(\Omega) \exists$ такое $T > 0$ (возможно, бесконечное), что обобщенное решение существует и единственно, причем если $T < \infty$, то $\lim_{t \rightarrow T^-} \|u\|_{H_0^1(\Omega)} = \infty$.

Для того чтобы оценить время разрушения, введем следующие обозначения:

$\Phi = \|u\|_{L_2(\Omega)}^2 + \|\nabla u\|_{L_2(\Omega)}^2, C_0 = \inf \left\{ C > 0 \|\|w\|_{L_4(\Omega)} \leq C \|w\|_{H_0^1(\Omega)} \forall w \in H_0^1(\Omega) \right\}$. С помощью

энергетических оценок и теоремы вложения $H_0^1(\Omega)$ в $L_4(\Omega)$ получим к следующие

неравенства:
$$\left(\frac{1}{\Phi(0)} - \frac{2\|u_0\|_{L_4(\Omega)}^4}{\Phi^2(0)} \int_0^t \mu(\tau) d\tau \right)^{-1} \leq \Phi \leq \left(\frac{1}{\Phi(0)} - 2C_0^4 \int_0^t \mu(\tau) d\tau \right)^{-1}.$$

Полагая $\alpha = \frac{1}{2C_0^4\Phi(0)}, \beta = \frac{\Phi(0)}{2\|u_0\|_{L_4(\Omega)}^4}, I = \int_0^\infty \mu(\tau) d\tau$, придем к следующему утверждению:

Теорема 2. Если $I \leq \alpha$, то решение существует глобально по времени. Если $I > \beta$, то решение разрушается за конечное время $T \in [T_1; T_2]$, где T_1 и T_2 определяются следующими уравнениями:

$$\int_0^{T_1} \mu(\tau) d\tau = \alpha \text{ и } \int_0^{T_2} \mu(\tau) d\tau = \beta.$$

Замечание. $\alpha \leq \beta$ при нетривиальных u_0 в силу теоремы вложения $H_0^1(\Omega)$ в $L_4(\Omega)$.

Литература:

1. А. Г. Свешников, А. Б. Альшин, М. О. Корпусов, Ю. Д. Плетнер. Линейные и нелинейные уравнения Соболевского типа. М., Физматлит, 2007.
2. А. Г. Свешников, А. Б. Альшин, М. О. Корпусов. Нелинейный функциональный анализ и его приложения к уравнениям в частных производных. М., Научный мир, 2008

Вариационное усвоение данных наблюдений в задаче об адекватности приливной модели

Ассовский Максим Владимирович, Ботвиновский Е.А.

Студент; м.н.с., к.ф.-м.н.

Московский физико-технический институт (ГУ), Долгопрудный, Россия;

Институт вычислительной математики РАН, Москва, Россия

bea@inm.ras.ru

Изучение адекватности сложных математических моделей геофизической гидродинамики является важной и актуальной задачей. Для её исследования и численного решения могут быть применены методы теории обратных задач [1], оптимального управления и сопряженных уравнений [2], современные методы вычислительной математики. В докладе рассматривается одна из таких задач – задача об адекватности модели динамики приливов в Мировом океане.

В модели динамики приливов [3] в качестве «дополнительного» источника рассматривается вектор-функция F . В качестве критерия того, что модель адекватно воспроизводит моделируемые физические процессы, принимается требование: совпадение функции уровня (отклонение свободной поверхности от её невозмущенного состояния) со спутниковыми данными наблюдений об уровне. Этот критерий можно свести к минимизации квадратичного функционала («стоимости») с целью вариационного усвоения данных наблюдений и вычисления F . После решения данной задачи оценивается норма F , и её значение берется в качестве меры адекватности исходной модели динамики приливов. В докладе представлены результаты однозначной разрешимости задачи минимизации функционала «стоимости» и доказывається, что минимум функционала равен нулю. Приводится итерационный алгоритм решения задачи минимизации, численный алгоритм решения задачи динамики приливов [4], и результаты численных экспериментов.

Авторы выражают благодарность д.ф.-м.н., профессору Агошкову В.И. за научное руководство и постановку задач.

Работа выполнена при поддержке РФФИ (грант 07-01-00714а).

Список литературы

3. Морозов В.А. Регулярные методы решения некорректных задач. – М: Наука, 1987.
4. Агошков В.И. Методы оптимального управления и сопряженных уравнений в задачах математической физики. – М.: ИВМ РАН, 2003.
5. Марчук Г.И., Каган Б.А. Океанские приливы. – Л: Гидрометеиздат, 1977.
1. Botvinovsky E.A. An algorithm for the solution of a tidal dynamics problem on a sphere // Russ. J. Numer. Anal. Math. Modelling, V. 23, No. 6, 2008, p. 523-536.

Генерация признаков при наличии артефактов для задачи распознавания личности по форме ладони¹

Бакина Ирина Геннадьевна

Аспирант

Московский Государственный Университет имени М. В. Ломоносова, Факультет
Вычислительной Математики и Кибернетики, Москва, Россия

irina_msu@mail.ru

В работе рассматриваются задачи генерации признаков и выбора метрики при распознавании человека по форме ладони [1, 2], которые бы учитывали следующие артефакты (рис. 1):

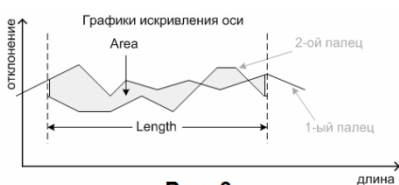
- На фотографии пальцы часто оказываются “склеенными”. Это делает невозможным определение длины пальца.
- Ошибочное нахождение ширины и длины пальца возможно при наличии длинных ногтей.
- Часть руки может быть прикрыта одеждой, поэтому невозможно определить параметры ладони, например, ее ширину.

В данной работе делается одно допущение – считается, что все ладони снимаются камерой с фиксированного расстояния.

В качестве вектора признаков предлагается рассматривать следующие характеристики пальца – график искривления серединной оси пальца относительно прямой и график изменения ширины пальца вдоль этой оси (используется модель скелета [3]). Таким образом, вектор признаков ладони содержит 10 компонент.

Вычисление признаков осуществляется для участка пальца, не подверженного влиянию артефактов. В работе предлагается способ нахождения этого участка – определение *точек основания и кончика пальца*. На рис.2 для каждого пальца жирной точкой отмечены найденные основание и конец. Вводится понятие *оси пальца* – прямой, проходящей через точки основания и кончика пальца.

Выделенные участки двух пальцев могут оказаться различными по



длине. В работе предлагается подход к нахождению общего участка пары пальцев для сравнения, основанный на минимизации отношения $Area/Length$, где $Area$ – площадь области, заключенной между двумя графиками искривления пальца относительно текущей оси, а $Length$ – длина этого участка (рис. 3). Минимизация основывается

на варьировании концевых точек пальцев. Соответственно, изменяется и ось пальца. В качестве меры сходства пары пальцев по одному из признаков рассматривается полученное оптимальное соотношение $Area/Length$.

Указанные признаки были вычислены для базы ладоней 70 человек, содержащей 2000 изображений. В работе приведены графики распределения значений признаков при сравнении ладоней одного человека и ладоней разных людей. Рассматривается возможность их применения для дальнейшего распознавания личности.

Литература

1. Chiang-Liang Su (2008) Index, Middle, and Ring Finger Extraction and Identification by Index, Middle, and Ring Finger Outlines // VISAPP (1) 2008, стр. 518-520.
2. Slobodan Ribaric, Damir Ribaric, Nicola Pavesic (2005) A Biometric Identification System Based on the Fusion of Hand and Palm Features // IEEE (27), стр. 1698 – 1709.
3. Mestetskii L.M. (2000) Fat Curves And Representation of Planar Figures // Computer & Graphics 24 (2000) 9 – 21.

¹ Работа поддержана грантами РФФИ 08-01-00670 и 08-07-00305.

Об асимптотическом поведении хроматического индекса случайных гиперграфов

Будников Юрий Александрович

Аспирант механико-математического факультета

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: y.budnikov@mail.ru

Пусть $k \geq 2$ - фиксированное натуральное число. Будем рассматривать гиперграфы $G = (V, E)$ - k -однородные, то есть каждое ребро которых содержит в точности k вершин. Упаковка в G – набор P непересекающихся ребер G . $\chi(G)$ - хроматический индекс – минимальное число упаковок, на которые можно разбить все ребра G .

Утверждение 1. Для любых $\delta > 0, \delta' > 0$, для любой возрастающей функции $g(n)$ можно построить класс гиперграфов $G(n)$ с возрастающим $k(n)$ - числом вершин на ребре – так, что $k(n) < g(n)$ для любых n , и $D(n)$ – максимальной степенью его вершин: при $n = n_k, k = 1, 2, \dots$ - возрастающая подпоследовательность натуральных чисел – будет выполнено: $\chi(G) = k(D(n) - 1) + 1$.

Теорема 1. Пусть константа $c > 1$. В модели случайного гиперграфа $G(n, p)$ для случая растущей $k(n)$ - числа вершин на ребре $G(n, p)$ - при стандартном выборе

$$p = \frac{D(n)}{C_{n-1}^{k-1}}, D(n) = o(C_{n-1}^{k-1}), n \rightarrow \infty:$$

1) Если $\lim_{n \rightarrow \infty} \frac{D(n)}{e^k} < 1$, то $P(G(n, p)$ содержит упаковку размера $\frac{n}{kc} \rightarrow 0, n \rightarrow \infty$

2) Если для некоторой константы $d > 1$ выполнено $\lim_{n \rightarrow \infty} \frac{D(n)}{de^k} > 1$, то

$$P(G(n, p) \text{ содержит упаковку размера } \frac{n}{kc} \rightarrow 1, n \rightarrow \infty$$

Теорема 2. В условиях Теоремы 1 для любой $c > 1$ $P(\chi(G(n, p)) \geq cD(n)) \rightarrow 1, n \rightarrow \infty$

Литература

1. Nicholas Pippenger and Joel Spencer, Asymptotic behavior of the Chromatic Index for Hypergraphs, Journal of combinatorial theory, Series A 51, 24-42(1989).

2. Будников Ю.А., "О мощности ребер гиперграфа"(с.70-72),
Материалы IX Международной конференции "Интеллектуальные системы и компьютерные науки", 2006

Ситуации равновесия в обобщённом классе игр многих лиц с личными и общественными критериями

Верхотурова Екатерина Евгеньевна

Студент

*Московский государственный университет имени М.В.Ломоносова,
Факультет Вычислительной Математики и Кибернетики, Москва, Россия*

E-mail: e.e.verkhoturova@gmail.com

В работе [1] был предложен класс игр многих лиц, в которых участники имеют векторные иерархические критерии. Основное внимание в [1] уделено анализу ситуаций равновесия, причем большинство результатов получено для случая, когда игроки распределяют скалярные ресурсы. Кроме того, в [1] отмечена сильная устойчивости ситуаций равновесия.

В работе [2] более подробно анализируется один важный класс моделей [1], в которых каждый игрок имеет только две компоненты критерия – личную и общественную и распределяет ресурс между ними. В случае единого для всех игроков общественного критерия была доказана сильная устойчивость ситуаций равновесия (см. [1,2]).

В данной работе рассматривается более широкий класс моделей [2]: допускается наличие различных общественных критериев у игроков. Это допущение выводит за рамки теории Вателя-Гермейера. Но в этом случае также удаётся при определённых условиях найти ситуации равновесия и показать их сильную устойчивость.

В качестве иллюстрирующего примера рассматривается одна игра трёх лиц на базе «Дилеммы заключенного».

Литература

1. Гермейер Ю.Б., Ватель И.А. Игры с иерархическим вектором интересов // Техническая кибернетика. 1974. №3. С.54-69.
2. Ватель И.А. Ядро в игре многих лиц с личными и общественными критериями // Автоматика и телемеханика. 1980. №1.
3. Гермейер Ю.Б. Введение в теорию исследования операций. М.: Наука. 1981.

Применение генетических алгоритмов в задаче классификации сигналов (приложение в BCI)¹

Власова Юлия Валерьевна

студент

Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия

e-mail: julia.vlasova@bk.ru

Проблема разработки эффективных методов классификации сигналов (конечных последовательностей измерений некоторой величины) является актуальной вследствие огромного числа приложений. Один из наиболее часто применяемых на практике подходов к решению этой проблемы состоит в синтезе информативного признакового описания сигналов и сведении проблемы к задаче классификации объектов в признаковом описании.

В данной работе предлагается воспользоваться аппаратом генетической оптимизации для построения качественного признакового описания сигнала. Признаком (в данном случае) называется вещественная функция от сигнала. Представим её в виде суперпозиции двух функций. Для этого введём два типа операторов обработки сигнала. Операторы первого типа по описанию сигнала $\tilde{x} = (x_1, x_2, \dots, x_l)$ получают ряд $\tilde{y} = (y_1, y_2, \dots, y_k)$, возможно, другой длины. Основное их назначение – преобразовать информацию (улучшить или перейти к иному описанию). Примеры операторов первого типа: фильтрация сигнала (уменьшение уровня шума), сглаживание сигнала (замена значения в точке средним значением в окрестности), взятие производной конечной разности, оператор взятия абсолютного значения и т.д. Операторы второго типа преобразуют сигнал в число, вычисляя его некоторую характеристику. Примеры операторов второго типа: среднее/минимальное/максимальное значение сигнала, число различных значений сигнала и т.д. Обозначим через B^* и C^* множества операторов первого и второго типа соответственно. Признак, описывающий сигнал, ищем в виде: $C_{i_0}(B_{i_k}(\dots(B_{i_1}(\tilde{x}))))$, где $k \in \{1, 2, \dots\}$, $B_{i_k}, \dots, B_{i_1} \in B^*$, $C_{i_0} \in C^*$. Для решения задачи поиска качественного признака используется генетический подход к задаче оптимизации. Функционалом качества отдельного признака является критерий AUC-ROC, а совокупности признаков – качество классификации метода kNN при применении скользящего контроля (при этом предлагается подход к выбору оптимальной метрики для метода kNN).

Предложенный метод тестировался при решении реальных задач классификации сигналов из области BCI (Brain-Computer Interface) [1, 2]. В этой области занимаются проблемой коммуникации человек-компьютер с помощью сигналов головного мозга. Сигналы регистрируются специальной аппаратурой. Предполагается, что, настроив классифицирующий алгоритм на обучающей выборке (сигналы, снятые во время известных ментальных действий), можно будет достаточно эффективно определять класс ментальных действий по сигналу. Приложения BCI очень разнообразны: контроль протезов и управление механизмами, организация общения с людьми с ограниченными физическими возможностями, компьютерные игры, исследование активности головного мозга человека в научных целях.

Литература

1. Blankertz B., Muller K.-R., Curio G., Vaughan T.M., Schalk G., Wolpaw J.R., Schlogl A., Neuper C., Pfurtscheller G., Hinterberger T., Schroder M., Birbaumer N. (2004) The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials // IEEE Trans. Biomed. Eng., 51(6):1044-1051.
2. http://ida.first.fraunhofer.de/projects/bci/competition_iii/

¹ Работа поддержана грантом РФФИ (проект № 08-07-00305-а).

Автоматизация построения кода медиаторов для функционального тестирования web-приложений по технологии UniTESK

Герлиц Евгений Анатольевич

аспирант

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия*

E-mail: gerlits@ispras.ru

В наши дни web-приложения стали одним из самых распространенных и востребованных видов программного обеспечения. Бурное развитие web-технологий способствовало росту функциональных возможностей современных web-приложений. Сегодня web-приложения могут представлять собой программные системы с довольно сложной функциональностью.

Технология функционального тестирования на основе моделей требований к функциональности тестируемой системы UniTESK [1] была разработана в Институте системного программирования Российской Академии Наук и хорошо себя зарекомендовала в проектах по тестированию различных классов сложных программных систем.

При исследовании существующих инструментов функционального тестирования web-приложений, было выделено три подхода к функциональному тестированию web-приложений: Capture and Playback, Data Driven, Keyword Driven. В основе этих подходов лежит логика построения тестов вручную. Ручное построение тестов уже не позволяет обеспечить высокое качество тестирования сложных программных систем, в том числе и сложных web-приложений.

Технология UniTESK автоматизирует процесс разработки функциональных тестов, позволяет качественно тестировать web-приложения. Но в ходе разработки тестов для web-приложений выяснилось, что значительная часть усилий тратится на создание кода медиаторов. Медиатор – компонент технологии UniTESK, который преобразует вызов модельных операций в последовательность воздействий на тестируемую систему и переводит реакцию тестируемой системы в модельное представление. В случае тестирования программных интерфейсов приложений реализация медиаторов обычно не является сложной задачей. В случае функционального тестирования web-приложений возникают трудности с переводом web-интерфейсов приложений в совокупность функций медиаторов.

Было принято решение реализовывать основную часть функциональности медиатора в отдельном компоненте - web-медиаторе. Классический медиатор UniTESK в этом случае использует методы web-медиатора для реализации оставшейся функциональности.

В результате исследований и решения ряда технических задач был разработан инструмент «UniTESK Web-mediator Maker», автоматизирующий построение кода web-медиаторов для функционального тестирования web-приложений по технологии UniTESK с использованием инструмента JavaTESK.

По итогам апробации инструмента на двух web-приложениях отмечено, что время построения кода web-медиатора составило 5 и 12 минут для первого и второго web-приложений соответственно. В случае разработки кода web-медиаторов вручную – 50 и 115 минут для первого и второго web-приложений соответственно. При использовании разработанного инструмента получили выигрыш по времени на порядок по сравнению с программированием web-медиаторов вручную.

Литература

1. Баранцев А.В., Бурдонов И.Б., Демаков А.В., Зеленов С.В., Косачев А.С., Кулямин В.В., Омельченко В.А., Пакулин Н.В., Петренко А.К., Хорошилов А.В. Подход UniTESK к разработке тестов: достижения и перспективы // Труды ИСП РАН № 5. М.: ИСП РАН, 2004. С. 121-156.

Реакция программного интерфейса на действия пользователя

Германов Константин Сергеевич¹

студент

*Московский авиационный институт (государственный технический университет)
имени С.Орджоникидзе, факультет систем управления, Москва, Россия*

E-mail: kgermanov@mail.ru

Большинство интерфейсов построено по одному принципу: программа ждет действия пользователя, затем на него реагирует. Если потребуются длительные по времени вычисления, то пользователь вынужден будет ждать завершения этого процесса. В такой модели программа загружает ресурсы неравномерно, только в определенные периоды времени. В данной работе мы попытались построить интерфейс по иному принципу.

Основываясь на предположении, что вычислительный процесс решения определенной задачи можно разбить на части, которые выполняются за короткий промежуток времени, был предложен способ, позволяющий интерфейсу более равномерно загружать имеющиеся ресурсы. Был написан планировщик, который запускается в отдельном потоке, одновременно с основным. При действиях пользователя в очередь планировщика заносятся заявка, которая содержит приоритет выполнения и признак возможности выполнения в отдельном потоке. Планировщик при рассмотрении очередной заявки с наибольшим приоритетом анализирует этот признак. Если возможно, то заявка выполняется в потоке планировщика, после чего она удаляется из очереди и посылается сообщение об ее завершении. Иначе планировщик понижает приоритет заявки и генерирует сообщение основному потоку, который, после ее выполнения посылает подтверждение об удачном выполнении. Если оно приходит, то заявка удаляется из очереди планировщика. Таким образом, если заявка не может завершиться, то ее приоритет с каждым разом понижается, что дает возможность выполнения другим заявкам. При последующих действиях пользователя в очереди может происходить либо добавление новой заявки, либо удаление из очереди заявок, уже не имеющих смысла для выполнения (такое может происходить, например, при изменении пользователем исходных данных), либо перераспределение приоритетов. Также, в очередь планировщика можно заложить типовую последовательность действий пользователя, что позволит автоматизировать рабочее место. При реализации для конкретной задачи оказалось удобным использование не одного, а нескольких приоритетов, что позволило управлять сразу группами заявок.

Проведенное испытание в воинских частях показало состоятельность данного принципа организации программного интерфейса. Так, среднее время ожидания пользователя сократилось с 3ч до 0,5ч (за 4ч работы) для одной и той же задачи. Нам кажется, что использование такого подхода к реализации интерфейсов сделает его более гибким и дружелюбным.

Литература

1. Д. Э. Кнут (2002) Искусство программирования. Том 1. Основные алгоритмы, Вильямс, 2002 г.
2. www.codenet.ru (Программирование на Visual C++).
3. www.trolltech.com (Руководство пользователя бесплатной библиотеки Qt).

¹ Автор выражает признательность доц. Орехову О.О. за помощь в подготовке статьи.

Алгоритм поиска дубликатов в базе видеопоследовательностей на основе сопоставления иерархии смен сцен

Глазистов Иван Викторович, Паршин Александр Евгеньевич

студент, аспирант

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: {aparshin, iglazistov}@graphics.cs.msu.ru

В начале XXI века получили значительное распространение большие электронные базы данных мультимедийного контента, такого как видео, изображения и аудио. Массовое использование этих баз требует эффективных алгоритмов сжатия, обработки и поиска мультимедийной информации. При наличии открытого доступа к базам видео через Интернет возникает потребность в алгоритмах поиска в базе похожих видеофрагментов. Эти алгоритмы могут быть использованы как для оптимизации хранения данных путём удаления дубликатов, так и для полуавтоматического модерирования новых видеофрагментов.

Предложенный алгоритм поиска похожих видеофрагментов основывается на сопоставлении дерева сцен фильма запроса и фильма в хранилище видео. Алгоритм состоит из следующих этапов:

1) Разбиение на сцены. При сравнении соседних кадров они разделились на блоки фиксированного размера. Для построения меры сходства использовалось сравнение гистограмм по блокам кадров по метрике L_1 с последующим усреднением по всем блокам. Несмотря на свою малую вычислительную сложность, этот метод показывает хорошие результаты [см. 1]. Нами были реализованы некоторые улучшения базового метода, повышающие стабильность метрики и качество определения смен сцены.

2) Построение дерева сцен. На основе найденных смен сцены строится двоичное дерево. Вершина дерева отвечает некоторому фрагменту фильма и хранит также величину и положение самой сильной (доминирующей) смены сцены в этом фрагменте. По этой смене сцены фрагмент разбивается на два фрагмента, которым соответствуют потомки рассматриваемой вершины.

3) Сравнение деревьев. Производится поиск в дереве фильма из базы данных вершин, близких по величине доминирующей смены сцены к соответствующему значению в корне дерева запроса. Для каждой такой вершины запрос позиционируется в фильме из базы: координата корневой смены сцены запроса сопоставляется с координатой смены сцены вершины. Затем производится поиск оставшихся смен сцены из запроса в фильме из базы. В зависимости от числа найденных смен сцены строится мера качества сопоставления пары сцен. Затем выбирается наилучшее сопоставление.

Результаты

На построение описания сотни коротких фрагментов общим размером около 3 Гбайт уходит ~25 минут, на поиск дубликатов в них уходит ~5 минут. Алгоритмическая сложность сравнения двух фильмов: $O(\ln(\max(L_1, L_2)))$, где L_1, L_2 - число смен сцены в сравниваемых фильмах. Размер описания 3-х гигабайтной базы данных составляет ~150 Кбайт. На этой базе мы добились точности 92% при 81% определенных пар дубликатов.

Благодаря использованию только временной информации о сменах сцены и величин разности кадров при смене сцены, алгоритм устойчив к большинству характерных искажений цифрового видео. Вместе с тем незначительное количество необходимых данных для сопоставления позволяет строить небольшие по размеру индексы видеофрагментов. Константная сложность сравнения двух фильмов и возможность создания иерархических индексов по фильмам базы данных позволяет говорить об эффективности предложенного алгоритма по скорости работы.

Литература

1. U. Gargi, R. Kasturi, S.H Trayer, "Performance characterization of video-shot-change detection methods", IEEE CSVT V. 10, I. 1, Feb'00, pp 1–13.

Построение расписания процедуры тестирования автомобильной комбинации приборов

Головенков Евгений Владимирович¹

студент

Курский государственный технический университет, факультет информатики и
вычислительной техники, Курск, Россия

E-mail: theaswert@yandex.ru

Автоматизация производственных процессов является актуальным направлением развития современной промышленности. В данной сфере все большее распространение получают опико-электронные системы для автоматического диагностирования электронных индицирующих устройств.

В докладе рассматривается задача построения расписания процедуры визуального контроля показаний автомобильной комбинации приборов (АКП). Диагностирование устройства осуществляется путем формирования эталонных тестовых воздействий на входных линиях АКП и последующего снятия показаний с передней панели АКП с помощью видеокамеры [1]. Особенностью алгоритма является возможность сокращения времени диагностирования за счет организации процедуры параллельного контроля индицируемых параметров АКП.

Пусть $G = \{g_1, \dots, g_n\}, n > 1, n \in N$ – множество индицируемых параметров. Каждому элементу $g_i, i = \overline{1, n}$ множества G поставлен в соответствие элемент $t_i, i = \overline{1, n}$ множества $T = \{t_1, \dots, t_n\}, n > 1, n \in N$ времен диагностирования индицируемых параметров. Имеется матрица $A[m \times m]$, в которой $a_{ij} = 1$, если параметр g_i возможно контролировать одновременно с параметром g_j , $a_{ij} = 0$ – в противном случае; $i, j = \overline{1, n}$. Для некоторых пар параметров заданы ограничения предшествования $g_i \rightarrow g_j; i, j = \overline{1, n}$. Необходимо провести диагностирование всех параметров $g_i, i = \overline{1, n}$ за наименьшее время t_D .

В докладе рассмотрен способ сведения данной задачи к классической задаче построения расписания проекта с учетом ограничения на ресурсы с запрещением прерывания процесса обслуживания (Resource-Constrained Project Scheduling Problem, RCPSP). Приведено решение задачи RCPSP по критерию минимизации времени выполнения проекта, представленное в виде набора моментов начала диагностирования параметров $S = (S_1, \dots, S_n)$ [2]. Получены положительные оценки эффективности применения данного решения в задаче автоматического диагностирования АКП.

Литература

1. Головенков Е.В. (2008) Опико-электронная система для автоматического диагностирования автомобильной комбинации приборов // Сб. конк. работ Всеросс. смотра-конкурса н.-т. творчества студентов вузов «Эврика-2008». – Новочеркасск: Лик, 2008. – С. 71 – 73.
2. Brucker, P. (1999) Complex Scheduling Problems // Zeitschrift Oper. Res.: Osnabruck, Germany, 1999.

¹ Автор выражает признательность д.т.н. Дегтяреву С.В. за помощь в подготовке тезисов.

Сравнение модификаций EM-алгоритма для декомпозиции волатильности финансовых временных рядов

Горшенин Андрей Константинович

аспирант

Московский Государственный Университет им. М. В. Ломоносова, Москва, Россия

E-mail: andygorshenin@gmail.com

При исследовании различных процессов возникает необходимость определить характеристики их изменчивости. В соответствии с этим в финансовой математике вводится понятие волатильности. В случае, когда распределение логарифмов приращений индекса или цены ценной бумаги описывается конечной сдвиг/масштабной смесью нормальных законов, волатильность данного процесса можно определить как вектор-функцию с тремя компонентами: веса (мера влияния данной компоненты на смесь), параметры сдвига и диффузии смешиваемых нормальных распределений (так называемые динамическая и диффузионная компоненты волатильности) [1]. Динамическая компонента волатильности характеризует локальные тенденции динамики финансовых временных рядов. Диффузионная компонента учитывает фактор наличия большого числа участников на рынке, каждый из которых реализует собственную стратегию. Статистическое оценивание весов и параметров сдвига и диффузии (декомпозиция смеси) позволяет выявить закономерности функционирования сложной системы (в данном случае финансового рынка), а также оценить силу влияния каждого выявленного фактора. Это особенно актуально в условиях финансового кризиса.

Часто для этих целей используется EM-алгоритм – итерационный метод для нахождения оценок максимального правдоподобия. Однако данный метод сильно зависит от начальных данных: в большинстве ситуаций он находит не глобальный максимум функции правдоподобия, а локальный, наиболее близкий к начальному приближению. Для борьбы с указанным недостатком были предложены различные модификации EM-алгоритма: медианная версия EM-алгоритма, стохастическая версия EM-алгоритма (SEM-алгоритм) и медианная версия SEM-алгоритма [2]. Отметим, что для смесей нормальных законов можно показать большую эффективность именно медианных оценок (по сравнению с выборочным средним) [3].

Указанные алгоритмы были реализованы в программных комплексах и подверглись многочисленным исследованиям как с применением реальных временных рядов с мировых финансовых рынков, так и с использованием смоделированных выборок с известными параметрами смесей (что позволило оценить правильность результатов, выдаваемых каждым алгоритмом и в оценке параметров, и в выделении компонент смесей). Значительное внимание уделялось как непосредственному анализу компонент волатильности, выделению характерных особенностей (например, приостановка торгов), так и временной эффективности каждого алгоритма. Наилучшим образом в данных исследованиях себя зарекомендовали медианная версия SEM-алгоритма для диффузионной компоненты и SEM-алгоритм для динамической компоненты.

Литература

1. Королёв В.Ю. Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. М., ИПИ РАН, 2007.
2. Королёв В.Ю. EM-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. М., ИПИ РАН, 2007.
3. Горшенин А. К., Королёв В. Ю., Турсунбаев А. М. Медианные модификации EM- и SEM-алгоритмов для разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых временных рядов. // Информатика и её применения, 2008. Т.2. Вып.4. С.12-47.

Управление внешней нормативно-справочной информацией в распределённых информационных системах

Гудков Кирилл Сергеевич

Аспирант факультета управления и прикладной математики

Московский физико-технический институт, Москва, Россия

E-mail: goudkovs@ultranet.ru

Почти перед каждым современным предприятием стоит задача создания эффективной системы управления нормативно-справочной информацией (НСИ) [5]. НСИ должна быть согласована в рамках информационной системы, полна для решения стоящих перед предприятием задач, актуальна и доступна в территориально-удалённых отделах предприятия. НСИ, не связанная напрямую со спецификой деятельности предприятия, должна базироваться на данных надёжных открытых справочников из внешних источников. Выполнение перечисленных требований желательно возложить на специально созданную автоматизированную систему.

