情報科学類専門語学 B

ソフトウェアサイエンス主専攻　201811333　楠田真大

指導教員　五十嵐康彦

提出日　2022 年 2 月 4 日

In the field of materials science, artificial intelligence is the key to finding attractive materials because of its high predictive performance[Butler et al., 2018]. The number of possible combinations for predicting polymeric materials is astronomical, making it difficult to make accurate predictions using simulation data or experimental data[Audus and de Pablo, 2017]. Thus, the development of AI that learns and predicts from those data is highly desirable to apply it to many practical problems involving polymeric materials. In a previous study, conductivity is regressed by Gaussian process regression based on features made from SMILES, which is a string of chemical formulas, via a graph neural network[Hatakeyama-Sato et al., 2020]. It has been reported that the features of the graph neural network used in this study can improve the prediction accuracy of performance values compared to conventional machine learning[Hatakeyama-Sato et al., 2020]. However, higher prediction accuracy is desired, and it has not been investigated what kind of neural network is the best learning method for the various graph neural networks that have been reported in recent years.

In this research, we aim to construct a model with higher prediction accuracy by examining graph neural networks and regression methods for the features obtained by them. We divide the experiment into two parts and aim to improve the prediction accuracy of conductivity based on the changes in each part: the first part is to create features from chemical formulas of compounds via graph neural networks. The number of data used for pre-training of the graph neural network, the optimization function, and the algorithm of the graph neural network is changed to create features that are more likely to contribute to prediction. The second part is to perform regression from the data obtained by combining the features created by the graph neural network with the features of the compounds. The regression algorithm is changed from Gaussian process regression to random forest and decision tree to achieve more accurate prediction. As a result, using graph convolutional neural network, in which uses different learnable parameters unlike the gated graph neural network which uses single learnable parameter through layers, and random forest regressor, we have improved the accuracy of predicting the physical properties of the target substance. We found descriptor made with graph convolutional neural network was 6% better than it with gated graph neural network in contribution to prediction. In the future, it will be necessary to proceed with the prediction of chemical functionalities other than conductivity and the analysis of the target material data, and to map them to graph neural networks suitable for the characteristics of each learning task.

参考文献

[Audus and de Pablo, 2017] Audus, D. J. and de Pablo, J. J. (2017). Polymer informatics: Opportunities and challenges. *ACS macro letters*, 6(10):1078–1082.

[Butler et al., 2018] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715):547–555.

[Hatakeyama-Sato et al., 2020] Hatakeyama-Sato, K., Tezuka, T., Umeki, M., and Oyaizu, K. (2020). Ai-assisted exploration of superionic glass-type li+ conductors with aromatic structures. *Journal of the American Chemical Society*, 142(7):3301–3305.