

Взаимосвязанные статистические модели для оценки констант равновесия и кислотности таутомеров

© Занков¹ Дмитрий Владимирович, Маджидов^{1*+} Тимур Исмаилович, Саттаров¹ Борис Борисович, Гимадиев¹ Тимур Рустемович, Нугманов¹ Рамиль Ирекович, Баскин² Игорь Иосифович и Варнек^{3*} Александр Алексеевич

¹ Кафедра органической химии. Химический институт им. А.М. Бутлерова.

Казанский (Приволжский) федеральный университет. ул. Кремлевская, 18. г. Казань, 420008.

Республика Татарстан. Россия. Тел.: (843) 233-73-71. E-mail: timur.madzhidov@kpfu.ru

² Физический факультет. Московский государственный университет им. М.В. Ломоносова. Ленинские горы, 1/2. г. Москва, 119991. Россия. Тел.: (495) 939-10-13. E-mail: igbaskin@gmail.com

³ Laboratory of Chemoinformatics. University of Strasbourg. B. Pascal, 4. Strasbourg. France.

Phone: +7 (843) 233-73-71. E-mail: varnek@unistra.fr

*Ведущий направление; +Поддерживающий переписку

Ключевые слова: моделирование «структура-свойство», гребневая регрессия, таутомерия, кислотность, константа равновесия, хемоинформатика, машинное обучение.

Аннотация

Логарифм константы прототропной таутомерии может быть выражен через разность в значениях показателей кислотностей соответствующих таутомеров. Для использования данной зависимости в предсказании константы таутомерного равновесия была предложена модификация метода гребневой регрессии, которая позволяет проводить минимизацию ошибки предсказания одновременно двух моделируемых свойств: константы равновесия реакций прототропной таутомерии и константы кислотности таутомеров, принимающих участие в равновесии.

На основе данного метода были построены модели совместного предсказания константы равновесия реакций прототропной таутомерии и показателей кислотности соответствующих таутомеров. В качестве обучающих выборок были использованы данные по 800 константам равновесия реакций прототропной таутомерии, взятые из опубликованных статей, а также данные по ~1300 константам кислотности органических молекул в различных растворителях, которые были нами специально собраны из литературы. Для построения моделей использовались локальные (помеченные) атом-центрированные фрагментные дескрипторы. Атомы, которые в молекулах таутомеров являются кислотными центрами, были выявлены с использованием подхода. Конденсированного графа реакции. Также были использованы дескрипторы, описывающие условия протекания реакций таутомерии и условия измерения кислотности органических молекул.

В ходе исследования с помощью процедуры перекрестного контроля с использованием значения коэффициента детерминации отобраны оптимальные модели совместного обучения. Для сравнения использовались индивидуальные модели, в которых константы кислотности (и константа таутомерного равновесия) предсказывались с использованием модели, полученной при использовании только данных по константам кислотности (таутомерного равновесия, соответственно) в обучении. Показано, что совместное обучение позволяет построить модель, которая предсказывает константы кислотности не хуже, чем индивидуальные модели, а константы таутомерного равновесия – даже лучше. В отличие от использования индивидуальных моделей, построенная совместная модель обеспечивает корректное соотношение значений предсказываемых свойств.