Экспертная система для интерпретации ИК спектров координационных соединений Ln3+

Автор: асп. 3 г/о ФНМ МГУ Кошелев Д.С.

[dan\_kosh@mail.ru](mailto:dan_kosh@mail.ru)

Экспертные системы создаются для интерпретации данных и служат для снижения рутинной нагрузки на экспертов-людей. Одним из рутинных анализов в химической науке является ИК спектроскопия. Это неразрушающий метод анализа, в основе которого лежит явление поглощения ИК излучения (с волновым числом 4000-400 см-1) материалом. При этом наибольшее поглощение наблюдается в случае резонанса, то есть совпадения энергии излучения с энергией какого-либо колебания химической связи в веществе. Существует ограниченное количество широко распространённых типов химических групп или связей с известными резонансными частотами колебаний. Эксперты по спектру ИК поглощения способны сказать о наличии или отсутствии той или иной группы в соединении, что может дать важную информацию для понимания состава или строения материала. Рутинность анализа и несовершенство человеческого восприятия подталкивает к использованию методов машинного обучения или нейронных сетей для выявления закономерностей между наличием химической связи в молекуле вещества с наличием полосы на ИК спектре.

Ранее в литературе были рассмотрены как примитивные методы машинного обучения для классификации подготовленных ИК спектров [1] [2] [3], так и более сложные модели, к примеру модель Inception V3 (TensorFlow) для распознания картинок ИК спектров и выявлению полос поглощения отдельных групп [4].

В моей работе предлагается использовать наработки предыдущих авторов и совместить методы машинного обучения и нейронных сетей для создания программного инструмента, который будет помогать в распознании рутинных ИК спектров учёным-химикам, совмещая и работу с табличными данными, и с картинками, по необходимости, или для повышения точности работы системы. Кроме того, предлагается сделать дообучение существующих или полученных моделей на спектрах координационных соединений (КС) лантанидов, что ранее в литературе не встречалось. Это проблема актуальна и интересна тем, что в данных соединениях существуют связи, колебания которых не являются повсеместно распространёнными. К тому же полосы колебаний даже распространённых групп в органическом лиганде в КС могут быть сдвинуты по энергии в связи со взаимодействием с ядром комплекса – ионом лантанида. Дополнительной информацией, которая может быть извлечена является расщепление полос из-за различного координационного взаимодействия данных групп. Эти особенности в совокупности должны привести к получению научно значимых данных, которые можно будет опубликовать.

Поскольку это представляется собой различные варианты задачи мультиклассовой классификации, то в качестве метрики можно применить balanced\_F1 или AUC-ROC. Физичными величинами ответов сетки будет положительная вероятность присутствия группы.

Список литературы:

[1] R.J. Fessenden, L. Györgyi, Identifying functional groups in IR spectra using an artificial neural network, J. Chem. Soc. Perkin Trans. 2. (1991) 1755–1762.

[2] R. Herges, R. July, Automatic Interpretation of Infrared Spectra: Recognition of Aromatic Substitution Patterns Using Neural Networks, 1 (1992) 723–731.

[3] J.A. Burns, G.M. Whitesides, Feed-forward neural networks in chemistry: mathematical systems for classification and pattern recognition, Chem. Rev. 93 (1993) 2583–2601.

[4] A.A. Enders, N.M. North, C.M. Fensore, J. Velez-alvarez, H.C. Allen, Functional Group Identi fi cation for FTIR Spectra Using Image-Based Machine Learning Models, (2021). https://doi.org/10.1021/acs.analchem.1c00867.