

УДК 541.13

**ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ
ДЛЯ ВОСПОЛНЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ
ХАРАКТЕРИЗАЦИИ СОЕДИНЕНИЙ МЕТОДАМИ РЕНТГЕНОВСКОЙ
СПЕКТРОСКОПИИ В ИССЛЕДОВАНИИ ЛИТИЙ-НАСЫЩЕННЫХ
СЛОИСТЫХ ОКСИДОВ**

Н.В. Киреева*, А.Ю. Цивадзе

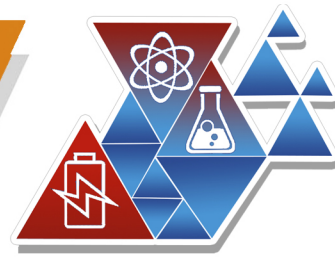
ФГБУН Институт физической химии и электрохимии им. А.Н. Фрумкина РАН,
Москва, Россия

*e-mail: kireeva@phych.ea.ru

Методы машинного обучения являются эффективным инструментом для поддержки исследований в различных областях химии и наук о материалах. Рациональный поиск соединений с заданными свойствами и оптимизация характеристик [1], прогнозирование структурного типа [2], анализ данных физико-химических методов характеристики соединений [3], микроструктурная информатика [4], компьютерный синтез/дизайн эксперимента [5] и идентификация новых наблюдений являются одними из наиболее быстро развивающихся направлений, связанных с исследованием новых материалов.

В данной работе представлена еще одна из возможностей методов машинного обучения, – возможность использования машинного обучения для восполнения экспериментальных данных (*Missing Value Imputation*) [6], являющаяся менее задействованной и известной в химии и науках о материалах. Использование физико-химических методов характеристики соединений одного и того же типа происходит несогласованно вследствие отсутствия одного принятого протокола исследования свойств получаемых соединений, что достаточно сложно ввести вследствие различной технической оснащенности лабораторий. Эта проблема, как и проблема недостаточной интегрированности методов машинного обучения в научно-исследовательский процесс, ограниченного количества эффективных инструментов междисциплинарного обмена большими объемами экспериментальных данных являются известными препятствиями в этой и смежных областях [7].

В данной работе используется один из представителей данной группы методов – модифицированный метод Гауссовских процессов – *Разреженные Вариационные Гауссовские Процессы* (*Sparse Variational Gaussian Processes*) [8]. Использование данной аппроксимации позволяет работать с большими данными в миллионы соединений, что является невозможным для оригинального подхода вследствие необходимости в ходе процесса обучения преобразовывать матрицу ковариаций размером $N \times N$. Метод реализован в виде задачи восстановления регрессии при решении проблемы состав-способ получения-структура-свойство,



где в ходе итерационного процесса иерархический ансамбль Гауссовских процессов на основе имеющейся матрицы вводимых данных размером $N \times M$ (где N - число соединений, M - число дескрипторов), имеющей пропущенные значения, где в ходе итерационного процесса построения вероятностной модели заменяет вводимые изначально средние по столбцу значения дескрипторов на значения прогноза, полученные на основе обучения на всех оставшихся данных. Столбцы для восполнения данных ранжируются по величине стандартного отклонения. После этого второй Гауссовский процесс используется для восполнения данных второго столбца, при наличии еще одного параметра с неизвестными значениями, при этом оперируя уже имеющимися результатами прогноза для первого. На Рисунке 1 представлена схема алгоритма.