Автором предлагается следующий подход к автоматизации процесса управления внешней НСИ. Вся НСИ сосредотачивается в специально выделенной центральной консолидированной базе данных нормативно-справочной информации (КБД НСИ). Наполнение внешней нормативно-справочной информацией осуществляется при помощи автоматизированной системы импорта данных внешних справочников. Эта система получает внешние справочники, проводит анализ произошедших в них изменений и на его основе производит модификацию КБД НСИ. Распространение изменений в таблицах КБД НСИ осуществляется при помощи гетерогенной системы репликации данных [1-3].

Автором был создан комплекс программ, реализующий данный подход на практике. Он включает в себя автоматизированную систему импорта данных внешних справочников [4] и гетерогенную систему репликации данных [3]. С его помощью было осуществлено внедрение в распределённую информационную систему справочника с иерархической нормативно-справочной информацией, созданной на основании слияния независимых внешних справочников разных форматов. Конкретная реализация автоматизированной системы импорта данных внешних справочников позволила обеспечить собственный контроль над актуальностью данных справочника, а организация самой КБД НСИ позволила поддерживать историю изменений справочника.

Литература

1. Бондаренко А.В., Лисицын А.В., Лисицын М.В., Арабаджиев С.Е. (2005) Синхронизация распределённых баз данных. Часть I. Сравнительный анализ систем репликации и синхронизации распределённых баз данных // Вестник компьютерных и информационных технологий.
2. Бондаренко А.В., Лисицын А.В., Лисицын М.В., Арабаджиев С.Е. (2005) Синхронизация распределённых баз данных. Часть II. Система сеансовой синхронизации распределённых баз данных (SS-синхронизация) // Вестник компьютерных и информационных технологий.
3. Гудков К.С. (2007) Решение проблемы готовности в рамках построения системы репликации баз данных // Труды 50-ой научной конференции МФТИ. Часть VII. Управление и прикладная математика, Том 2.
4. Гудков К.С. (2008) Консолидация нормативно-справочной информации в распределённых информационных системах // Труды 51-ой научной конференции МФТИ. Часть VII. Управление и прикладная математика, Том 3.
5. Захарушкин В.Ф. (2003) Особенности создания информационного обеспечения корпорации. // Электронный журнал «Исследовано в России».

О нижних оценках числа независимых множеств в некоторых классах графов¹

Дайняк Александр Борисович

Аспирант

Московский государственный университет имени М.В.Ломоносова,

Факультет ВМК, Москва, Россия

E-mail: dainiak@gmail.com

Рассматриваются простые неориентированные графы (основные определения см. в [1]). Независимым (н.м.) называется подмножество попарно не смежных вершин графа. Вопросы, связанные с размером и числом н.м. в различных классах графов, возникают в задачах раскраски графов, алгоритмических задачах теории графов, задачах комбинаторики суммирования конечных подмножеств абелевых групп (см., например, [2], [3]). Как правило, исследуются верхние оценки для числа независимых подмножеств вершин в графах из различных классов (регулярные графы, графы с известной максимальной мощностью н.м. и др.) Важную роль играют также максимальные по включению н.м. (м.н.м.). В то же время вопрос нижних оценок числа н.м. и м.н.м. в графах с заданными параметрами практически не исследовался. Нами получены следующие результаты.

Теорема 1. Пусть G – граф, имеющий минимальное число н.м. среди всех графов на n вершинах с m ребрами, у которых степени вершин не превосходят 2. Тогда G представляет собой при $m=n$ объединение $n/3$ циклов длины 3, если $n=0(\text{mod } 3)$, объединение $(n-4)/3$ циклов длины 3 и одного цикла длины 4, если $n=1(\text{mod } 3)$, объединение $(n-5)/3$ циклов длины 3 и одного цикла длины 5, если $n=2(\text{mod } 3)$. При $m \leq n/2$ G является объединением паросочетания на $2m$ вершинах и $(n-2m)$ изолированных вершин. При $n/2 < m < n$ G является объединением паросочетания на $2(n-m)$ вершинах и $(2m-n)/3$ циклов длины 3, если $(2m-n)=0(\text{mod } 3)$, объединением одной цепи длины 3, паросочетания на $2(n-m+1)$ вершинах и $(2m-n-1)/3$ циклов длины 3, если $(2m-n)=1(\text{mod } 3)$. При $n/2 < m < n$ и $(2m-n)=2(\text{mod } 3)$ граф G является объединением паросочетания на $2(n-m+1)$ вершинах и либо $(2m-n-2)/3$ циклов длины 3 и одной цепи длины 4, либо $(2m-n+1)/3$ циклов длины 3 и одной изолированной вершины.

Теорема 2. Пусть G – граф, имеющий минимальное число м.н.м. среди всех графов диаметра d , где $d \geq 4$. Тогда множество вершин G можно разбить на d подмножеств V_1, \dots, V_d так, что выполнены следующие свойства. При любом k , таком, что $2 \leq k \leq d$, и любом j , таком, что $0 \leq j \leq d-k$, в графе G отсутствуют ребра вида $\{u, u'\}$, где $u \in V_j$, $u' \in V_{j+k}$. Для всякого j , $0 \leq j < d$, подграф графа G , порожденный множеством $V_j \cup V_{j+1}$, является полным двудольным с долями V_j и V_{j+1} .

Литература

1. Дистель Р. (2002) Теория графов (пер. с англ.). Новосибирск: Изд-во Ин-та математики, 2002. – 336 с.
2. Сапоженко А. А. (2003) Доказательство гипотезы Камерона-Эрдёша о числе множеств, свободных от сумм // в сб. «Матем. вопросы киберн. Вып. 12». М.:Физматлит, 2003.
3. Eppstein D. (2003) Small maximal independent sets and faster exact graph coloring // J. Graph Algorithms and Applications (special issue for WADS'01) 7(2):131-140, 2003.

¹ Тезисы доклада основаны на материалах исследований, проведенных в рамках гранта РФФИ №07-01-00444.

Восстановление управляющих конструкций языка высокого уровня из исходной программы на языке ассемблера

Деревенец Егор Олегович, Долгова Катерина Николаевна

Студент, Ассистент

Московский государственный университет имени М.В. Ломоносова, Москва, Россия

E-mail: yegord@ispras.ru, katerina@ispras.ru

Декомпиляция – процесс восстановления программы на языке высокого уровня из программы на языке низкого уровня (ассемблера или машинных кодов). Ее можно условно разделить на следующие подзадачи:

1. выделение функций в потоке инструкций,
2. восстановление структурных конструкций языка высокого уровня,
3. выявление обращений к объектам языка высокого уровня: локальным переменным, аргументам функции, массивам, структурам и т.д.,
4. определение типов данных объектов, найденных на предыдущем этапе.

Так как целью декомпиляции является представить восстанавливаемые программы, исходный код которых недоступен, в виде более понятном для анализа, то восстановление структурных конструкций должны быть максимально полным. Ограничиться при восстановлении только одним типом цикла типа while и оператором ветвления if-then нельзя: в этом случае программа будет перегружена непривычными для программиста конструкциями, что сведет на нет эффект от декомпиляции.

Работа посвящена методу восстановления структурных конструкций, который реализован в декомпиляторе *TyDec*, разрабатываемом в Институте системного программирования РАН.

Метод основан на построении графа потока управления программы[1]. В графе потока управления выделяются регионы, где регион – это набор базовых блоков, допускающий только одну входную и только одну выходную дуги. После чего размеченный регионами граф перестраивается в модифицированный граф потока управления, где узлами в отличие от исходного являются не базовые блоки, а регионы. После того, как в графе выделены регионы, выполняется следующая разметка графа: для вершин графа вычисляется множество доминаторов, а дуги классифицируются на прямые, обратные и косые. Далее на модифицированный граф потока управления накладываются шаблоны, соответствующие восстанавливаемым структурным конструкциям программы: if-then, if-then-else, while, do-while, switch в порядке приоритета. Шаблон конструкции if-then имеет высший приоритет, дальше накладываются шаблоны для циклов. Циклом считается множество вершин, доступных из данной по обратному ребру и удовлетворяющих дополнительным условиям, определяющим тип цикла. В последнюю очередь восстанавливается оператор выбора switch. Для него восстанавливается таблица переходов.

Метод работает итеративно. Восстановленная структурная конструкция преобразовывается в новый регион, поле чего на обновленный граф снова накладываются шаблоны структурных конструкций. Выполнение завершается, когда на следующем этапе ни один из структурных шаблонов не подходит.

Также в работе представлен метод восстановления конструкций обработки исключительных ситуаций языка Си++. Исключительные ситуации восстанавливаются посредством анализа переходов между стековыми фреймами.

Литература

1. Muchnick, S., *Advanced Compiler Design and Implementation*. Morgan Kaufmann Publishers 1997.

Исследование и разработка декомпилятора в язык Си

Долгова Катерина Николаевна

Ассистент

Московский государственный университет имени М.В. Ломоносова, Москва, Россия

E-mail: katerina@ispras.ru

Декомпиляция – одна из задач обратной инженерии. Область применения декомпиляторов на практике обширна: от поддержки унаследованных программ до анализа программ на уязвимости, недокументированные возможности и т.д. [1].

Декомпилятор – это инструмент, относящийся к семейству. Декомпилятор переводит программу из языка низкого уровня в язык высокого уровня. Трудность разработки алгоритмов декомпиляции состоит в том, что при декомпиляции нужно выполнить преобразование, обратное компиляции. При компиляции программы много высокоуровневой информации утрачивается. Например, разрушается информация об управляющих конструкциях. Так же утрачивается информация о высокоуровневых типах данных, так как ассемблер бестиповый язык, утрачиваются интерфейсы между подпрограммами, и требуют восстановления протоколы обмена как внутренние, так и внешние.

Так как задача автоматической декомпиляции сама по себе алгоритмически не разрешима, то построение полностью автоматического декомпилятора представляется неразрешимой задачей, что увеличивает важность приближенных эвристических методов ее исследования. Качество восстанавливаемой программы можно оценивать по степени близости восстановленного кода к тому, который мог быть написан программистом вручную. Наличие артефактов трансляции сводит на нет эффект от декомпиляции.

Восстановленная программа должна содержать минимум артефактов декомпиляции таких, как явные приведения типов данных, замена производных типов набором базовых, в частности, структур – набором переменных, а массивов – указателем на первый элемент. Так же должны восстанавливаться высокоуровневые управляющие конструкции: операторы ветвления, циклы (while, do-while и обязательно оператор for), а также оператор множественного выбора switch.

Так как полностью автоматизировать процесс декомпиляции для большинства программ невозможно, то декомпилятор должен поддерживать развитый графический интерфейс, позволяющий пользователю полностью управлять процессом декомпиляции. Такой интерфейс поддерживается декомпилятором TyDec.

Декомпилятор TyDec, разрабатываемый автором, восстанавливает программы на языке ассемблера в программы на языке Си. Декомпилятор содержит модули для восстановления внутренних и внешних интерфейсов обмена. В модуле для восстановления управляющих конструкций реализован метод восстановления высокоуровневых управляющих конструкций, основанный на модификации графа потока управления посредством итеративного выделения регионов. Также реализовано восстановление оператора выбора, основанное на восстановлении таблицы переходов. В модуле восстановления типов данных реализованы методы восстановления как базовых типов, так и производных. Модуль восстановления типов данных декомпилятора TyDec поддерживает как статические, так и динамические методы анализа программы.

Апробация работы декомпилятора проводилась на наборе тестовых примеров, состоящем из программ с открытым исходным кодом. Тестирование показало состоятельность разработанных и реализованных методов.

Литература.

1. «Почему декомпиляция»: <http://www.program-transformation.org/Transform/WhyDecompilation>

Разрешимость и регулярность задач «нечёткой» разметки точечных конфигураций

Дорофеев Николай Юрьевич

Студент

*Московский государственный университет имени М.В.Ломоносова, факультет
вычислительной математики и кибернетики, Москва, Россия*

E-mail: stc.nick@gmail.com

Рассматривается задача построения обучаемых алгоритмов классификации точек в плоских конфигурациях. Постановка задачи дана в [1]. Требуется каждой точке конфигурации сопоставить элемент из некоторого множества, называемого словарём разметки. В работе [2] были рассмотрены вопросы разрешимости и регулярности задач выделения трендов. Указанные задачи были сведены к задаче классификации точек в плоских точечных конфигурациях. Там же были даны определения и получены критерии локальной разрешимости и локальной регулярности этих задач.

Полученные критерии опирались на понятие сдвиг-эквивалентности конфигураций и сдвиг-эквивалентности окрестностей, в котором конфигурации и окрестности считались эквивалентными, если они совпадали с точностью до сдвига. Неразрешимыми в этом случае считались задачи, в которых эквивалентным конфигурациям и окрестностям соответствовали различные разметки. В случае указанного отношения сдвиг-эквивалентности достаточно минимального изменения любой точки окрестности, например, как следствие шумов в данных, и отношение сдвиг-эквивалентности теряется.

Разумной представляется идея потребовать, чтобы «похожим» окрестностям сопоставлялись похожие метки. Для оценки близости окрестностей и разметок предлагается ввести соответствующие метрики. Между метриками над окрестностями и разметками устанавливается соответствие с помощью масштабирующего параметра таким образом, что для окрестностей, находящихся на расстоянии меньше порогового, метки должны отличаться не больше чем на значение порога, умноженное на масштабирующий параметр.

Задача разметки точечной конфигурации фактически задаётся своим набором прецедентов. Под разрешимостью такой задачи понимается существования алгоритма, размечающего конфигурацию в соответствии с набором прецедентов и удовлетворяющего указанному ограничению. Располагая указанными метриками можно сформулировать понятие противоречивого набора прецедентов – набора, в котором близкие окрестности имеют далёкие с поправкой на масштабирующий коэффициент метрики. Это в свою очередь, позволяет сформулировать необходимые и достаточные условия разрешимости поставленной задачи. Разрешимая задача называется регулярной, если разрешима любая регулярная задача из некоторой её окрестности. Эта окрестность формируется с помощью произвольного изменения разметки набора прецедентов.

В докладе ставятся задачи синтеза локального алгоритма «нечёткой» разметки точечных конфигураций. Локальность понимается в работе в двух значениях. С одной стороны локальный алгоритм использует для разметки окрестность точки, в общем случае являющуюся подмножеством всей конфигурации. С другой алгоритм даёт содержательный ответ лишь для объектов, находящихся в некоторой окрестности множества прецедентов. На остальных конфигурациях алгоритм выдаёт отказ от классификации.

В докладе представлены критерии разрешимости и регулярности задач нечёткой разметки точечных конфигураций.

Литература

1. Чехович Ю.В. (2002) Об обучаемых алгоритмах выделения трендов // Искусственный интеллект (научно-теоретический журнал НАН Украины). № 2, с. 298-305.
2. Рудаков К.В., Чехович Ю.В. (2003) Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // Доклады Академии наук. Том 388, № 1, с. 33-36.

Оценка мимической динамики движения челюсти в процессе жевания по трёхмерному видеоряду¹

Дышкант Наталья Федоровна

Аспирант

Московский государственный университет имени М.В. Ломоносова, факультет
Вычислительной математики и кибернетики, Москва, Россия

E-mail: nfd3001@gmail.com

Задача оценки мимической динамики движения челюсти человека в процессе жевания актуальна для исследований в области челюстно-лицевой хирургии и ортодонтии. Входными данными в задаче являются трёхмерная модель S лица, полученная при нейтральном выражении лица снимающегося, и трёхмерный видеоряд – последовательность трёхмерных моделей D_1, D_2, \dots, D_n лица, полученная при съемке жующего человека. Требуется получить описание траектории движения челюсти в процессе жевания.

Съемка для данного исследования производилась трёхмерным сканером компании «Artec Group» [1]; полученные трёхмерные модели лица заданы в виде облаков точек.

Для сравнения двух снимков из видеоряда предлагается подход, заключающийся в

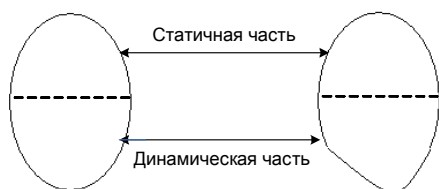


Рис. 1. Разделение модели лица на две части

определении статичной (верхней) части лица методом подгонки и описании движения динамической (нижней) части лица относительно верхней. В первоначальной модели S и в каждой модели D_i выделяются статичная и подвижная части (см. Рис.1). Идея подхода заключается в построении преобразований (движений) модели D_i , при которых наилучшим образом совпадают по отдельности статичная и динамическая части

лица. Подгонка двух поверхностей производится методом, предложенным в [2], и состоит в нахождении такого движения, при котором мера различия между поверхностями, описывающими лица, минимальна.

Для верхней и нижней частей лица строятся собственные местные системы декартовых координат. Для модели S верхняя и нижняя системы координат совпадают, а для моделей D_i они различаются. Задача состоит в регистрации этих различий. Преобразование нижней системы координат в верхнюю для каждого кадра D_i и описывает динамику движения нижней челюсти. Формальное описание движения представляется в виде матриц преобразования нижних координат в верхние. Результат может быть визуализирован в виде анимации, показывающей движение базисных векторов нижней системы координат при фиксированном положении базисных векторов верхней системы.

Предложенный метод позволяет сегментировать трёхмерную модель на статичную и динамическую части и описывать динамику движения этих частей относительно друг друга.

Литература

1. www.artec-group.com (Artec Group – технологии 3D сканирования).
2. Дышкант Н.Ф., Местецкий Л.М. «Сравнение однолистных поверхностей полученных при 3D сканировании» // Proceedings of 18th International Conference on Computer Graphics and Vision «GraphiCon'2008», 2008, С.270-277.

¹ Работа поддержана грантами РФФИ (проекты 08-01-00670-а, 08-07-00305-а).

Использование мультифрактальных показателей при оценивании параметров гауссовского и мультипликативного шума на изображениях

Емец Юрий Владимирович¹

Аспирант

Одесский национальный политехнический университет, кафедра прикладной математики и информационных технологий в бизнесе, Одесса, Украина

E-mail: emetsuv@rambler.ru

Для подавления гауссовского и мультипликативного шума существуют особые методы предварительной обработки изображений. Неточная оценка помеховой ситуации или параметров шума могут привести к некачественной предварительной обработке. Недооценка уровня мультипликативной помехи вызывает большие разрывы контуров объектов изображений при решении задачи контурной сегментации. Это, в свою очередь, влияет на ошибку распознавания этих объектов. Завышенная оценка уровня аддитивной гауссовской помехи ведет к размыванию контуров объектов, что также отражается на ошибке распознавания. Поэтому актуальной является проблема оценка параметра гауссовского и мультипликативного шума на изображениях.

Целью данной работы являлось снижение погрешности оценивания отношения сигнал/шум для гауссовского и мультипликативного шума путем использования мультифрактальных показателей.

Приведены определения структурной функции и сингулярной меры, которые позволяют вычислять мультифрактальные показатели H_1 и C_1 для одномерных данных. Для обработки изображений, представленных матрицами дискретных отсчетов интенсивности в [1] введены показатели H_1 , C_1 по направлениям x , y : H_x , C_x , H_y , C_y , H_{xy} , C_{xy} соответственно. Также в [1] получены графики зависимостей значений мультифрактальных показателей от отношения сигнал/шум Q для гауссовского и мультипликативного шума.

В данной работе на основе этих зависимостей осуществлялось оценивание параметра гауссовского и мультипликативного шума. Получены уравнения регрессии, описывающие зависимость отношения сигнал/шум для гауссовского и мультипликативного шума Q от значений показателей H_x , H_y , H_{xy} , C_x , C_{xy} . В качестве регрессионной зависимости выбрана полиномиальная. Оценены коэффициенты полинома регрессии методом наименьших квадратов, который используется для приближения функций, заданных числовым массивом. Качество уравнений регрессии отношения сигнал/шум для гауссовского и мультипликативного шума на значения H_x , H_y , H_{xy} , C_x , C_{xy} определялось с помощью коэффициента детерминации.

Из шести исследованных зависимостей выбрана та, для которой коэффициент детерминации максимален. В нашем случае – это зависимость Q от C_{xy} , которая аппроксимируется полиномом 8-й степени:

$$Q = a_0 + a_1 C_{xy} + a_2 C_{xy}^2 + a_3 C_{xy}^3 + \dots + a_8 C_{xy}^8, \quad (1)$$

где a_0, a_1, \dots, a_8 – коэффициенты полинома регрессии.

С учетом этой формулы метод оценивания отношения сигнал/шум Q для гауссовского и мультипликативного шума формулировался следующим образом: для исходного изображения необходимо оценить C_{xy} ; с использованием формулы (1) вычислить отношение шума сигнал/шум для гауссовского и мультипликативного шума.

Литература

1. Полякова М.В., Крылов В.Н. Мультифрактальный метод автоматизированного распознавания помех на изображении // Автоматика. Автоматизация. Электротехнические комплексы и системы. Межвузовский

¹ Автор выражает признательность профессору, д.т.н. Крылову В.Н. и к.т.н. Поляковой М.В. за помощь в подготовке тезисов.

журнал. – Херсон, 2006. - № 1 (17). – С. 60 - 69.

Эффективная работа с данными при решении задач коллаборативной фильтрации.

Заборовский Никита Владимирович

Студент

Московский физико-технический институт (государственный университет)

e-mail: turnik@mail.ru

С начала 90-х годов создаются рекомендательные системы для решения задач коллаборативной фильтрации (*collaborative filtering*), позволяющие делать автоматические прогнозы интересов конкретного пользователя относительно его оценок интересов других пользователей. Большое применение такие системы нашли в интернет-магазинах (Amazon, Ozon), системе проката фильмов Netflix. Использование такой системы для музыкальных и видео сайтов сильно повышает их привлекательность с точки зрения конечного пользователя.

При реализации методов решения задач коллаборативной фильтрации, возникает необходимость работать с очень большим количеством данных. Чтобы понять масштабы задачи, можно рассмотреть, к примеру, такие проекты как livejournal или ВКонтакте. В каждом из этих проектов число пользователей может быть оценено несколькими десятками миллионов, число интересов – порядка миллиона. Если составить матрицы пользователи – интересы или пользователи – пользователи, получится, что эти матрицы - сильно разрежены (*sparse*): из миллиона интересов пользователь отмечает в среднем тридцать.

Для того чтобы обеспечить возможность работы с несколькими матрицами такого размера даже с учётом способов, обеспечивающих компактное хранение данных в разреженных матрицах, необходима реализация менеджера памяти. К менеджеру памяти (далее МП) предъявляются требования:

1. Функциональность МП может быть использована для полного замещения стандартного функционала работы с памятью в ОС Windows.
2. Реализация МП должна предусматривать масштабируемость и различные политики работы с виртуальной памятью и с диском.
3. Код МП не должен быть тяжеловесным, должен быть простым и эффективным в смысле производительности.
4. МП должен позволять оперировать объёмами данных, превышающими объём виртуальной памяти, выделенной на процесс (для этого должны быть введены виртуальные указатели и своппинг).
5. Архитектура компонента МП должна предусматривать наиболее простой переход от указателей в уже реализованной библиотеке работы с разреженными матрицами к виртуальным указателям МП, используя подход объектно-ориентированного программирования.
6. В будущем возможно написание специальных драйверов для более эффективной работы с памятью и диском.

Литература

1. Стивен Дьюхерст. С++ Священные знания. СПб.: Символ-Плюс, 2007
2. Альфред В.Ахо, Джон Э. Хопкрофт, Джеффри Д. Ульман. Структуры данных и алгоритмы. Издательский дом "Вильямс" Москва 2000
3. Руссинович М., Соломон Д. Внутреннее устройство Microsoft Windows: Windows Server 2003, Windows XP и Windows 2000. Мастер-класс. М.: "Русская редакция".

Семейство разностных схем с дискретными прозрачными граничными условиями для уравнения Шрёдингера на полуоси

Злотник Илья Александрович

Студент

Московский энергетический институт, Москва, Россия

E-mail: ilya.zlotnik@gmail.com

Рассматривается начально–краевая задача для обобщенного уравнения Шрёдингера на полуоси

$$i\hbar\rho\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}\frac{\partial}{\partial x}\left(B\frac{\partial\psi}{\partial x}\right) + V\psi \text{ при } x > 0, t > 0,$$
$$\psi|_{x=0} = 0, \psi(x, t) \rightarrow 0 \text{ при } x \rightarrow \infty \text{ для всех } t > 0,$$
$$\psi|_{t=0} = \psi^0(x) \text{ при } x > 0.$$

с искомой комплекснозначной функцией $\psi = \psi(x, t)$. Здесь i – мнимая единица, $\hbar > 0$ – постоянная, коэффициенты ρ, B, V вещественны и зависят от x , причем ρ и B строго положительны. Предполагается также, что при $x \geq X_0$ коэффициенты постоянны, а начальная функция ψ^0 равна 0 (для некоторого $X_0 > 0$).

Для численного решения этой задачи строится семейство двухслойных симметричных разностных схем с трехточечным усреднением по пространству, зависящим от параметра θ ; оно включает ряд известных схем. Для семейства схем ставится абстрактное приближенное прозрачное граничное условие (ПГУ). В связи с данной проблематикой см. обзор [1]. Предлагается способ такого сочетания схем с абстрактным приближенным ПГУ, который позволяет достаточно просто анализировать их устойчивость, а для дискретного ПГУ приводит к вычислительно устойчивой форме записи.

Доказываются две равномерные по времени оценки, выражающие устойчивость решения в весовой сеточной норме L_2 и в энергетической норме по отношению к начальным данным и свободным членам в уравнении и приближенном ПГУ, при соответствующих условиях на оператор приближенного ПГУ.

Для всех схем семейства выводится дискретное ПГУ и доказывается выполнение для его оператора условий устойчивости (при определенном условии на θ). Предлагается и анализируется вычислительно устойчивая форма записи дискретного ПГУ.

Семейство разностных схем с дискретным ПГУ программно реализовано в среде **MATLAB**. Выполнены серии численных экспериментов, демонстрирующих как устойчивость вычислений, так и полное отсутствие отражений от искусственной границы.

Подробно результаты представлены в [2,3]. Работа выполнена при частичной финансовой поддержке РФФИ, проект 08-01-90009-Бел.

Литература

1. **X. Antoine, A. Arnold, C. Besse, M. Ehrhardt, A. Schädle.** A review of transparent and artificial boundary conditions techniques for linear and nonlinear Schrödinger equations // Commun. Comp. Phys. 2008. V. 4. № 4. P. 729–796.
2. **А.А. Злотник, И.А. Злотник.** Об устойчивости семейства разностных схем с приближенными прозрачными граничными условиями для уравнения Шрёдингера на полуоси // Вестн. МЭИ. 2008. № 6.
3. **B. Ducomet, A. Zlotnik, I. Zlotnik.** On a family of finite–difference schemes with discrete transparent boundary conditions for a generalized Schrödinger equation // Kinetic and Related Models. 2009. V. 2. № 1. P. 151–179.

Моделирование работы систем резервного копирования данных¹

Казаков Виталий Гайясович²

Аспирант

*Мордовский Государственный Университет имени Н.П. Огарева, факультет
электронной техники, Саранск, Россия*

E-mail: vitalykg@gmail.com

Повышение эффективности процессов резервного копирования и восстановления путем разработки новых схем, а также создание методов прогнозирования основных характеристик данных процессов, ведут к необходимости разработки методов математического и компьютерного моделирования работы систем резервного копирования.

Основная идея разрабатываемого метода моделирования заключается в рассмотрении резервного копирования как последовательности элементарных копий, производимых в процессе резервирования. Каждая элементарная копия, несет информацию об изменении между двумя состояниями данных для резервирования. Элементарным копиям присваиваются сигнатуры, однозначно идентифицирующие периоды изменений. Выбор формата записи сигнатур определяется удобством записи схем резервного копирования в общем виде, удобством введения операции объединения элементов.

Работу схем резервного копирования удобно представлять в виде ориентированных мультиграфов. Понятие пути восстановления в такой интерпретации становится наглядным и совпадает с понятием пути в орграфах. При работе сложных алгоритмов создания резервных копий, появляется потребность в нахождении оптимального пути восстановления. Согласно проведенным исследованиям гибридный алгоритм [1] в составе с алгоритмом Йена вполне подходит для данной задачи.

Используя записи алгоритмов в общем виде [2] можно составить совокупность хранимых элементов на каждый момент времени. Поставив в соответствие каждому элементу репозитория число, обозначающее его объем, и просуммировав их все, мы получим оценку места, требуемого для хранения резервных копий, необходимого для работы рассматриваемой схемы резервного копирования. Для этого нужно иметь информацию о характере роста и изменения данных для резервного копирования, полученную, к примеру, на основе имеющейся статистике работы системы.

Для компьютерного моделирования была разработана программная система резервного копирования, средствами которой возможно проведение тестирования разных аспектов работы процессов резервирования и восстановления, как на моделируемых данных, так и на реальных. В системе реализован ряд схем резервного копирования, среди которых: полное; инкрементное; дифференциальное; мультиуровневое; схема А.М. Костелло et al.; Z scheme и др. В систему внедрены инструмент аналитического прогнозирования, механизм отыскания оптимальных путей восстановления. Средствами системы была проведена серия экспериментов, показавшая работоспособность предложенного подхода.

Литература

1. Pascoal M.M.B. (2006) Implementations and empirical comparison for K shortest loopless path algorithms // The Ninth DIMACS Implementation Challenge: The Shortest Path Problem.
2. Казаков В.Г., Федосин С.А. Метод моделирования алгоритмов резервного копирования для получения оценок объема репозитория.// Инфокоммуникационные технологии, Т.3, 2008. С.126-132.

¹ Работа поддержана Фондом содействия развитию малых форм предприятий в научно-технической сфере в рамках программы «УМНИК» (№08-2-8446); работа также поддержана в рамках «Аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы» (№ 2.1.2/2485).

