

Секция «Математика и механика»

Двухфазная схема решения задачи «структура – свойство»

Прохоров Евгений Игоревич

Аспирант

Московский государственный университет имени М.В. Ломоносова,

Механико-математический факультет, Москва, Россия

E-mail: eugeny.prokhorov@gmail.com

Работа посвящена методам поиска количественных соотношений «структура – свойство» для прогнозирования активности химических соединений. Ключевой особенностью задачи «структура – свойство» по сравнению с классическими задачами классификации является ее ориентированность на предсказание активности новых неизученных соединений. Для достижения этой цели автором предложено использование ограниченной допустимости для распознающих моделей [1,2]. Двухфазная схема решения задачи «структура – свойство» заключается в следующем. Пусть исходная задача классификации состоит в том, чтобы построить распознающую модель, которая по описанию структуры химического соединения в виде векторов дескрипторов относит его к одному из заданных классов активности. Предполагаем, что обучение модели происходит на некоторой выборке соединений, активность которых известна. Пусть также построена распознающая модель, решающая исходную задачу классификации, назовем её моделью первого уровня. И в ходе процедуры скользящего контроля на обучающей выборке вычислен её функционал качества φ_1 (отношение количество верно спрогнозированных соединений к общему числу осуществленных прогнозов). Сформулируем задачу классификации второго уровня. В первый класс соединений новой задачи классификации войдут те соединения обучающей выборки, прогноз активности которых моделью первого уровня осуществлен верно, а во второй класс – те, соединения, активность которых предсказана неверно. Пусть для задачи второго уровня также построена распознающая модель и её функционал качества равен φ_2 . Определим результирующую распознающую модель. Результирующая модель решает исходную задачу классификации, но в отличие от модели первого уровня, результирующая модель обладает опцией отказа от прогноза. Результирующая модель осуществляет отказ от прогноза данного соединения тогда, когда модель второго уровня относит данное соединение ко второму классу и осуществляет прогноз активности моделью первого уровня в противном случае. Обозначим функционал качества результирующей модели через φ_0 , а через Reject – количество отказов от прогноза. Тогда верна следующая оценка качества результирующей модели.

$$\varphi_0 = \frac{(\varphi_1 + \varphi_2)N - \text{Reject}}{2(N - \text{Reject})}.$$

В частности если $\varphi_2 \geq \varphi_1 > 1/2$, тогда если $\text{Reject} > 0$ имеем $\varphi_0 > \varphi_1$. Таким образом, доказано улучшение качества прогноза при использовании нетривиальных правил отказа от прогноза. Также в ходе работы проведены многочисленные испытания, подтверждающие практическую эффективность разработанного подхода.

Литература

1. E.I. Prokhorov, L.A. Ponomareva, E.A. Permyakov and M.I. Kumskov Fuzzy classification and fast rules for refusal in the QSAR problem // Pattern Recognition and Image Analysis. 2011. Volume 21. № 3. 542–544.
2. Прохоров Е.И. Нейронные сети для построения ограничений допустимости в задаче «структура – свойство» // Нейрокомпьютеры: разработка, применение. 2012. № 10. 46–56.

Слова благодарности

Работа выполнена при финансовой поддержке гранта РФФИ № 10-07-00694.