Macromolecules

pubs.acs.org/Macromolecules Article

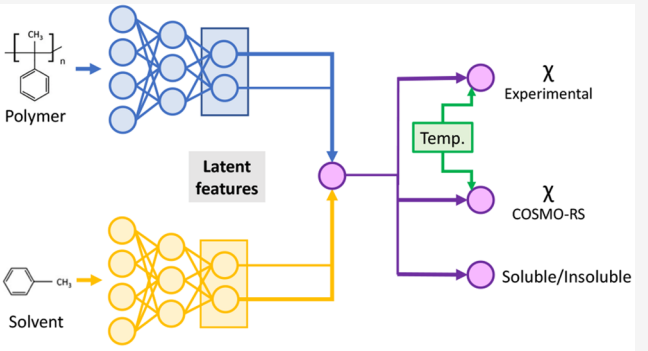
Multitask Machine Learning to Predict Polymer–Solvent Miscibility Using Flory–Huggins Interaction Parameters

Yuta Aoki,[‖] Stephen Wu,[‖] Teruki Tsurimoto, Yoshihiro Hayashi, Shunya Minami, Okubo Tadamichi, Kazuya Shiratori,* and Ryo Yoshida*

Cite This: https://doi.org/10.1021/acs.macromol.2c02600 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predicting and understanding the phase equilibria or phase separation in polymer–solvent solutions represent unresolved fundamental problems in polymer science. The phase behavior and thermodynamics of polymer miscibility depend on the inter- and intramolecular interactions of a polymer with a certain molecular weight distribution mixed with a solvent. Here, we develop a machine-learning framework to achieve highly generalized and robust prediction of Flory–Huggins χ parameters for polymer–solvent solutions. The model was trained using experimentally observed temperature-dependent χ parameters for 1190 samples, comprising 46 unique polymers and 140 solvent species. However, the difficulty was that the data set was quantitatively limited and qualitatively biased owing to technical issues in determining the Flory–Huggins χ parameters. To overcome these limitations, we produced an in-house data set of χ parameters obtained from quantum chemical calculations for thousands of polymer–solvent pairs and a large list of soluble and insoluble polymer–solvent pairs. Using these three data sets, we conducted multitask machine learning that simultaneously performed the "soluble/insoluble" classification and quantitative evaluation of both experimental and calculated χ parameters. Consequently, we obtained a highly generalized model applicable to a wide range of polymer solution spaces. In this paper, the predictive power and physicochemical implications of the model are demonstrated, along with quantitative comparisons with existing methods.



INTRODUCTION

Generally, the dissolution of polymers in a selected solvent occurs during various material development processes, such as plastic recycling,[1,2] polymer synthesis,[3,4] processing,[5−7] purification,[4,8,9] blending,[10,11] and painting and coating.[12,13] However, predicting and understanding the phase equilibria or phase separation of polymer–solvent binary systems represent an unresolved fundamental issue in polymer chemistry. According to the Flory–Huggins polymer solution theory, the thermodynamic properties of polymer solutions, such as miscibility or swelling equilibria, can be expressed in terms of a polymer–solvent interaction parameter called the Flory–Huggins χ parameter.[14−17] For a given value of the χ parameter, the polymer–solvent phase space can be described as a function of the following controllable variables: temperature, volume fractions, and the length of the molecular chain. However, experimentally measuring the χ parameters is technically difficult and costly.[18−20]

Although various models have been developed to computationally predict the χ parameter,[21] an empirical model based on the distance between the polymer and solvent solubility parameters is the most widely used.[22−25] For example, the Hansen solubility parameter (HSP) represents the potential solubility of a given molecule as a three-dimensional vector consisting of dispersion (van der Waals forces), polarity (dipole moments), and hydrogen-bonding components. The polymer–solvent solubility is determined based on the distance between the HSP vectors. Solubility parameters and their derivatives have been experimentally obtained for various molecules.[24,26] For molecules with undetermined solubility parameters, empirical methods such as the group contribution methods have been extensively studied and widely used.[27−29] However, the prediction performance of such methods is empirically known to be quite low. As alternative methods for predicting the solubility, highly parametrized free-volume models, such as UNIFAC-FV[30] and entropic-FV,[31,32] and advanced statistical-associating fluid theory (SAFT) are

Received: December 30, 2022
Revised: May 30, 2023

available.[33,34] However, as with HSP, such methods are not generally applicable because their calculations require experimental or empirical determinations of material-specific parameters. Alternatively, more general-purpose methods based on atomistic simulations can also be applied. The COSMO-RS method based on quantum chemistry has been applied to a wide range of polymer−solvent systems.[35−37] Molecular interactions between segments can be evaluated using quantum chemistry calculations[38] or classical force-field calculations.[39] Solvation free energies can also be calculated by combining molecular dynamics simulations with energy representation methods.[40] However, these atomistic simulations are computationally expensive.