² Автор выражает признательность профессору, к.т.н. Федосину С.А. за помощь в оформлении тезисов.

Критерий k -сингулярности системы точек и оптимальное разбиение на подсистемы¹

Карпович Павел Алексеевич

аспирант

*Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия,
факультет Вычислительной математики и кибернетики
e-mail: pkarпович@mail.ru*

Система q точек в конечномерном пространстве называется k -сингулярной, если размерность пространства значений полиномов (с адамаровым умножением) от столбцов матрицы попарных расстояний точек системы меньше q . Например, 1-сингулярные системы – системы с вырожденной матрицей попарных расстояний. Исследования этих систем вызваны приложениями в теории интерполяции [1], [2] и в алгебраическом подходе к решению задач распознавания образов [3]. Вопросы возможности точной интерполяции и корректности алгебраических замыканий сводятся к выяснению, является ли система точек k -сингулярной. Причём особый интерес представляет изучение k -сингулярности в метрике l_1 , для которой более 50 лет (начиная с работы [4]) не удавалось получить геометрический критерий вырожденности матрицы попарных расстояний [1].

В докладе излагаются результаты, связанные с некоторыми свойствами k -сингулярных систем: критерий k -сингулярности, разделение 1-сингулярной системы точек на минимальное количество непересекающихся подмножеств, каждое из которых не обладает свойством 1-сингулярности.

Теорема. Система точек $S = \{\mathcal{S}_i\}_{i=1}^q$ пространства \mathbf{R}^m является k -сингулярной тогда и только тогда, когда существует ненулевой вектор (c_1, \dots, c_q) такой, что для всех $\mathcal{T} \in \mathbf{R}^m$ справедливо равенство

$$\sum_{i=1}^q c_i \rho^k(\mathcal{T}, \mathcal{S}_i) = 0$$

где ρ – метрика Хэмминга или l_1 метрика.

Этот результат является прямым обобщением критерия 1-сингулярности, полученного в [3]. Кроме того, предложен алгоритм разделения 1-сингулярной системы точек на минимальное число непересекающихся не 1-сингулярных подмножеств. Получена верхняя оценка на число таких подмножеств.

Литература

1. Reid L., Sun X. Distance matrices and ridge function interpolation // Canadian Journal of Mathematics. – 1993. – V. 45. – pp. 1313–1323.
2. Baxter B.J.C. Conditionally positive functions and p -Norm distance matrices // Constr. Approx. – 1991. – №7. – Pp.427–440.
3. Дьяконов А.Г. Критерии корректности алгебраических замыканий модели алгоритмов вычисления оценок // Доклады Академии наук, 2008, Т. 420, №6, С. 732–735.
4. Schoenberg I.J. Metric spaces and completely monotone functions // Ann. Math. – 1938. – №39. – Pp.811–841.

¹ Работа поддержана Российским фондом фундаментальных исследований, грант № 08-07-00305-а.

Надстройка над системой пакетной обработки задач пользователя MAUI с целью управления динамическими приоритетами.

Князев Николай Александрович

Студент

*Московский государственный университет имени М.В.Ломоносова,
Факультет Вычислительной Математики и Кибернетики, Россия, Москва*

E-mail: irumata@gmail.com

При работе с многопроцессорными вычислительными системами используются Системы Управления Пакетной Обработки(СПО).

Через СПО пользователь добавляет задачи в очередь на вычислительной системе, далее система планирования задач (Scheduler) принимает решение, какие ресурсы необходимо для этого задания выделить и в какой момент.

Для того, чтобы более важные задания выполнялись вперед менее важных, во многих планировщиках существует возможность при запуске задания указывать её приоритет. Проблемой является то, что пользователь, который может распределить приоритеты между своими задачами, не обладает достаточной компетенцией, чтобы определить приоритет своей задачи по сравнению с другими.

Предлагаемая система, являющиеся надстройкой над планировщиком “MAUI”, решает эту проблему использованием невозобновляемых “фишек приоритета”.

Каждому пользователю выдаётся набор "фишек", администратором или автоматически, возможно на коммерческой основе.

Эти фишки пользователь может назначать на свои задания при постановке в очередь. На основе числа фишек и объема задачи предлагаемое средство рассчитывает приоритет задачи и передает планировщику. Таким образом, пользователь может запускать много заданий с небольшим приоритетом или одно с высоким.

Как и любая СПО, разрабатываемая система разделена на клиентскую часть, работающую на front-end и серверную. Пользователь, минуя клиентскую часть MAUI, ставит задачу в очередь через клиент, затем клиент, находящийся на машине front-end, отправляет задание и количество использованных фишек программе, находящейся на сервере, которая работает в фоновом режиме. Эта программа, получая задание, и на основе израсходованных фишек создаёт задачу и приоритет и передает её планировщику MAUI, с которым находится на одной машине. Доступ же к MAUI из front-end напрямую невозможен. Таким образом, не допускается запускать свои задачи, “вручную” предоставляя приоритеты.

Благодаря тому, что планировщик MAUI может работать совместно с большинством используемых СПО, предлагаемое решение также может использоваться в большинстве современных СПО. В частности использование MAUI в составе MBS-1000 через интерфейс Wiki было продемонстрировано в работе [1].

Литература:

1. А.В. Баранов, Д.М. Голинка “Исследование возможности использования планировщика Maui в составе СУПЗ MBS-1000”
2. А.В. Баранов О. Лацис С.В. Сажин М.Ю. Храпцов “Руководство пользователя системы MBS-1000/RSC4.”
3. IBM Corporation “Tivoli Workload Scheduler LoadLeveler Using and Administering” USA
4. Коваленко В.Н., Орлов А.В. (2002) “Управление заданиями в распределенной среде и протокол резервирования ресурсов” М. ИПМ им. М.В.Келдыша РАН
5. Cluster Resources Inc. www.clusterresources.com Документация системы Maui Cluster Scheduler.

О методах трёхмерного моделирования фарлей-бунемановской неустойчивости на современных многопроцессорных вычислительных комплексах

Ковалёв Дмитрий Владимирович¹

аспирант

*Московский государственный университет им. М.В.Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: dmitry.kovalev@mail.ru

Фарлей-бунемановская неустойчивость - это двухпоточковая неустойчивость, наблюдаемая в плазме E-области ионосферы Земли [1,2]. Для описания неустойчивости используется гидродинамическое уравнение для электронной плотности, кинетическое уравнение для ионов и уравнение Пуассона [1,3]. В отличие от большинства предыдущих работ моделирование проводится с помощью численных методов, применённых к решению уравнений в частных производных, а не методом частиц, что позволяет не увеличивать электронную массу для проведения расчетов [2,4].

Из-за недостатка вычислительных мощностей сложные нелинейные моделирования фарлей-бунемановской неустойчивости в основном проводились в двумерном пространстве. Появление высокопроизводительных вычислительных комплексов позволило использовать более продвинутые модели неустойчивости. Предполагается, что новые моделирования прольют свет на механизмы аномального нагрева электронов и на особенности насыщения неустойчивости в трёхмерном пространстве.

Данная работа посвящена описанию методов эффективного решения уравнений, входящих в трёхмерную модель неустойчивости, на современных многопроцессорных комплексах. Рассматриваются методы распараллеливания алгоритмов решения набора четырёхмерных (в пространстве координат и скоростей) кинетических уравнений, нелинейных трехмерных конвекционно-диффузионных уравнений и уравнения Пуассона. Проведено сравнение различных алгоритмов решения трёхмерного уравнения диффузии.

Расчеты, представленные в данной работе, проводились на вычислительных комплексах IBM Blue Gene/P (факультет ВМК МГУ) и СКИФ МГУ Чебышёв (МГУ). Для обоих комплексов исследована зависимость времени выполнения тестовой задачи от используемой вычислительной мощности (в Gflops). Результаты показывают, что при наличии интенсивного обмена данными между узлами, комплекс IBM Blue Gene/P имеет преимущество до 10% во времени выполнения для одинаковой задействованной вычислительной мощности.

Литература

1. Kovalev D.V., Smirnov A.P., Dimant Y.S. (2008) Modeling of the Farley-Buneman instability in the E-region ionosphere: a new hybrid approach // *Annales Geophysicae*, № 26(9), p. 2853-2870.
2. Oppenheim M.M., Dimant Y., Dyrud L.P. (2008) Large-scale simulations of 2-D fully kinetic Farley-Buneman turbulence // *Annales Geophysicae*, № 26(3), p. 542-553.
3. Ковалёв Д.В. (2008) Моделирование фарлей-бунемановской неустойчивости с использованием четырехмерного кинетического уравнения // *Математическое моделирование*, №20(12), с. 89-104.
4. Ковалёв Д.В., Смирнов А.П., Димант Я.С. (2009) О влиянии изменения электронной массы в численных расчетах фарлей-бунемановской неустойчивости // *Вестник Московского университета. Серия Вычислительная математика и кибернетика*, №33 (1), в печати.

¹ Автор выражает признательность своему научному руководителю доценту к.ф.м.н. Смирнову А.П. за помощь в подготовке тезисов.

Прогнозирование временных рядов при несимметричной функции потерь: прогнозирование плотности распределения и квантильная регрессия

Коваль Артур Сергеевич

Студент 6 курса

*Московский физико-технический институт (государственный университет),
Долгопрудный, Россия*

E-mail: archi87@rambler.ru

Во многих прикладных задачах прогнозирования, в частности, в задаче прогнозирования потребительского спроса, вместо стандартной квадратичной функции потерь используют несимметричную функцию потерь. Задача построения прогноза при несимметричной функции потерь занимает скромное место в обширной литературе по прогнозированию временных рядов по сравнению с классическим методом наименьших квадратов. Стандартный метод для построения прогнозов при кусочно-линейной функции потерь, представляющей собой несимметричную L_1 -норму – это метод квантильной регрессии (quantile regression) [1]. Главным её недостатком является численная неэффективность, особенно на длинных временных рядах. В данной работе предлагается более эффективный метод, основанный на оценивании плотности распределения ошибок по скользящему контролю.

За основу может быть взят любой стандартный метод точечного прогнозирования с квадратичной функцией потерь, в частности, один из адаптивных методов [2]. Строится временной ряд ошибок точечных прогнозов в режиме скользящего среднего, затем по нему строится эмпирическая плотность распределения ошибок (density forecast) [3]. Для получения прогноза плотности в заданный момент времени эмпирическая плотность распределения ошибок сдвигается на величину точечного прогноза в данный момент времени. Наконец, с помощью свёртки эмпирической плотности с несимметричной функцией потерь вычисляется квантильный прогноз и величина потерь в данный момент времени. Главным преимуществом предложенного метода (по сравнению с квантильной регрессией) является его вычислительная эффективность. Второе преимущество — возможность использовать произвольные функции потерь, а не только несимметричную L_1 -норму.

Для сравнения предложенного метода и квантильной регрессии были проведены вычислительные эксперименты на реальных временных рядах объемов продаж в сети супермаркетов. Всего было рассмотрено более 40 рядов. Вид авторегрессионной модели был одинаков для обоих методов. Эксперименты показали, что предложенный метод не только строит прогнозы на порядок быстрее квантильной регрессии, но в ряде случаев является более точным по качеству прогнозов.

Литература

1. Постникова Е. (2000) Квантильная регрессия. НГУ.
2. Лукашин Ю.П. (2003) Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика.
3. Tay A. S., Wallis K. F. (2000) Density forecasting: A Survey // Journal of forecasting, №19, p. 235–254.

Компьютерная технология исследования потребительского спроса с помощью обобщенного непараметрического метода

Кондраков Иван Александрович

аспирант

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия*

e-mail: ivankondrakov@mail.ru

Актуальность построения экономических индексов цен и потребления для случаев массово собираемых “кассовых” торговых статистик связана с необходимостью оперативно отслеживать динамику предпочтений потребителей товаров и услуг, гибко управлять процессом ценообразования, количественно прогнозировать поведение потребителей при изменениях цен.

Цель и задачи исследования. Целью работы было совершенствование методологии исследования структуры спроса и построения экономических индексов цен и потребления.

Предметом исследования явилась методология анализа структуры потребительского спроса и расчета экономических индексов, основанная на обобщенном непараметрическом методе (ОНМ), и проблема ее развития и совершенствования для узкоспециализированных рынков. Основой методологии служит теорема Африата-Веряна о рационализированности потребительского спроса [1-2]. ОНМ базируется на процедуре выявления групп товаров, связанных естественными отношениями взаимозаменяемости-взаимодополняемости таким образом, что совокупный спрос потребителей на товары этой группы можно описать как результат максимизации некой линейно однородной функции объемов продаж – функции полезности.

Вычисляемые по торговой статистике с помощью ОНМ индексы являются индексами Конюса-Дивизиа.

Результаты. 1. Разработаны предложения по совершенствованию подготовки временных рядов торговой “кассовой” статистики для случаев отсутствующих продаж в магазине по причинам недопоставки товара, отсутствия покупательского спроса, выбытия устаревшей, появления новой номенклатуры товара.

2. Эмпирически (для статистики продаж безалкогольных напитков в 643-х магазинах Москвы за 12 месяцев 2007 г.) доказано существование реальных конкурентных узкоспециализированных рынков, объединяющих магазины мегаполиса (Москвы).

3. Сделан вывод: что для товаров одного бренда существует единый московский рынок, на котором можно определить единую рыночную цену (индекс).

4. Эмпирически выявлено рациональное поведение потребителей, заключающееся в ярко выраженном предпочтении покупки товара нужного бренда в любом магазине, но не в замещении отсутствующего в магазине бренда другим брендом, продаваемым в этом же магазине. Другими словами, было показано, что гораздо лучше агрегируется один товар по всем магазинам, чем все товары одного магазина.

Технология поддерживается программным комплексом, разработанным на языке C# на платформе .NET в среде MS VS 2005.

Литература

3. Afriat S.N. On a system of inequalities in demand analysis and extension of the classical method. // International economic review. 1973. V. 14. № 2. P. 460-472.
4. И.А. Кондраков, Поспелова Л.Я., Усанов Ю.А., Шананин А.А. (2007) Разработка технологии и инструмента исследования потребительских рынков с помощью обобщенного непараметрического метода. // Сообщения по прикладной математике. М.:ВЦ РАН, 2007, 52 стр.
5. В. А. Гребенников, А. А. Шананин, “Обобщенный непараметрический метод: закон спроса в задачах прогнозирования” // Матем. моделирование, 20:9 (2008), 34–50.

Система генерации тестовых программ с использованием ограничений ТЕСЛА

Корныхин Евгений Валерьевич

Аспирант

*Московский государственный университет имени М.В.Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия
E-mail: kornevgen@gmail.com*

Системное функциональное тестирование микропроцессоров является хорошо зарекомендовавшей себя техникой тестирования микропроцессоров. Оно выполняется с использованием большого числа программ на языке ассемблера (тестовых программ). Для сложных современных процессоров таких программ должно быть много, что делает актуальной задачу автоматической генерации тестовых программ. В работе [1] предложена технологическая цепочка построения тестовых программ на основе модели микропроцессора. В рамках этой цепочки сначала тестовые программы систематически строятся в абстрактном виде (в виде *тестового шаблона*) – без конкретных параметров инструкций. Тестовые шаблоны описывают последовательность инструкций, параметры инструкций с указанием ограничений на параметры и происходящих событий – *тестовой ситуации* (например, переполнение, промахи или попадания в кэш-памяти). В качестве параметров инструкции могут быть указаны регистры и константы. Чтобы получить тестовую программу по заданному тестовому шаблону, достаточно найти начальные значения регистров и той части кэш-памяти и других подсистем, с которыми работают инструкции шаблона. Эти начальные данные называются *тестовыми данными*, а задача их построения – задачей *генерации тестовых данных*. По тестовым данным строится набор инструкций инициализации микропроцессора (загрузка значений в регистры, кэш и т.д.), который добавляется в начало тестового шаблона.

Для решения этой задачи разработана система ТЕСЛА, включающая в себя императивный язык описаний тестовых ситуаций (для описания поведения инструкций микропроцессора) и генератор тестовых данных. Описание тестовой ситуации представляет собой последовательность присваиваний, операторов утверждений (assert) и вызовов специальных процедур (осуществляющих загрузку и сохранение данных в памяти и трансляцию адресов). Генератор тестовых данных использует технологию разрешения ограничений (constraint). При этом (в отличие от других инструментов) весь шаблон с тестовыми ситуациями транслируется в одну задачу на разрешение ограничений, что допустимо для шаблонов из [1] и помогает использовать зависимости между инструкциями для более эффективного разрешения ограничений. Кроме того, генератор тестовых данных использует новый алгоритм трансляции тестовых ситуаций в кэш-памяти и буфере трансляции адресов, который позволяет значительно уменьшить сложность получающихся ограничений по сравнению с кодированием последовательности изменений состояния кэш-памяти.

Преимуществом использования системы ТЕСЛА является низкая трудоемкость подготовки начальных данных (описаний тестовых ситуаций), так как открытые стандарты архитектуры микропроцессоров уже содержат описания тестовых ситуаций. Кроме того, из-за наличия похожей функциональности инструкций разных микропроцессоров высок процент переиспользования описаний тестовых ситуаций. Тестирование микропроцессоров с помощью тестовых программ позволяет проводить его еще на стадии его проектирования, что удешевляет процесс исправления ошибок в микропроцессоре. Прототип системы ТЕСЛА выполняется для архитектур микропроцессоров стандарта MIPS64.

Литература

1. Камкин А.С. (2008) Генерация тестовых программ для микропроцессоров // Труды ИСП РАН. Том 14(2). С.23-64.

Определение параметров источника света по фотографиям

Коробченко Дмитрий Александрович, Ильин Андрей Алексеевич

студент, аспирант

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия
e-mail: lklabs@gmail.com, ailyin@graphics.cs.msu.ru*

Для многих задач компьютерной графики требуется автоматическое определение параметров источника света реальной сцены, в частности его положения в системе координат сцены. Без специальных фотометрических устройств измерить координаты источника с необходимой точностью достаточно трудоёмко. Существует алгоритм, позволяющий зарегистрировать источник света в сцене путем анализа фотоизображения, сделанного из позиции этого источника. Но при данном подходе требуется присутствие калибровочного объекта в сцене, специального пакета калибровки, дополнительное фото, а также становится невозможна съемка сцены на видео.

В данной работе был предложен и реализован алгоритм автоматического определения положения источника света по набору фотоизображений. Данный алгоритм отличается от описанных выше меньшими затратами при съёме данных, так как не требует дополнительных измерений и объектов в сцене. На вход алгоритму подается несколько фотоизображений некоторого объекта содержащих блики от источника света с разных ракурсов. Так же на вход подается геометрическая модель исследуемого объекта, по которой можно создать несколько текстур объекта, где каждой точке объекта на текстуре соответствует нормаль и положение в системе координат сцены. Чтобы найти области бликов, из изображений, содержащих блики, попиксельно вычитается изображение без бликов [1] (рис. 2). Зачастую при использовании фотоаппарата возникает проблема с шумом на фотоизображении, которая может быть частично решена применением некоторых алгоритмов шумоподавления [2]. Пиксели, принадлежащие блику, находятся методами машинного зрения [3]. Для каждого блика ищется его геометрический центр. По координатам центра блика, положению камеры и нормали в данной точке, находится направление на источник из этой точки. (рис. 1) Если бликов на изображении несколько, производится усреднение по ним. По совокупности лучей, полученных с разных изображений, вычисляется положение источника. В частности, исследуемым объектом может быть пластинка. Предложенный алгоритм способен восстановить направление на источник света лишь по одной фотографии пластинки, содержащей блик. Координаты вычисляются с использованием дополнительно измеренного расстояния от блика до источника.

Апробация алгоритма проводилась в системе реконструкции BRDF [1], и на практике были получены результаты с приемлемой точностью.

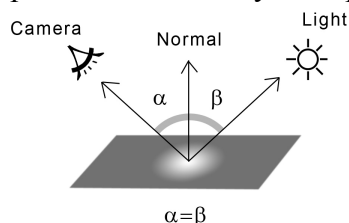


Рис.1. Закон отражения



Рис.2. Нахождение блика

Литература

6. A. Ilyin, A. Lebedev, V. Sinyavsky, A. Ignatenko "The System for the Acquisition, Processing and Material Rendering from Images" Proc. of Graphicon'2008, pp. 134-141, Moscow, Russia, June 2008.
7. A. Lukin "A multiresolution Approach for Improving Quality of Image Denoising Algorithms" ICASSP-2006, Toulouse, France. 2006.
8. G. Medioni, S. B. Kang "Emerging Topics in Computer Vision" Prentice Hall, 2005.

Внутренняя эллипсоидальная оценка пересечения двух concentрических эллипсов.

Косенкова Лидия Викторовна

студентка

*Московский государственный университет имени М.В.Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: kosenkova.lida@gmail.com

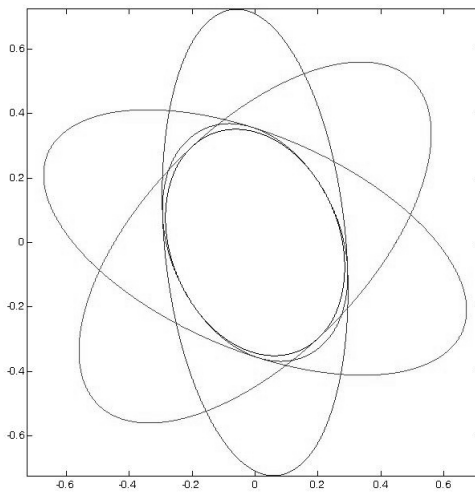
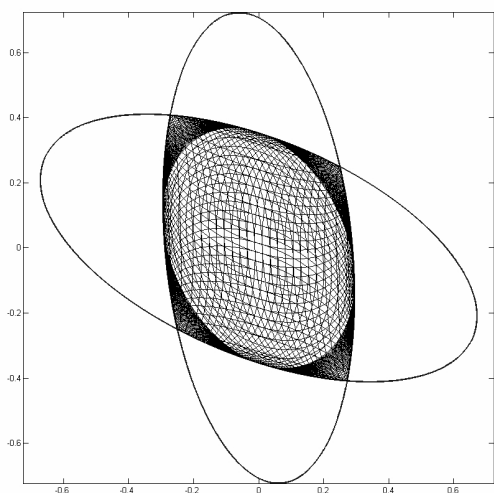
Исследование динамических систем с неопределенностью является одним из востребованных направлений современной математической теории. Особое внимание уделяется проблемам оценивания в задачах управления и наблюдения. Но в приложениях статистическое описание не является полным или адекватным. По этой причине, в рамках теории гарантированного оценивания развивается альтернативный подход, связанный с предположением об ограниченности неопределенных параметров модели множествами известного вида.

В рамках направления гарантированного оценивания удастся получить точные аналитические описания множеств возможных состояний изучаемой системы. Однако, они могут иметь достаточно сложную структуру. Поэтому интерес вызывает задача построения для них гарантированных аппроксимаций в некотором классе множеств. В качестве таковых в работе используются эллипсоиды, обладающие удобством численной реализации.

Набор оцениваемых операций над эллипсоидами преимущественно связан с задачами, возникающими в линейных системах с эллипсоидальными или симметричными ограничениями на неизвестные параметры. В связи с этим более востребованными операциями являются алгебраическая сумма, геометрическая разность и пересечение эллипсоидов.

Но, в то время, как задачи внешнего и внутреннего оценивания суммы и геометрической разности эллипсоидов к настоящему моменту полностью решены и соответствующие методы оценивания давно и успешно используются в различных приложениях, вопросы внутренних аппроксимаций пересечения эллипсоидов разработаны недостаточно.

В данной работе рассматривается внутренняя эллипсоидальная оценка пересечения concentрических эллипсов. Каждый эллипс, принадлежащий полученному семейству оценок, является недоминируемым. Само семейство является параметрическим. Параметр множества имеет область изменения – отрезок, границы которого явно выражаются из параметров исходных эллипсов. Семейство обладает свойством полноты, т.е. в объединении по всем значениям параметра получаем все пересечение эллипсов. Полученная схема оценивания допускает эффективную численную реализацию.



Литература

1. Важенцев А.Ю. (2004) Внутренние эллипсоидальные оценки в задачах динамики и управления. Москва.
2. Гантмахер Ф.Р. (2004) Теория матриц. М.: Физматлит.
3. Kurzanski A.B., Valyi I. (1997) Ellipsoidal Calculus for Estimation and Control, Boston-Basel-Berlin: Birkhaeuser.

Защита web-приложений

Кочетков Вячеслав Владимирович

*Студент факультета электроники и вычислительной техники
Волгоградский государственный технический университет, Волгоград, Россия*

E-mail: centurionman@mail.ru

В современное время большую распространенность получили web-приложения и приложения, использующие их. Web-приложения находят свое применение в науке, областях электронной коммерции, индустрии развлечений и рекламы. В связи с этим, нарушение работоспособности web-приложения или вывод его из строя приводит к финансовым или информационным потерям, поэтому необходимо разрабатывать web-приложения повышенной защищённости.

Данному вопросу уделялось немало внимания специалистов. Тем не менее, по статистике, собранной членами Web Application Security Consortium и отечественными специалистами, видно, что большое число web-приложений подвержено взлому из-за недостаточной защищенности приложения на уровне кода, которая возникает при разработке.

Не имеется точного руководства, которое следует использовать при разработке web-приложений для защиты от возможных атак и повышения защищённости приложения.

Актуальной остается задача проектирования безопасных и защищённых web-приложений и защиты существующих приложений.

Необходимым является повышение защищенности и улучшение технологии проектирования web-приложений, а также разработка рекомендаций по их проектированию. Существующие рекомендации, описания уязвимостей, которым может быть подвержено приложение, и отдельные замечания по разработке приложений не изменяют ситуацию.

В работе проводится анализ и классификация атак на web-приложения. Выделяются самые распространённые и опасные виды атак.

Выделены наиболее актуальные виды атак:

- SQL-инъекции;
- PHP-инклюддинг;
- XSS (межсайтовое выполнение сценариев).

Проводится анализ архитектуры и процесса функционирования типового web-приложения. Рассматривается трёхзвенная архитектура web-приложения.

Предлагается алгоритм модификации архитектуры приложений для повышения их защищенности. Вырабатываются рекомендации по разработке web-приложений повышенной защищенности.

В ходе работы разрабатывается модуль для повышения защищенности web-приложений, выполняющий функции экрана между приложением и внешней, по отношению к нему, средой Интернет. Разрабатываемый программный продукт представляет собой модуль, подключаемый к существующему web-приложению и выполняющему функции обеспечения защищённости приложения от внешних угроз.

Результаты, полученные в ходе работы, и разрабатываемая технология могут применяться при разработке web-приложений для повышения защищенности программных продуктов и предотвращения появления возможных уязвимостей.

Система параллельной конвейерной обработки данных

Кривов Максим. Андреевич, Притула Михаил Николаевич

студент, студент

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия*

E-mail: m_krivov@cs.msu.su, pritmick@yandex.ru

В связи с широким распространением многоядерных процессоров было разработано множество библиотек и диалектов языков, позволяющих достаточно легко создавать многопоточные приложения. Большинство из них используют декомпозицию по данным и задачам, что позволяет решить практически любую проблему. Однако существует класс задач, решение которых намного проще искать с учётом их конвейерной структуры. К таким проблемам можно отнести обработку потоковых аудио и видео данных, современные компьютерные игры и вычисление некоторых разностных схем. Подобные задачи характеризуются внутренней неоднородностью и зависимостями между различными подсистемами, что затрудняет их решение популярными методами.

В настоящей работе рассматривается подход, максимально учитывающий конвейерный параллелизм. Для этого исходная задача разбивается на независимые блоки с входами и выходами, каждый из которых получает некоторый набор данных, обрабатывает его и отправляет дальше. Требуемая функциональность реализуется соединением полученных блоков особым образом, в результате чего и образуется конвейер.

Отличительной особенностью предложенного подхода от аналогов (таких как Intel TBB и Microsoft CCR) является предположение о наличии некоторой априорной информации о каждом блоке. Как показал анализ типовых задач, при известных входных данных можно быстро определить приблизительную сложность их обработки в некоторых абстрактных единицах. После этого система выполнения может оценить ожидаемое время обработки, учитывая отношения временной сложности к абстрактной на предыдущих шагах. Данная информация позволяет равномерно загрузить все вычислительные ресурсы, что особенно важно в приложениях реального времени. Другим отличием является абстрагированность от конкретной архитектуры. Так некоторые блоки могут уметь обрабатывать одни и те же данные с помощью различных технологий, что позволяет использовать гибридные вычислительные системы типа CPU + GPU.

Рассматриваемый подход был реализован в виде библиотеки для языка Visual C++ и набора утилит, позволяющих визуализировать работу конвейера. Поддержка низкоуровневой многопоточности обеспечивалась через интерфейсы WinAPI и Intel TBB. Полученная библиотека была использована при создании технологической демо-игры, реализованной с помощью конвейера из 19 звеньев. В тестах с высокодетализированными физическими моделями загрузка процессора Intel Core 2 Quad Q9300 держалась равной 90-95%.

В дальнейшем планируется создание коллекции готовых часто используемых блоков, а также среды, позволяющей интерактивно создавать нужный конвейер из элементов данной коллекции. Другим направлением развития является реализация поддержки параллелизма на уровне блока.

Литература

1. Developing Multithreaded Applications: A Platform Consistent Approach, Intel Corporation, February 2005.
2. Bruce Dawson. Lockless Programming Consideration for Xbox 360 and Microsoft Windows, Microsoft Corporation, February 2007.
3. Ресурсы <http://software.intel.com/en-us/articles/open-source-game-development/>.

