Identifying important molecular fragments for property predictions by graph neural networks with explainable AI

Jan G. Rittig,a Dominik Goldstein,a Manuel Dahmen,b Alexander Mitsosc,a,b

aRWTH Aachen University, Process Systems Engineering (AVT.SVT), Aachen, Germany

bForschungszentrum Jülich GmbH, Institute for Energy and Climate Research IEK-10: Energy Systems Engineering, Jülich, Germany

cJARA-CSD, Aachen, Germany

amitsos@alum.mit.edu

Abstract

Graph neural networks (GNNs) have shown great potential for predicting molecular properties. We herein utilize explainable artificial intelligence (XAI) methods to identify molecular fragments, e.g., functional groups, decisive for the GNN prediction of a particular property. We consider both pure component and mixture properties. We systematically search for molecular fragments that are frequently marked as important. We find that the identified molecular fragments comply with available chemical knowledge, therefore providing valuable insights into molecular structure-property relationships.

**Keywords**: machine learning, molecular modeling, structure-property relationships

* 1. Introduction

Explainable artificial intelligence (XAI) provides promising methods for extracting chemical knowledge from molecular property data. Machine learning methods, particularly graph neural networks (GNNs), have recently been applied for predicting a variety of molecular properties that are relevant to process systems engineering such as fuel ignition qualities (Schweidtmann et al., 2020), activity coefficients (Sanchez Medina et al., 2023; Rittig et al., 2023), and solubility (Vermeire et al., 2022). While the developed GNNs achieve high prediction accuracies, they typically do not provide explanations for the predicted property values due to their black-box characteristic. Therefore, gaining insight into property predictions made by GNNs is of high interest and is actively investigated, cf. overviews in (Yuan et al, 2022; Wellawatte et al., 2023), with new explainability methods emerging from the field of XAI. However, the focus in molecular applications to date has primarily been on explaining predicted property values for individual molecules, e.g., which molecular fragment of a specific molecule is most influential to the corresponding prediction. Herein, we are also interested in molecular fragments that are frequently marked as important in a diverse collection of molecules. Thereby, we identify generalizable relations of the molecular structure to a property, i.e., structure-property relationships, from molecular property predictions by GNNs.

* 1. Explainable AI for Graph Neural Networks

GNNs learn properties directly from a graph representation of molecules, with atoms as nodes and bonds as edges, by encoding the molecular graph through convolutional layers into a molecular fingerprint vector which is then mapped to the property of interest. To find explanations for GNN predictions, we herein investigate two recently introduced XAI methods, the GNNExplainer (Ying et al, 2019) and the Molecular Model Agnostic Counterfactual Explanations (MMACE) method (Wellawatte et al, 2022). The GNNExplainer marks atoms and bonds that are influential for the prediction. The MMACE identifies minimal structural changes that exert a large impact on the prediction. To systematically identify structure-property relationships, we apply the two XAI methods to the molecules in the respective training and testing data sets of the GNN and search for frequently occurring molecular fragments that have a high impact on the predicted property values. Specifically, we aim at the explainability of two of our previously developed GNNs. First, we investigate a GNN for predicting the research octane number (RON) of pure components (Schweidtmann et al., 2020), i.e., a well-established measure for rating the knock resistance of a fuel. Secondly, we consider a GNN for mixture property prediction by the example of activity coefficients of solutes in ionic liquids (Rittig et al., 2023).

* 1. Results and Discussion

Fig. 1 shows the top molecular fragments identified by the MMACE. On the left, the most occuring molecular fragments are illustrated that lead to increases in RON when added (green) or removed (red) to a fuel molecule. Analogously, on the right, the molecular fragments are shown that result in a higher activity coefficient in ionic liquids on addition to or removal from a solute. In both cases similar structure-property relationships are identified by the GNNExplainer method (not shown here).

For the knock resistance (Fig. 1, left), the MMACE suggests that decreasing the carbon chain length results in a higher RON, which complies with chemical intuition. Likewise, the addition of hydroxyl and carbonyl groups is identified to yield a higher RON. Both groups are well-known to exert strong effects on fuel autoignition and thus knock resistance.