In recent years, with the increasing availability of big data and high-performance computational resources, machine learning has emerged as a promising approach for achieving high-speed predictions. However, it requires a data set of χ parameters to train the model. In this regard, numerous techniques have been proposed to obtain experimental χ parameter values. However, such techniques are technically difficult and costly.[18−20] Moreover, each method has limited applicability: the vapor pressure method[20,41] requires measurement under high vapor pressure; the osmotic pressure measurements[42,43] can only be applied when the polymer and solvent are in a molecularly miscible state; and inverse gas chromatography[44,45] requires that a polymer should be soluble in its solvent that can be evaporated to prepare a uniform film on the silica capillary. Hence, the resulting data are quantitatively limited and distributionally biased toward certain types or states of systems, such as soluble polymers or solvents that can be examined under high vapor pressures. This makes it difficult to create highly generalizable predictive models of χ parameters that can be applied to a wide variety of systems.

Currently, little progress has been made in studying polymer miscibility using machine learning.[46−53] Chandrasekaran et al.[46] derived a binary classifier using a list of soluble and insoluble polymer−solvent pairs recorded in the polymer properties database PoLyInfo.[54,55] However, temperature and processing conditions were neglected because they were missing in the database. Thus, the model cannot be used to draw phase diagrams. Ethier et al.[56] worked on predicting the cloud point of copolymers. To the best of our knowledge, this is the first study that uses machine learning to predict the polymer solution phase space. However, note that the chemical space spanned by the sample set was quite limited as only 97 polymer−solvent pairs, consisting of 21 polymers and 61 solvents, were used for machine learning. Their model was developed primarily to predict changes in the cloud point as a function of the copolymer composition ratio and temperature for a given polymer−solvent pair, and its generalization performance over a large material space is limited. More recently, Nistane et al.[50] presented a Gaussian process regression to predict temperature-dependent χ parameters, which were trained with a total of 1586 data points corresponding to 58 polymers and 140 solvents for varying temperatures. In our view, the applicability domain of their model is limited because the training data were quantitatively limited.

In this paper, we present a machine-learning framework to achieve a highly generalizable prediction of temperature-dependent χ parameters. Our model describes χ parameters as a function of the chemical structure of the polymer and solvent learned using a sample of 1190 experimentally observed χ parameters with 766 unique polymer−solvent pairs, consisting of 46 polymers and 140 solvents.[19] As in previous studies, this data set has limited chemical diversity; furthermore, as noted above, the experimental χ parameters are biased. To overcome these limitations, two auxiliary data sets were created. We extracted an auxiliary data set from PoLyInfo[54,55] to provide a list of 29 777 soluble and insoluble polymer−solvent pairs. In addition, we produced an in-house data set of χ parameters for 9575 polymer−solvent pairs using quantum chemistry calculations with COSMO-RS. Using these three data sets, we conducted multitask machine learning[57,58] based on deep neural networks to simultaneously perform binary classification of polymer miscibility and quantitative prediction of χ parameters in the real and model systems. The input variable for the neural network included a polymer−solvent pair with its chemical structures along with the temperature of mixing. The output layer branches into three different tasks: predicted values for the experimental and calculated χ parameters and a classification probability indicating whether the given polymer−solvent pair is miscible. The mapping from input to output was modeled using a multilayer neural network retaining an analogy with Hansen's solubility sphere. While the HSP distance is defined in the space of the three different forces relevant to dispersion, polarity, and hydrogen bonding, our model was designed to autonomously create generalized, extended solubility spheres by embedding polymer−solvent chemical features into a 10−40-dimensional latent space. The embedded features and their relevant dimensionality were learned autonomously based on the observed data.

The trained model showed considerably good generalization performance. The predictive power surpassed that of the prediction using COSMO-RS and the predictor using Hansen's solubility sphere. The data set of the experimental χ parameters provided limited training samples because the structural diversity of the polymer−solvent species was insufficient. Therefore, within the scope of ordinary single-task machine learning, the applicability domain of the trained model is restricted to the inside or slightly outside of the narrow data distribution. Here, it is shown that by jointly learning with the two additional large data sets, the applicability domain of the model can be successfully expanded. Recently, a similar idea of integrating a large number of COSMO-RS χ parameters in a multitask-learning framework to create a fundamental machine-learning model was proposed to improve the solubility prediction of small molecules.[59] In this study, we created a fundamental model that simultaneously solves three closely related tasks for polymer miscibility.