Об одном подходе к организации системы распознавания таблиц, содержащих статистические данные¹

Кудинов Павел Юрьевич

Аспирант

Вычислительный центр им. А. А. Дородницына РАН, Москва, Россия

E-mail: pkudinov@gmail.com

В настоящее время в Российской Федерации не существует системы, интегрирующей в себе в той или иной степени значительную базу статистических показателей. Существующие базы, такие как, например, [1] или [2], хоть и содержат некоторые разделы статистических сборников Росстата, но, тем не менее, обладают рядом недостатков. Основным недостатком таких систем является необходимость в ручном переводе статистических таблиц в формат базы данных, что впоследствии отражается на полноте и актуальности статистического материала.

Задача настоящего исследования заключается в разработке системы, позволяющей эффективно обрабатывать большие объемы статистических таблиц, производить поиск по названиям показателей и строить пространственно-временные ряды статистических показателей [3].

Входными данными для системы являются таблицы в формате Excel, HTML или любом другом, распознанные с помощью OCR или созданные вручную. Для каждого найденного в таблице числа строится множество $D = \{a_i \mid i = 1, \dots, k\}$, где a_i – атрибуты (значения ячеек, образующих описание ячейки с данными). После этого происходит поиск соответствия атрибутов из построенного описания имеющимся в базе данных системы, таким как, например, ОКВЭД, ОКП, ОКАТО и пр. В результате образуется множество троек $\hat{D} = \{(c_i, \hat{a}_i, p_i) \mid i = 1, \dots, k\}$, где c_i – классификатор из БД, \hat{a}_i – наиболее близкое к a_i значение атрибута, $p_i = \rho(a_i, \hat{a}_i) \in [0, 1]$, ρ – некоторая функция близости атрибутов. Имея некоторый порог $\tau \in [0, 1]$, можно задать условие успешной

классификации: $Q = \prod_{i=1}^k p_i > \tau$, Q – мера качества распознавания ячейки. При условии

успешного прохождения классификации построенное описание вместе со значением записывается в базу данных и становится доступным для поиска. Настройка параметров системы, влияющих на качество, производится на основе прецедентной информации, формируемой экспертам. Описанный подход реализован в прототипе системы и протестирован на наборе статистических таблиц простой структуры (строковых матриц) с корректными значениями ячеек. Однако на практике возникает множество проблем, негативно влияющих на качество распознавания и, вообще говоря, требующих вмешательства эксперта.

В ближайшем будущем планируется развить методы построения описания ячеек для таблиц произвольной структуры, усовершенствовать функцию близости для атрибутов, расширить базу атрибутов и провести масштабные вычислительные эксперименты.

Литература

9. www.gks.ru (*Федеральная служба государственной статистики России*).
10. www.uisrussia.msu.ru (*Университетская информационная система РОССИЯ*)
11. А.В. Богомолова, Н.Ф. Дышкант, О.И. Карасев, П.Ю. Кудинов, Т.Н. Юдина «Интегрированная база данных по социально-демографической статистике: ресурс и пользовательские сервисы для поддержки гуманитарных исследований» // Труды XI Всероссийской объединенной конференции «Интернет и современное общество», 2008, стр. 58-59.

¹ Работа поддержана грантами РГНФ (проект №08-02-12104в) и РФФИ (проект №08-07-00305-а)

Генетический подход к проблеме предсказания сроков сдачи ПО в эксплуатацию

Кульдин Сергей Павлович¹

Студент

Московский государственный университет им. М.В.Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия

E-mail: develoooper@gmail.com

Необходимо определить с приемлемой точностью функцию времени до релиза проекта $R(t)$, зависящую от текущего момента времени t , отсчитываемого от начала работы над проектом. Причем $R(0) = R_0$ – общее первоначальное плановое время работы над проектом. Искомую функцию предлагается искать в виде $R(t) = R_{\text{plan}}(t) + R_{\text{bugs}}(B(t))$, где

1.) $R_{\text{plan}}(t)$ – идеальное плановое время до релиза, рассчитанное на текущий момент времени t , без учета возможных дефектов на всех этапах разработки. $R_{\text{plan}}(t)$ в общем случае нелинейно зависит от t .

2.) $R_{\text{bugs}}(B(t))$ – время, необходимое для устранения имеющихся дефектов и для достижения необходимого показателя качества ПО. $B(t)$ – имеющаяся на момент времени t база данных дефектов.

Т.е. по сути, смысл подхода в разделении исходной задачи на “четкую” и “нечеткую” составляющие.

Первая часть работы посвящена исследованию применимости имеющихся методов расчета планового времени в идеальном случае, т.к. для этого разработано большое количество методик, основанных на различных метриках разрабатываемого ПО.

Вторая часть – непосредственно рассмотрение генетического подхода применительно к расчету $R_{\text{bugs}}(B(t))$. Генетический подход предполагает рассмотрение базы данных ошибок в качестве хромосомной популяции. Одной ошибке ставится в соответствие одна хромосома. На множестве хромосом популяции Ω задается оценочная функция $Durability(\omega)$, которая рассчитывается динамически на основании макроскопических характеристик популяции. В данной интерпретации оценочная функция характеризует время жизни “хромосомы-дефекта” на основании ее характеристик (например: дата и время, когда была обнаружена проблема, серьезность (критичность) проблемы и приоритет её решения, какие-либо характеристики сложности ее решения и т. п.). К популяции применяются генетические операторы скрещивания $C(\Omega)$, мутации $M(\Omega)$ и отбора $S(\Omega)$, использующие оценочную функцию, эмулируя эволюцию популяции (процесс возникновения новых дефектов и устранения старых) через некоторые фиксированные (в простейшем случае) промежутки времени. Алгоритм использует макропоказатели базы данных $B(t)$ ($N(t)$ – количество ошибок, $N'(t)$ – скорость их появления и др.) таким образом, чтобы правильно рассчитать параметры генетических операторов, обеспечив правильную скорость сходимости размера популяции к 0. Как только $|\Omega| < \epsilon$, считаем, что система достигла заданного качества и процесс можно дальше не продолжать. Гипотетическое время этого процесса эволюции популяции и есть $R_{\text{bugs}}(B(t))$.

В работе особое внимание уделяется нахождению путей альтернативного решения проблем, возникавших в предыдущих методах ее решения (сложность применимости разработанных методов на практике, потребность в высоком уровне формализма всех этапов жизненного цикла ПО и др.).

Литература

1. Norman E. Fenton, Niclas Ohlsson. "Quantitative Analysis of Faults and Failures in a Complex Software System". Version 2.0, 22 January 1998
2. Martin Neil, Norman E. Fenton. "A Critique of Software Defect Prediction Models". IEEE transactions on software engineering, VOL. 25, NO. 3, May/June 1999

¹ Автор выражает благодарность научному руководителю, к.ф.-м.н. Луковникову И.В. и компании Parallels за спонсирование и предоставленные материалы.

Непараметрическая модель динамики срочной структуры процентных ставок

Лапшин Виктор Александрович¹

аспирант

Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия

e-mail: lapsh@rambler.ru

Срочная структура процентных ставок, более известная как кривая доходности, в развитых странах рассматривается как главный и наиболее информативный индикатор состояния финансового рынка, один из важнейших макроэкономических параметров и эталон для оценки ценных бумаг в других секторах рынка инструментов фиксированной доходности. В связи с этим особую важность имеет задача нахождения кривой доходности по рыночным данным.

До сих пор используемые модели определяли либо всю кривую в один момент времени, работая с «моментальным снимком» рынка, либо временную стохастическую динамику одной точки кривой (обычно её левого конца, который имеет особый экономический смысл). В [1] показано, что ни одна из существующих параметрических моделей кривой доходности не может быть снабжена никакой стохастической динамикой при условии отсутствия арбитражных возможностей. В [2] был полностью описан класс параметрических «моделей моментального снимка», допускающих нетривиальную динамику своих параметров, причём этот класс оказался слишком бедным для использования на практике.

Для преодоления этих ограничений возможно использование непараметрических моделей. Их сложность, а также трудоёмкость оценки параметров таких моделей приводят к тому, что на текущее время не известно ни одной модели подобного рода. В настоящей работе предложена первая практическая модель непараметрической стохастической динамики кривой доходности, дающая экономически разумные формы кривой, имеющая нетривиальную динамику и не допускающая арбитражных возможностей. Кроме того, в предельном случае при отсутствии информации о прошлом, модель сводится к одному из известных «методов моментального снимка» [3].

В рамках модифицированного для практического использования подхода из [4] предложена конкретная непараметрическая модель процентных ставок и их динамики, построены численные алгоритмы оценки неизвестных параметров модели по наблюдаемым и историческим рыночным данным, реализованные в виде трёхуровневого программного комплекса, часть которого позволяет проводить оценку текущего значения кривой доходности по мере поступления новой рыночной информации в реальном времени.

Литература

12. T.Bjork, B.J. Christensen (1999) Interest Rate Dynamics and Consistent Forward Rate Curves // *Mathematical Finance*, vol. 9, issue 4, 323 – 348.
13. D.Filipović, J.Teichmann (2003) Existence of invariant manifolds for stochastic equations in infinite dimension // *Journal of Functional Analysis*, vol.197, issue 2, 398 – 432.
14. В.А.Лапшин (2006) О задачах, связанных с определением срочной структуры процентных ставок // *Вестник молодых ученых «Ломоносов»*. Выпуск III, 66 – 71.
15. D.Filipović (2001) Consistency Problems for Heath-Jarrow-Morton Interest Rate Models // *Lecture notes in mathematics*, Springer-Verlag, vol. 1760.

¹ Автор выражает признательность Смирнову Сергею Николаевичу, доценту, к.ф.-м.н., за постановку задачи.

О задачах распределения ресурсов и проверки устойчивости для систем информационного мониторинга

Лебедев Анатолий Анатольевич¹

аспирант

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

E-mail: lebedev_aa@rambler.ru

Технология информационного мониторинга ([2]) была разработана для анализа сложных, слабоформализованных проблем (процессов) на основе всей доступной информации, построения прогнозов их развития и выработки рекомендаций по управлению их развитием.

В настоящей работе технология информационного мониторинга формализуется с использованием классического аппарата дискретной математики – схем функциональных элементов и функций k -значной логики ([1]). В этой формализации решаются две ключевые задачи технологии информационного мониторинга – проверка устойчивости модели и задача оптимального распределения ресурсов.

Задача проверки устойчивости

Формальная постановка задачи имеет следующий вид:

$F(x_1, \dots, x_N)$ – функция k -значной логики, заданная схемой функциональных элементов над базисом, состоящим из всех функций от n и менее переменных. Для заданного $1 \leq A \leq k - 2$ необходимо проверить, удовлетворяет ли F следующему условию (назовём его A -устойчивостью):

$$\forall \alpha, \beta \in E_k^N \max_{i=1 \dots N} |\alpha_i - \beta_i| \leq A \Rightarrow |F(\alpha_1, \dots, \alpha_N) - F(\beta_1, \dots, \beta_N)| \leq A$$

Задача распределения ресурсов

Итак, формальная постановка задачи выглядит следующим образом:

$F(x_1, \dots, x_N)$ – функция k -значной логики, заданная схемой функциональных элементов над базисом, состоящим из всех функций от n и менее переменных, a_1, \dots, a_N – начальные значения переменных, $C_i(x)$ – стоимость присвоения i -ой переменной значения x (из текущего состояния). Для заданного C необходимо максимизировать $F(x_1, \dots, x_N)$ при ограничении $\bigsqcup_i C_i(x_i) \leq C$, где \bigsqcup – бинарная операция,

удовлетворяющая аксиомам коммутативности, ассоциативности и монотонного неубывания, которую мы будем называть функционалом стоимости.

Для обеих задач была доказана их труднорешаемость в общем случае и выделены частные случаи моделей, допускающие применение быстрых алгоритмов. Также был получен алгоритм, проверяющий принадлежность произвольной модели этим частным случаям.

Литература

1. Яблонский С.В. Основные понятия кибернетики // Проблемы кибернетики. 1959. Вып. 2. С. 7-38.
2. Рыжов А. Basic principles and foundations of information monitoring systems. In: Monitoring, Security, and Rescue Techniques in Multi-agent Systems. Springer, 2005. p.147–160.

¹ Автор выражает признательность В.Б. Кудрявцеву и А.П. Рыжову за обсуждение работы и ценные рекомендации

Нелинейная аппроксимация функций отражательной способности материалов

Лебедев Андрей Сергеевич, Ильин Андрей Алексеевич

студент, аспирант

Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия
e-mail: alebedev@graphics.cs.msu.ru, ailyin@graphics.cs.msu.ru

Для получения реалистичных моделей объектов помимо текстуры материала необходимо уметь восстанавливать его отражательные свойства. Существует несколько способов подобной реконструкции, различающихся по следующим пунктам: физическая корректность, удобство получения необходимых данных об объекте (в том числе использование сложной аппаратуры [1]), класс восстанавливаемых материалов, информация о геометрии объекта, возможность интерактивной визуализации материала.

В настоящей работе по сравнению с [2] предлагается несколько сузить класс восстанавливаемых материалов, ослабив при этом требования на входные данные: не требуется полигональная модель объекта, используется меньшее число фотографий. В качестве модели освещения мы используем двулучевую функцию отражательной способности (ДФОС), зависящую только от углов между нормалью и направлениями на камеру и на источник света, игнорируя зависимость от азимутальных углов. Для реконструкции материала необходимо сфотографировать плоскую поверхность объекта под разными углами к плоскости так, чтобы на фотографии был виден блик. Далее производится вычисление ДФОС. Для полученного облака точек строится ряд срезов по углу падения света на плоскость. В каждом срезе облако точек приближается нелинейной кривой при помощи алгоритма Левенберга-Марквардта (см. Рис. 1). При использовании обычных фотографий теряется информация о блике материала из-за ограниченности диапазона яркостей (см. Рис. 2), поэтому необходимо получить HDR изображение, сделав несколько фотографий с разной выдержкой. На последнем этапе материал интерактивно визуализируется с использованием графического процессора. При визуализации данные со срезов интерполируются.

Проведен анализ корректности восстановления материалов путем сравнения исходных фотографий и восстановленных изображений с исходными положениями камеры и источника света при помощи метрики PSNR. На Рис. 3 приведена зависимость PSNR от угла падения света при реконструкции по различному числу снимков. Видно, что 5 снимков вполне достаточно для корректного восстановления материала. Приведенный метод позволяет восстанавливать довольно широкий класс изотропных материалов и обеспечивает интерактивную, физически корректную экранизацию.

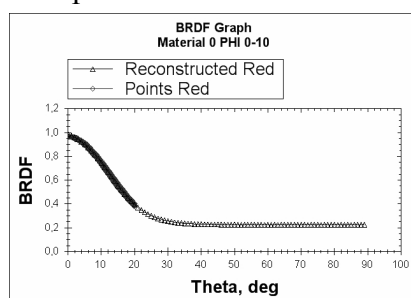


Рис.1

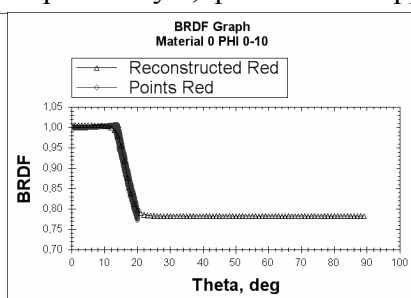


Рис.2

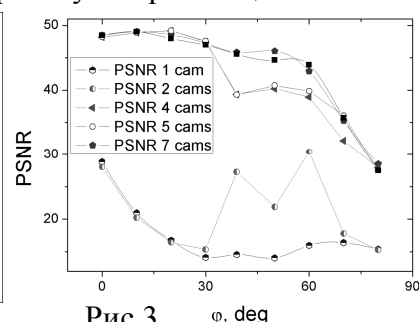


Рис.3

Литература

1. Волобой А.Г., Галактионов В.А., Ершов С.В., Летунов А.А., Потемин И.С. Аппаратно-программный комплекс для измерения светорассеивающих свойств поверхностей "Информационные технологии и вычислительные системы", № 4, 2006.
2. A. Ilyin, A. Lebedev, V. Sinyavsky, A. Ignatenko "The System for the Acquisition, Processing and Material Rendering from Images" Proc. of Graphicon'2008, pp. 134-141, Moscow, Russia, June 2008.

Адаптация онтологий для повышения эффективности рассуждений в приложениях семантического web

Левшин Дмитрий Владимирович

аспирант

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия
e-mail: levshin@nicevt.ru*

Семантический web был предложен Тимом Бернерс-Ли (Berners-Lee, 2001) для обеспечения интеллектуальной машинной обработки информации в Интернет. В настоящее время семантические технологии реализуются в практических системах нового поколения Всемирной паутины, Web 3.0 (Hendler, 2009).

Одной из центральных технологий являются онтологии на языке OWL (Bechhofer, 2004), позволяющие описывать понятия из различных областей деятельности. Описания используются для вывода новых фактов, например, для связывания ресурсов различных систем. В условиях Интернет с огромными объемами информации, ключевым моментом для реализации рассуждений является возможность обрабатывать большие онтологии и наборы данных. Для достижения этой цели автором был предложен (Левшин, 2007) метод выполнения запросов к RDF-данным с учетом их семантики, основанный на использовании средств реляционных СУБД. Метод предполагает разбиение вложенных OWL-описаний на наборы простых описаний для хранения их в отдельных таблицах и реализации логического вывода с помощью SQL-запросов.

В докладе представлен алгоритм очищения онтологий для повышения эффективности рассуждений, выполняемых с помощью данного метода. Отдельные описания могут являться составной частью нескольких более сложных описаний или встречаться в различных утверждениях и тем самым дублировать друг друга. Алгоритм позволяет обнаружить и удалить все такие повторные описания, переписывая соответствующим образом составные описания. В результате выполнения алгоритма для хранения онтологий и логически выведенных утверждений требуется меньшее дисковое пространство, рассуждения относительно концепций (TBox) и индивидуумов (ABox) выполняются более эффективно без потери полноты и корректности ответов. Используемая схема базы данных позволяет генерировать OWL-документы на основе хранящейся информации; поэтому алгоритм может использоваться для генерации документов эквивалентных данным, но меньших размеров. Алгоритм реализован как хранимая процедура в PostgreSQL и продемонстрирован на примере ряда web-онтологий различной выразительности и размеров. Эксперименты показали, что до 50% описаний в онтологии могут быть избыточными, и их удаление позволяет сократить время выполнения рассуждений более чем в 2 раза.

Литература

3. Berners-Lee T., Hendler, J., and Lassila, O. (2001) The Semantic Web. // Scientific American, № 284(5), p. 35-43.
4. Hendler, J. (2009) Web 3.0 Emerging. // IEEE Computer, № 42(1), p. 111-113.
5. Bechhofer, S.; van Harmelen, F.; Hendler, J.; Horrocks, I.; McGuinness, D. L.; Patel-Schneider, P. F., and Stein, L. A. (2004) OWL Web Ontology Language Reference, W3C Recommendation. (<http://www.w3.org/TR/owl-ref/>).
6. Левшин Д.В. (2007) Об Интеграции Semantic Web с PostgreSQL. // Сборник статей молодых ученых факультета ВМиК МГУ, выпуск № 4, стр. 92-100.

Численная оптимизация обеспечения для портфеля производных финансовых инструментов с учетом пакета активных заявок

Лукина Анна Юрьевна, Долматов Андрей Сергеевич,

студент, главный эксперт, к.э.н.

Московский Государственный Университет им. М.В. Ломоносова, Москва, Россия;
ЗАО ЮниКредит Банк

e-mail: lukina.a.u@gmail.com, Andrey.Dolmatov@unicreditgroup.ru

Либерализация финансового сектора экономики стимулировала развитие рынка производных финансовых инструментов, позволяющих участникам рыночных отношений хеджировать свои инвестиции от рисков возможных потерь. Был разработан целый ряд методик, предназначенных для работы с портфелями деривативов и учитывающих специфику данных инструментов. Стандартом индустрии в этой области стала система «Стандартный портфельный анализ риска» («Standard Portfolio Analysis of Risk» - SPAN), разработанная в 1988 году на Чикагской товарной бирже (Chicago Mercantile Exchange - CME) и используемая более чем 50 биржами и клиринговыми организациями, а также более чем 1500 крупных финансовых институтов по всему миру.

Система SPAN моделирует несколько типовых сценариев изменения базисных активов и волатильностей в течение одного торгового дня, чтобы определить максимальные ожидаемые потери стоимости портфеля за один день и установить покрывающий эти потери необходимый залог.

В соответствии со стандартной биржевой методикой расчет риска производится для фиксированного портфеля производных финансовых инструментов. Однако в случае частичного исполнения торговой стратегии для опционов величина залогового требования может существенно отличаться от вычисленной для фиксированного портфеля. В связи с этим возникает интерес к решению задачи расчета начального залога для портфеля с учетом активных заявок. Другими словами, портфель фьючерсов и опционов, относящихся к одному комбинированному товару, можно представить в виде вектора $X = [x_1, \dots, x_n]^T$ заявок на покупку и/или продажу фьючерсов и опционов и коротких и длинных позиций по ним, причем, $0 \leq x_i \leq n_i \forall i \in B$ - множество заявок на покупку, $-n_i \leq x_i \leq 0 \forall i \in S$ - множество заявок на продажу, $-n_i \leq x_i \leq -n_i \forall i \in P_s$ - множество всех коротких позиций в портфеле, $n_i \leq x_i \leq n_i \forall i \in P_l$ - множество всех длинных позиций в портфеле, где n_i - объем заявок, $n_i > 0 \forall i$.

Для поставленной задачи функция расчета начального залога на основе методики SPAN, состоящая из ряда слагаемых, представляет собой кусочно-линейную функцию, которую необходимо оптимизировать. В работе область допустимых значений данной функции разбивается на подобласти, в которых функция линейна, а задача оптимизации сводится к совокупности задач линейного программирования. При этом ограничения в каждой подзадаче задаются в зависимости от возможного варианта исполнения заявок. Решением общей задачи является наибольшее значение критерия, полученное при решении каждой из подзадач. В процессе расчетов было написано инструментальное средство в среде MatLab, самостоятельно рассчитывающее величину начального залога для портфеля фьючерсов и опционов с учетом пакета активных заявок. При достаточно большом размере портфеля данная задача имеет высокую вычислительную сложность.

Литература

1. А. С. Долматов. Математические методы риск-менеджмента - М.: Издательство «Экзамен», 2007. – 319, [1] с.
2. CME SPAN Risk Parameter File Layouts
3. <http://www.cme.com>
4. Hull J. C. Options, Futures and Other Derivatives. Fifth Edition. – Prentice Hall.

Нейросетевые и статистические методы обнаружения сигналов

Ляликова Виктория Геннадиевна

Аспирант

Воронежский Государственный Университет, Воронеж, Россия

e-mail: vikalg@yandex.ru

В последнее время много внимания уделяется технологиям искусственных нейронных сетей. При этом утверждается, что интеллектуальные системы на основе обучаемых нейронных сетей позволяют с успехом решать проблемы распознавания образов, выполнения прогнозов, оптимизации, обработки сигналов, управления и др. Естественное желание использовать преимущества искусственных нейронных сетей и в задачах обнаружения сигналов.

Целью работы является проведение сравнительного анализа характеристик нейросетевого и статистического алгоритмов применительно к задаче обнаружения сигнала, наблюдаемого на фоне случайной помехи.

В качестве статистического алгоритма использовался алгоритм Байеса, а в качестве нейросетевых обнаружителей были выбраны нейронные сети Хемминга, Кохонена, двухслойный персептрон и нейронная сеть на основе радиально-базисных функций активации (РБФ).

Рассматривались сигналы вида: $S = A \cdot \exp[-(t - t_0)/2]$, где A -амплитуда сигнала, t_0 - момент появления сигнала, t -текущий момент времени. В общем случае процесс имел вид: $X = S + \xi$, где ξ - реализация аддитивной гауссовской помехи с нулевым математическим ожиданием и единичной дисперсией. Момент появления сигнала t_0 является случайной величиной, распределенной по равномерному закону на интервале времени T .

На основании указанных выше алгоритмов были построены математические модели байесовского обнаружителя и нейросетевых обнаружителей сигналов. При реализации алгоритмов шум моделировался методом скользящего суммирования, сигнал как случайная величина с равномерным распределением времени появления и постоянной амплитудой для 100 предъявлений.

В результате вычислительного эксперимента был проведен сравнительный анализ качества обнаружения сигнала с помощью нейронных и статистических алгоритмов. Проведенный анализ показал, что нейронные сети Хемминга, Кохонена и РБФ по сравнению с алгоритмом Байеса показывают более высокие значения обнаружения сигнала лишь при малом отношении сигнал/шум (от 0 до 1). Однако нейронные сети Хемминга и Кохонена не обеспечивают фиксированного низкого значения вероятности ложного обнаружения в отличие от статистического алгоритма и алгоритма сети РБФ. Кроме того, двухслойный персептрон не выдерживает конкуренции с остальными рассмотренными алгоритмами. Исследование показало, что из всех рассмотренных нейросетевых алгоритмов только нейронная сеть РБФ, обученная по алгоритму предложенным автором, обеспечивает значения качества обнаружения сигнала сравнимые с алгоритмом Байеса.

Литература

1. Круглов В.В. Искусственные нейронные сети. Теория и практика/В.В.Круглов, В.В.Борисов – М.: Горячая линия – Телеком, 2002.-352с.
2. Левин Б.Р. Теоретические основы статистической радиотехники/Б.Р. Левин- М.: Радио и связь,1989.-656 с.
3. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика /Ф. Уоссермен - М.: Мир, 1992.
4. Хайкин С. Нейронные сети: полный курс/С. Хайкин - М.: Издательский дом "Вильямс", 2006.-1104с.

Расчет времени деградации лазерной мишени

Малинина Елена Александровна

студент

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: malininaea@gmail.com

Задача использования для практических целей энергии термоядерного синтеза, которая выделяется при слиянии легких ядер: дейтерия (D) и трития (T), D и D, D и ^3He была сформулирована учеными полвека тому назад. Существует два основных подхода к решению поставленной задачи. Это магнитный термоядерный синтез (МТС) и инерциальный термоядерный синтез (ИТС).

Мы рассматриваем ИТС и прямое облучение мишени лазером. Лазерная мишень (ЛМ), которую используют в управляемом термоядерном синтезе, представляет собой стеклянную или полистироловую сферическую оболочку. На ее внутренней стенке выморожены твердые изотопы водорода в виде шарового криогенного слоя. Эта начинка из ядерного топлива является шаровым слоем лишь в идеале. Но и геометрически идеальная ЛМ при доставке ее в зону горения реактора подвергается механическим и тепловым нагрузкам, которые частично разрушают топливный слой. Неустойчивость процесса сжатия ЛМ лазерным пучком накладывает высокие требования к геометрическим и физическим параметрам топливного слоя.

Во время инжекции в фокус лазерного пучка мишень короткое время находится в рабочей камере реактора, неравномерно горячие стенки которого излучают электромагнитные волны, нагревающие мишень. За это время топливный слой успевает частично испариться, и его геометрические свойства изменяются. Задача состоит в том, чтобы оценить вид и время этих изменений.

Аналитическое решение модели почти безнадежно, так она принадлежит к классу дифференциальных функциональных уравнений с частными производными. Однако можно ввести разумные упрощения, после которых проблема изменения безразмерной толщины слоя $\bar{w}(t)$ сводится к решению нелинейного обыкновенного дифференциального уравнения

$$\frac{d\bar{w}(\varphi, \theta, t)}{dt} = -a_0(\varphi, \theta, \xi_1) \frac{\sqrt{1 + \beta t} - 1}{\bar{w}(1 - \delta \bar{w})} - a_1 \frac{\bar{w}}{(1 - \delta \bar{w})} + a_2 \frac{\bar{w}^{-2}}{(1 - \delta \bar{w})} - \frac{a_3 e^{-\lambda^2 t}}{(1 - \delta \bar{w})} \left\{ \cos(\pi \bar{w}) + \frac{\delta \sin(\pi \bar{w})}{\pi(1 - \delta \bar{w})} \right\}$$

с начальным условием $\bar{w}(t = 0) = 1$

Численный анализ показал, что время деградации превосходит время доставки ЛМ в фокус лазера, что позволяет использовать известные технологии инжекции ЛМ [1].

Литература:

1. И.В.Александрова, А.А.Белолипецкий, Е.Р.Корешева. *К решению проблемы сохранения параметров криогенной мишени в процессе ее доставки в зону термоядерного горения.* // Ж. Вопросы атомной науки и техники, 2007, вып. 3, с. 27 – 47.
2. А.А.Белолипецкий. *Об одной сингулярно возмущенной задаче Стефана, описывающей разрушение топливного слоя в лазерной мишени.* // Вестник МГУ, серия 15, вычисл. матем. и кибернетика, 2008, № 1, с. 10-18.

Автоматизированный поиск повторяющихся элементов на ректифицированных изображениях фасадов городских зданий¹

Мизин Иван Сергеевич, Якубенко Антон Анатольевич

Студент, аспирант

*Московский государственный университет им. М.В. Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: ivan.mizin@rambler.ru, toh@graphics.cs.msu.ru

Необходимость повышения качества восприятия визуальной информации о пространстве логичным образом ведет к созданию трехмерных геоинформационных систем. Одним из перспективных подходов создания трехмерных карт городов является трехмерное моделирование по фотографиям. Одной из актуальных подзадач является автоматизация анализа изображений фасадов зданий. Полученная с помощью методов компьютерного зрения семантическая информация о фасаде здания позволяет в том числе эффективно реконструировать геометрию фасада и, определяя загромождающие здание объекты переднего плана, восстанавливать текстуру. Целью данной работы является автоматизация поиска повторяющихся фасадных элементов на ректифицированном изображении фасада с учетом априорной информации о регулярности частей фасадов зданий.