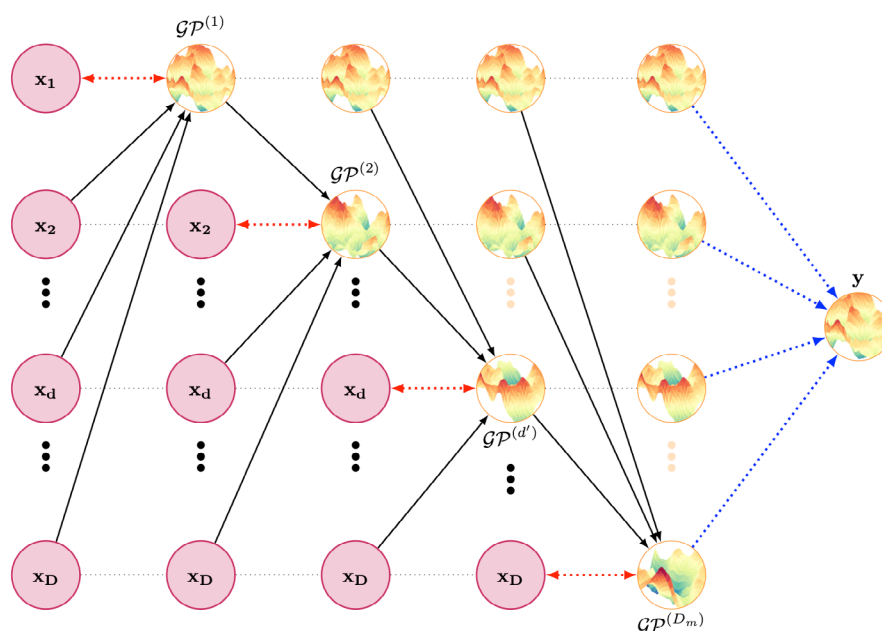
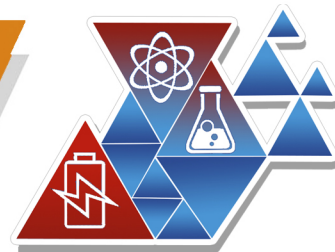


Рисунок 1. Схема алгоритма восполнения данных метода разреженных вариационных Гауссовских процессов. Воспроизведено с разрешения Elsevier (2023) [8].

В качестве объектов исследования в работе были использованы данные по литий-насыщенным слоистым оксидам, исследуемым как катодные материалы аккумуляторов с улучшенными характеристиками плотности энергии. В число целевых характеристик, являющихся объектами для методов восполнения данных, вошли постоянная решетки для предварительной оценки циклируемости соединений, значения отношения интенсивности пиков $I(003)/I(104)$ XRD спектра для оценки степени разделения катионов лития и никеля в структуре и данные фотоэлектронной спектроскопии (XPS) по катионам марганца и никеля для оценки состояния поверхности катодного материала. Последующее моделирование включало прогноз трех основных функциональных характеристик литий-



насыщенных слоистых оксидов: исходную разрядную емкость, Кулоновскую эффективность и циклируемость. Для дополнительной оценки эффективности предложенного подхода полученные результаты были сопоставлены со статистическими параметрами моделей, полученных в более ранних работах [9, 10].

Список литературы

1. *Chen C., Zuo Y., Ye W., Li X., Deng Z., Ong S. P.* // *Adv. Energy Mater.* 2020. V. 10 (8). P. 1903242–1903242.
2. *Graser J., Kauwe S. K., Sparks T. D.* // *Chem. Mater.* 2018. V. 30 (11). P. 3601–3612.
3. *Torrise S. B., Carbone M. R., Rohr B. A., Montoya J. H., Ha Y., Yano J., Suram S. K., Hung L.* // *npj Computational Materials.* 2020. V. 6(1). P. 109–120.
4. *Wodo O., Broderick S., Rajan K.* // *MRS Bulletin.* 2016. V. 41(8). P. 603–609.
5. *Young S. R., Maksov A., Ziatdinov M., Cao Y., Burch M., Balachandran J., Li L., Somnath S., Patton R. M., Kalinin S. V., Vasudevan R. K.* // *J. Applied Physics.* 2018. V. 123 (11). P. 115303-1–115303-11.
6. *Little R. J., Rubin D. B.* *Statistical analysis with missing data.* V. 793. John Wiley & Sons, 2019.
7. *Schneider G.* // *Nature Rev. Drug Discov.* 2010 V. 9(4). P. 273–276.
8. *Jafrasteh B., Hernandez-Lobato D., Lubian-Lopez S.P., Benavente-Fernandez I.* // *Knowledge-Based Systems.* 2023. V. 273 (3). P. 110603–110615.
9. *Kireeva N., Pervov V.S.* // *Batteries & Supercaps.* 2020. V. 3. P. 427–438.
10. *Kireeva N., Pervov V.S., Tsivadze A.Yu.* // Submitted to *Comput. Mater. Science.*