METHODS

Preliminary. The phase behaviors and miscibility of a binary polymer mixture are determined by the free energy of mixing. According to the Flory−Huggins theory, the change in Gibbs's free energy ΔG_{mix} upon mixing polymer p with solvent s is expressed by the following equation:

$$\frac{\Delta G_{mix}}{kT} = \frac{\Phi_p}{N_p}\log \Phi_p + \frac{\Phi_s}{N_s}\log \Phi_s + \Phi_p\Phi_s\chi \quad (1)$$

where k, T, and Φ_i denote the Boltzmann's constant, absolute temperature, and volume fraction of component $i \in \{p, s\}$, respectively. Further, N_p and N_s denote the lengths of the molecular chains of p and s, respectively. In this study, the reference volume was set as the molecular volume of the solvent, that is, $N_s = 1$. The first two terms represent the change in the combinatorial entropy of p and

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

s, respectively, which is related to the number of possible conformational states in the mixture. The third term includes the Flory–Huggins interaction parameter, χ. This parameter is a critical dimensionless metric that represents the difference in the noncombinatorial entropy and the strength of the pairwise interaction energies between p and s in the mixture. It involves the energy difference between the contact of p–p and s–s and that of p–s in the binary mixture system. As in eq 1, the two combinatorial entropy terms can be calculated given the volume fraction of p and s and the length of the molecular chain. Therefore, to analyze the solution phase behaviors of p and s, we only need to observe or estimate the temperature-dependent value of the χ parameter: the smaller the value of the χ parameter, the smaller ΔG_{mix} and the more likely the mixing of p and s.

For a given χ parameter, the polymer–solvent phase space can be entirely described as a function of the volume fractions, the length of the molecular chain, and the temperature. Various experimental techniques have been studied to determine the χ parameters; however, as mentioned earlier, they are technically difficult and costly.[18,19] Therefore, simple statistical models using solubility parameters have been widely applied to obtain an estimate. For example, with the Hildebrand solubility parameters, δ_p and δ_s for p and s, the approximate value of the χ parameter is obtained by considering their distance as $\hat{\chi}(T) = \frac{V_s}{kT}(\delta_p - \delta_s)^2$, where V_s is the molecular volume of s used as the reference volume.[16] However, by definition, the Hildebrand solubility parameter distance measure is only applicable to nonpolar and non-hydrogen-bonding solvents.[16] An alternative method is to use the three physicochemical parameters of HSP: the dispersion force (d_p, d_s), dipole interaction energy (p_p, p_s), and hydrogen-bonding energy (h_p, h_s). The HSP distance assesses the relative energy difference between p and s as follows:

$$R = \sqrt{4(d_p - d_s)^2 + (p_p - p_s)^2 + (h_p - h_s)^2} \quad (2)$$

With this representation, the solubility is estimated as follows:

$$\text{solubility} = \begin{cases} R < R_p \;\Rightarrow\; \text{soluble} \\ R > R_p \;\Rightarrow\; \text{insoluble} \end{cases} \quad (3)$$

where R_p is the interaction radius of p,[24,28,51,60] which is estimated experimentally or statistically. Furthermore, an estimate of the χ parameter can be obtained as $\hat{\chi}(T) = \frac{V_s}{4kT}R^2$. Here, if the three parameters are closer in the Hansen space, p and s are estimated as being more compatible. According to the group contribution method, (d_i, p_i, h_i) for p or s is estimated based on the number of prescribed atomic groups and their contributions.

Note that the HSP distance merely implies that p and s with similar cohesive energy densities are compatible. In addition to the three cohesive energies, other intermolecular interactions, such as ion-induced dipoles, metallic bonding, and electrostatic interactions, as well as molecular size, are also involved in the determination of solubility. For these reasons, estimation using solubility parameters performs poorly in several practical applications. The proposed deep neural network generalizes the HSP-based predictor to autonomously encode hidden physicochemical features into a higher-dimensional latent space in a completely data-driven manner.

Task and Data. The model input consists of three variables: the chemical structure X_p of the repeat unit of polymer p, chemical structure X_s of solvent s, and temperature t. The primary output variable is the χ parameter $Y_\chi(t)$ at temperature t. Further, experimental values of χ parameters for 766 polymer–solvent pairs were compiled from a supplementary table of Orwoll and Arnold,[19] including 46 unique polymers and 140 solvent molecules (the digitized data are provided as Supplementary Data). The data set also recorded temperatures of mixing. The total sample size was 1190. Hereafter, this data set is denoted by D_χ. It is important to note that the data are distributed over a rather limited region of the overall chemical space. Consequently, the applicability domain of the learned model is restricted to a neighboring region of the data distribution. For example, as discussed above, for an immiscible polymer–solvent system where no single phase appears, experimentally determining χ parameters is rather difficult. Thus, the data distribution is significantly biased.