Пользователь выделяет прямоугольник одного из окон фасада здания. Для каждого пикселя изображения, как потенциального центра окна, рассчитывается нормализованная кросс-корреляция области вокруг него с соответствующей областью вокруг исходного окна. На полученной карте определяются локальные пики, часть которых фильтруется по порогу. Вычисляются расстояния по вертикали и по горизонтали попарно между всеми пиками. Значения, близкие к кратным другим значениям, удаляются. К оставшимся значениям добавляются значения с небольшой субпиксельной разницей, так как расстояния между окнами в реальности могут быть дробными. Пробегаясь по всем возможным парам потенциальных векторов смещения, в предположении полностью регулярной сетки подсчитывается, сколько окон голосует за данную пару векторов. Выбирается победитель, задающий шаг и смещение регулярной сетки. Все отсутствующие ячейки, в ряду или колонке которых уже есть изначально найденный элемент, заполняются. Алгоритм повторяется для оставшихся ячеек, пока каждая новая итерация добавляет новые окна. Оставшиеся окна считаются выбросами и исключаются. Для получения более точного результата для каждой строки и каждого столбца итеративно выбирается небольшое смещение по вертикали или горизонтали соответственно так, чтобы сумма значений ранее построенной карты корреляции в центрах окон такой сетки была максимальной. Параметры, пороги, цветовое пространство и мера сравнения были выбраны эмпирически как дающие суммарно лучший результат на ручную размеченной базе изображений.

Разработанный автоматизированный метод позволяет с помощью простого взаимодействия с пользователем более точно и надежно находить регулярные сетки повторяющихся фасадных элементов на изображениях фасадов зданий, чем существующие алгоритмы.

Литература

7. Thommen Korah, Christopher Rasmussen “Analysis of Building Textures for Reconstructing Partially Occluded Facades”, ECCV 2008.
8. P. Muller, G. Zeng, P. Wonka, and L. V. Gool. “Image-based Procedural Modeling of Facades”, SIGGRAPH 2007.

¹ Тезисы доклада основаны на материалах исследований, проведенных в рамках грантов Российского фонда фундаментальных исследований (грант №№ 08-01-00883-а, 09-01-92470-МНКC_а).

Исследование многопроцессорного расписания

Мирошниченко Роман Викторович

Соискатель

*Ярославский государственный университет имени П.Г.Демидова,
факультет информатики и вычислительной техники, Москва, Россия*

E-mail: markus96@ya.ru

Пусть есть N процессоров, скорости выполнения задач которых задаются константами $v_i > 0$ ($1 \leq i \leq N$) ($\forall i = 1..N-1: v_i \leq v_{i+1}$). Также имеется N работ, где $ob_i > 0$ ($1 \leq i \leq N$) ($\forall i = 1..N-1: ob_i \leq ob_{i+1}$) - их объемы. В любой момент времени каждую из задач может выполнять только один механизм (процессор). Необходимо определить расписание: какому процессору какую работу назначить и, при необходимости, моменты передачи работ между механизмами. Цель - минимизировать время выполнения всего комплекса работ и количество передач любой работы между процессорами.

$$T = \max_m \left\{ \left(\sum_m^N ob_i \right) / \left(\sum_m^N v_i \right) : m = 1..N \right\} = T_m \quad (*)$$

Определение. Критический набор (КДТ) – набор процессоров и работ, на котором достигается время T по формуле (*).

Определение. Свободный набор – набор процессоров и работ, не входящих в состав критического.

Утверждение 1. Все процессоры из критического набора должны работать постоянно.

Следствие 1. За T в критическом наборе будет выполнен нужный объем.

Утверждение 2. Все работы из свободного набора должны выполняться процессорами из свободного набора.

Утверждение 3. Весь план работ можно разбить на два подмножества: критический набор и свободный набор - без их пересечения (свободный набор может быть пустым).

Утверждение 4. Существует расписание критического набора, при котором процессоры работают время T без помех друг другу.

Теорема 1. Любую задачу, обладающую свободным набором, можно разбить на несколько непересекающихся подзадач, состоящих только из критических наборов.

Определение. Минимальный набор (минКДТ) – набор процессоров и работ, на котором достигается T и в котором не существует двух подмножеств индексов работ IO

и механизмов IM , таких, что $\left(\sum_{IO} ob_i \right) / \left(\sum_{IM} v_i \right) = T$.

Определение. Расписание $(N-1) \times 2 + 1$ - это расписание минимального набора, при котором каждая из $(N-1)$ работ выполняется своей парой процессоров.

Теорема 2. Любой критический набор, не являющимся минимальным набором, можно разбить на несколько минимальных наборов с тем же временем выполнения работ.

Теорема 3. Для любого минимального набора существует $(N-1) \times 2 + 1$ расписание.

Данный труд рассматривает динамическую задачу многопроцессорного расписания. Решение любой задачи данного класса сводится к нахождению расписания в атомарных подзадачах с интересными свойствами. На их основе формулируются и доказываются вспомогательные утверждения. Сформулирована и доказана теорема о существовании решения за минимально возможное время, при котором минимизируется количество передач между процессорами. На основе исследования был разработан алгоритм, зарегистрированный в ОФАП (номер ОФАП 9001 от 27.08.2007, номер ВНИИЦ 50200701911 от 06.09.2007.).

Оптимизация блочного умножения разреженных матриц на вектор для графических акселераторов*

Монаков Александр Владимирович, Кравец Алексей Сергеевич

аспирант, студент

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

e-mail: {amonakov,kayrick}@ispras.ru

Современные графические акселераторы являются программируемыми многоядерными вычислительными устройствами, по производительности превосходящими процессоры за счёт высокого параллелизма работы. В связи с этим они начинают использоваться для ускорения вычислений общего назначения. Так, для акселераторов Nvidia была разработана модель программирования CUDA [2], позволяющая писать программы для акселератора на языке Си с расширениями.

Для эффективного переноса вычислений на акселератор требуется выделить достаточное количество параллельных потоков (на акселераторах Nvidia необходимо запускать несколько тысяч одновременно выполняющихся потоков) и организовать эффективную работу с памятью акселератора. В связи с этим написание эффективных реализаций типичных задач, таких как перемножение матриц и FFT, является нетривиальной задачей.

В ряде численных задач, решаемых на компьютерах, используются операции с разреженными матрицами. Разреженные матрицы характеризуются очень малой долей ненулевых элементов; соответственно, для их хранения используются специальные структуры данных, хранящие только значения и координаты ненулевых элементов [3].

Целью нашей работы является разработка эффективной реализации умножения разреженной матрицы на вектор. Эта операция, как правило, требует наибольших временных затрат в методах, работающих с разреженными матрицами.

Производительность при работе с разреженными матрицами сильно зависит от выбранного формата данных. В работе [1] произведён обзор производительности реализаций, работающих с различными неблочными форматами.

Поскольку на графических акселераторах Nvidia производительность умножения разреженной матрицы на вектор часто ограничена пропускной способностью памяти, использование блочных форматов позволяет улучшить производительность за счёт сокращения объёма данных, требующихся для записи координат ненулевых элементов.

В работе описывается реализация блочного метода умножения разреженной матрицы на вектор. Обсуждаются способы поднятия производительности за счёт использования гибридного формата данных и переупорядочивания матрицы. Также приводятся результаты тестирования на ускорителе Geforce GTX280, показывающие ускорение на 15% – 70% по сравнению с работой [1].

Литература

- a. N. Bell, M. Garland. Efficient Sparse Matrix-Vector Multiplication on CUDA.
<http://mgarland.org/files/papers/nvr-2008-004.pdf>
- b. CUDA 2.1 Programming Guide.
http://developer.download.nvidia.com/compute/cuda/2_1/toolkit/docs/NVIDIA_CUDA_Programming_Guide_2.1.pdf
- c. Richard Wilson Vuduc. Automatic performance tuning of sparse matrix kernels. Technical report, 2003.

* Исследования поддержаны грантами РФФИ и Royal Society

Численная реализация метода внесённой границы для моделирования вязкой несжимаемой жидкости в областях сложной конфигурации

Мортиков Евгений Валерьевич

студент

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия*

E-mail: takeit8ezi@gmail.com

Во многих задачах вычислительной гидродинамики встречаются области со сложной геометрией. Часто для решения подобных задач используют криволинейные сетки, соответствующие границам области. Такой подход требует больших вычислительных затрат для генерации сетки, особенно в задачах с движущимися или деформируемыми границами. В настоящее время, в области вычислительной гидродинамики растущее внимание уделяется методам решения задач о течении в сложных областях на декартовых сетках. К преимуществам такого подхода можно отнести легкость и экономичность построения прямоугольных сеток, а также относительную наглядность конечно-разностных схем, получающихся в результате дискретизации. Основной сложностью является представление краевых условий на криволинейной границе с необходимой точностью.

Первые попытки повысить точность аппроксимации краевых условий на криволинейной границе на декартовых сетках появились в конце пятидесятих годов XX века. В настоящее время все большее распространение получают различные варианты метода внесённой границы, позволяющие решать задачи течения вязкой несжимаемой жидкости в областях со сложной геометрией, применяя декартовы сетки. Одной из важных причин появления подобных методов является растущее количество задач со сложной геометрией и движущимися границами, для которых приближение границы ступеньками оказывается грубым, а представление расчетной области криволинейными сетками будет избыточным для предъявляемых требований к точности получаемого решения.

В докладе рассматривается численная реализация одного из вариантов метода внесённой границы (метода фиктивных ячеек) и его применение для решения двухмерных задач об обтекании кругового цилиндра, о течении над ступенькой, об обтекании нескольких круговых цилиндров в различной конфигурации, а также трехмерной задачи о течении вокруг сферы. Полученные результаты используются для демонстрации возможностей метода и для сравнения с экспериментальными данными и с методами, основанными на построении криволинейной сетки, соответствующей границам области.

Литература

1. R. Mittal, G. Iaccarino. Immersed boundary methods. Annual Review of Fluid Mechanics 2005 37: 239-261. 2005.
2. J. Mohd-Yosuf. Combined immersed boundary/B-spline methods for simulation of flow in complex geometries. Annual Research Briefs, Center for Turbulence Research 317-328. 1997.

Метод автоматического выделения терминалей дофаминергических нейронов на изображениях фронтальных срезов стриатума¹

Мягков Артем Александрович, Яшина Вера Владимировна

Студент; научный сотрудник

Московский Государственный Университет имени М. В. Ломоносова;

Вычислительный центр им. А. А. Дородницына РАН, Москва, Россия

e-mail: aem.istranet@gmail.com, werayashina@gmail.com

Данная работа посвящена описанию метода автоматизации получения экспериментальных данных, необходимых для наполнения модели преklinической стадии болезни Паркинсона (БП) (Ogawa, 1987). БП характеризуется прогрессирующей дегенерацией дофаминергических (ДА-ергических) нейронов в компактной части черной субстанции, приводящей к снижению уровня дофамина в стриатуме, в результате чего теряется способность контролировать движения. Модель БП представляет различия характеристик ДА-ергических нейронов в "опытной" и "контрольной" группах. Использование разработанного метода позволяет количественно оценить дегенерацию ДА-ергических нейронов в черной субстанции и их аксонов в стриатуме у мышей под влиянием специфического нейротоксина, а также функциональное состояние ДА-ергических нейронов и аксонов, сохранившихся после воздействия нейротоксина.

Исходными данными являются цифровые изображения окрашенных срезов различных областей головного мозга. Экспериментальные данные были получены на изображениях дистальных отделов аксонов (терминалей). Предлагаемый метод: 1) основан на следующих операциях математической морфологии: открытие, тоновая реконструкция (Vincent, 1993), закрытие, преобразование "дно шляпы", морфологический градиент, морфологический водораздел; 2) позволяет: сглаживать неоднородный сложный фон, выделять мелкие объекты на изображениях в зависимости от заданных размеров и значений яркости, устранять объекты, находящиеся не в фокусе, разделять близко находящиеся объекты, вычислять характеристики выделенных объектов; 3) предназначен для автоматического выделения терминалей дофаминергических нейронов на изображениях фронтальных срезов стриатума.

С помощью разработанного метода был проведен анализ десяти срезов головного мозга (пять изображений из "опытной" группы и пять изображений из "контрольной" группы), основными задачами которого являлись: сравнение результатов автоматического и ручного выделения терминалей; подсчет характеристик терминалей; выявление различий в "опытной" и "контрольной" группах. Экспериментальные исследования подтвердили возможность и целесообразность автоматизации обработки и анализа изображений срезов с помощью разработанного метода.

Литература

1. N. Ogawa, K. Mizukawa, Y. Hirose et al. (1987) Mptp-induced parkinsonian model in mice: biochemistry, pharmacology and behavior. // Eur Neurol., Vol.26, № 1. p.16–23.
2. Vincent, L. (1993) Morphological grayscale reconstruction in image analysis, Applications and efficient algorithms. // IEEE Transactions on Image Processing, Vol.2 p.176-201

¹ Данная работа была выполнена при частичной поддержке Российского Фонда Фундаментальных Исследований гранты №№ 07-07-13545 и 08-01-90022 и проекта программы президиума РАН "Фундаментальные науки - медицине", 2009 год.

Модификация многоуровневого метода Монте-Карло

Нагапетян Тигран Артурович

студент кафедры Системного анализа.

Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия

e-mail: nagapetyan@gmail.com

При моделировании траекторий стохастических дифференциальных уравнений методом Монте Карло, широко используемым в финансовой математике, целью является вычисление цены опциона, который является некоторым функционалом от решения соответствующего стохастического дифференциального уравнения. Стандартные методы Монте Карло, решающие данную задачу с точностью $O(\varepsilon)$ при использовании схемы Эйлера, имеют сложность порядка $O(\varepsilon^3)$. В работе [1] был предложен многоуровневый метод Монте-Карло, идеи которого схожи с идеями, используемыми в многосеточных методах решения систем линейных алгебраических уравнений. Данный метод имеет сложность $O(\varepsilon^2(\log \varepsilon)^2)$ при достижении той же точности.

В данной работе исследовалось применение данного метода к различного рода опционам и был получен класс задач, на которых прямое использование данного метода не приводило к верному результату. Данная проблема была изучена и был предложен метод, решающий ее, основанный на изменении способа выбора начальной сетки. Кроме того, была показана связь между изменениями уровня сгущения начальной сетки и условиями задачи.

Литература

1. Giles M. «Multi-level Monte Carlo path simulation» Operations Research, 56(3):607-617, 2008.
2. P.E. Kloeden and E. Platen. «Numerical Solution of Stochastic Differential Equations» Springer-Verlag, Berlin, 1992.
3. P. Glasserman. «Monte Carlo Methods in Financial Engineering» Springer-Verlag, New York, 2004.

Применение сеточных методов к задаче разделения смесей вероятностных распределений.

Назаров Алексей Леонидович, Гапонова Маргарита Олеговна

студент, студент

*Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия
e-mail: nazarov.vmik@gmail.com, margarita.gaponova@gmail.com*

Конечные смеси нормальных законов находят самое широкое применение как модели распределения многих величин, наблюдаемых на практике. В связи с этим часто возникает задача разделения смесей. Для решения данной задачи авторами был использован сеточный метод. Его главная идея заключается в приближении истинного распределения некоторой смесью нормальных распределений с заранее заданным числом компонент. При этом параметры компонент (математическое ожидание и дисперсия) известны и соответствуют узлам сетки, «наброшенной» на область их возможных значений. Корректность данного подхода объясняется тем, что при увеличении числа узлов сетки найдутся параметры, практически неотличимые от истинных. Конечные смеси нормальных законов обладают свойством идентифицируемости, поэтому найденные веса будут близки к искомым.

Для оценки весов исследуемой смеси используется метод максимального правдоподобия, заключающийся в поиске вектора, на котором функция правдоподобия достигает максимума. Область возможных значений весов является выпуклым компактом, функция правдоподобия на этом множестве - гладкая и вогнутая, поэтому данный вектор существует и единственен.

На практике поиск данного вектора может быть осуществлен с помощью «усеченного» EM-алгоритма или метода условного градиента. Особенностью этих подходов является то, что итерационный процесс может быть записан в явном виде.

Оба метода были реализованы авторами на языке C++. Для исследования их работы использовалась задача получения портретов волатильности финансовых индексов, решаемая с помощью скользящего разделения смесей. В качестве исходных данных были взяты минутные логарифмические приращения биржевых индексов AMEX, NASDAQ, NIKKEI, S&P40. В силу того, что временные ряды логарифмических приращений лишены явной трендовой составляющей, и их среднее близко к нулю, математическое ожидание компонент принималось нулевым.

Анализ полученных результатов показал, что оба метода дают приблизительно одинаковые значения как оцениваемых параметров, так и скорости работы. Сравнение результатов работы сеточных методов с аналогичными классическими дает право сделать вывод о том, что данная модель является адекватной для поставленной задачи.

Литература

1. Королев В.Ю. Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. // М.: Издательство ИПИРАН, 2007
2. Васильев Ф.П. Численные методы решения экстремальных задач. // М.: Издательство Наука, 1988.
3. Королев В.Ю., Непомнящий Е.В., Рыбальченко А.Г., Виноградова А.С. Сеточные методы разделения смесей нормальных распределений и их применение к декомпозиции волатильности финансовых индексов. // Информатика и ее применения, 2008, Т.2 Вып. 2 С. 3-18.

Усиленный закон больших чисел для случайных процессов

Наумов Алексей Александрович¹

Студент

Московский государственный университет имени М.В. Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия

E-mail: naumov_ne@inbox.ru

Для мартингалов с непрерывным параметром в работе доказаны аналоги усиленных законов больших чисел Колмогорова, Зигмунда-Марцинкевича и Брунка-Прохорова. Классические законы больших чисел нашли применения в рамках метода Монте-Карло, например, для вычислений интегралов большой размерности. Предложенные аналоги законов больших чисел могут оказаться полезными для тех же целей, а также для приближенного вычисления континуальных интегралов. Автором доказана

Теорема 1. Пусть даны сепарабельный мартингал $\{M_t, t \in \mathbb{P}_+\}$ относительно некоторой фильтрации $\{\Phi_t, t \in \mathbb{P}_+\}$ и неограниченно возрастающая положительная функция $f(t), t \in \mathbb{P}_+$. Если $\int_1^\infty \frac{dE |M_t|^\alpha}{f^\alpha(t)} < \infty$ для некоторого $\alpha \geq 1$, то

$$\lim_{t \rightarrow \infty} \frac{\sup_{0 \leq s \leq t} M_s}{f(t)} = 0 \text{ п.в.}$$

В статьях [1] и [2] приведены обобщения теоремы Брунка-Прохорова. Следующая теорема представляет собой новое обобщение теоремы Брунка-Прохорова.

Теорема 2. Пусть даны мартингал $\{M_n, n \in \mathbb{N}\}$ относительно фильтрации $\{\Phi_n, n \in \mathbb{N}\}$ и неограниченно возрастающая последовательность положительных чисел $b_n, n \in \mathbb{N}$. Положим $M_0 = 0$. Если

$$(1) \quad \sum_{n=1}^{\infty} \frac{n^{\alpha-1} E |M_n - M_{n-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty \text{ и}$$
$$\sum_{n=1}^{\infty} \frac{n^{\alpha-2} \sum_{k=1}^n E |M_k - M_{k-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty$$

для некоторого $\alpha \geq 1$, то $\lim_{n \rightarrow \infty} \frac{M_n}{b_n} = 0$ п.в. и $\lim_{n \rightarrow \infty} E \left| \frac{\max_{1 \leq k \leq n} M_k}{b_n} \right|^{2\alpha} = 0$.

В качестве нормирующих постоянных $b_n, n \in \mathbb{N}$ в теореме 2 можно взять произвольные положительные числа при условии, что они образуют неограниченно возрастающую последовательность. Общность нормирующих постоянных достигается ценой дополнительного условия (1) на случайные величины. Это условие в ряде случаев автоматически выполняется. В частности, оно выполняется в рамках оригинальной теоремы Брунка-Прохорова.

В работе также доказаны аналоги усиленных законов больших чисел для сепарабельных однородных случайных процессов с независимыми приращениями.

Литература

1. Круглов В.М. (2002) Обобщение усиленного закона больших чисел Брунка-Прохорова // Теория вероятн. и ее примен., т. 47, № 2, с. 347 – 349.
2. Fazekas I., Klesov O. (2000) A general approach to the strong law of large numbers // Теория вероятн. и ее примен., т. 45, № 3, с. 568 – 583.

¹ Автор выражает признательность профессору, д.ф.м.н. Круглову В.М.

Решение некоторой проблемы обработки сигналов ЯМР с использованием обобщенного спектрально-аналитического метода¹

Новикова Дарья Андреевна

Аспирант

*Московский государственный университет им. М.В. Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: pionerk@gmail.com

Предлагается рассмотреть одну из задач, возникающих в области обработки сигналов ЯМР и методы ее решения с использованием обобщенного спектрально-аналитического метода (ОСАМ).

ОСАМ основан на адаптивном разложении исходных массивов в функциональном базисе из числа классических алгебраических систем полиномов и функций (многочленов Якоби, Чебышева, Лагранжа, Лагерра, Гегенбауэра и др.). Обработка данных происходит в пространстве коэффициентов разложения начальных данных, что ускоряет и делает более точным процесс вычисления. Данный подход объединяет в себе аналитические и цифровые процедуры обработки данных и фактически является универсальной комбинированной технологией обработки информационных массивов.

Рассматривается следующая задача. Из эксперимента ЯМР опытным путем может быть получен спектр $G(w)$ объекта, представляющего собой порошок (совокупность хаотически расположенных молекул). Теоретический интерес представляет спектр $F_0(w)$ ориентированного образца, который представляет собой раствор (совокупность упорядоченных молекул). Спектр $F_0(w)$ можно вычислить путем решения уравнения, представляющего собой зависимость между такими спектрами. Уравнение имеет вид

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F_0(x) \sqrt{\frac{\pi}{3xt}} \left[C_2\left(\frac{3xt}{2}\right) + iS_2\left(\frac{3xt}{2}\right) \right] \exp\left(-\frac{ixt}{2}\right) dx \quad (1)$$

где $C_2(z), S_2(z)$ - интегралы Френеля[5].

Данная зависимость представляет собой уравнение Фредгольма первого рода, не имеет аналитического решения.

В работе рассматриваются основные пути решения поставленной задачи, а также вопросы применения ОСАМ для получения решения задачи.

Литература

1. M. Bloom, J.H. Davis, A. L. Mackay, Direct determination of the oriented sample NMR spectrum from the powder spectrum for systems with local axial symmetry// Chemical physics letters, volume 80, pages 198-202 1981.
2. E. Sternin, M. Bloom, A. L. Mackay, De-pake-ing of NMR spectra// Journal of magnetic resonance, volume 55, pages 274-282, 1983.
3. K. P. Whittall, E. Sternin, M. Bloom, A. L. Mackay, Time- and frequency- domain "De-pake-ing" using inverse theory// Journal of magnetic resonance, v.84, pp. 64-72, 1989.
4. H. Schafer, B. Madler, F. Volke, De-pake-ing of NMR powder spectra by nonnegative least-squares analysis with Tikhonov Regularization// Journal of magnetic resonance, v. 116, pp.145-150, 1995.
5. M. Alan McCabe, S. R. Wassal, Rapid deconvolution of NMR powder spectra by weighted fast Fourier transformation// Solid state nuclear magnetic resonance, v.10, pp. 53-62, 1997.
6. Дедус Ф.Ф., Куликова Ф.И., Панкратов А.Н., Тетуев Р.К. «Классические ортогональные базисы в задачах аналитического написания и обработки информационных сигналов»- Москва, 2004.

¹ Работа выполнена при поддержке проекта РФФИ № 06-07-89303.

Программная поддержка языка лексико-синтаксических шаблонов¹

Носков Алексей Анатольевич

*Студент факультета вычислительной математики и кибернетики
Московский государственный университет имени М.В.Ломоносова, Москва, Россия
E-mail: alno87@mail.ru*

В последнее время в области создания систем автоматической обработки текстов на естественном языке активно развиваются средства [1], позволяющие специфицировать и находить в тексте различные языковые конструкции. К таким средствам относится язык LSPL [2], предложенный для описания конструкций русского языка (в первую очередь именных словосочетаний) и учитывающий его особенности, в частности, высокую флективность. Язык позволяет описывать языковые конструкции в виде лексико-синтаксических шаблонов, задающих словоформы, лексемы и их морфосинтаксические характеристики. Важной особенностью языка LSPL является возможность указывать грамматическое согласование между составными частями описываемых языковых конструкций. Например, шаблон $A N <A.g=N.g, A.n=N.n, A.c=N.c>$ описывает сочетание из прилагательного (A) и следующего за ним существительного (N), согласованного с ним в роде (g), числе (n) и падеже (c). Язык LSPL также позволяет использовать ранее описанные шаблоны, что дает возможность сначала подготовить набор шаблонов, описывающих простые конструкции, а затем на их основе описывать более сложные конструкции.

Целью представленной в докладе работы было создание программной системы для поиска в тексте на русском языке конструкций, описанных в виде шаблонов на языке LSPL. При этом система должна была допускать простую интеграцию с другими программными средствами обработки текста, а также использоваться в виде отдельного приложения, поддерживающего работу пользователя-лингвиста по анализу текста.

В ходе реализации системы были выделены три основных компонента:

1. ядро системы, осуществляющее частичный синтаксический анализ текста и поиск в нем конструкций, описанных в виде LSPL-шаблонов;
2. программный интерфейс (API) для использования ядра из приложений на языке Java;
3. графический пользовательский интерфейс для задания шаблонов, загрузки анализируемого текста и поиска в нем нужных конструкций.

Для более эффективной (с точки зрения производительности и использования памяти) реализации ядра использовался язык C++. Пользовательский интерфейс реализован на языке Java с использованием кроссплатформенной библиотеки SWT.

Ядро системы использует специальное графовое представление анализируемого текста, опирающееся на разметку его фрагментов с использованием аннотаций. Похожий подход для разметки текста был применен в системе GATE [1], однако предлагаемое графовое представление специализировано для эффективного поиска в нем конструкций на основе LSPL-шаблонов. Для оптимизации поиска используются заранее построенные индексы слов разных частей речи и индексы на основе возможных префиксов слов.

Литература

1. Bontcheva K., et. al, Developing Reusable and Robust Language Processing Components for Information Systems using GATE // Proc. 13th Intern. Workshop on Database and Expert Systems Applications. IEEE Computer Society, Washington, DC, 2002, pp. 223-227
2. Большакова Е.И. и др. Лексико-синтаксические шаблоны в задачах автоматической обработки текстов // Труды межд. конф. Диалог '2007 – М.: Издательский центр РГГУ, 2007, с. 70-75.

¹ Работа выполнена при поддержке гранта РФФИ № 06-01-00571.

Алгоритмы машинной графики в задачах геопространственного моделирования

Овсянников Михаил Сергеевич

*аспирант факультета Информатики,
Томский государственный университет, Томск, Россия
E-mail: Michael.Ovsyannikov@gmail.com*

С момента появления графических устройств вывода задачам обработки изображений уделялось особое внимание. Особым классом задач является представление информации для вывода на экран. Алгоритмы машинной графики, реализующие задачу визуализации, характеризуются высокой степенью оптимизации по затратам памяти и процессорному времени. Кроме того, для многих алгоритмов реализовано аппаратное ускорение на уровне графического адаптера. В то же время, задачи геопространственного (ГИС) моделирования также предполагают обработку информации, совмещенную с процессом визуализации 2D и 3D сцен. Таким образом, при сведении отдельных подзадач ГИС-моделирования к традиционным задачам машинной графики появляется возможность упростить и ускорить процесс вычислений.

Среди многих задач ГИС-моделирования можно выделить задачу определения области видимости. Суть задачи в том, что необходимо определить область на карте, отвечающую критерию прямой видимости до некоторой точки. При добавлении некоторых ограничений, данная задача сводится к применению панорамного Z-буфера. Таким образом, последовательно перебирая объекты на карте и формируя список расстояний до них, становится возможным определить границы видимости (рисунок 1). Ещё одной особенностью модифицированного алгоритма является то, что размер Z-буфера не фиксирован, а увеличивается по мере появления новых интервалов, хранимых в виде секторов с заданными углами границ.

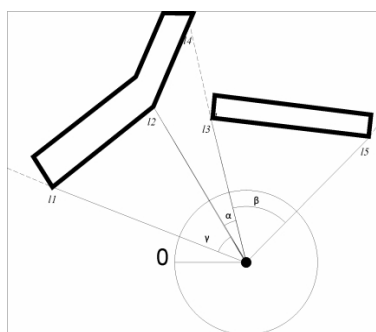


Рисунок 1. Панорамный Z-буфер.

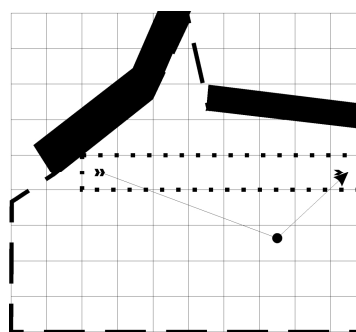


Рисунок 2. Построчный обход области

Другая задача, широко востребованная при гис-моделировании загрязнения окружающей среды, является оценка прямого воздействия источника на заданную область. В случае, когда данная задача решается на ячеистой модели данных, определенную сложность представляет обход всех ячеек в выбранной зоне без необходимости прямого перебора [1]. Одним из решений является применение адаптированного алгоритма закраски[2], используемого при заполнении областей экрана. При использовании алгоритма построчного заполнения используется информация об ограничивающем многоугольнике, определяющем область загрязнения (рисунок 2). Главное отличие от традиционного алгоритма заливки состоит в том, что в отличие от используемого в машинной графике, при ГИС-моделировании для всех ячеек записывается не одно и то же значение, а производится расчет на основании расстояния от источника, вида загрязнения и ряда прочих факторов.

Литература

1. Овсянников С.Н. Овсянников М.С. Расчет эквивалентных уровней шумового загрязнения селитебной территории методом обратной трассировки на растре // Вестник Том. гос. ун-та., 2008. - № 1(2). - С. 50-56.
2. Роджерс Д. Алгоритмические основы машинной графики / Пер. с англ. – М.: Мир, 1989.– 512 с.

