

Graph convolutional neural network applied to the prediction of normal boiling point[☆]

Chen Qu, Anthony J. Kearsley, Barry I. Schneider, Walid Keyrouz, Thomas C. Allison^{*}

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD, 20899, USA

ARTICLE INFO

Keywords:

Normal boiling point
Machine learning
Graph neural network
Deep learning

ABSTRACT

In this article, we describe training and validation of a machine learning model for the prediction of organic compound normal boiling points. Data are drawn from the experimental literature as captured in the NIST Thermodynamics Research Center (TRC) SOURCE Data Archival System. The machine learning model is based on a graph neural network approach, a methodology that has proven powerful when applied to a variety of chemical problems. Model input is extracted from a 2D sketch of the molecule, making the methodology suitable for rapid prediction of normal boiling points in a wide variety of scenarios. Our final model predicts normal boiling points within 6 K (corresponding to a mean absolute percent error of 1.32%) with sample standard deviation less than 8 K. Additionally, we found that our model robustly identifies errors in the input data set during the model training phase, thereby further motivating the utility of systematic data exploration approaches for data-related efforts.

1. Introduction

Determination of the normal boiling point has an interesting history dating back hundreds of years [1]. Predictive models for normal boiling point were developed as early as 1842 by Kopp who studied the correlations of molecular volume to various chemical and physical properties and noted a 19 °C increment in temperature per carbon group in a homologous series of hydrocarbons [2–5]. In 1905, Young summarized and extended the work of the previous century, and proposed a more complicated formula for the boiling point increment for a series of related compounds [6]. In 1937, Aten noted some of the difficulties of Young's approach and reverted to an expression published by Boggio-Lera in 1899 that demonstrated a linear relationship between the square of the boiling point of a compound in a homologous series and the number of CH₂ groups [7,8].

In 1947, Wiener published a seminal paper introducing a “path number,” now known as the Wiener index, to aid in the determination of boiling points for alkanes [9]. In creating his index, Wiener leveraged the work of Taylor, Pignocco, and Rossini who had previously reported predictions of boiling points of alkane and alkene substances [10]. Structural indices have been widely applied and extended in property prediction. For example, the “topological index” defined by Hosoya was

cited as correlating particularly well with the boiling point [11].

More recent work takes advantage of a larger database of measured boiling points and makes predictions for a wider variety of chemicals. Notably, Stein and Brown used a database of 4,426 organic chemical compounds to fit a group contribution model with a mean absolute error of 15.5 K (3.2%) [12]. When their model was tested against an additional 6,584 compounds not used in fitting, the error increased to 20.4 K (4.3%). This model was incorporated into the United States Environmental Protection Agency's (EPA) EPI Suite program that predicts a number of properties for a variety of compounds using molecular structure information [13].

Quantitative Structure-Property Relationship (QSPR) models have been used extensively in making predictions of a wide range of chemical properties including the normal boiling point. Some of the early work in this area has been reviewed by Dearden, so we will cite only a relatively small number of articles of interest here [14]. Toropov and coworkers used an atomic orbital graph basis to fit a one-parameter equation and predict normal boiling points for haloalkanes with a standard error of 9.2 °C. Duchowicz, Castro, and Toropov used a correlation weighting scheme to predict normal boiling points for a set of carbonyl compounds [15]. Ribeiro and Ferreira reported QSPR predictions of boiling point, octanol–water partition coefficient, and retention index for a series of

[☆] Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.

^{*} Corresponding author.

E-mail addresses: chen.qu@nist.gov (C. Qu), anthony.kearsley@nist.gov (A.J. Kearsley), barry.schneider@nist.gov (B.I. Schneider), walid.keyrouz@nist.gov (W. Keyrouz), thomas.allison@nist.gov (T.C. Allison).

<https://doi.org/10.1016/j.jmglm.2022.108149>

Received 30 September 2021; Received in revised form 19 January 2022; Accepted 2 February 2022

Available online 4 February 2022

1093-3263/Published by Elsevier Inc.

polycyclic aromatic compounds (PAHs) [16]. González, Toropov, and coworkers employed atomic orbital graphs in a study of 200 diverse organic compounds [17]. The standard errors for their models were large (35 K–60 K), presumably due to the diversity of molecules within their training and testing sets. Ghavami et al. studied a set of 394 compounds, using topological indices to predict normal boiling points, obtaining a root-mean-square (RMS) error of 6.17 K for their best model [18]. Toropov and coworkers employed QSPR based on the 66 features from the molecular SMILES representation to predict normal boiling points for 90 cyclic and acyclic hydrocarbons, reporting a standard error of 11.8 °C [19]. Yi-min et al. used an electronegativity topological descriptor to predict normal boiling points for a set of 215 organic compounds, reporting standard errors of about 7 K for selected compounds within their testing set [20]. Saaidpur et al. used a QSPR model based on descriptors derived from the results of semiempirical quantum chemistry calculations employing the AM1 model [21]. Their set consisted of 216 liquid amine compounds which was divided into sets of primary, secondary, and tertiary amines, yielding standard errors of 11 K–17 K for their QSPR models. Arjmand and Shafiei studied a set of 227 alcohols and phenols using a variety of topological indices, with their best model yielding a standard error of 7.16 K [22].