To overcome the aforementioned barriers, we utilized two additional data sets: the polymer properties database PoLyInfo and an in-house data set of the COSMO-RS calculations. PoLyInfo includes 18 329 and 11 448 samples of soluble and insoluble polymer–solvent pairs, respectively, neglecting structured data of mixing temperatures and other processing conditions (see the Supporting Information). Herein, this data set is denoted as D_c. In this data set, 6455 polymers and 326 solvents in D_c are distributed over a wider chemical space than those in D_χ (Figure S1). Note that D_c includes negative samples of insoluble polymer mixtures. In addition, we produced an in-house data set of χ parameters, denoted as D_s, for 9129 polymer–solvent pairs in D_χ and D_c using quantum chemistry calculations with COSMO-RS (see the Supporting Information). Using these three data sets, we jointly trained the model by mapping from the chemical structures of p and s to the experimental and computational χ parameters and the binary class labels indicating polymer–solvent miscibility. This method is called multitask learning, where different related tasks that share common underlying mechanisms are learned simultaneously using a unified model. Under the limited data supply for the χ parameters in the primary task, multitask learning can boost predictive performance by leveraging and transferring feature representations learned from two auxiliary tasks with sufficient data provided to the target. This idea is beneficial for expanding the applicability domain of the resulting predictive model.

In addition to the polymer–solvent structural bias as exemplified in Figure S2, there is another significant bias in the χ parameter experimental data set. As mentioned in the Introduction section, owing to technical limitations of the experimental measurement methods, a part of the observable χ parameters would be given only for polymer–solvent molecules in a miscible state. We investigated the existence of this bias based on the observed data. Here, we extracted 429 polymer–solvent pairs that were included in both the PoLyInfo soluble/insoluble data set and the χ parameter experimental data set. Figure 1 displays the groupwise distributions of the χ parameters for the soluble and insoluble sample sets. Evidently, the majority in the experimental χ parameter data set consists of soluble samples.

As mentioned above, PoLyInfo did not record structured data on mixing temperatures or processing conditions. Therefore, we constructed a model by treating these conditions as a black box.

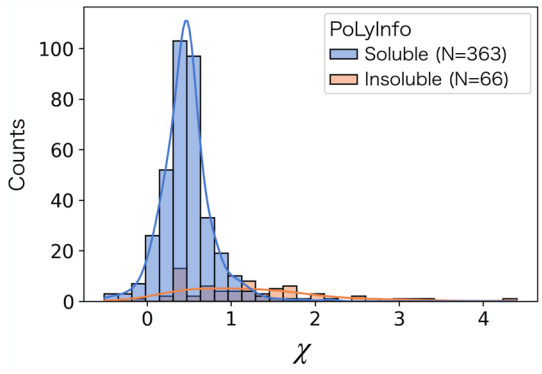


Figure 1. Polymer–solvent pairs in the experimental data set D_χ of the χ parameters were mapped to the soluble/insoluble labels in PoLyInfo, and the groupwise histograms are displayed for the two groups as a function of χ parameter values. The total amount of polymer–solvent pairs for each group is included in the legend, which shows a significant bias to obtain χ parameters for soluble polymer–solvent pairs.