Информационные графы с автоматными функциями

Пивоваров Александр Павлович¹

Аспирант

Московский государственный университет имени М.В.Ломоносова,

механико-математический факультет, Москва, Россия

E-mail: pivizz@gmail.com

В информационно-графовой модели поиска информации [1] задача информационного поиска представляет из себя тройку $I = \langle X, V, \rho \rangle$, где X – множество запросов, V – библиотека, являющаяся конечным подмножеством множества всех возможных записей Y , а ρ – отношение поиска заданное на $X \times Y$. При этом содержательно задача I состоит в том, чтобы для любого произвольного запроса $x \in X$ перечислять те и только те записи из V , которые находятся в отношении ρ с x . Реализация алгоритма поиска с помощью информационного графа предполагает, что требуемые записи находятся по одной и отправляются на выход алгоритма.

Данная модель хорошо подходит для описания именно такой ситуации, когда от алгоритма требуется перечислить все записи из библиотеки, удовлетворяющие поступившему запросу. Однако в ряде ситуаций требуется получить не сам ответ как множество записей, но результат вычисления некоторой функции на нем. Например, нам требуется узнать количество записей, состоящих в отношении ρ с запросом x . В данном исследовании такая задача рассматривалась на примере одномерного интервального поиска: база данных V представляет собой конечное подмножество отрезка $[0,1]$, а множество запросов – множество всех возможных отрезков, лежащих внутри отрезка $[0,1]$. Запись $y \in [0,1]$ удовлетворяет запросу $x = [u, v] \subseteq [0,1]$ в том случае, если $y \in [u, v]$. Для данной задачи существует удобное разложение на две подзадачи: нахождение числа записей библиотеки, принадлежащих отрезку $[0, v]$ и нахождение числа записей библиотеки, принадлежащих полуинтервалу $[0, u)$. После чего, для получения ответа остается вычесть из первого числа второе. При наличии достаточно широкого базового множества, каждая из указанных подзадач по отдельности может быть эффективно решена с использованием алгоритмов, описанных в [1]. При этом затраты памяти на каждый из них будут линейными. Однако объединить их в один информационный граф оказывается затруднительным. Более того, оказывается, что для любого информационного графа, являющегося деревом, затраты памяти на реализацию любого алгоритма будут как минимум квадратично расти при увеличении базы данных.

В связи с этим для задач подобного рода предлагается модифицировать определение информационного графа путем подсоединения к нему конечного автомата. Листьям будем приписывать некоторые символы входного алфавита этого автомата. В процессе обхода графа символы, приписанные обходимым листьям, будем подавать на вход автомата, а ответом на запрос будем считать состояние автомата к моменту конца обхода графа. При таком определении понятия информационного графа оказывается возможным реализовать описанный выше алгоритм с затратами памяти, растущими линейно при увеличении базы данных.

Литература

1. Гасанов Э.Э., Кудрявцев В.Б. (2002) Теория хранения и поиска информации. М.: ФИЗМАТЛИТ, 2002.

¹ Автор выражает признательность профессору, д.ф.-м.н. Гасанову Э.Э. за помощь в подготовке тезисов.

Программная реализация метода разделения области FETI-DP для массивно-параллельной вычислительной системы IBM Blue Gene/P¹

Позднеев Александр Валерьевич

*Аспирант факультета вычислительной математики и кибернетики
Московский государственный университет имени М.В. Ломоносова, Москва, Россия
E-mail: pozdneev@cs.msu.su*

Работа посвящена разработке параллельного кода, реализующего численное решение эллиптических уравнений методом конечных элементов на основе метода разделения области FETI-DP [1]. Установлено, что в методе FETI-DP число обусловленности κ предобусловленного оператора задачи относительно переменных, обеспечивающих непрерывность решения на границах между подобластями, подчиняется условию $\kappa \leq C(1 + \log(H/h))^2$, где C — константа, не зависящая от характерного размера H подобластей, их числа и характерного размера h конечных элементов [2]. Эта теоретическая оценка обеспечивает возможность масштабирования метода при увеличении размера области и числа процессоров [1].

Десять лет назад научные группы из университета штата Колорадо и лабораторий Сандия изучали поведение методов предыдущего поколения семейства FETI на суперкомпьютере ASCI Option Red [3]. Целью данной работы является реализация метода FETI-DP на современной массивно-параллельной системе IBM Blue Gene/P и исследование его масштабируемости при использовании до 8192 процессорных ядер.

В настоящей работе предлагается применить подход «несколько подобластей на один процессор», что позволяет использовать одно и то же разбиение области при расчетах на разном числе процессоров. Кроме того, на многоядерных системах становится возможным распараллелить вычисления по подобластям в рамках одного вычислительного узла с помощью OpenMP и воспользоваться преимуществами общей памяти, уменьшив объем вычислений. Межпроцессорные обмены осуществляются посредством механизма MPI. Для решения локальных задач, операций с плотными и разреженными матрицами и векторами применяются стандартные последовательные математические библиотеки.

Требование работы с большим числом процессорных узлов и возможность обработки каждым процессором нескольких подобластей привело к необходимости разработки специального формата файла с исходными данными (координаты узлов, ячейки сетки), который мог бы быть эффективно считан параллельно. Параллельные операции ввода-вывода реализованы с помощью коллективных функций MPI-IO.

В работе предполагается исследовать параллельные характеристики кода при решении уравнения Пуассона в зависимости от различных параметров (размер области, размер и число подобластей и конечных элементов) при разном числе процессоров. Разработанная библиотека должна послужить основой для написания программ моделирования с использованием метода разделения области.

Литература

1. C. Farhat, M. Lesoinne, P. LeTallec, K. Pierson, D. Rixen «FETI-DP: a dual-primal unified FETI method—part I: a faster alternative to the two-level FETI method» // Int. J. Numer. Meth. Engng. — 2001. — 50. — p. 1523–1544
2. A. Klawonn, O.B. Widlund «Dual-primal FETI methods for linear elasticity» // Comm. Pure Appl. Math. — 2006. — 59. — p. 1523–1572
3. M. Bhardwaj «Application of the FETI method to ASCI problems—scalability results on 1000 processors and discussion of highly heterogeneous problems» // Int. J. Numer. Meth. Engng. — 2000. — 47. — p. 513–535

¹ Тезисы доклада основаны на материалах исследований, проводимых в рамках гранта Российского Фонда Фундаментальных Исследований (проект № 08-07-12081-офи)

Оценка компактности задач распознавания образов

Потепалов Дмитрий Николаевич

Студент (специалист)

*Московский государственный университет имени М.В. Ломоносова,
Факультет вычислительной математики и кибернетики, Москва, Россия*

E-mail: potep@mail.ru

Все методы решений задач распознавания образов (РО) неявно основаны на истинности некоторых эмпирических предположений относительно обучающей выборки. В частности, гипотеза компактности (ГК), предполагает, что «классам соответствуют компактные области в пространстве выбранных свойств» [1]. Ясно, что для практического применения данная формулировка ГК требует уточнения - необходимо чётко определить, что понимать под компактностью.

На основании анализа уже существующих формулировок ГК для некоторых частных случаев [2, 3, 4] было предложено измерять компактность расположения прецедентов в

метрических пространствах признаков с помощью оценки $K_n = \frac{1}{N_X^2} \sum_{i=1}^{N_X} m_i$, где N_X -

число прецедентов в классе X , а m_i - номер ближайшего соседа из чужого класса для i -того прецедента класса X (среди соседей из всех классов). Построен также непрерывный аналог оценки K_n , позволяющий поставить задачу оптимизации компактности. Обе эти характеристики дают математическую интерпретацию, естественным представлениям о «скупенности», «сгруппированности», и позволяют получить оценку качества текущего признакового пространства. Они также позволяют дать рекомендации относительно сложности модели, в рамках которой необходимо решать поставленную задачу. Улучшение этих характеристик может служить одним из критериев для построения нового, более удобного для распознавания, пространства признаков.

В докладе рассмотрены варианты параметризации метрики признакового пространства, позволяющие поставить и решать задачу оптимизации его компактности. Было исследовано поведение предложенных оценок компактности на ряде модельных задач в пространствах различной размерности (не более 100 признаков). Генерировались прецеденты, подчиняющиеся различным вероятностным распределениям, а также построенные по законам невероятной природы. Предложенные характеристики компактности позволили отличать задачи РО, в которых возможно построить классификатор с высокой обобщающей способностью, от задач, где этого сделать не удаётся. Рассмотрены возможности применения характеристик компактности при решении задач РО композицией распознающих операторов, а также нахождения преобразования исходного признакового пространства, улучшающего компактность.

Литература

1. Аркадьев А.Г., Браверман Э.М. Обучение машины распознаванию образов. - М: Наука, 1964.
2. Донской В.И. О метрических свойствах кратчайших конъюнктивных эмпирических закономерностей // Учёные записки Таврического национального университета им. В.И. Вернадского. Сер. «Математика. Механика. Информатика и Кибернетика», 2003, № 2. – С. 143-14.
3. Загоруйко Н.Г. Гипотезы компактности и l-компактности в методах анализа данных // Сибирский журнал индустриальной математики. Изд. ИМ СО РАН. Том 1, № 1, 1998. - С. 114-126.
4. Воронцов К.В., Колосков А.О. Профили компактности и выделение опорных объектов в метрических алгоритмах классификации. // Искусственный Интеллект. — 2006. — С. 30–33

Кластеризация схем поведения пользователей веб-ресурсов

Проскурнев Андрей Александрович

*Аспирант факультета вычислительной математики и кибернетики
Московский государственный университет имени М.В.Ломоносова, Москва, Россия
E-mail: ctpahhuk@mail.ru*

Журналы посещений веб-ресурсов представляют собой источник статистических данных о посетителях ресурса и о характере посещений, необходимых для оптимизации работы как самого веб-ресурса, так и сопутствующей сетевой инфраструктуры. Одной из проблем анализа журналов посещений является необходимость выявления и фильтрации "мусорных" данных, а именно, записей о посещениях ресурса программами-роботами.

Для автоматического выявления записей о программах-роботах был разработан метод, основанный на гипотезе, что схемы поведения людей и программ-роботов на веб-ресурсе существенно отличаются. На первом шаге все запросы посетителя к серверу сортируются по времени, и вычисляются временные интервалы между соседними по времени запросами, множество интервалов для одного пользователя называется схемой его поведения. Для каждой схемы определяется ряд параметров, которые в дальнейшем используются в эвристической оценке посетителя. К числу наиболее важных параметров относятся: количество обращений к серверу, процент запросов, содержащих пустое значение заголовка Referer протокола HTTP.

На следующем шаге проводится кластеризация полученных схем поведения пользователя в метрическом пространстве, в качестве меры взят критерий согласия Колмогорова-Смирнова для двух статистик [1], при этом статистиками служат сами схемы. При кластеризации учитывается только центроид класса, представляющий собой статистику, полученную объединением всех входящих в кластер схем. Считается, что схема поведения принадлежит кластеру, если критерий Колмогорова-Смирнова для центроида и схемы не превышает определенного порогового значения, которое выбирается заранее и задает дисперсность получаемого множества кластеров. Шаг кластеризации состоит в добавлении нового элемента (схемы поведения) к уже существующим кластерам, элемент добавляется в самый массивный из тех кластеров, к которым он может быть отнесен согласно критерию Колмогорова-Смирнова. Если элемент не может быть отнесен ни к одному из существующих кластеров (или кластеров еще нет), то он образует новый кластер.

Полученные таким образом кластеры оцениваются с помощью ряда эвристик для разделения их на посетителей-людей и посетителей, представляющих собой программы-роботы. Схемы посетителей, относящиеся к роботам, исключаются из расчета итоговой статистики посещений веб-ресурса. Применяемые эвристики учитывают уже вычисленные параметры входящих в кластер схем поведения.

На базе описанного метода разработан комплекс программ для анализа журналов посещений веб-ресурса, включающий в себя средства выявления программ-роботов в статистике посещений веб-ресурса, механизмы контроля и оценки хода кластеризации схем поведения пользователей, а так же средства визуализации полученных кластеров. Был проведен ряд экспериментов по оценке работоспособности комплекса и подбору коэффициентов для эвристических оценок, подтвердивших верность исходной гипотезы о различиях схем поведения людей и программ-роботов на веб-ресурсе.

Литература

1. Кендалл М., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973. 899 С.

Поиск похожих пользователей в социальных сетях

Пустовойтов Никита Юрьевич

Студент

Московский физико-технический институт (государственный университет)

pustovoytov@mycolornet.com

В настоящее время социальные сети становятся все более популярными. При этом не только увеличивается число пользователей, но и сами пользователи становятся все более активными. Как правило, в социальной сети пользователь $u \in U$ может указать, помимо социально-демографических данных, свои интересы (ключевые слова) $w \in W$, характеризующие его, опубликовать записи $p \in P$ в блоге, назначить записи P теги $t \in W$, прокомментировать $c \in C$ чью-то (возможно, свою) запись в блоге. Будем считать, что множества W и T совпадают. Кроме того, доступна информация об отношениях «дружбы» (заинтересованности, friendsheep) между пользователями. Любое из перечисленных множеств, отличное от множества пользователей U , будем называть множеством ресурсов R .

Будем считать, что задана матрица UU (User-User) и набор матриц вида UR (User-Resource) и R_1R_2 (Resource₁-Resource₂). На множестве пар индексов матриц определена функция привлекательности $A(i, j) \in [-1; 1]$, элементами матриц являются значения функции привлекательности. Значение $A(i, j) > 0$ означает, что объекту i интересен объект j интересен (например, если рассматривается матрица UW , то пользователь u_i указал ключевое слово w_j). Значение $A(i, j) = 0$ соответствует отсутствию информации, отрицательные значения означают явное отсутствие интереса или связи. Значения функции привлекательности известны лишь на очень малом количестве пар (обучающей выборке), все исходные матрицы являются сильно разреженными. Требуется восстановить функцию привлекательности (т.е. заполнить все матрицы), минимизировав среднеквадратичное отклонение восстановленной функции привлекательности от заданной на обучающей выборке (функционал RMSE).

Для решения этой задачи предлагается алгоритм, основанный на идее скрытых (латентных) переменных. Элемент любого множества можно описать с помощью элементов множества ключевых слов W . На предварительном этапе выполняется кластеризация множества ключевых слов с использованием специально адаптированных и разработанных метрик и одного из методов кластеризации. Каждый полученный кластер будем называть топиком (темой [1]). Вектор, составленный из оценок близостей топикам, является профилем, а сами оценки – значениями скрытых переменных. Элементу каждого множества можно сопоставить профиль, благодаря этому можно определять сходство (привлекательность) между элементами различных множеств. Кроме того, скрытые переменные легко интерпретируются экспертами. Ещё одним преимуществом является избавление от «проклятия размерности» вследствие перехода к профилям – длина профиля гораздо меньше, чем размерности исходных матриц.

Литература

1. *Leksin V. A., Vorontsov K. V.* The overfitting in probabilistic latent semantic models // Pattern Recognition and Image Analysis: new information technologies (PRIA-9-2008): Nizhni Novgorod, Russian Federation, 2008. Vol 1. Pp. 393–396.
2. *Potter G.* Putting the collaborator back into collaborative filtering // Proceedings of the 2nd Netflix-KDD Workshop, 2008.

Перестановочные критерии для одновыборочных задач проверки гипотез

Рябенко Евгений Алексеевич

Студент

*Московский государственный университет имени М.В.Ломоносова,
факультет Вычислительной математики и кибернетики, Москва, Россия*

E-mail: riabenko.e@gmail.com

Перестановочные методы статистического анализа представляют собой непараметрические процедуры проверки гипотез, опирающиеся на информацию о распределении статистики на всех возможных перестановках данных. Теоретическая основа перестановочных методов была заложена ещё в работах Фишера 1935 года, однако бурное развитие они получили только в последние десятилетия в связи с появлением достаточных для их применения вычислительных мощностей. Современные перестановочные методы позволяют решать многие задачи классической теории проверки гипотез, зачастую с большей точностью, а также, что более ценно, некоторые задачи, приемлемое решение которых в рамках традиционных методов получить затруднительно. Среди последних можно выделить задачи проверки гипотез для малых выборок, решение которых рассмотрено нами на примере данных нескольких медицинских экспериментов.

По сути, перестановочные критерии представляют собой условные статистические процедуры с обусловленностью по пространству перестановок данных, возможных в условиях справедливости нулевой гипотезы. Критерии удовлетворяют принципу условности Фишера, согласно которому в том случае, когда в минимальной достаточной системе статистик (T_r, T_s) для распределений (P_k, P_l) подсистема T_s не зависит от P_k , для получения выводов о P_k нам нужно только условное распределение $T_r|T_s$. Для выполнения этого условия достаточно допустить внутригрупповую (слабую) взаимозаменяемость данных при справедливости нулевой гипотезы. Это допущение является легко проверяемым, достаточно слабым по сравнению с предположениями классических статистических процедур, а для многих популярных моделей возникает естественным образом и выполняется автоматически. На примере реальных медицинских задач нами исследован перестановочный критерий, являющийся условно и безусловно несмещённым и состоятельным. Критерий также является точным, в то время как традиционные параметрические критерии обладают лишь свойством консервативности.

Известно, что в задаче проверки однородности для любой простой альтернативы наиболее мощный критерий может быть найден в семействе критериев перестановок. Для рассмотренного перестановочного критерия мы исследовали его функцию мощности, её оценки с помощью условной мощности, а также оценки по методу Монте-Карло и бутстреп-оценки. Показано, что в рассмотренных задачах мощность критерия превосходит мощность t-критерия Стьюдента, критерия Манна-Уитни-Вилкоксона, критерия Мак-Нимара.

Литература

1. Pesarin, F. (2001) *Multivariate Permutation Tests With Applications in Biostatistics*. Chichester: John Wiley & Sons, Ltd.
2. Кендалл, М., Стьюарт, А. (1973) *Статистические выводы и связи*. М.: Наука.
3. Hájek, J., Šidak, Z., Sen, P.K. (1999) *Theory of Rank Tests*. San Diego: Academic Press.
4. De Martini, D., Rapallo, F. (2000) Calculating the Power of Permutation Tests: a Comparison between Nonparametric Estimators // *Journal of Applied Statistical Science*, №11(2), p. 111-122.
5. Deutsch, S.J., Schmeister, B.W. (1982) The Power of Paired-Sample Tests // *Naval Research Logistics Quarterly*, №29(4), p. 635-649.
6. Berger, V.W. (2000) Pros and Cons of Permutation Tests in Clinical Trials // *Statistics in Medicine*, №19, p. 1319-1328.

LSPL-шаблоны для решения задачи автоматизированного построения онтологий

Седова Яна Анатольевна

аспирант

Астраханский государственный технический университет, Астрахань, Россия

E-mail: yanasedova@rambler.ru

Онтологии – современный эффективный способ формализации знаний. Сегодня трудоемкий процесс разработки онтологий осуществляется вручную, поэтому актуальна задача его автоматизации, в частности, этапов сбора и анализа данных с подготовкой для эксперта чернового варианта онтологии. В разрабатываемой автором автоматизированной системе текстовый материал для онтологии собирается информационно-поисковыми модулями системы с веб-ресурсов по заданным ключевым словам. Нетривиальна задача извлечения из текстов на естественном (русском) языке утверждений для построения онтологии.

Можно говорить о существовании общих схем, по которым строятся многие утверждения, а также выделить схемы, характерные для текстов конкретной предметной области. Для записи схем обоих типов в разрабатываемой системе был выбран язык LSPL-шаблонов регулярных языковых конструкций [1]. Для автоматической интерпретации LSPL-шаблонов были построены конечный автомат (рис.1) и грамматика (фрагмент представлен на рис.2). Классы входных символов: a – латинские буквы, b – русские буквы, c – кавычки, d – цифры, e – пробел, f – другие символы. Классы лексем: S1 (Id) – идентификаторы, S2 (RusWord) – русские слова, S3 (String) – строки, S4 (Number) – числа, S5 – другие символы. S0 – начальное состояние.

Лексема	a	b	c	d	e	f
S0	1	2	3	4	0	-5
S1	1	-1	-1	1	-1	-1
S2	-2	2	-2	-2	-2	-2
S3	3	3	-3	3	3	3
S4	-4	-4	-4	4	-4	-4
S5	-5	-5	-5	-5	-5	-5

Рис.1. Матрица переходов

Expr->Expr<Attr>	Signs->Sign
Expr->Id	Sign->Sname=Sval
Expr->String	Sign->Element=Equal
Attr->RusWord; Signs	Equal->Element=Equal
Attr->RusWord	Equal->Element
Attr->Signs	Element->Id.sname
Signs->Sign, Signs	Element->Id

Рис.2. Грамматика

При ручном анализе ряда текстов естественного языка автором были выделены 17 общих LSPL-шаблонов, по которым независимо от предметной области строятся утверждения, а также специальные шаблоны для предметной области «виноделие».

Примеры общих шаблонов (синтаксис LSPL описан в работе [1], после каждого шаблона следует пример и предполагаемый факт, NP – именная группа):

AN=A1 [“и” A2] N <A1=A2=N> (N) («ароматическое и вкусовое воздействие»): A1, A2 – характеристики N.

NP [A<который; A=NP>] V<измеряться; t=pres, p=3> “в” N<n=plur, c=pre> («длительность послевкусия измеряется в каудалях»): N – единица измерения NP.

N1 “–“ {A} N2 («брожение – химический процесс»): N1 – разновидность N2.

Пример специального шаблона:

NP1 V<изготавливаться; t=pres, p=3> “из” NP2<c=gen> («розовое вино изготавливается из красного винограда»): NP2 – ингредиент NP1.

Общие шаблоны должны задаваться в программе по умолчанию, специальные – составляться и добавляться пользователем в зависимости от задачи. Следующий этап – автоматический поиск в тексте утверждений в соответствии с заданными шаблонами и трансляция их в OWL-утверждения онтологии, которую будет редактировать эксперт.

Литература

1. Большакова Е.И., Баева Н.В. и др. Лексико-синтаксические шаблоны в задачах автоматической обработки текста. // Труды Международной конференции Диалог-2007. М., 2007, с.70-75.

Об одном приближенном способе решения задач оптимального подвижного управления для систем гиперболического типа

Сейидова Улвия Вахаб

Магистр

Сумгаитский Государственный Университет, Аз.Республика

E-mail: revane-seyidova@mail.ru

В представленном докладе рассматривается построение приближенного решения задачи оптимального подвижного управления для систем описываемых следующими дифференциальными уравнениями

$$\frac{\partial^2 y}{\partial t^2} = Ly(x, t) + u(t)\delta(x - v(t)), \quad (1)$$

с граничными

$$u(0, t) = 0, u(l, t) = 0 \quad (2)$$

и начальными

$$u(x, 0) = \varphi(x), u'_t(x, 0) = \psi(x) \quad (3)$$

условиями, где, $Ly(x, t) = \frac{\partial}{\partial x} \left[p(x) \frac{\partial y}{\partial x} \right] + q(x)y$, $\varphi(x), \psi(x)$ заданные достаточно гладкие

функции на $[0, l]$, $u(t), v(t)$ управляющие параметры из $L_2[0, t]$ удовлетворяющие

условиям $\int_0^1 u^2(t)dt \leq L, 0 \leq v(t) \leq l, \delta$ - дельта функция Дирака .

Требуется из множества допустимых управлений найти $(u(t), v(t))$ -такое, которое при решениях задачи (1)-(3) доставляет наименьшее возможное значение функционалу

$$J[u, v] = \int_0^1 \{ [y(x, T) - y_0(x)]^2 + [y'_t(x, T) - y_1(x)]^2 \} dx + \int_0^T [\alpha u^2(t) + \beta v^2(t)] dt, (\alpha^2 + \beta^2 > 0)$$

Сначала доказывается существование и единственность поставленной задачи. Далее рассматривается вопрос о построении приближенного решения как смешанной задачи, так и задачи оптимального подвижного управления. Приближенное решение задачи (1)-(3) определяется с помощью проблемы моментов в гильбертовом пространстве. Кроме того построенное приближенное решение задачи (1)-(3) используется для определения решения задачи оптимального подвижного управления.

Литература

1. Бутковский А.Г., Пустыльников Л.М. (1975) Теория подвижного управления системами с распределенными параметрами. - М.: Наука.
2. Мамедов А.Д. (1989) Задачи оптимального подвижного управления для процесса теплопроводности. // Дифференциальные уравнения, т.25, №6.

Задачи граничного слоя для уравнения колебаний сферического слоя.

Сергеев Сергей Андреевич

Аспирант

Московский государственный университет им. М.В. Ломоносова, факультет

Вычислительной Математики и Кибернетики, Москва, Россия

E-mail: SergeevSel@yandex.ru

Рассматриваются колебания сферического слоя при условии радиальной симметрии, описываемые уравнением

$$u_{rr}(r,t) - \frac{1}{r^2}[r^2 u(r,t)]_r = 0 \quad (1)$$

в области $\{0 < R_1 < r < R_2\} \times \{0 < t < T\}$, где R_1 и R_2 - радиусы сфер, ограничивающих рассматриваемый слой. Финальный момент времени T равен $2m\delta + \Delta$, $\delta = R_2 - R_1$, $m \in \mathbb{N}$ и $\Delta \in [0, 2\delta]$ - произвольное число.

Известны начальное $\{\varphi(r), \psi(r)\}$ и финальное $\{\varphi_1(r), \psi_1(r)\}$ состояния слоя. На границе слоя ведется управление либо первым краевым условием

$$u(R_1, t) = \mu(t), \quad u(R_2, t) = \nu(t), \quad (2)$$

либо специальным третьим

$$u_r(R_1, t) + \frac{1}{R_1} u(R_1, t) = \mu(t), \quad u_r(R_2, t) + \frac{1}{R_2} u(R_2, t) = \nu(t). \quad (3)$$

Функции управления $\mu(t)$ и $\nu(t)$ принадлежат классу $W_2^1[0, T]$ в случае (2) и классу $L_2[0, T]$ в случае (3).

Ставится задача отыскания управления $\{\mu(t), \nu(t)\}$, переводящего слой из начального состояния в финальное за время T . Существует бесконечно много таких функций, поэтому, помимо нахождения управления, требуется его оптимизация. Критерием оптимальности служат следующие условия:

$$\int_0^T \{a^2 [R_1 \mu'(t)]^2 + b^2 [R_2 \nu'(t)]^2\} dt \rightarrow \min, \quad \text{в случае (2)}$$

и

$$\int_0^T \{a^2 [R_1 \mu(t)]^2 + b^2 [R_2 \nu(t)]^2\} dt \rightarrow \min, \quad \text{в случае (3)}.$$

Константы a и b - произвольные, отличные от нуля, числа.

В данной постановке удается явным образом выписать формулы для управления. Из этих формул видно, что в случае (2) функции управления являются периодическими на сегменте $[0, T]$, а в случае (3) в этих функциях можно выделить линейную и периодическую части. Кроме того, в случае (3) сумма функций $\mu(t)$ и $\nu(t)$ является на этом же сегменте периодической функцией.

Рассмотренные здесь задачи актуальны в физике, так как описывают колебания сферического осциллятора. Помимо этого, в работах рассмотрены более общие критерии оптимальности, в отличие от тех, которые рассматривались до сих пор.

Литература

1. Ильин В.А., Моисеев Е.И. (2008) Оптимизация управления на двух концах струны упругими граничными силами за любой достаточно большой промежуток времени. // Дифференциальные уравнения, 2008, т. 44, N 1, с. 89-110.

Матирование видео на основе оптического потока

Синдеев Михаил Сергеевич, Конушин В.С.

Студент, аспирант

*Московский государственный университет им. М.В. Ломоносова;
Институт прикладной математики им. М.В. Келдыша, Москва, Россия
E-mail: msindeev@graphics.cs.msu.ru, vadim@graphics.cs.msu.ru*

Матированием видео называется отделение объекта переднего плана от фона в видеопоследовательности. Объект должен выделяться с учетом полупрозрачных областей, что делает его выглядящим естественно при наложении на новый фон. Широкое применение матирование нашло в кино: изначально неподвижные или подвижные (покадровые) маски («маты») объекта переднего плана рисовались на стеклянных панелях [1] и предотвращали экспонирование фона на пленку, куда затем отдельным проходом экспонировался новый фон на основе инвертированной маски.

В конце XX – начале XXI века, с развитием цифровых технологий записи изображения, появились понятия «цифровой кинематограф» и «цифровое видео». В плане матирования цифровое видео дает возможность более детальной проработки отснятой сцены и практически неограниченного увеличения числа слоев; снимаются ограничения на движение камеры. Однако до сих пор большая часть работы по матированию сложных сцен на видео выполняется вручную, путем ротоскопирования (рисования масок через каждые 5-10 кадров с их интерполяцией на остальные кадры).

Существующие алгоритмы матирования видео либо выполняют жесткую сегментацию (т.е. не учитывают прозрачность и выдают бинарную маску), либо применяют покадрово некоторый алгоритм матирования изображений. Последнее приводит к тому, что информация о движении не учитывается и получающиеся маски для соседних кадров могут иметь различную структуру.

Предлагаемый алгоритм получает на вход мягкую маску первого кадра (например, полученную алгоритмом матирования изображения) и выполняет ее отслеживание по последующим кадрам. Используется формула композиции (в каждом пикселе): $C = \alpha F + (1 - \alpha)B$, где C – цвет изображения, F , B – цвета объекта и фона соответственно, α – коэффициент смешивания ($\alpha \in [0, 1]$). Изображение коэффициентов α образует маску объекта.

При последовательном передвижении по кадрам маска деформируется оптическим потоком, который вычисляется путем минимизации функционала энергии, измеряющей соответствие маски изображению (что отличается от простого вычисления оптического потока по кадрам без учета маски). Предлагаются 2 функционала энергии: основанный на алгоритме [2] (оценивает на сколько маска подходит к изображению, измеряя локальную корреляцию прозрачности и цвета), и, для случая известного фона, на ошибке рекомпозиции объекта на известный фон с учетом его маски.

Для минимизации функционала энергии применяется дискретный метод QRBO с альфа-расширением, обработка ведется иерархически (по размеру изображения).