One of the first applications of a neural network was reported by Cherqaoui and Villemin who studied a series of alkanes up to decane [23]. Their model was able to predict the boiling point of these compounds with a mean absolute error of 3.09 K and a standard deviation of 1.99 K. Their analysis showed superior performance to contemporary QSPR models, with a slightly larger correlation coefficient and a substantially smaller standard deviation. Goll and Jurs reported the use of multilinear regression, neural networks, and a genetic algorithm to predict boiling points for 104 compounds [24]. Their best neural network model had an RMS error of 11.72 K, while their genetic algorithm result has an RMS error of 8.59 K. Gharagheizi et al. employed QSPR with a feed-forward neural network to predict normal boiling points for a set of 17,768 chemical compounds, reporting an RMS error of 21 K in their test set [25]. Their article contains an interesting review of historical and contemporary QSPR modeling studies of normal boiling point. Jin and Bai studied 240 acyclic oxygen-containing organic compounds using a radial basis function artificial neural network approach, obtaining an RMS error of 3.93 K for their testing set [26]. A subsequent study by the same authors using a similar approach on 432 oxygen-containing organic molecules gave an RMS error of 9.45 K [27]. These authors also studied normal boiling points of a set of 168 hydroxyl compounds using their radial basis network approach, finding an RMS error of 5.33 K [28]. Zang et al. studied a wide variety of compounds using a support vector regression model and obtaining an RMS error of 19.72 K. Fissa et al. studied pure hydrocarbon compounds with an artificial neural network model and obtained an RMS error value of 1.37 K for 44 compounds after training on a set of 179 compounds [29]. They noted that the performance of their neural network model was superior to that of a related multilinear regression model. Groven et al. used molecular simulation and an artificial neural network approach to predict boiling points and critical points for a small set of PAH compounds, obtaining a mean absolute error of 26.85 K with a standard deviation of 13.55 K [30]. Finally, we note a recent study of normal boiling points of 160 compounds that explores a number of machine learning methods [31].

Normal boiling points may also be determined using first principles calculations. Mewes and Smits used density functional theory combined with thermodynamic integration and thermodynamic perturbation theory to predict boiling points for atomic liquids, reporting accuracy within a few percent of the best available data [32].

Today, the normal boiling point is routinely used as a measure of purity of chemical substances sold as reagents or after synthesis procedures. Prediction of normal boiling points using a variety of methods continues to be of significant interest.

In this article, we describe the application of a graph neural network

(GNN) model to the prediction of experimental normal boiling point data. We have previously found that a graph neural network model performs well in predicting Kováts retention indices [33]. Given the linear relationship between retention times and boiling points, we expect that the same model will perform well on the data set described herein. Further, we feel that in order to produce models with greater accuracy than those reported to date, new approaches, such as the methodology described here, are needed. The data used in this study was extracted from the literature as captured by the NIST Thermodynamics Research Center (TRC) SOURCE Data Archival System [34,35]. This database contains more than 12,000 experimental determinations of the normal boiling point for a wide variety of molecules, providing a robust data set for training a new predictive model.

2. Data and methods

2.1. Data preparation

The TRC data was organized into a set of molecules and a set of property values. The molecule set was processed prior to consideration of the normal boiling point values to ensure that the data set can adequately represent the chemical functionalities and normal boiling point data with the goal of ensuring a robust model. Each of the filters described below is designed to ensure that the data set contains a sufficient number of molecules so that the training procedure is able to learn from a larger number of molecules as opposed to fitting many disparate cases. The number of occurrences of a given atom type in the set of molecules was counted. If a particular atom occurred in less than 100 molecules, then those molecules were excluded from the set. Molecules containing one or more of the atoms C, H, O, N, Cl, F, Br, S, Si, P, and I were retained. Monoatomic compounds were removed as were molecules that did not contain a C atom (as there were very few of these). Application of these criteria removed fewer than 100 molecules from the set. Molecules with mass less than 50 amu or greater than 600 amu were similarly removed from the set. This filter removed 122 molecules. The molecules were supplied in a 2D MolFile format. This format includes fields that describe features of atoms and bonds, e.g. a stereocenter label, bond order. Importantly, this format does not contain any 3D information about the molecules from which property data such as bond distances might be derived. The bonding geometric information is thus limited to whether two atoms are bonded and what type of bond exists between them. We removed molecules with feature codes that occurred infrequently in the MolFile. From the original set of 12,679 molecules, 11,057 were retained (note: many of these did not have normal boiling point data).