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX–XXX

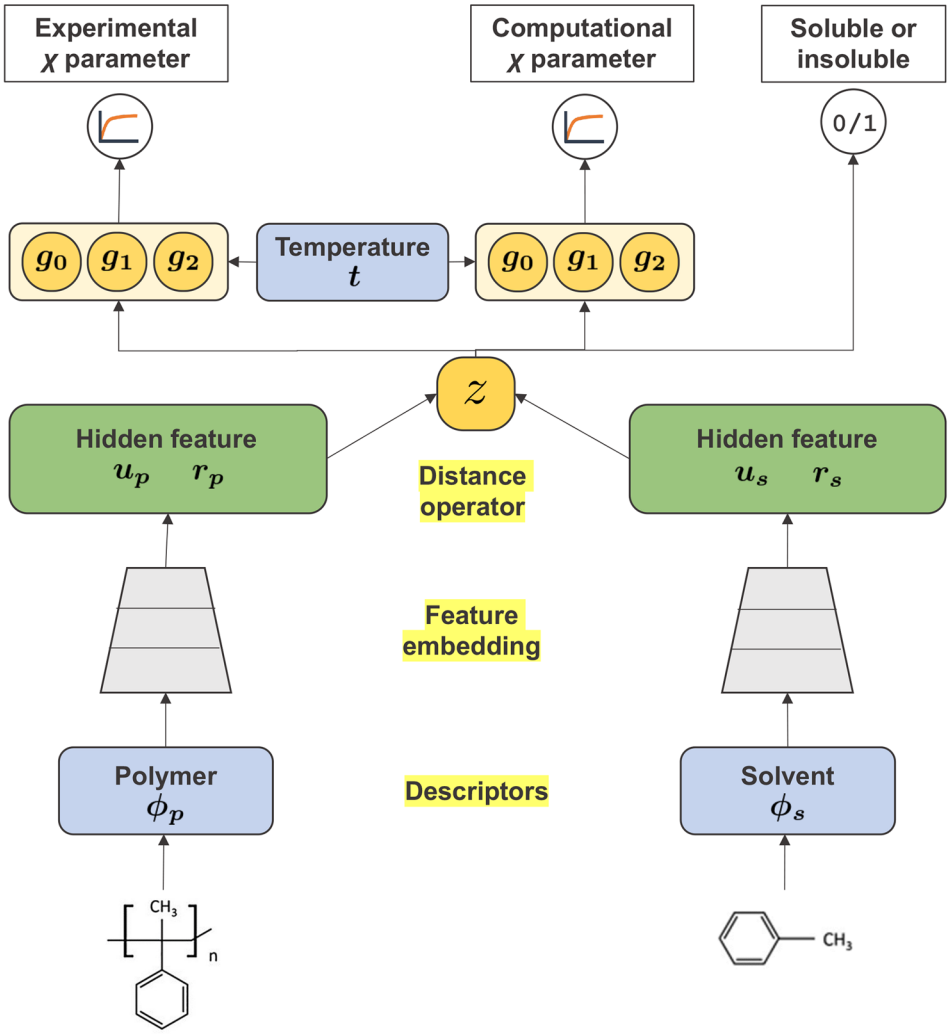


Figure 2. Neural network architecture for the prediction of the Flory–Huggins χ parameter of polymer p exposed to solvent s. Compositional and structural features of p and s are embedded into the latent feature space, and their similarity is decoded back to the temperature-dependent χ parameter with the three coefficient parameters (g_0, g_1, g_2) in the second-order equation in eq 5 through a set of hidden variables z. The decoder also branches out to the prediction of the computational χ parameter and the auxiliary classification task discriminating whether p and s are soluble.

However, additional data in the form of textual annotations about the mixing temperatures were partially provided. We processed this information and classified the partially annotated data set D_c into three categories: room temperature, higher, or lower than room temperature (see the Supporting Information for the frequency distribution). The number of polymer–solvent pairs for which class labels were recorded in the different temperature ranges was 814, of which 207 samples had different labels. Consequently, very few polymers had different class labels at different temperatures; thus, no special treatment for multiple labels was applied.

Here, we also provide a remark on the treatment of molecular weights. Owing to the lack of data, we modeled the experimental χ parameters without including the molecular weight as an input descriptor. However, it will be shown later that our model successfully predicted the χ parameter, which is inherently molecular weight dependent. We discuss this result in the Results and Discussion section.

Modeling. As shown in Figure 2, the neural network architecture defines a mapping from the chemical structure of (p, s) and temperature t to the experimental χ parameter $f_\chi(p, s, t)$, computational χ parameter $f_s(p, s, t)$, and classification probability $f_c(p, s)$ of observing a binary class label that takes a value of one or zero depending on whether p and s are soluble or insoluble. The compositional, structural, and physicochemical features of p and s are encoded into 397-dimensional descriptor vectors, $\phi_p \in \mathbb{R}^{397}$ and $\phi_s \in \mathbb{R}^{397}$, respectively. To represent a polymer, a decametric oligomer was created by linearly connecting its repeat unit 10 times, and a virtual infinite-length polymer was computationally generated by connecting two junction points in the simplified molecular input line entry system (SMILES) string.[61] The descriptor vector was computed based on the cyclic representation by concatenating two different descriptors: the 190-dimensional force-field kernel mean descriptor from the RadonPy library[62] and the 207-dimensional descriptor from the Python library RDKit[63] (for details, see the Supporting Information). The same descriptor set was used for solvent molecules. The vectorized molecules ϕ_p and ϕ_s were mapped into two different k-dimensional latent variables, u_p, $r_p \in \mathbb{R}^k$ and u_s, $r_s \in \mathbb{R}^k$, respectively, by passing through three stacking hidden layers, with each unit comprising an affine transformation layer, batch normalization, and sigmoid activation function.[64] To avoid a scaling issue of batch normalization in multitask learning,[65] we trained the batch normalization layers using only D_c, which contains the most diversified polymer–solvent pairs.

In this study, the number of neurons in each hidden layer was designed to decrease linearly from the input dimension to $k \times 2$. Similar to the three-dimensional HSP vector, the latent variables (u_p,

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX–XXX

r_p) and (u_s, r_s) are modeled as unknown factors that regulate the solubility of p and s. Motivated by the HSP distance, the solubility in a two-component system is modeled by the discrepancy between p and s as follows:

$$z_i = (u_{p,i} - u_{s,i})^2 - r_{p,i}^2 - r_{s,i}^2 \ (i = 1, ..., k) \tag{4}$$

The first term calculates the elementwise squared difference of the k-dimensional latent vectors, u_p and u_s. The second and third terms denote the polymer- and solvent-specific discount factors, respectively, which are analogous to the interaction radius of the HSP sphere. The discount factor determines the range of p or s that is soluble. Polymers that are soluble in a wider variety of solvents have a larger r_p, and solvents that dissolve more diverse polymers have a larger r_s.