Эксперименты показали, что данный алгоритм имеет преимущества над существующими аналогами, выдавая более точную маску для прозрачных областей и обеспечивая ее стабильность на соседних кадрах, а также позволяет учитывать известный фон (например, при съемке неподвижной камерой).

Литература

1. Hanson B. Matte Painting: Art in Film Special Effects [PPT] (<http://www-graphics.stanford.edu/courses/cs99d-00/projects/BrookeHanson-specialfx.ppt>)
2. Levin A., Lischinski D., Weiss Y. A Closed Form Solution to Natural Image Matting // Proc. of IEEE CVPR, 2006. P. 61–68.

Построение виртуальных сцен реального мира в системе реконструкции отражательных свойств материалов

Синявский Виталий Александрович, Ильин Андрей Алексеевич

студент, аспирант

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

E-mail: vsinyavsky@graphics.cs.msu.ru, ailyin@graphics.cs.msu.ru

Основной задачей расширенной реальности [1] является встраивание виртуальных объектов в сцену реального мира. В системе реконструкции отражательных свойств материалов по фотоизображениям [2] виртуальная сцена представляется набором виртуальных камер, источников света и трехмерной моделью исследуемого объекта. Для определения виртуальной сцены требуется по фотографиям исследуемого объекта, некоторому числу входных параметров определить положения и ориентации камер, положения источников света и положение исследуемого объекта в единой системе координат. Данная работа посвящена разработке алгоритмов нахождения данных параметров, обеспечивая возможность получения точных данных при генерации сеточных ДФОС в системе [2].

В системе виртуальные сцены классифицируются на два типа: с телесными и плоскими исследуемыми объектами.

В случае исследования телесных объектов подразумевается, что известна их геометрическая информация (mesh). Положения и ориентации камер находятся при помощи стороннего калибровочного пакета, при условии присутствия в реальной сцене калибровочного объекта (рис.1). В рамках данной работы решается задача встраивания виртуального объекта в сцену. Первоначально производится грубое нахождение положения объекта на основе информации о центрах объектов на фото. В дальнейшей более точной подгонке положение и ориентация объекта подгоняется при помощи линейных преобразований, и критерием остановки является полное совпадение объектов на каждой фотографии (рис.2). Сходимость итерационного процесса достигается методом градиентного спуска при помощи минимизации специально разработанного функционала ошибок. В ходе апробации и верификации работы алгоритма на синтетических и реальных данных была получена высокая скорость и точность сходимости (ошибка в 0-2% площади объекта на изображении от идеального совпадения).

Вторым типом исследуемых сцен являются сцены с плоским объектом. В данных сценах калибровочным объектом может выступать исследуемый объект. Задавая внутренние параметры камеры, и решая P4P задачу по четырем соответствиям точек, находим внешние параметры камер. В силу того, что решение этой задачи обладает

погрешностью (из-за неточности входных данных), был реализован алгоритм уточнения положений источников света и объекта. Суть его заключается в варьировании положений источников света для выполнения физических законов (угол падения равен углу отражения) в области блика.

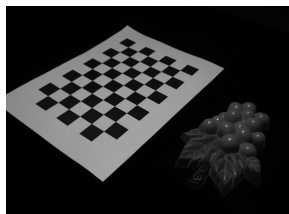


Рис.1

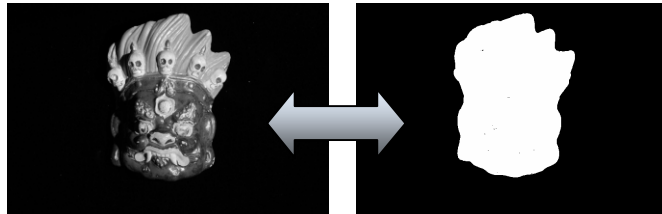


Рис.2

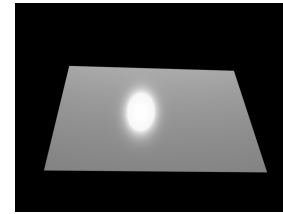


Рис.3

Литература

1. R.Raskar, G.Welch, H.Fuchs "Spatially Augmented Reality" Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, U.S.A.
2. P.Sikachev, A.Ilyin, A.Ignatenko "User-Assisted Acquisition, Processing and Rendering of Materials from Images" Proc. of Graphicon'2007, pp. 131-134, Moscow, Russia, June 2007.

Экспериментальное исследование свойств методов индуктивного формирования знаний¹

Смагин Сергей Владимирович²

Ведущий инженер-программист

Институт автоматизации и процессов управления, Владивосток, Россия

E-Mail: smagin@iacp.dvo.ru

Индуктивное формирование знаний (ИФЗ) лежит в основе многих направлений исследований, таких как машинное обучение, распознавание образов, прогнозирование и т.д., каждое из которых характеризуется собственным подходом к проблеме ИФЗ, собственными постановками задач и методами их решения. На сегодняшний день предложен ряд проблемно-независимых методов ИФЗ, которые решают задачу формирования базы знаний – в том или ином представлении – на основе выборки. При решении прикладных задач возникает необходимость выбора наиболее подходящего метода ИФЗ и формирование обучающей выборки достаточного объема. Такой выбор должен осуществляться исходя из условий и ограничений задачи, а также из известных значений свойств методов ИФЗ.

В рамках данного исследования разработан подход к организации компьютерных экспериментов для изучения свойств методов индуктивного формирования знаний на модельных выборках разного объема. Идея подхода состоит в том, что на основе онтологии предметной области генерируется случайная база знаний. На основе этой базы генерируются случайные выборки данных – обучающие и контрольные. Применяя исследуемый метод к обучающим выборкам, создаются индуктивно формируемые базы знаний, внешнее качество которых может быть оценено на контрольной выборке, а внутреннее – при сравнении их с исходной, случайно сгенерированной базой знаний.

Предложенный подход реализован на примере экспериментального исследования свойств проблемно-ориентированного метода ИФЗ (метода случайной расстановки границ периодов динамики) для онтологии медицинской диагностики. С точки зрения временных затрат метод является очень быстрым и время работы при увеличении объемов обучающих выборок растет почти линейно в достаточно широком диапазоне. Внешняя оценка метода получилась очень хорошей и к тому же с увеличением объемов обучающих выборок очень устойчивой. Используя только внешнюю оценку, можно сказать, что в этих условиях применения исследованный метод дает очень хорошие результаты. Когда решается задача ИФЗ с полной информацией, внутренняя оценка оказывается очень хорошей. Когда решается задача с неполной информацией, внутренняя оценка оказывается очень плохой. Если рассматривать метод не с точки зрения практического применения, а с точки зрения выяснения того, как на самом деле устроена природа, то этот метод решает задачу плохо. В общем случае проведенное экспериментальное исследование показывает, что хорошие внешние оценки метода ИФЗ могут сочетаться с его плохими внутренними оценками.

Литература

1. Клещев А.С., Смагин С.В. (2008) Организация компьютерных экспериментов по индуктивному формированию знаний // НТИ. Сер. 2. – 2008. – №1. – С. 16-24.
2. Клещев А.С., Смагин С.В. Компьютерный эксперимент по исследованию свойств метода случайной расстановки границ периодов динамики. Владивосток: ИАПУ ДВО РАН, 2009, 44 с. [<http://iacp.dvo.ru/is/publications/2009-Kleshev,Smagin-ExperOne.pdf>]

¹ Тезисы доклада основаны на материалах исследований, проведенных при финансовой поддержке ДВО РАН в рамках Программы №2 фундаментальных исследований Президиума РАН “Интеллектуальные информационные технологии, математическое моделирование, системный анализ и автоматизация”, проект “Развитие систем управления базами знаний с коллективным доступом”.

² Автор выражает признательность профессору, д.ф.-м.н. Клещеву А.С. за помощь в подготовке тезисов.

Применение технологии UniTESK для функционального тестирования инфраструктурного ПО Грид

Смолов Сергей Александрович

студент

Московский Физико-Технический Институт, Москва, Россия

ssedai@ispras.ru

Грид (англ. grid – решетка, сеть) – это согласованная, открытая и стандартизованная компьютерная среда, которая обеспечивает гибкое, безопасное и скоординированное использование вычислительных ресурсов и данных. Термин “Грид” появился в 1990-х гг. [1] как метафора о такой же легкости доступа к вычислительным ресурсам, как и к электрической сети (англ. power grid). Грид-системы в настоящий момент являются одним из приоритетных направлений в вычислительной технике. Потому первоочередной задачей становится эффективное использование их преимуществ, что неразрывно связано с проблемой переносимости программного обеспечения Грид-систем. Одним из наиболее простых решений этой проблемы является введение стандартов на программные комплексы. Следовательно, соответствие стандарту представляется одним из основных требований, предъявляемых к системе, а потому его выполнение должно быть подтверждено с высокой достоверностью.

Существующие тестовые наборы для Грид-систем [3,4] проверяют исключительно применимость последних в типовых сценариях использования – постановке вычислительной задачи в очередь, выполнении задачи на одном из вычислительных узлов, доставке результатов вычисления клиенту. Проверка совместимости с другими инфраструктурными ПО (ИПО) ими не проводится.

В данной работе рассматривается задача тестирования соответствия реализации стандарту систем управления распределенными ресурсами WSRF 1.2. Данный стандарт был проанализирован и на его основе составлен его каталог требований. На основе полученных данных был разработан тестовый набор с применением технологии тестирования UniTESK (JavaTESK) [2] и проведено тестирование реализации ИПО ГС Globus Toolkit 4.2. Тестирование показало, что данная реализация соответствует стандарту WSRF 1.2, хотя выявило отсутствие выполнения ряда необязательных требований, как семантических, так и синтаксических.

В качестве направления для дальнейшей работы рассматривается разработка тестового набора, проверяющего специфические требования к сервисам ИПО Грид, таким как сервисы передачи больших массивов данных, сервисы создания и управления вычислительными задачами и др.

Список литературы

1. Ian Foster The Grid: Blueprint for a New Computing Infrastructure. — Morgan Kaufmann Publishers. — ISBN 1-55860-475-8
2. А.В. Баранцев, И.Б. Бурдонов, А.В. Демаков, С.В. Зеленов, А.С. Косачев, В.В. Кулямин, В.А. Омельченко, Н.В. Пакулин, А.К. Петренко, А.В. Хорошилов. Подход UniTESK к разработке тестов: достижения и перспективы. Труды Института системного программирования РАН, №5, 2004.
3. S. Schulze. Achieving Grid Interoperability: The ETSI Approach. The 20th Open Grid Forum - OGF20/EGEE 2nd User Forum. Manchester, UK. May 7 - 11, 2007
4. E.Slabospitskaya, V.Petukhov, and G.Grosdidier. Glite I/O Storm Testing in EDG-LCG Framework. Nuclear Electronics and Computing 2005. Varna, Bulgaria, 12-18 September 2005.

О конструктивной характеристике пороговых функций

Соколов Андрей Павлович

младший научный сотрудник

Московский государственный университет имени М.В.Ломоносова

механико-математический факультет, Москва, Россия

E-mail: sokolov@intsys.msu.ru

Пороговые функции алгебры логики представляют интерес в связи с простотой технической реализации, а также благодаря своим вычислительным возможностям.

Средством задания пороговых функций являются линейные формы вида $x_1 w_1 + \dots + x_n w_n - \sigma$. Вводится понятие сигнатуры пороговой функции f как набор знаков коэффициентов некоторой линейной формы, задающей f . Оказывается, что отношение равенства сигнатур разбивает множество существенных пороговых функций на 2^n взаимно непересекающихся равномоощных классов, одним из которых является класс монотонных пороговых функций.

В работе исследуется сложность преобразования одной пороговой функции, заданной линейной формой, к другой, путем последовательного изменения коэффициентов линейной формы. В качестве меры сложности принимается изменение коэффициента или свободного члена линейной формы на единицу. Для характеристики сложности обучения в худшем случае вводится шенноновская функция $\rho(n)$. Она говорит о том, сколько минимально достаточно выполнить единичных модификаций исходной линейной формы от n переменных для задания желаемой пороговой функции. В работе показывается, что при стремлении n к бесконечности величина $\log \rho(n)$ растет по порядку как $n \log n$.

Для любой пороговой функции существует бесконечное множество задающих ее линейных форм. Линейные формы назовем эквивалентными, если они задают одну и ту же пороговую функцию. Множество всех существенных линейных форм с целочисленными коэффициентами и свободным членом, задающих пороговую функцию f , обозначим $U(f)$. Легко видеть, что множество $U(f)$ замкнуто относительно операции сложения линейных форм. В работе доказано, что всякое множество $U(f)$ содержит единственный базис относительно операции сложения линейных форм, который является счетным и разрешимым, а также описан алгоритм, который строит данный базис.

Литература:

1. Hastad J. - On the size of weights for threshold gates, SIAM Journal on Discrete Mathematics, 1994.

Модель вычислительной платформы для динамического анализа защищенного бинарного кода

Соловьев Михаил Александрович

Аспирант

*Московский государственный университет имени М. В. Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия*

E-mail: eyescream@ispras.ru

В выполнении исследований программного обеспечения, оформленного в виде готового к работе бинарного кода, не сопровождаемого исходными текстами, заинтересованы многие организации, решающие задачи сертификации ПО и проблемы отладки своих разработок, а также их совместимости с другими программами и системами. Всем им требуется проводить анализ бинарного кода различных программ, как по отдельности, так и в комплексе со всей средой исполнения компьютера, иногда вплоть до самого низкоуровневого кода операционной системы. Целью таких исследований является получение информации об особенностях реализации алгоритмов, восстановление реализаций этих алгоритмов и их представление в понятном аналитику виде, восстановление протоколов обмена информацией и форматов данных, а так же поиск «недокументированных» возможностей, ошибок и уязвимостей.

Многие из применяемых для анализа алгоритмов базируются на обходе графа динамических зависимостей по данным исследуемой программы. Это относится, например, к алгоритмам слайсинга, восстановления состояния буферов памяти в определенной точке выполнения и генерации листинга. В большинстве случаев такие алгоритмы не требуют информации об анализируемой вычислительной платформе, выходящей за рамки описания зависимостей по данным между инструкциями. Таким образом, оказывается возможным предоставить спецификацию вычислительной платформы в достаточно сжатом виде, избавив сами алгоритмы анализа от необходимости учитывать специфику платформы.

Институтом системного программирования РАН разрабатывается инструмент для комплексного динамического анализа бинарного кода TrEx2, в рамках которого был реализован и опробован на практике описанный подход. Спецификации в предлагаемом виде были реализованы для вычислительных платформ Intel 64 под управлением WinNT и MIPS64 под управлением Cisco IOS. Принципиальное различие между указанными платформами указывает на применимость предлагаемого способа спецификации для широкого класса вычислительных платформ. Реализация разработанных спецификаций позволила для каждого из существующих в системе алгоритмов анализа ограничиться лишь одной, достаточно компактной, реализацией, в противовес необходимости отдельной версии для каждой из целевых платформ, что существенно ускорило и упростило разработку.

Литература

1. Weiser, M. (1981) Program slicing // Proceedings of the 5th International Conference on Software Engineering, p. 439-449. IEEE Computer Society Press, 1981.
2. Korel, B., Laski, J. (1988) Dynamic program slicing // Information Processing Letters, vol. 29, issue 3, p. 155-163.
3. Smith, R., Korel, B. (2000) Slicing event traces of large software systems // Automated and Algorithmic Debugging.
4. Cifuentes, C., Sendall, S. (1998) Specifying the semantics of machine instructions // 6th International Workshop on Program Comprehension, p. 126-133. IEEE Computer Society Press, 1998.
5. Тихонов А.Ю., Аветисян А.И., Падарян В.А. (2008) Извлечение алгоритма из бинарного кода на основе динамического анализа // Труды XVII общероссийской научно-технической конференции «Методы и технические средства обеспечения безопасности информации», с. 109. Санкт-Петербург, 2008.

Семантическая сегментация дорожного покрытия

Судаков Сергей Владимирович

Студент

*Московский государственный университет имени М.В.Ломоносова,
4факультет вычислительной математики и кибернетики, Москва, Россия
E-mail: SVSudakov@gmail.com*

Для оценки состояния автомобильных дорог и планирования ремонтных работ сейчас широко используются системы видео-паспортизации дорог (СВПД). Подобные системы, например 1.a.i.1, представляют собой комплекс видеокамер и других датчиков, размещенных на автомобиле. Поскольку использование видео для обнаружения разметки и дефектов обладает рядом недостатков 1.a.i.2, в качестве исходных данных используются изображения-развертки дорожного покрытия, которые получаются с помощью перспективного преобразования плоскости, примененного к области дороги на каждом кадре видеопоследовательности. В [3] рассматривался метод автоматического обнаружения разметки и дефектов дорожного покрытия, который встраивается в систему СВПД. В данной работе представлены улучшения метода, заключающиеся во встраивании в систему фазы предобработки изображений и добавлении новые признаков.

Изображения, подающиеся на вход системе автоматического распознавания разметки и дефектов дорожного покрытия, сильно различаются по яркости и цвету, что существенно снижает качество распознавания. Поэтому была встроена фаза предобработки, на которой проводится коррекция яркости изображения модифицированным методом Retinex [4], коррекция цвета изображения и подавление шумов с помощью билатеральной фильтрации.

Также, для улучшения качества работы системы, были встроены новые признаки сегментов, такие как текстонные гистограммы [5]. С помощью этих признаков учитывается текстурная информация о сегменте, что дает прирост производительности распознавания. Текстонные признаки являются общепризнанными дескрипторами объектов для распознавания и используются практически во всех современных системах семантической сегментации. Также были добавлены разнообразные цветовые и геометрические статистики сегментов.

Проведенные исследования показали, что приведение входных данных к унифицированному виду с помощью фазы предобработки, а так же добавление новых признаков, учитывающих дополнительную информацию о сегментах дает увеличение эффективности работы системы распознавания классов.

Литература

1. НПО «Регион», <http://www.nporegion.ru>
2. С. Судаков, А. Семашко, О. Барина, А. Конушин, В. Киншаков, А. Крылов (2008) “Алгоритмы детектирования разметки и дефектов дорожного покрытия” // Труды конференции GraphiCon'2008.
3. Сергей Судаков, Ольга Барина, Александр Велижев, Антон Конушин (2008) Семантическая сегментация дорожного покрытия на основе каскада классификаторов // РОАИ'2008.
4. Land, E., McCann, J. “Lightness and retinex theory” // Journal of the Optical Society of America, vol. 61, issue 1, p.1.
5. M. Varma and A. Zisserman. (2005) “A statistical approach to texture classification from single images” // IJCV, 62(1–2):61–81.

Методы уменьшения количества временных узлов, создаваемых при выполнении запросов на языке XQuery

Таранов Илья Сергеевич

Аспирант

Институт системного программирования РАН, Москва, Россия

E-mail: epsilon@socio.msu.ru

Язык XML является на сегодняшний день одним из самых популярных средств представления слабоструктурированных данных. Технология управления XML данными предусматривает, декларативный язык запросов XQuery. В настоящее время накоплено гораздо меньше опыта в оптимизации XQuery запросов, чем скажем в области оптимизации SQL-запросов. Конструирующие выражение языка XQuery являются важной, но в то же время очень ресурсоёмкой операцией. Семантика конструирующих выражений требует во многих случаях лишнего копирования вложенных узлов.

В ходе работы разработаны три метода оптимального вычисления конструирующих выражений. Приведены утверждения о применимости каждого из методов к некоторому классу XQuery выражений. В терминах одной из существующих XML-алгебр проведено доказательство каждого из утверждений. В рамках работы данные методы оптимизации реализованы для СУБД Sedna. Показано, что оптимизированные варианты конструкторов дают существенный прирост производительности при выполнении некоторых классов запросов.

Литература

1. M. Grinev, P. Pleshachkov (2005). Rewriting-based optimization for XQuery Transformational Queries. // Proceedings of IDEAS 2005.
2. Ghelli et al (2008): XML query optimization in the presence of side effects // SIGMOD 2008
3. L. Novak, A. V. Zamulin (2006). An XML Algebra for XQuery. // Proceedings of ADBIS 2006.
4. R. Kaushik et al. (2002) Updates for structure indexes. // Proceedings of VLDB 2002.
5. Ioana Manolescu et al. Towards micro-benchmarking XQuery. // Experimental Evaluation of Data Management Systems 2006.
6. XQuery 1.0 and Xpath 2.0 Data Model. W3C Recommendation // <http://www.w3.org/TR/xpath-datamodel/>

Выделение лиц на изображениях в условиях искажающих факторов

Тарасова Дарья Алексеевна, Шаров Андрей Николаевич

Студент, студент

*Ярославский государственный университет имени П.Г. Демидова,
физический факультет, Ярославль, Россия*

E-mail: dariya.tarasova@piclab.ru

Система автоматического обнаружения лиц решает следующую задачу: по произвольному изображению на входе системы определить имеются ли на этом изображении лица, и если да, то указать, где находится каждое лицо и каков его размер. Алгоритмы выделения лиц находят применение в системах технического зрения, компьютерной графике, робототехнике, системах видеонаблюдения и контроля доступа, в интерфейсах взаимодействия человек-компьютер. Основными требованиями, которые предъявляются к подобному классу алгоритмов, являются: высокое качество распознавания, работа в режиме реального времени, робастность по отношению к внешним факторам.

За последние несколько лет было предложено множество алгоритмов обнаружения лиц, использующих различные подходы. Наиболее эффективными из них считаются методы на основе обучения. В данной работе проводится исследование устойчивости некоторых из них к различным видам искажений. Проводится исследование трех современных алгоритмов, основанных на обучении классификаторов с помощью набора тренировочных изображений. Первый из описанных алгоритмов использует процедуру обучения, основанную на бустинге [1]; второй – обучающую сеть SNoW [2] (Sparse Network of Winnows); третий – обучение на базе машины опорных векторов (МОВ) [3]. Было проведено сравнение описанных алгоритмов между собой.

Для проведения экспериментов была составлена база данных из 50 цветных изображений, разрешения 768×576 пикселей, суммарно содержащая 213 лиц. При проведении тестирования использовались искажения следующих видов: гауссов шум, импульсный шум, размытие и сжатие JPEG. Эффективность работы алгоритмов оценивалась по уровню выделения (detection rate), который показывает процент обнаруженных лиц.

Алгоритм SNoW оказался наиболее чувствителен к шумам и сжатию. Это связано с тем, что данные искажения наиболее сильно влияют на SMQT признаки, используемые в данном алгоритме. Однако при размытии алгоритм показывает наилучшие результаты благодаря тому, что процедура квантования с порогом равным среднему значению в области изображения 3×3 пикселя, используемая для вычисления SMQT признаков, способствует частичному восстановлению информации о структуре этой области.

Наибольшую устойчивость к искажениям всех типов показал алгоритм, использующий для обучения процедуру бустинга. Он использует в своей работе простые прямоугольные признаки, значения которых мало изменяются при внесении данных искажений. Так для гауссова шума изменение значений признаков определяется разностью средних значений шума в двух прямоугольных областях.

Алгоритм на базе МОВ показал результаты схожие с алгоритмом на базе бустинга, однако при внесении шума уровень выделения убывает несколько быстрее.

Литература

1. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2001. № 1. P. 511-518.
2. Nilsson M., Nordberg J., Claesson I. Face Detection Using Local SMQT Features and Split Up SNoW Classifier // Proceedings of IEEE Int. Conf. ICASSP, 2007. V. 2. P. 589-592.
3. Kienzle W., Bakir G., Franz M., Schölkopf B. Face Detection – Efficient and Rank Deficient // Adv. in Neural Inf. Proc. Systems, 2005. V.17. P. 673-680.

Построение композитно-иерархических скрытых марковских моделей для анализа поведения

Темлянец Александр Валерьевич

студент

*Московский Государственный Университет имени М.В. Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия
e-mail: alexander.temlyantsev@gmail.com*

Успехи современной экспериментальной биологии и нейропсихологии обуславливают все возрастающую потребность в эффективных методах автоматического анализа поведения. Так, представляет интерес задача качественного определения реакции подопытного животного на те или иные стимулирующие воздействия [1], решаемая на сегодняшний день лишь с привлечением квалифицированных экспертов. Автоматизация процесса решения позволила бы придать исследованиям промышленные масштабы, существенно расширяя границы творческой свободы экспериментатора. Однако существующие методы анализа поведения в большинстве своем носят эвристический характер, не принимая во внимание глубинной структуры предметной области и, как следствие, обеспечивают недопустимо низкую точность распознавания.

Задаваясь целью по исходному признаковому описанию деятельности объекта наблюдения (например, временной развертке характеристик его движения) построить адекватную абстракцию его психической картины, мы обретаем возможность простой и естественной формализации эмпирических законов и имеем в перспективе богатый, наглядно интерпретируемый результат.

Предложенный в данной работе подход состоит в использовании формализма байесовских сетей [2] для учета структуры причинно-следственных зависимостей между поведением особи и ее эмоциональным состоянием. Выделение отдельных эмоциональных контуров и специфика их совместного влияния на активность наблюдаемого объекта обусловили усиление принятой модели дополнительными ограничениями композитности, составившими в итоге ее существенную особенность.

Исследование ведется в рамках одного из наиболее актуальных и бурно развивающихся направлений в машинном обучении – структурного распознавания [2], средствами которого успешно решаются такие задачи, как выделение фонем, распознавание рукописных текстов, сегментация изображений и прочие. Структурный подход не только предоставляет универсальный метод построения и эффективного решения сложных вероятностных моделей, но и располагает широкими средствами приближенных вычислений, существенно снижающими сложность итогового алгоритма. Применение новых методов графического вывода (inference) позволяет осуществлять решение поставленной задачи в реальном времени, открывая широкие перспективы дальнейшего внедрения метода как в научных, так и в коммерческих проектах.

Литература

1. Spruijt B.M., DeVisser L.(2006) Advanced behavioral screening:automated home cage ethology. Drug Discovery Today: Technologies Vol. 3, No. 2 2006, pp.231-237
2. Bishop C. M. (2006) Pattern recognition & mashine learning. Ch. 8-13. Springer

**Метод байесовского оценивания параметров.
Алгоритмы «Марковская цепь Монте Карло»**

Ушакова Анастасия Николаевна

к.ф.-м.н.

Норвежский университет естественных и технических наук (NTNU)

e-mail: anastasi@math.ntnu.no

Байесовские методы оценивания параметров приобрели широкую популярность в современной статистике. В данном контексте байесовский подход представляет собой рассмотрение неизвестного параметра как случайной величины, имеющей вероятностное распределение, которое можно описать в рамках заданной модели с точностью до нормализующей константы, и получение выборки из этого распределения для исследования его свойств.

Методы «Марковская цепь Монте Карло» (MCMC – Markov Chain Monte Carlo) представляют собой класс алгоритмов выборки из заданного вероятностного распределения, основанных на построении цепи Маркова, которая имеет желаемое распределение в качестве своего равновесного распределения. На основе этой выборки и производится оценка искомых параметров. Детальное описание методики представлено в книге Gilks et al. (1995).

Данный метод оценивания применим для широкого круга параметрических моделей, описывающих самые различные явления. В частности, байесовские методы оценивания используются в финансовом моделировании, эпидемиологических, физических и биологических моделях, и во многих других областях. В настоящей работе применение метода продемонстрировано на примере модели взаимодействия иммунной системы пациента с вирусом ВИЧ-инфекции.

Рассматриваемая модель является широко известной в исследованиях динамики ВИЧ-инфекции, впервые введена Nowak & May (2000) и описывается системой обыкновенных дифференциальных уравнений:

$$\begin{aligned}\frac{dX}{dt} &= \lambda - \rho X - \beta(1 - v_t)XV; \\ \frac{dY}{dt} &= \beta(1 - v_t)XV - aY; \\ \frac{dV}{dt} &= Ka(1 - \eta_t)Y - uV,\end{aligned}$$

где фазовые переменные X , Y и V обозначают популяции здоровых CD4 клеток, инфицированных CD4 клеток и вирусных частиц, соответственно; здоровые CD4 клетки производятся иммунной системой со скоростью λ , умирают естественной смертью со скоростью ρX и инфицируются со скоростью $\beta(1 - v_t)XV$, где параметр v_t соответствует эффективности лечения: так, $v_t = 0$ для периода, когда пациент не принимает терапию и

$0 < v_i < 1$, если пациент принимает терапию. Далее, инфицированные клетки появляются со скоростью $\beta(1-v_i)XV$ и умирают со скоростью aY , производя при этом $K(1-\eta_i)$ вирусных частиц каждая, где $\eta_i=0$ без терапии, и $0 < \eta_i < 1$ на терапии. Вирусные частицы разрушаются со скоростью uV .

В данной работе проводится оценка параметров λ и β для трех пациентов, периодически прерывающих антиретровирусную терапию, на основе измерений величин $X+Y$ и V в некоторые моменты времени; используется байесовский подход оценивания параметров. Для остальных параметров используются значения, ранее полученные из других трудов, в частности, Mohri et al. (1998), Putter et al. (2002), Adams (2005).

Литература

1. Adams, B. M. (2005) Non-parametric parameter estimation and clinical data fitting with a model of HIV infection. PhD thesis. Raleigh, North Carolina.
2. Gilks, W.R., Richardson, S., Spiegelhalter, D. (1995) *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*. CRC Press LLC.
3. Mohri, H., Bonhoeffer, S., Monard, S., Perelson, A. S., Ho, D. D. (1998) Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. *Science* **279**, 1223--1227.
4. Nowak, M. A. and May, R. M. (2000) *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press, Oxford.
5. Putter, H., Heisterkamp, S. H., Lange, J. M., and de Wolf, F. (2002) A Bayesian approach to parameter estimation in HIV dynamical models. *Stat. Med.* **21**, 2199--2214.

Локальные методы прогнозирования временных рядов¹

Федорова Валентина Павловна

студент

Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия

e-mail: alushaf@gmail.com

Работа выполнена в рамках участия в серии ежегодных соревнований «Forecasting Competition for Neural Networks & Computational Intelligence», целью которых является верификация нестатистических методов машинного обучения на реальных финансовых временных рядах [1]. Исследуются так называемые локальные методы прогнозирования временных рядов. Эти методы отказываются от нахождения представления временного ряда в классе заданных функций от времени. Вместо этого прогноз осуществляется на основе данных о каком-то участке временного ряда (используется локальная информация).