Next, we processed the normal boiling point data. We started with a set of 22,935 data points taken from a large number of literature sources (as captured in the NIST/TRC SOURCE Data Archival System) [34,35]. In many cases there were multiple measurements of the normal boiling point for a single molecule. When we matched the data to the set of molecules described in the previous section, we were left with 3,876 molecules with 20,544 determinations of the normal boiling temperature. For molecules with a single determination of the normal boiling point, the single value was used; otherwise, the normal boiling point was determined as the mean of the set of values. For sets with three or more data points, the Grubbs 2-tail outlier test was applied. In this test, the datum with the largest deviation from the mean is identified and used to compute the quantity

$$G = \max_i \left| \frac{Y_i - \bar{Y}}{s} \right|, \quad (1)$$

where Y_i is an element of the data set with mean \bar{Y} and s is the sample standard deviation of the data set. If the quantity G satisfies the relationship

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{(t_{\alpha/(2N), N-2})^2}{N-2 + t_{\alpha/(2N), N-2}}} \quad (2)$$

where N is the number of points in the data set and $t_{\alpha/(2N), N-2}$ is the critical value of the t distribution with $(N-2)$ degrees of freedom and a significance level of $\alpha/(2N)$ with $\alpha = 0.05$, then the data point is rejected from the set. Although it is possible to apply this test more than once to a set of data, we did not do so. This procedure removed outlier data points from 403 molecules. Finally, 22 compounds with boiling point values less than 200 K and greater than 700 K were removed from the set as they were not representative of the data set. An additional 4 compounds with normal boiling temperatures suspected to be in error, as discussed in the next section, were also removed from the data set, leaving a set of 3,850 molecules for training and testing the machine learning model. A histogram plot showing the distribution of boiling points in the final data set is presented in Fig. 1.

2.2. Machine learning model

Our machine learning model is based on the materials graph network (MEGNet) approach developed by Chen et al. [36,37] This model has been tested for a variety of chemical properties on both molecular and crystalline systems. We have previously used this approach to predict Kováts retention indices with excellent results [33]. The MEGNet methodology incorporates a graph network architecture that captures molecular structure in a very natural way, providing a powerful framework for machine learning of chemical properties. In a graph neural network (GNN), the atomic centers correspond to vertices in the graph, and chemical bonds correspond to graph edges.

The input data to the model is obtained from a 2D MolFile representation of the molecule. This format contains information about the atoms and about the chemical bonds, but does not provide any 3D structural information. As our goal is to facilitate rapid prediction of normal boiling point, using MolFile input that is easily created using chemical structure drawing software or obtained via name to structure conversion allows us to develop an efficient workflow for property prediction.

The MEGNet methodology captures molecular information at the level of atoms, chemical bonds, and whole molecule, with the chemical structure captured in the structure of the graph representation. At the atom level, information about the nuclear charge was included. At the bond level, bond order information was used. Finally, global properties such as the molecular mass were added as inputs to the model.

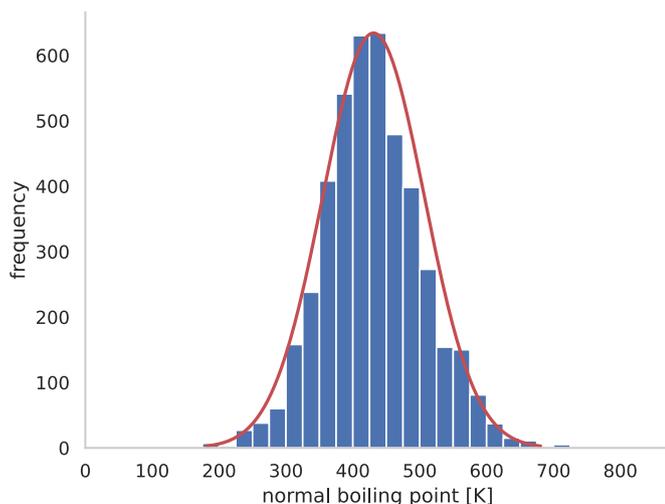


Fig. 1. Histogram showing the distribution of the normal boiling points used in this study.

Our MEGNet model incorporates 3 atom features (encoded as 18 one-hot variables), 3 edge features (encoded as 7 one-hot variables), and 3 global features. The atom features are the atomic number (11 one-hot variables representing C, H, O, N, F, Si, P, S, Cl, Br, and I), the hybridization of the atom (6 one-hot variables, s , sp , sp^2 , sp^3 , sp^3d , and sp^3d^2), and the formal charge (1 integer variable) on the atom. The hybridization is calculated from the information in the MolFile using RDKit. The edge features are the bond type (i.e., no bond, single, double, triple, or aromatic, 5 one-hot variables), whether the atoms were in the same ring (a single 0/1 integer variable), and a graph distance (1 integer variable). The graph distance is calculated as the smallest number of edges that have chemical bonds between the atoms in the pair. Recall that an edge here refers to any pair of atoms, not just those that are formally chemically bonded. However, our model does not encode the complete graph of the molecule as we do not consider pairs of atoms with a graph distance higher than 5 edges, leading to a considerable reduction in the memory required to train the model. Global features are the number of heavy (non-hydrogen) atoms in the molecule, the molecular mass divided by the number of heavy atoms, and the number of chemical bonds divided by the number of heavy atoms. Importantly, only 2D representations of the molecule are needed for the present model.

The features used in the model were selected through a trial and error process, i.e. the initial model contained a number of features that were removed when it was found that they did not significantly improve the performance of the model. For example, a feature that encoded ring size was eliminated when it was found that a single bond feature indicating whether the bond was in a ring produced results of similar quality. Other features that have been included in MEGNet models, stereochemical center information and bond lengths, were not included as there was insufficient information to include these features since the MolFile contained a 2D representation of the molecule (as opposed to some equilibrium 3D geometry) and stereochemistry was not consistently labeled because this information is not frequently included in the source literature from which the compound data was taken.