The similarity $z = (z_1, ..., z_k)$ on this embedding space branches out to the three separate tasks: predictions of the experimental χ parameter, computational χ parameter, and the binary class label of solubility. In the branching path to the binary classification, z is converted to a scalar value by passing through a fully connected affine layer, and a softmax function converts it into the classification probability $f_c(p, s) \in [0, 1]$ such that p and s are soluble. On each of the other branches to $f_s(p, s, t)$ and $f_\chi(p, s, t)$, z is transformed through a fully connected layer into three latent variables, $g_s = (g_{s,0}, g_{s,1}, g_{s,2})$ and $g_\chi = (g_{\chi,0}, g_{\chi,1}, g_{\chi,2})$, respectively. These variables represent the coefficients in a second-order equation[66,67] for the temperature dependence of $f_\chi(p, s, t)$ as follows:

$$f_\chi(p, s, t) = g_{\chi,0} + \frac{g_{\chi,1}}{t} + \frac{g_{\chi,2}}{t^2} \tag{5}$$

The computational χ parameter $f_s(p, s, t)$ is also modeled in the same form. The first term, $g_{\chi,0}$ or $g_{s,0}$, defines the baseline for the χ parameter, while the coefficients of the first- and second-order terms, ($g_{\chi,1}$, $g_{\chi,2}$) or ($g_{s,1}$, $g_{s,2}$), determine the gradient and curvature of the temperature-dependent curve. When the gradient of eq 5 is negative, the polymer−solvent system becomes more soluble at higher temperatures, resulting in an upper critical solution temperature (UCST) phase diagram with the critical point located above the two-phase region. Conversely, if the first derivative is positive, the phase diagram is of the lower critical solution temperature (LCST) type. Note that, in an application shown later, a linear model excluding the second-order terms was used. This is because the data does not contain enough information to support a higher-order temperature dependence model, which will be discussed in more detail in the Results and Discussion section. However, in the Python code distributed,[68] users can optionally specify both first- and second-order models.

Learning. The model training in multitask learning is formulated as an empirical risk minimization problem as follows:

$$\begin{aligned}\min_{f_\chi, f_s, f_c} & \frac{(1-\lambda_s)}{|D_\chi^{\text{train}}|} \sum_{i \in D_\chi^{\text{train}}} (y_\chi(t_i, p_i, s_i) - f_\chi(t_i, p_i, s_i))^2 \\ & + \frac{\lambda_s}{|D_s^{\text{train}}|} \sum_{i \in D_s^{\text{train}}} (y_s(t_i, p_i, s_i) - f_s(t_i, p_i, s_i))^2 \\ & - \lambda_c \sum_{i \in D_c^{\text{train}}} (y_c(p_i, s_i) \log f_c(p_i, s_i) + (1 - y_c(p_i, s_i)) \\ & \log(1 - f_c(p_i, s_i)))\end{aligned} \tag{6}$$

The first term quantifies the squared loss between an observed experimental χ parameter $y_\chi(t_i, p_i, s_i)$ and the model output $f_\chi(t_i, p_i, s_i)$ for sample i of (t, p, s) in the training set D_χ^{train}. The second term quantifies the squared loss between an observed computational χ parameter $y_\chi(t_i, p_i, s_i)$ and the model output $f_\chi(t_i, p_i, s_i)$ for sample i of (t, p, s) in the training set D_s^{train}. The third term is the logistic loss for evaluating the discrepancy between an observed binary class label $y_c(p_i, s_i) \in \{0, 1\}$ and the classification probability $f_c(p_i, s_i) \in [0, 1]$ for each instance in the training set D_c^{train}. The hyperparameters λ_s and λ_c to be controlled in the validation step determine the importance of the primary and auxiliary tasks in model training. In particular when $\lambda_s = \lambda_c = 0$, the learning scheme transforms into a conventional single-task machine-learning scheme for the experimental χ parameters.

In the model training, gradient descent was performed in a conventional setting: maximum epoch of 50, minibatch size of 5−20, and use of an Adam optimizer with the AMSGrad variant.[69] The learning rate was set to decay by 0.5 from its initial value after every 10 epochs. See the PyTorch code available at GitHub[68] for further details. The set of hyperparameters to be tuned is comprised of the task importance parameters (λ_s and λ_c), initial learning rate, and dimension of the latent space k. The basic structure of the neural network was fixed in advance as described above.