В данной работе подробно исследован следующий метод (обобщение классического «ближайшего соседа»). Пусть $\tilde{f} = (f_1, \dots, f_n)$ – временной ряд, требуется продолжить его до ряда $\tilde{g} = (f_1, \dots, f_n, f_{n+1}, \dots, f_{n+t})$. Предполагается, что такое продолжение определяется «предысторией» (f_{n-l+1}, \dots, f_n) , т.е. в ряде нужно найти часть (f_{k-l+1}, \dots, f_k) , которая после некоторого преобразования A «становится похожа» на (f_{n-l+1}, \dots, f_n) :

$$B(A(f_{k-l+1}, \dots, f_k), (f_{n-l+1}, \dots, f_n)) \rightarrow \min_{k, \tilde{\alpha}},$$

здесь $\tilde{\alpha}$ – параметры преобразования A , B – функция близости двух отрезков временного ряда (например, евклидово расстояние между отрезками или взвешенное евклидово расстояние, которое «последним» точкам приписывает больший вес). Определив k , мы находим ближайшего соседа к нашей предыстории. Искомое продолжение запишется в виде $A(f_{k+1}, \dots, f_{k+t})$. В общем случае ищем несколько ближайших соседей. Продолжение запишется в виде их линейной комбинации

$$\sum_r c_r A(f_{k_r+1}, \dots, f_{k_r+t}), \quad \sum_r c_r = 1.$$

В данной работе предложены специальные функции близости, учитывающие особенности реальных временных рядов: наличие выбросов, пропуски в значениях ряда. Например, для решения первой проблемы вычисляем близость между отрезками два раза. После первого – из отрезков удаляются точки, устранение которых существенно уменьшает значение функции близости («подозрительные на выброс»). Близостью считаем результат второго вычисления (на отрезках «без выбросов»). Исследовались также различные функции близости и проблема выбора параметра l и числа соседей.

Алгоритмы тестировались на данных соревнования [1] и данных, описывающих поведение хаотических динамических систем [2], при этом показали неплохие результаты на широком классе временных рядов и заняли пятое место (из 44) на соревновании «NN3 2006/07 Forecasting Competition for Neural Networks & Computational Intelligence» [1].

Литература

1. <http://www.neural-forecasting-competition.com/>
2. J. McNames, "Innovations in local modeling for time series prediction," Ph.D. Thesis, Stanford University, May 1999.
3. J. McNames, "A nearest trajectory strategy for time series prediction," Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, Katholieke Universiteit Leuven, Belgium, pp. 112-128, July 1998.

¹ Работа поддержана грантом РФФИ (проект № 08-07-00305-а).

Аппроксимация полигональной сетки головы человека с помощью параметризованной антропометрической модели

Федюков Максим Александрович

Аспирант

*Московский государственный университет имени М.В.Ломоносова,
факультет вычислительной математики и кибернетики, Москва, Россия*

E-mail: mfedjukov@graphics.cs.msu.ru

С развитием многопользовательских виртуальных сред все более актуальной становится задача параметрического моделирования форм головы и тела человека. В данной работе рассматривается задача преобразования полигональной сетки, описывающей геометрическую форму головы человека, в параметризованную антропометрическую модель, представляющую собой набор числовых характеристик, описывающих особые точки и черты лица. На сегодняшний день самой распространенной [1, 2] параметризованной антропометрической моделью является модель телосложения человека SL [3]. В ней, в частности, содержится описание черт лица человека, таких как ширина скул, высота лба, разрез глаз и т. п. Каждый такой параметр представляет собой вещественное число от 0 до 1. С каждым параметром связан набор N вершин полигональной сетки и набор N векторов смещения для каждой вершины. При решении задачи получения параметризованной модели, описывающей черты лица, полагаем, что у нас есть внешняя система, устройство которой неизвестно, но известна реакция на сигналы («черный ящик») Ξ , которая берет на вход K параметров и выдает соответствующую им полигональную сетку. Мы же решаем обратную задачу: по полигональной сетке получаем описывающий ее набор из K параметров для системы Ξ .

Таким образом, общая схема предлагаемого решения выглядит следующим образом. На вход внешней системы Ξ подаются нулевые параметры. Сгенерированную таким образом сетку будем называть оригинальной полигональной сеткой O . На вход разработанной в рамках данной работы системы подается полигональная сетка M , геометрия и топология которой соответствует O . Проводится последовательный перебор наборов параметров методом градиентного спуска. В каждом наборе собраны параметры, относящиеся только к одной черте лица и не влияющие на другие. При этом для каждой черты лица есть по два набора: параметры, влияющие на форму, и параметры, формирующие композицию черт лица. Значения перебираемых параметров подаются на вход внешней системы Ξ . Вершины получаемой в результате полигональной сетки параметризованной модели P , относящиеся к текущей черте лица, сравниваются с соответствующими вершинами исходной сетки M , и аргументы, соответствующие минимальному отклонению принимаются за искомый набор параметров $\alpha_1.. \alpha_K$. Отклонение вычисляется с помощью взвешенного функционала невязки $F(\alpha_1.. \alpha_K) = \sum_{i=1}^N v_i \varphi_i \rho_i$, где N — количество вершин полигональных сеток, v_i — вес, учитывающий эмпирическую оценку важности данной вершины для восприятия, φ_i — вес, отражающий зависимость вершины от набора параметров $\alpha_1.. \alpha_K$, а ρ_i — расстояние между радиус-векторами координат i -ой вершины исходной сетки M и полигональной сетки P .

Литература

1. V. Fernandez-Carbajales, J.-M. Martinez, and F. Moran. (2006) High-Level Description Tools for Humanoids. Multimedia Content Representation, Classification and Security.
2. N. Magnenat-Thalmann, D. Thalmann. (2004) Handbook of Virtual Humans, Standards for Virtual Humans. Wiley.
3. А.В. Герасимов. (2007) Структурно-параметрическая модель визуализации персонажа в образовательных виртуальных мирах. Educational Technology & Society 10(4).

Применение обобщенного спектрально-аналитического метода в задаче анализа биологических данных

Филиппов Виталий Владимирович, Горчаков Михаил Александрович

Аспирант, аспирант

Московский Государственный Университет им. М.В. Ломоносова, Москва, Россия

E-mail: skiffctmc@mail.ru

Актуальной задачей является поиск самоподобных фрагментов в биологических данных. Данный доклад рассматривает её с точки зрения использования обобщенного спектрально-аналитического метода (ОСАМ). Новизна подхода заключается в том, что использование ОСАМ позволяет существенно упростить анализ данных (в т.ч. и благодаря хорошим перспективам его реализации с помощью новых вычислительных технологий), что обеспечивает возможность обработки больших объемов данных за меньшее время даже на обычных ПК. Были поставлены следующие задачи: реализация алгоритма поиска самоподобных фрагментов в больших объемах данных (в частности, в описании генома), поиск оптимальных базисов для этой цели и создание алгоритма автоматического подбора оптимальных значений параметров, или модуля самообучения.

В ходе решения этих задач и тестировании алгоритма на реальных данных было выяснено, что наиболее интересные результаты достигаются при использовании ортонормированной полиномиальной системы Чебышева 2-го рода. Это соответствует ожиданиям в начале работ, т.к. вид весовой функции этой системы позволяет наиболее точно локализовывать самоподобные фрагменты в данных, в то же время гарантируя отсутствие у алгоритма проблемы «переобучения».

В дальнейшем были проведены работы по созданию модуля автоматической оптимизации параметров алгоритма с использованием различных оптимизационных методов, например, метода градиентного спуска.

Литература

1. Ф.Ф. Дедус, Л. И. Куликова, А. Н. Панкратов, Р. К. Тетуев «Классические ортогональные базисы в задачах аналитического описания и обработки информационных сигналов»
2. Ф.Ф.Дедус, А.Ф.Дедус, С.А.Махортых, М.Н.Устинин «Обобщённый спектрально-аналитический метод обработки информационных массивов», М., Машиностроение, 1999 г.
3. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery “Numerical Recipes in C”
4. “Intel Integrated Performance Primitives for Intel Architecture. Reference Manual”

Применение плоских кривых для программирования поверхностных эскизов

Фомкина Мария Сергеевна¹

студент

Сургутский государственный университет, Сургут, Россия

E-mail: msfomkina@rambler.ru

В настоящее время актуальной проблемой является создание высококачественных дизайнерских эскизов, отвечающих разнообразным целям производства и рекламы. Однако далеко не каждый заказчик имеет возможность воспользоваться услугами профессионального дизайнера, а библиотеки готовых изображений не всегда соответствуют его требованиям. Решением данной проблемы может быть использование графиков плоских кривых в качестве художественных объектов.

Целью данной работы является написание графического редактора, позволяющего создавать дизайнерские эскизы, включающие узоры, построенные на основе графиков плоских кривых. [1]

Основная задача заключается в предоставлении возможности строить графики плоских кривых. Для этого написан специальный модуль *GnuplotForLazarus*, который обеспечивает использование свободного программного средства *Gnuplot* для построения графиков как по заданным в редакторе, так и по введённым пользователем уравнениям. Для поворота и переноса графика используются соответствующие преобразования координат.

Следующая задача – организация хранения и отображения рисунка. Каждому редактируемому изображению соответствует два объекта: типа *TBitmap*, в котором находится весь рисунок, и типа *TImage*, который содержит видимый пользователю фрагмент с учётом масштаба отображения. Все операции производятся над первым объектом и немедленно проецируются на второй.

Ещё одна задача – организация удобного пользовательского интерфейса, включающего, в том числе, как классические графические инструменты (линия, карандаш, ластик и т.д.), так и нестандартные средства комбинирования и редактирования изображений (паркет, плитка, наложение, смена цветов и т.д.). Графический редактор с рабочим названием *Artist-Geometer* написан на языке программирования *FreePascal* в среде *Lazarus*, имеет *MDI*-интерфейс, позволяет работать на двух языках, русском и английском, имеет справку в формате *.CHM*.

Кроме основных графических форматов, редактор позволяет хранить изображения в специальном векторном формате *AGI*, содержащем описания входящих в изображение объектов, что достигается путём заполнения в процессе создания рисунка специальной динамической структуры с последующей её оптимизацией и сохранением. В отличие от стандартных форматов, данный имеет расширенный набор базовых объектов и функций, что даёт возможность хранить более сложные рисунки в файлах меньшего объёма. В дальнейшем предполагается развитие данного формата.

Редактор также используется в образовательном процессе с целью демонстрации геометрического смысла математических формул. В настоящее время он внедрён в учебный процесс в сургутской гимназии «Лаборатория Салахова».

Литература

1. Фомкина М. С. Средства создания дизайнерских поверхностных эскизов, содержащих графики различных плоских кривых // Материалы XLVI Международной научной студенческой конференции «Студент и научно-технический прогресс»: Информационные технологии / Новосиб. гос. ун-т. Новосибирск, 2008. 236 с. – С. 14-15.

¹ Автор выражает признательность доценту, к. ф.-м. н. Моргуну Д.А. за помощь в подготовке тезисов.

Применение базиса бинарных изображений в рентгеновских аппаратах сканирующего типа.

Хасанов Ренат Ришатович, Хасанов Марат Ришатович

Аспирант, студент

Уфимский государственный авиационный технический университет, Уфа, Россия

E-mail: re2003@mail.ru

Наиболее характерной чертой современной интроскопии и медицинской рентгенодиагностики в частности является дигитализация (переход на цифровую обработку информации) и компьютеризация (мощная математическая обработка цифрового сигнала) всех средств визуализации. На цифровые технологии переходит и нерентгеновская лучевая диагностика - ультразвуковые аппараты, магнито-резонансная томография. Отсюда следует, что широкое применение методов цифровой диагностики – это не просто дань моде, а совершенно необходимый уровень техники в современной промышленности и медицине. Все эти процессы вызваны с всевозрастающей ролью качества, быстроты, экономической эффективности, а главное быстротой и точностью диагностики и лечения.

В России внедрение цифровых систем в медицинскую практику пока находится на стадии начального развития. Процесс перехода от традиционной технологии исследований к цифровой долгое время замедлялся недостатком качественной и доступной по цене отечественной техники. По этим причинам возникает необходимость конкурирующих разработок отечественных цифровых систем рентгенографии, в которых все требования к основным характеристикам являются предельными и не последнее слово в этой конкуренции – цена отечественных разработок. Она ниже импортных аналогов.

В результате проведенных исследований была разработана функциональная структура устройства регистрации и обработки рентгеновского изображения обладающая рядом отличительных возможностей. Применение специальных приборов оптической электроники и поликристаллических сенсоров (сцинтилляторов) чувствительных к рентгеновскому излучению для визуализации и дискретизации регистрируемых образов объектов исследований расширяет функциональные возможности и характеристики устройства, повышается чувствительность к рентгеновскому излучению и обеспечивается высокое пространственное разрешение. Техническим результатом является повышение разрешающей способности (20 пар лин/мм и более) и общего быстродействия всей системы. Техническую реализацию основных элементов устройства предлагается осуществлять в едином технологическом цикле на одной, общей подложке. Также разработан способ регистрации образов исследуемого объекта основанный на использовании базиса бинарных изображений. Использование этих методов позволяет расширить динамический диапазон преобразования интегральной интенсивности оптического сигнала, регистрируемого фотоприемником, а также существенно сократить скорость передачи массивов данных в ЦВМ в цифровой (бинарной) форме.

Литература

1. Сойфер В.А. Компьютерная обработка изображений. Часть 2. Методы и алгоритмы. // Соросовский образовательный журнал, №3 1996.
2. <http://www.matlab.ru/imageprocess/book2/index.asp> Журавель И.М. "Краткий курс теории обработки изображений". – Консультационный центр MATLAB компании SoftLine. (15 фев. 2003)

Оптимизация граничного управления смещением на одном конце струны с условием закрепления на другом конце в пространстве W_2^2 для задачи возбуждения

Холмеева Анна Андреевна

Аспирант

Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия

E-mail: kholomeyeva@hotmail.com

Данная работа является продолжением работ В.А. Ильина и Е.И. Моисеева, посвященных задачам оптимального граничного управления, и попыткой найти решение одной из классических задач этой тематики в пространстве с повышенной гладкостью.

Рассмотрим колебания упругой струны, удовлетворяющие волновому уравнению $u_{tt}(x,t) - u_{xx}(x,t) = 0$ в прямоугольнике $Q_T = [0 < x < l] \times [0 < t < T]$, причем рассмотрение ведется в течение промежутка времени T , кратного удвоенной длине струны, т.е. $T = 2l(n+1)$, $n = 1, 2, \dots$. Правый конец струны закреплен, т.е. $u(l,t) = 0$, на другом конце проводится управление смещением $u(0,t) = \mu(t)$. Задача состоит в отыскании управления, которое переводит процесс колебаний струны из начального состояния покоя $u(x,0) = 0$, $u_t(x,0) = 0$, $0 \leq x \leq l$ в произвольно заданное финальное состояние $u(x,T) = \varphi(x)$, $u_t(x,T) = \psi(x)$, $0 \leq x \leq l$. При этом на финальные функции накладываются ограничения, обусловленные условием закрепления правого конца струны.

Поскольку при $T > 2l$ управлений, удовлетворяющих поставленным условиям, существует бесконечное число, ставится вопрос об отыскании так называемого оптимального граничного управления. В общей постановке это задача об отыскании среди всех решений $\mu(t)$ из некоторого класса тех, которые минимизируют некоторый функционал при выполнении некоторых условий связи, обусловленных граничными, начальными и финальными условиями. Ранее в работах [1], [2] и других была развита теория решения задач оптимального граничного управления в смысле обобщенного решения из класса $\hat{W}_2^1(Q_T)$, в частности, в этом классе была полностью решена поставленная выше задача, где оптимальное граничное управление из класса $L_2[0, T]$ доставляло минимум функционалу $\int_0^T (\mu'(t))^2 dt$.

В данной работе эта задача решена в явном виде в пространстве $\hat{W}_2^2(Q_T)$, то есть найдено оптимальное граничное управление $\mu(t) \in W_2^1[0, T]$, доставляющее минимум интегралу $\int_0^T (\mu''(t))^2 dt$. Кроме того, для некоторых частных случаев предьявляется в явном аналитическом виде решение задачи для пространства $\hat{W}_p^2(Q_T)$, $p \geq 1$. Для случая $p = 1$ доказывается неединственность решения, которая имела место и в пространстве $\hat{W}_1^1(Q_T)$ (см. [3]).

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (проект 08-01-00313) и программы поддержки ведущих научных школ (проект НШ - 2726.2008.1).

Литература

1. Ильин В.А., Дифференциальные уравнения. 2000. Т.36, номер 11. с. 1513 – 1528.
2. Ильин В.А., Моисеев Е.И., Дифференциальные уравнения. 2006. Т.42, N 11. с. 1558 – 1570.
3. Моисеев Е.И., Дифференциальные уравнения. 2008. Т.44, N 5. с. 709 – 711.

Информационный критерий оценки качества цветных цифровых изображений

Хорунжий Михаил Дмитриевич

аспирант

Санкт-Петербургский Государственный Университет Кино и Телевидения
Факультет приборов и систем кино и телевидения, Санкт-Петербург, Россия
E-mail: horungij@gmail.com

В настоящий момент система цифровой технологии создания фильма и средства визуализации контента представляют собой важную область для научно-исследовательских работ. Необходимо отметить, что первостепенным значением для адекватности зрительного восприятия цифровых цветных изображений, а также видеопоследовательностей обладает задача установления критериев к качеству экранных изображений в особенности требований к цветопередаче. В представленном докладе анализируются современные аспекты оптимизации качества цветных цифровых изображений с учетом информационного содержания, и предлагается авторский подход, приведенный на рис. 1, с построением сенсорной зависимости по ощущаемому контрасту цветного изображения [Хорунжий, 2008].

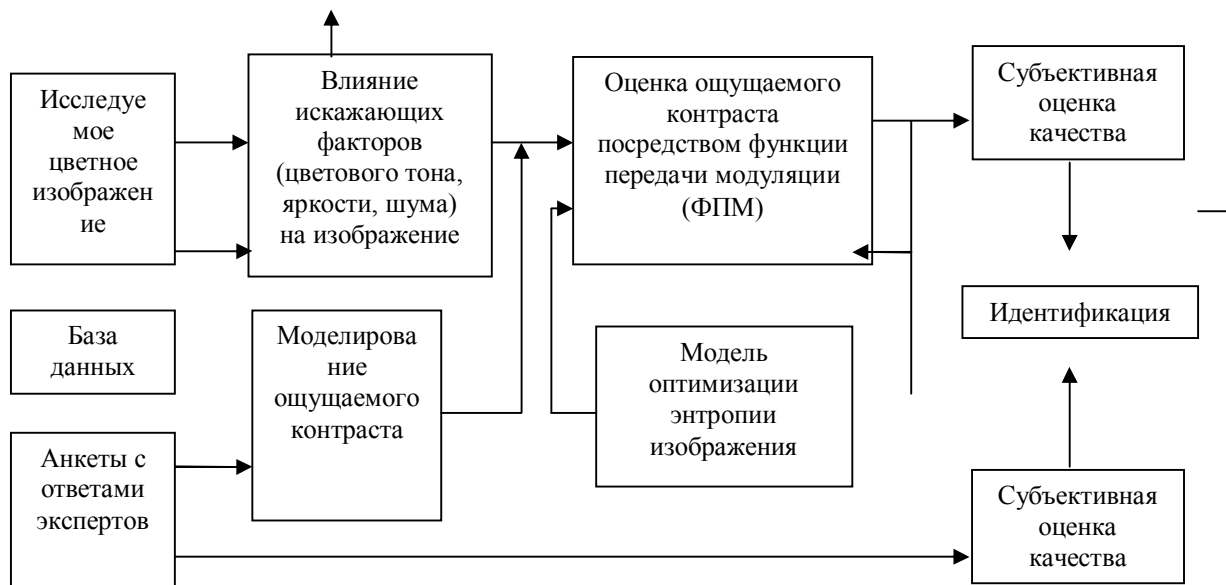


Рис. 1 Общая структурная схема оценки качества цветопередачи

Результаты проведенного исследования используются в лабораторной установке по дисциплине “Основы записи и воспроизведения информации” в СПбГУКиТ.

Литература

1. М.Д. Хорунжий. Метод количественной оценки цветов различий при восприятии цифровых изображений. // Вестник НГУ. Серия “Информационные технологии”. – Новосибирск, 2008. – том 6, вып. 1 с. 80-88.
2. X.M. Zhang and B.A. Wandell. Color image fidelity metrics evaluated using image distortion maps, Signal Processing, 70, 201-214 (1998).

Использование алгоритмов онлайн обучения при разработке системы семантической сегментации изображений

Шаповалов Роман Викторович, Барина Ольга Вячеславовна

студент, аспирантка

*Московский Государственный Университет имени М.В. Ломоносова,
факультет Вычислительной Математики и Кибернетики, Москва, Россия*

E-mail: overrider@cmcspec.ru, olga.barinova@gmail.com

Системы семантической сегментации изображений применяются в различных областях, например, см. [1]. Часто в основе таких систем лежат методы машинного обучения, поэтому для настройки системы обычно требуется разметить большую базу изображений, что представляет собой довольно трудоёмкий процесс. Мы поставили перед собой цель облегчить его, сделав процесс обучения интерактивным. В данной работе предлагается использовать при разметке базы изображений классификатор, обученный на уже размеченной части данных, таким образом, вместо полной разметки изображения пользователю необходимо лишь исправить ошибки классификации.

Такая постановка задачи накладывает ряд ограничений на используемый метод обучения. Каждый поступающий на обучение объект может быть использован только один раз, причём время обучения не должно зависеть от количества уже использованных при обучении объектов. Методы, обладающие таким свойством, называют методами онлайн обучения. Один из онлайн классификаторов, *On-line Costing*, способный хорошо обучаться даже в случае несбалансированных классов, описан в [2].

Как правило, объекты интереса занимают небольшую часть площади изображения. Поэтому в разработанной нами системе используется каскад классификаторов, как это описано в [1]. Во время классификации на первых слоях каскада отбрасываются сегменты, которые очевидно являются фоном, а на последнем классифицируются объекты интереса. Обучение каскада онлайн классификаторов – отдельная интересная задача. В разработанной системе классификатор на каждом слое обучается на той части изображения, которую он классифицировал (то есть, которую ему оставили предыдущие слои). Но если позволить классификаторам на первых слоях неправильно обучиться, в дальнейшем невозможно будет отнести неправильно классифицированные объекты к их классу. Для того чтобы избежать ошибок с отнесением объекта целевого класса к фону на первых слоях каскада, к обучающей выборке добавляются *все* объекты целевых классов. На последнем слое каскада работает многоклассовый классификатор.

Поскольку система разметки является интерактивной, необходимо предоставить пользователю широкий набор инструментов для переразметки классифицированного на предыдущем этапе изображения. На ранних стадиях обучения классификатор может выдавать результаты, смещённые в сторону одного из классов. Поэтому в интерфейс вынесены элементы управления, позволяющие регулировать пороги между классами. Таким образом, работа пользователя по переразметке изображения сводится к минимуму.

Эксперименты на реальных базах изображений показали, что разработанная система позволяет обучать классификатор в режиме реального времени, сокращая затраты труда на обучение в 2-3 раза.

Литература

1. Semantic segmentation of road images based on cascade classifiers. Sergey Sudakov, Olga Barinova, Alexander Velizhev, Anton Konushin. // Proceedings of PRIA'2008.
2. Шаповалов Р.В., Барина О.В., Велижев А.Б., Конушин А.С. Онлайн бэггинг, чувствительный к стоимостям ошибок на разных классах. // Труды 51-й научной конференции МФТИ. Часть VII. Управление и прикладная математика. – Москва-Долгопрудный, 2008. Том 3. Стр. 79-82.

Кластеризация интересов пользователей социальной сети

Шейнин Владимир Валерьевич

Студент

*Московский физико-технический институт (государственный университет)
факультет управления и прикладной математики
sheynin@mycolornet.com*

В настоящее время социальные сети становятся все более популярными. Как правило, описывая себя в социальной сети, пользователь заполняет список своих интересов – ключевых слов. В качестве интересов часто указывают не только слова, но и словосочетания. Из-за этого число различных интересов существенно превышает число пользователей. С другой стороны, для выявления предпочтений пользователей с целью построения рекомендательных сервисов было бы удобно выделить некоторое подмножество «опорных» интересов, и все дальнейшие операции проводить именно с ним, а не с полным множеством.

В данной работе рассматривается задача кластеризации интересов пользователей социальной сети. Будем считать, что задана матрица M , $m_{ij} \in [0,1]$, строкам которой соответствуют пользователи, а столбцам – интересы (items). Элемент m_{ij} означает вероятность принадлежности классу 1. Класс 1 означает, что пользователь отметил интерес. Большая часть элементов матрицы равна 0, что соответствует ситуации, когда пользователь никак не связан с этим интересом.

Для решения задачи предлагается использовать только ту информацию, которую можно извлечь непосредственно из матрицы M . Благодаря этому наш подход легко переносится для кластеризации других ресурсов.

На первом этапе выделяются основные кластеры интересов. Для этого используется L-вариант модифицированного метода островной кластеризации [2] с измененным выражением для степени корреляции. Роль документов играют пользователи, роль термов — ключевые слова. При переборе всех существующих связей интересов рассматриваются только такие связи, где степень корреляции меньше некоторого порога. Это позволяет строить кластеризацию, основываясь только на самых значимых связях интересов. В результате первого шага получается некоторое количество кластеров. Интересы, кластеризованные на первом этапе назовём «опорными», они примечательны тем, что наилучшим образом характеризуют соответствующие кластеры со смысловой точки зрения. Второй этап заключается в определении степени принадлежности каждого интереса к каждому найденному кластеру.

В результате действия алгоритма каждому интересу сопоставляется профиль – нормированный на единицу вектор в пространстве размерности, равной числу найденных кластеров. Благодаря этому, для сравнения любых двух интересов можно использовать обычное евклидово расстояние. Кроме того, при решении задач коллаборативной фильтрации наличие профилей у интересов позволяет индуцировать профили у пользователя, что позволит эффективно искать пользователей со схожими интересами.

Литература

1. *Sahami M.* Mining the Web to Determine Similarity Between Words, Objects, and Communities — FLAIRS Conference, 2006.
2. *Киселев М.В., Пивоваров В.С., Шмулевич М.М.* Метод кластеризации текстов, учитывающий совместную встречаемость ключевых терминов, и его применение к анализу тематической структуры новостного потока, а также ее динамики // *Международ. сб. науч. раб.: Интернет-математика 2005: автоматическая обработка веб-данных.* — М.: Яндекс, 2005.

Оценка артефактов блочности JPEG-изображениях

Шмаглит Лев Александрович

Студент

Ярославский государственный университет имени П.Г. Демидова, Ярославль, Россия

E-mail: lev.shmaglit@piclab.ru

В большинстве случаев человек является конечным получателем визуальной информации. Поэтому измерение артефактов блочности с учетом известных характеристик зрительной системы человека (ЗСЧ) является целесообразным для оценки видимости артефактов блочности. Структурная маскировка, возникающая на изображениях с большим количеством мелких деталей, и яркостная маскировка – это две хорошо изученные характеристики ЗСЧ, которые в значительной степени влияют на восприятие артефактов блочности [1]. Предположим, что общая детализированность каждого блока – это просто сумма детализированностей по всем пространственным частотам, причем амплитуда каждой составляющей этой суммы равна соответствующему коэффициенту дискретного косинусного преобразования. Поскольку артефакты блочности возникают только в двух конкретных направлениях (вертикальном и горизонтальном), определим две по-разному ориентированные детализированности (вертикальную A_v и горизонтальную A_h) следующим образом [2]:

$$A_v = \sum_{v=1}^7 v \sum_{u=0}^7 |R(u, v)|, \quad A_h = \sum_{u=1}^7 u \sum_{v=0}^7 |R(u, v)|.$$

Для вертикальных артефактов блочности преобладает эффект маскировки за счет горизонтальной детализированности, поэтому общая детализированность определяется по формуле $A_{total}^v = A_h + \alpha A_v$. Общая же детализированность для горизонтальных артефактов блочности определяется аналогично: $A_{total}^h = A_v + \alpha A_h$. Здесь $\alpha = 0,8$.

С учетом обоих эффектов маскировки итоговое значение видимости артефактов блочности на границе двух соседних блоков η определяется следующим образом [2]:

$$\eta = \frac{|\beta|}{(1 + A_{total}^h) \left(1 + (\mu/\mu_0)^\gamma \right)},$$

где μ и β определяются из математических выражений, моделирующих эффект блочности на границе двух смежных блоков, A_{total}^h – это общая горизонтальная детализированность блока \hat{b} , а $\mu_0 = 150$ и $\gamma = 2$.

Результатом работы описанного выше алгоритма является вычисление видимости артефактов блочности на границе двух соседних блоков η для каждой пары соседних блоков. Локальные значения блочности могут быть объединены для получения численного значения блочности для всего изображения с целью оценки его качества.

$$\text{НИБ} = \sqrt[p]{\frac{1}{N} \sum_{k=1}^N \eta_k^p},$$

где N – это общее число границ между всеми блоками изображения, НИБ – это глобальное значение блочности всего изображения, а показатель степени $p = 4$.

Литература

1. Ahumada A., Jr., Peterson H. Luminance-model-based DCT quantization for color image compression // Proc. SPIE Human Vision, Visual Processing and Digital Display III, Aug. 1992. V. 1666, P. 365-374.
2. Саутов Е.Ю. Уменьшение артефактов блочности в JPEG-изображениях // Матер. 15-ой межд. науч.-тех. конф. «Проблемы передачи и обработки информации в сетях и системах телекоммуникаций», Рязань, 2008. Т. 2, С. 132-133.