Our MEGNet model was used with 3 MEGNet blocks. These blocks are composed of two layers of densely-connected multilayer perceptrons (MLP) and a graph neural network layer in which each of the attributes is successively updated. The dense layers used 64 and 32 units, respectively. The MEGNet block steps are followed by a ‘Set2Set’ layer in which the output of the bond and atom attributes are mapped to the appropriate vector quantities. This is followed by a concatenation step and two densely-connected MLPs (32 and 16 units) as we used previously [33].

2.3. Model training and validation

Training was carried out for up to 2,000 epochs, with early stopping employed with restoration of the best weights if the value of the loss function for the validation set did not improve for 300 epochs. These limits were found to be reasonable as a result of empirical testing during the training phase of the project. The average number of epochs in training was approximately 1,100. A batch size of 32 was used during training. The Adam optimizer was used in fitting the model, with a learning rate of 2×10^{-4} . The mean absolute error (MAE) was used as the loss function. The rectified linear unit (ReLU) activation function was used in the MLP layers. The experimental values of the normal boiling point taken from the TRC library were used as target values. These were normalized by converting them to a z-score

$$BP_{\text{norm}} = \frac{BP - \mu}{s} \quad (3)$$

where BP represents the normal boiling point, μ is the mean of the boiling point data set, and s is its sample standard deviation.

In order to facilitate the eventual model validation, the data set was divided into 10 equally-sized “folds” of randomly selected data. During

model training, 80% of the data was used as the training set with an additional 10% of the data used as a validation set. Briefly, the validation set is used to update the parameters of the machine learning model during training and is helpful in avoiding issues with over fitting. The remaining 10% of the data is used as a testing set.

The training history for a prototypical training run is depicted in Fig. 2 where the mean absolute and mean squared errors are depicted for the training and validation sets. In this figure, it is evident that over fitting starts to become significant after about 200 epochs. The training error continues to fall, while the error in the validation set remains approximately constant. This is a sign that our model has reached the limits of its ability to predict new structures more accurately. It is also clear in the figure that there is considerable noise in the validation

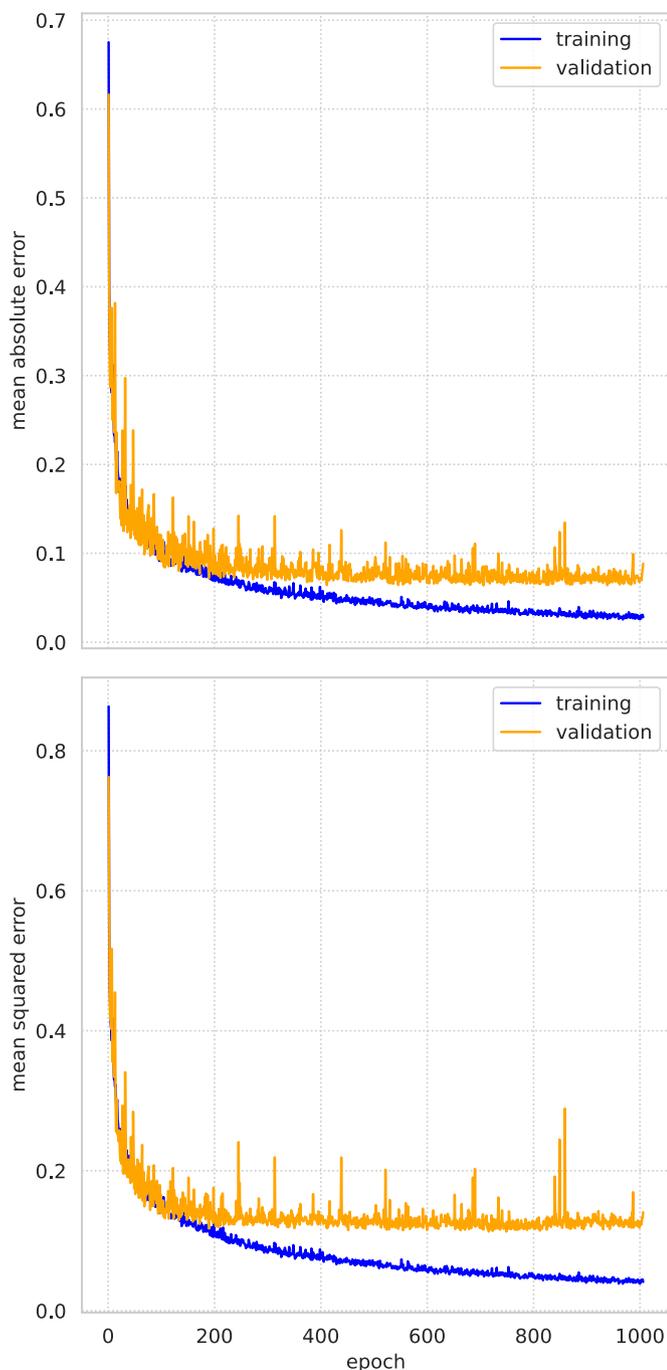


Fig. 2. Plots of the MAE and MSE values for the training and validation data sets during training of the neural network model.

history. This seems to be due to the relatively small size of the validation set (385 molecules).