For each of the data sets, D_χ, D_s, or D_c, approximately 80% of the polymer−solvent species were randomly assigned to the training set. Here, the data partitioning was performed according to the polymer−solvent species to avoid multiple profiles of the same polymer−solvent pair leaking into the test sets. Then, fivefold cross-validation based on the training set of D_χ was performed to determine the best-performing hyperparameters among 100 randomly sampled hyperparameter sets with $k \in \{3, ..., 40\}$, $\lambda_s \in [0, 1]$, $\lambda_c \in [0, 100]$, and a learning rate between 0.001 and 0.01. Note that only the root-mean-square error (RMSE) with respect to the experimental χ parameter was considered in the evaluation of the validation performance. For each combination of the hyperparameters, the mean-squared errors (MSEs) of the five validation sets were calculated and averaged. The hyperparameters with the lowest average MSE were chosen as the best. Finally, we retrained the model with the selected hyperparameters 10 times using different parameter initializations and random data splitting. Here, the validation data were used to choose the best parameters during the gradient descent process for 50 epochs. The performance was evaluated based on the ensemble mean and variability of the 10 trained models.

RESULTS AND DISCUSSION

Based on the cross-validation, the dimensionality of the latent space was chosen as $k = 34$. The latent dimension was considerably larger than the three-dimensional HSP sphere; the implications of the latent space will be discussed later. The task importance parameters were selected to be $\lambda_s = 0.51$ and $\lambda_c = 1.00$.

It should be noted that, as shown in Figure S2, the distribution of the validation MSE showed a wide variation as a function of the hyperparameter values. Thus, there was no statistically significant difference between the selected hyperparameters and some of the other values. For the dimensionality selection of the latent space, the distributions of the validation MSEs were comparable in the range of $k \in [15, 35]$. For the task importance parameters, there was no statistically significant difference in the distribution of the validation MSEs in the range of $\lambda \in [0.1, 0.6]$. This implies the following: the machine-learning algorithm found it advantageous to simultaneously use the additional data set of the computational χ parameters, without overweighting, to improve the prediction performance of the experimental χ parameters.

The prediction performance metrics of the experimental χ parameter for the test samples were $R^2 = 0.834$, MAE = 0.199, and RMSE = 0.311 (Figure 3a); the performance metrics of the computational χ parameter were $R^2 = 0.763$, MAE = 0.468, and RMSE = 0.775 (Figure 3b); and the performance metrics of the solubility classification were accuracy = 0.857, F_1 score = 0.892, and area under curve (AUC) associated with the receiver operating characteristic (ROC) curve of 0.919 (Figure 3c−d). The classification performance is comparable to that reported in the previous studies: Chandrasekaran et al.[46] reported an AUC of 0.980, while Kern et al.[49] reported an

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

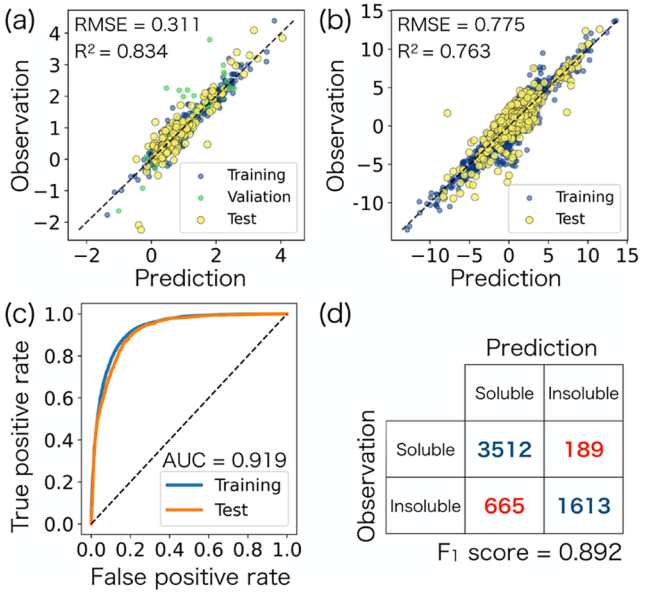


Figure 3. Prediction of χ parameters and binary classification of solubility. (a) Parity plot of predicted and observed experimental χ parameters (horizontal and vertical axes). (b) Prediction of computational χ parameters. (c) ROC curve for the binary classification of solubility. (d) Confusion matrix summarizing the prediction outcome of the test samples classified into soluble or insoluble.