*Figure 1. Molecular fragments identified by MMACE resulting in higher predicted values for the knock resistance (left) and the activity coefficient in ionic liquids (right).*

In case of the activity coefficient (Fig. 1, right), the polarity of the molecular fragments added to or removed from the solute is recognized to be of high importance for the GNN predictions. For example, adding non-polar groups, such as methyl and methylene groups, yields higher activity coefficients. Since the polarity is one of the main influences on the type and strength of intermolecular forces (Pfennig, 2004), it is directly related to the activity coefficient and has also been included as input descriptor to activity coefficient prediction models, cf. (Lazzaroni et al., 2005; Sanchez Medina et al., 2023).

We thus find that XAI allows us to identify molecular fragments that provide meaningful chemical insights into structure-property relationships in case of both applications.

* 1. Conclusion

We utilize XAI methods to extract chemical insights about structure-property relationships from GNNs for molecular property prediction, herein, for research octane number and activity coefficient prediction. We find that the identified molecular fragments comply with available structure-property knowledge. For example, hydroxyl and carbonyl groups as well as increasing lengths of carbons chains are known to exert strong effects on fuel knock resistance. Likewise, for the activity coefficient predictions, the polarity of substructures is known to have a high importance. The systematic explanation of structure-property relationships learned by GNNs thus successfully relates to chemical knowledge and provides promising insights for future investigations. As a next step, it would be interesting to consider interactions of multiple molecular fragments.

References

M. D. Boot, M. Tian, E. J. Hensen, & S. M. Sarathy, 2017, Impact of fuel molecular structure on auto-ignition behavior – Design rules for future high performance gasolines. Prog. Energy Combust. Sci., 60, 1-25.

M. J. Lazzaroni, D. Bush, C. A. Eckert, T. C. Frank, S. Gupta, & J. D. Olson, 2005, Revision of MOSCED parameters and extension to solid solubility calculations. Ind. Eng. Chem. Res., 44(11), 4075–4083.

F. vom Lehn, B. Brosius, R. Broda, L. Cai, & H. Pitsch, 2020, Using machine learning with target-specific feature sets for structure-property relationship modeling of octane numbers and octane sensitivity. Fuel, 281, 118772.

A. Pfennig, 2004, Thermodynamik der Gemische, Springer Berlin, Heidelberg, Germany.

J. G. Rittig, K. Ben Hicham, A. M. Schweidtmann, M. Dahmen, & A. Mitsos, 2023, Graph neural networks for temperature-dependent activity coefficient prediction of solutes in ionic liquids. Comput. Chem. Eng., 171, 108153.

E.I. Sanchez Medina, S. Linke, M. Stoll, & K. Sundmacher, 2023, Gibbs–Helmholtz graph neural network: capturing the temperature dependency of activity coefficients at infinite dilution. Digital Discovery, 2, 781–798.

A.M. Schweidtmann, J. G. Rittig, A König, M. Grohe, A. Mitsos, & M. Dahmen, 2020, Graph neural networks for prediction of fuel ignition quality, Energy Fuels, 34(9), 11395-11407

F. H. Vermeire, Y. Chung, & W. H.Green, 2022, Predicting solubility limits of organic solutes for a wide range of solvents and temperatures, J. Am. Chem. Soc., 144(24), 10785-10797.

G. P. Wellawatte, A. Seshadri, & A. D. White, 2022, Model agnostic generation of counterfactual explanations for molecules. Chem. Sci., 13(13), 3697-3705.

G. P. Wellawatte, H. A. Gandhi, A. Seshadri & A. D. White, 2023, A perspective on explanations of molecular prediction models, J. Chem. Theory Comput., 19(8), 2149-2160.

Z. Ying, D. Bourgeois, J. You, M. Zitnik, & J. Leskovec, 2019, GNNExplainer: Generating explanations for graph neural networks. NeurIPS, 32.

H. Yuan, H. Yu, S. Gui, & S, Ji, 2022, Explainability in graph neural networks: A taxonomic survey. IEEE Trans. Pattern Anal. Mach. Intell., 45(5), 5782-5799.