The model performance was assessed using a 10-fold cross validation procedure. In this scheme, the data set is trained 10 times. In each training run, different validation and testing sets are used. In this way, the entire data set is used to test the predictive power of the machine learning model. This procedure has been shown to be a robust measure of model performance. Final statistics for the model are computed as averages of the corresponding individual statistics of the 10 testing sets.

3. Results and discussion

The performance of our GNN model is shown in Fig. 3. In the plots, the red shaded line has a width of $2s$ (where s is the sample standard deviation of the absolute difference between the predicted and experimental value of the normal boiling point). The plots clearly show that the error is larger in the validation and testing sets than in the training set as expected. The errors in the validation and testing sets are similar, however. This is an indication that the validation set was effective during training, resulting in a model with good predictive capabilities.

The error performance of the machine learning model is depicted in Fig. 4. Again it is seen that the error in the training set is smaller, whereas the validation and testing sets exhibit longer “tails” in their error distributions. The fact that the latter error distributions are so similar is an indication that the model is performing well.

In order to understand the performance of the model more clearly, the details of the 10-fold cross validation procedure are given in Table 1. In this table, the mean absolute error and the sample standard deviation of the absolute errors are given (units are K). By examining the mean errors of the testing set, it is immediately evident that our model has an overall error in the 5 K–6 K range with a slightly larger sample standard deviation. The errors are considerably larger than those of the training set, an indication of over fitting in the model. The overall statistics of our model are shown in Table 2, with a mean absolute error of 5.77 K and a sample standard deviation of 7.81 K.

To further demonstrate the performance of the model, we examine detailed statistics for a single training run as shown in Table 3. It is seen that the model can make predictions that have significant errors ($\approx \pm 60$ K). However, the median absolute error of 3.17 K indicates that these outliers are few in number. The mean absolute percentage error for our model is 1.32% with a sample standard deviation of 1.95%, another indication that the model is working well.

We also characterize the performance of our model by comparison to the method of Stein and Brown as implemented in the EPI Suite package [12,13]. The mean absolute error of the Stein and Brown method over all compounds in our data set is 11.84 K with a median error of 7.99 K and a standard deviation of 12.84 K. These values are in line with the accuracy of the experimental data reported by Stein and Brown and demonstrate that our model is performing about a factor of two better.

During model training, we noticed that predictions for certain molecules consistently exhibited large errors. We took note of these and went back to the source literature. This procedure identified a number of errors in the input data set, i.e. we found our model to be useful in detecting various types of literature and data abstraction errors. In some cases, there was not enough information to unambiguously identify a datum as “bad,” and these were retained in the data set.

Our experience with repeated cycles of training and tracking molecules with consistently large errors, combined with the lack of information (usually another experimental measurement of the normal boiling point) leads us to believe that the performance of our model is ultimately limited by the error present in the input data. As many of the data in the input set were taken from literature sources where the boiling point was determined as a measure of the purity of the final product in a synthesis procedure, it may be expected that the errors will be larger than for a set of experiments specifically designed to measure the normal boiling point with the highest possible accuracy. Though it is certainly

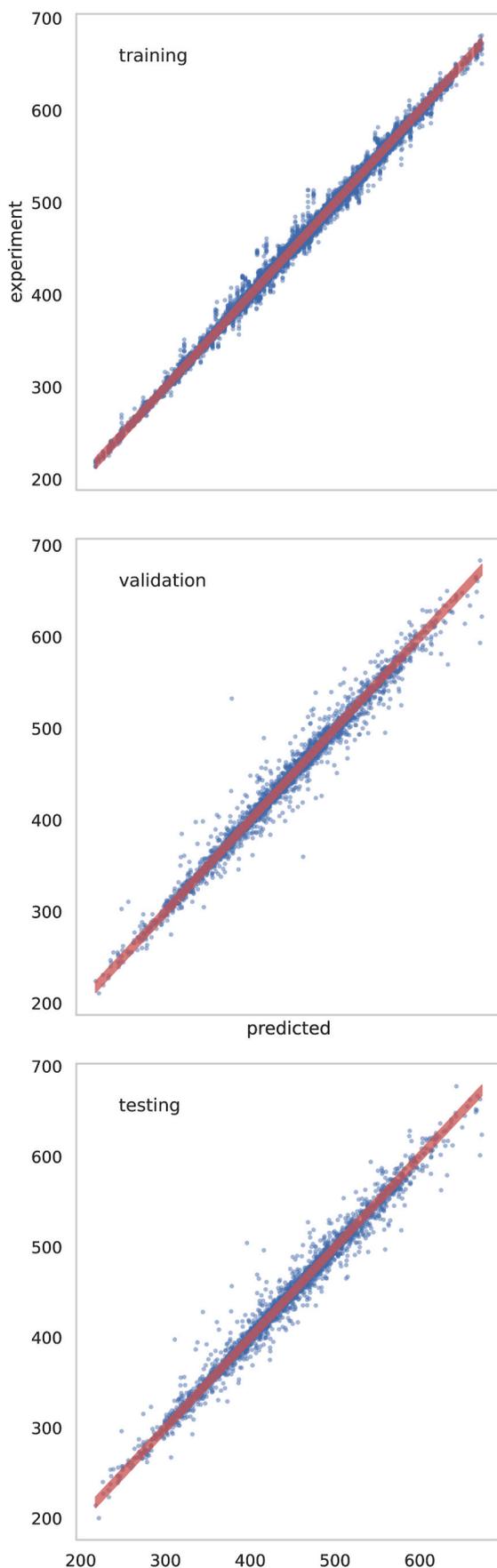


Fig. 3. Plots of experimental versus predicted value of the normal boiling point for the training, validation, and testing data sets.

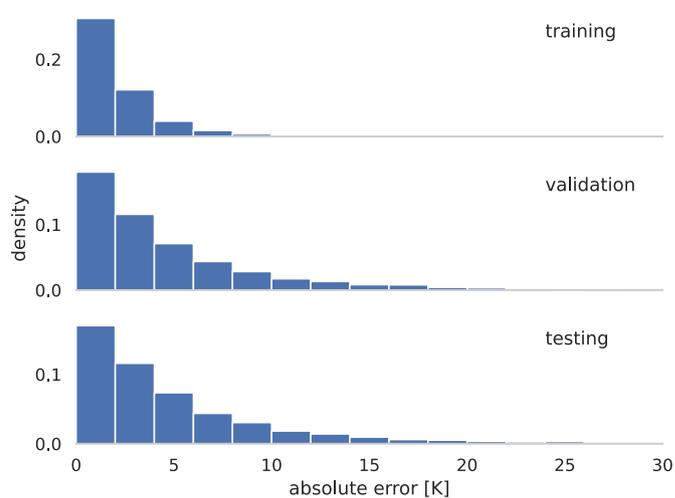


Fig. 4. Histogram plots of the absolute errors for the training, validation, and testing data sets.

Table 1

Detail statistics for each run of the 10-fold cross validation procedure. The mean value and standard deviation of the mean absolute error (MAE) is given for each of the 3 data sets used.

Training		Validation		Testing	
μ	s	μ	s	μ	s
2.55	3.11	5.68	7.70	6.34	9.84
3.06	3.50	6.43	11.15	5.91	7.97
2.23	2.73	4.88	5.21	5.92	7.50
2.08	2.84	5.58	7.33	6.14	8.25
1.85	2.32	6.14	8.51	5.38	6.56
2.53	3.22	5.17	6.49	5.27	6.73
1.64	1.92	5.29	6.74	5.86	8.21
2.81	3.40	5.32	7.90	5.09	7.13
2.20	2.60	5.04	7.27	5.78	8.22
2.01	2.37	6.11	9.51	5.98	7.65

Table 2

Summary statistics of the 10-fold cross validation procedure. The mean value and standard deviation of the mean absolute error (MAE) and the root mean square error (RMSE) over 10 runs is given for each of the 3 sets used.

Set	MAE		RMSE			
Training	2.30	±	0.44	2.80	±	0.51
Validation	5.56	±	0.52	7.78	±	1.66
Testing	5.77	±	0.40	7.81	±	0.94

Table 3

Statistics describing the graph convolutional network model performance in predicting the normal boiling point for the training, validation, and testing sets. The error is calculated as $\varepsilon = BP_{\text{experiment}} - BP_{\text{predicted}}$. The sample standard deviation, s , is calculated for the signed and unsigned errors.

Quantity	Training	Validation	Testing
n	3080	385	385
mean ε	-0.54	-0.58	0.37
RMSE	3.40	8.84	10.05
min ε	27	63	61
max ε	-37	-54	-59
median $ \varepsilon $	1.53	3.00	3.17
mean $ \varepsilon $	2.20	5.04	5.78
$s(\varepsilon)$	2.60	7.28	8.24
median $ \%e $	0.35	0.70	0.78
mean $ \%e $	0.50	1.15	1.32
$s(\%e)$	0.58	1.62	1.94
max $ \%e $	9.28	15.79	19.40

possible that another methodology might produce lower error statistics when applied to the present data set, our experience indicates that we are at or near the limit of model accuracy.

This conclusion is partially supported by the work of Stein and Brown who noted an estimated error in the experimental source data for their model of approximately 12 K that was similar to the error of the training set [12]. The limit on accuracy is also supported by the work of Zang et al. who used an experimental database taken from the EPI Suite collection and noted that the larger error produced by their support vector regression (SVR) model may have been due to experimental errors in the source data [13,38].

In order to improve and test the robustness of our approach, we used a synthetic data technique to expand the sizes of the training, validation, and testing sets by an order of magnitude. We did this by replicating each molecule in the set 10 times, each time scrambling the order of the atoms. As our model is sensitive to the order in which the atoms are input, this provided a critical test of how well predictions could be made when identical structures that differed only in the atomic ordering were input. We found that predictions using varying atomic ordering were different, with a mean span (defined as the difference between the largest and smallest predictions of the normal boiling point) of 1.22 K (standard deviation = 2.82 K) for the testing set. These values were derived from a 10-fold cross-validation procedure as described above. The mean absolute error of the predictions was 5.37 K with a median absolute error of 3.08 K and a standard deviation of 0.36 K. Outliers could be quite large (our largest outlier was an error of 70.81 K), but were rare. Nevertheless, this points to an area in which the current model could be improved to make it insensitive to the order in which atoms are input.

4. Conclusion

In this article, we have described the application of a graph neural network model to fit normal boiling point data for a diverse collection of molecules. Our model reproduces the data well, with mean absolute errors of less than 6 K produced by a 10-fold cross-validation procedure. During our study, we found that the neural network model itself was capable of detecting errors in the source data, and we used this to make corrections to the original data set. Based on our studies, we feel that more accurate models will require a larger set of source data with even better accuracy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the assistance of Dr. Chris Muzny, Dr. Demian Riccardi, and Dr. Eugene Paulechka, all of the NIST Thermodynamics Research Center for providing access to their data set and for assistance in checking and correcting errors in the data.

References

- H. Chang, The myth of the boiling point, *Sci. Prog.* 91 (3) (2008) 219–240, <https://doi.org/10.3184/003685008X360632>, pMID: 18853575. arXiv:<https://doi.org/10.3184/003685008X360632>, doi:10.3184/003685008X360632. URL.
- H. Kopp, Ueber die vorausbestimmung einiger physikalischen eigenschaften bei mehreren reihen organischer verbindungen, *Justus Liebigs Ann. Chem.* 41 (1842) 79–89, <https://doi.org/10.1002/jlac.18420410108>.
- H. Kopp, Ueber die vorausbestimmung einiger physikalischen eigenschaften bei mehreren reihen organischer verbindungen, *Justus Liebigs Ann. Chem.* 41 (1842) 169–189, <https://doi.org/10.1002/jlac.18420410209>.
- H. Kopp, Ueber den zusammenhang zwischen der chemischen constitution und einigen physikalischen eigenschaften bei flüssigen verbindungen, *Justus Liebigs Ann. Chem.* 50 (1844) 71–144, <https://doi.org/10.1002/jlac.18440500105>.
- H. Kopp, Ueber die siedepunkte einiger isomerer verbindungen, und über siedepunktregelmäßigkeiten überhaupt, *Justus Liebigs Ann. Chem.* 55 (1845) 166–200, <https://doi.org/10.1002/jlac.18450550204>.
- S. Young, I. on the boiling-points of homologous compounds, *The Lond. Edinb. Dublin Philos. Mag. J. Sci.* 9 (49) (1905) 1–19, <https://doi.org/10.1080/14786440509463252>, arXiv:<https://doi.org/10.1080/14786440509463252>, doi:10.1080/14786440509463252. URL.
- E. Boggio-Lera, *Gazz. Chim. Ital.* 29 (1) (1899) 441.
- A.H.W. Aten, On the boiling points of organic compounds. i. the formula of boggi-lera, *J. Chem. Phys.* 5 (1937) 260–263, <https://doi.org/10.1063/1.1750016>.
- H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* 69 (1) (1947) 17–20, <https://doi.org/10.1021/ja01193a005>.
- W.J. Taylor, J.M. Pignocco, F.D. Rossini, Method for calculating the properties of hydrocarbons and its application to the refractive indices, densities, and boiling points of the paraffin and monoolefin hydrocarbons, *J. Res. Natl. Bur. Stand.* 34 (1945) 413–434.
- H. Hasoya, Topological index. a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* 44 (9) (1971) 2332–2339, <https://doi.org/10.1246/bcsj.44.2332>.
- S.E. Stein, R.L. Brown, Estimation of normal boiling points from group contributions, *J. Chem. Inf. Comput. Sci.* 34 (1994) 581–587, <https://doi.org/10.1021/ci00019a016>.
- United States Environmental Protection Agency, Washington, DC, USA, *Estimation Programs Interface Suite 4*, 2021, 11.
- J.C. Dearden, Quantitative structure–property relationships for prediction of boiling point, vapor pressure, and melting point, *Environ. Toxicol. Chem.* 22 (8) (2003) 1696–1709.
- P. Duchowicz, E. A. Castro, A. Toropov, Qsqr modeling of normal boiling point of aldehydes, ketones and esters by means of nearest neighboring codes correlation weighting, *Science Direct Working Paper No S1574-0331(04)70226-X*. URL <https://ssrn.com/abstract=2969228>.
- F.A. de Lima Ribeiro, M.M.C. Ferreira, Qsqr models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons, *J. Mol. Struct.-Theochem* 663 (2003), <https://doi.org/10.1016/j.theochem.2003.08.107>, 1–126.
- M.P. González, A.A. Toropov, P.R. Duchowicz, E.A. Castro, Qsqr calculation of normal boiling points of organic molecules based on the use of correlation weighting of atomic orbitals with extended connectivity of zero- and first-order graphs of atomic orbitals, *Molecules* 9 (2004) 1019–1033.
- R. Ghavami, A. Najafi, B. Hemmateenejad, Qsqr studies on normal boiling points and molar refractivities of organic compounds by correlation-ranking-based pcr and pc-ann analyses of new topological indices, *Can. J. Chem.* 87 (2009) 1593–1604, <https://doi.org/10.1139/V09-109>.
- A.A. Toropov, A.P. Toropova, E. Benfenati, Qsqr modelling of normal boiling points and octanol/water partition coefficient for acyclic and cyclic hydrocarbons using smiles-based optimal descriptors, *Cent. Eur. J. Chem.* 8 (5) (2010) 1047–1052, <https://doi.org/10.2478/s11532-010-0072-5>.
- D. Yi-min, Z. Zhi-ping, C. Zhong, Z. Yue-fei, Z. Ju-lan, L. Xun, Prediction of boiling points of organic compounds by qsqr tools, *J. Mol. Graph. Model.* 44 (2013) 113–119, <https://doi.org/10.1016/j.jmkgm.2013.04.007>.
- S. Saaidpour, A. Bahmani, A. Rostami, Prediction the normal boiling points of primary, secondary and tertiary liquid amines from their molecular structure descriptors, *CM* 21 (4) (2015) 201–210, <https://doi.org/10.12921/cmst.2015.21.04.004>.
- F. Arjmand, F. Shafiei, Prediction of the normal boiling points and enthalpy of vaporizations of alcohols and phenols using topological indices, *J. Struct. Chem.* 59 (3) (2018) 748–754, <https://doi.org/10.1134/S0022476618030393>.
- D. Cherqaoui, D. Villemin, Use of a neural network to determine the boiling point of alkanes, *J. Chem. Soc., Faraday Trans.* 90 (1) (1994) 97–102, <https://doi.org/10.1039/FT9949000097>.
- E.S. Goll, P.C. Jurs, Prediction of the normal boiling points of organic compounds from molecular structures with a computational neural network model, *J. Chem. Inf. Comput. Sci.* 39 (1999) 974–983, <https://doi.org/10.1021/ci9900711>.
- F. Gharagheizi, S.A. Mirkhani, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, Determination of the normal boiling point of chemical compounds using a quantitative structure–property relationship strategy: application to a very large dataset, *Fluid Phase Equil.* 354 (2013) 250–258, <https://doi.org/10.1016/j.fluid.2013.06.034>.
- L. Jin, P. Bai, Qsqr study on normal boiling point of acyclic oxygen containing organic compounds by radial basis function artificial neural network, *Chemometr. Intell. Lab. Syst.* 157 (2016) 127–132, <https://doi.org/10.1016/j.chemolab.2016.07.007>.
- P.B. Liangjie Jin, Prediction of the normal boiling point of oxygen containing organic compounds using quantitative structure–property relationship strategy, *Fluid Phase Equil.* 427 (2016) 194–201, <https://doi.org/10.1016/j.fluid.2016.07.015>.
- L. Jin, P. Bai, Modelling of normal boiling points of hydroxyl compounds by radial basis networks, *Mod. Chem.* 4 (2) (2016) 24–29, <https://doi.org/10.11648/j.mc.20160402.12>.
- M.R. Fissa, Y. Lahiouel, L. Khaouane, S. Hanini, Qsqr estimation models of normal boiling point and relative liquid density of pure hydrocarbons using mlr and mlp-ann methods, *J. Mol. Graph. Model.* 87 (2019) 109–120, <https://doi.org/10.1016/j.jmkgm.2018.11.013>.

- [30] J.D. Steven, D. Groven, C. Desgranges, Prediction of the boiling and critical points of polycyclic aromatic hydrocarbons via wang-landau simulations and machine learning, *Fluid Phase Equil.* 484 (2019) 225–231, <https://doi.org/10.1016/j.fluid.2018.11.030>.
- [31] K. Hatakeyama-Sato, K. Oyaizu, A Generative Model for Extrapolation Prediction in Materials Informatics, last access: Sep. 2, 2021, 2021. arXiv:2103.00157. URL, <https://arxiv.org/abs/2103.00157>.
- [32] J.-M. Mewes, O.R. Smits, Accurate elemental boiling points from first principles, *Phys. Chem. Chem. Phys.* 22 (2020) 24041–24050, <https://doi.org/10.1039/d0cp02884c>.
- [33] C. Qu, B.I. Schneider, A.J. Kearsley, W. Keyrouz, T.C. Allison, Predicting kováts retention indices using graph neural networks, *J. Chromatogr. A* 1646 (2021) 462100, doi:10.1016/j.chroma.2021.462100, <https://www.sciencedirect.com/science/article/pii/S0021967321002247>.
- [34] A. Kazakov, C. Muzny, K. Kroenlein, V. Diky, R. Chirico, J. Magee, I. Abdulagatov, M. Frenkel, Nist/trc source data archival system: the next-generation data model for storage of thermophysical properties, *Int. J. Thermophys.* 33 (2012) 22–33, <https://doi.org/10.1007/s10765-011-1107-7>.
- [35] Nist thermodynamics research center, last accessed: Sep. 23, 2021, <https://www.trc.gov/>, 2021.
- [36] C. Chen, W. Ye, Y. Zuo, C. Zheng, S.P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.* 31 (9) (2019) 3564–3572, <https://doi.org/10.1021/acs.chemmater.9b01294>.
- [37] MEGNet: MatErials graph network, last accessed: Sep. 23, 2021 (2020). URL, <https://github.com/materialsvirtuallab/megnet>.
- [38] Q. Zang, K. Mansouri, A.J. Williams, R.S. Judson, D.G. Allen, W.M. Casey, N. C. Kleinstreuer, In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning, *J. Chem. Inf. Model.* 57 (2017) 36–49, <https://doi.org/10.1021/acs.jcim.6b00625>.