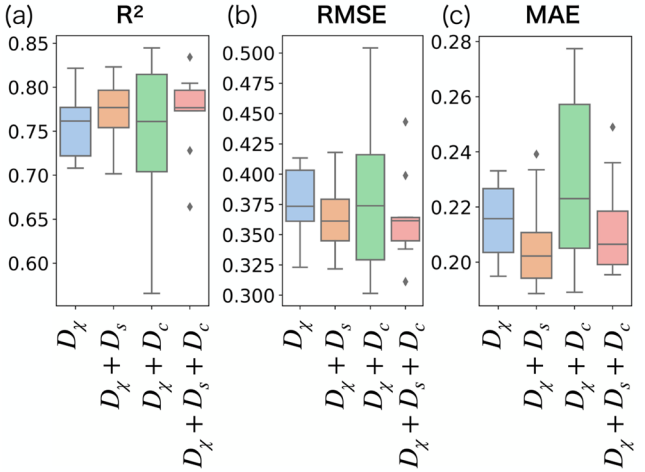


Figure 4. Comparison of prediction performance of experimental χ parameters between single-task learning (D_χ), two-task learning of experimental and computational χ parameters ($D_\chi + D_s$), two-task learning with solubility classification ($D_\chi + D_c$), and three-task learning ($D_\chi + D_s + D_c$). The boxplots represent the distributions of (a) R^2, (b) RMSE, and (c) MAE for 10 independently trained models.

accuracy of 0.832. Although polymer−solvent pairs in the three different tasks were distributed in distinct regions, as shown in Figure S1, the machine-learning algorithm was able to learn hidden common features, resulting in high prediction accuracy. It is particularly interesting to note that the binary classifier demonstrated high prediction accuracy even though experimental conditions, such as mixing temperatures, were entirely ignored in the modeling. This observation suggests that the chemical pattern of polymer−solvent is the dominant factor in the determination of solubility and that processing conditions play only an auxiliary role.[46,49]

As summarized in Figure 4, the effectiveness of multitask learning was assessed by comparing the prediction performance of (1) three-task joint learning with that of (2) two-task joint learning with the experimental and computational χ parameters ($\lambda_c = 0$), (3) two-task joint learning with the experimental χ parameter and the binary classification of solubility ($\lambda_s = 0$), and (4) single-task learning ($\lambda_c = \lambda_s = 0$). Here, we selected the best hyperparameters for each task separately. Cases 2 and 3 of the two-task joint learning showed better prediction performance than the single-task learning on average for the 10 different models in terms of the RMSE; however, the performance index values exhibited high variability. By contrast, three-task joint learning (1) consistently outperformed single-task learning. The trend of multitask learning improving the prediction performance from single-task learning was not sensitive to the random selection of test data (see the Supporting Information).

Furthermore, the prediction performance of the three-task model was compared with that of the quantum chemical calculations using COSMO-RS and the empirical model based on the HSP distance. The prediction performance metrics of COSMO-RS were $R^2 = 0.620$, MAE = 0.335, and RMSE = 0.476, in which the volume fractions were set to the actual values reported in the experimental data set.[19] Although the predicted and experimental values agreed to a certain extent, their discrepancy was considerably larger than that of the proposed method (Figure 5a). Regarding the computation

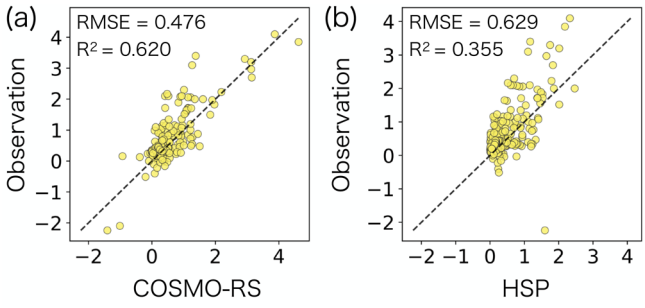


Figure 5. Prediction of experimental χ parameters by (a) quantum chemistry calculations using the COSMO-RS method and (b) the HSP distance predictor.

time, it took a total of 4129 s to perform the quantum chemistry calculation and create COSMO files for 47 polymers and 138 solvents on a conventional server. In addition, it took 732 s to calculate the χ parameters from the COSMO files for the 1190 pairs. Therefore, it took $(4129 + 732)/1190 \simeq 4.1$ s per polymer−solvent pair. This is nearly 40 times slower than the execution time of the neural network, which is approximately 0.11 s per polymer−solvent pair including the descriptor calculations. However, the HSP prediction, as detailed in the Supporting Information, showed a large deviation from the experimental χ parameters. The performance metrics of the HSP distance were $R^2 = 0.355$, MAE = 0.421, and RMSE = 0.629 (Figure 5b). Based on these observations, we concluded that it is extremely difficult to predict χ parameters based on the three physicochemical HSP parameters alone.

As previously described, our multitask-learning approach can be interpreted as an extension of HSP-based solubility prediction through a fully data-driven framework. Representa-

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

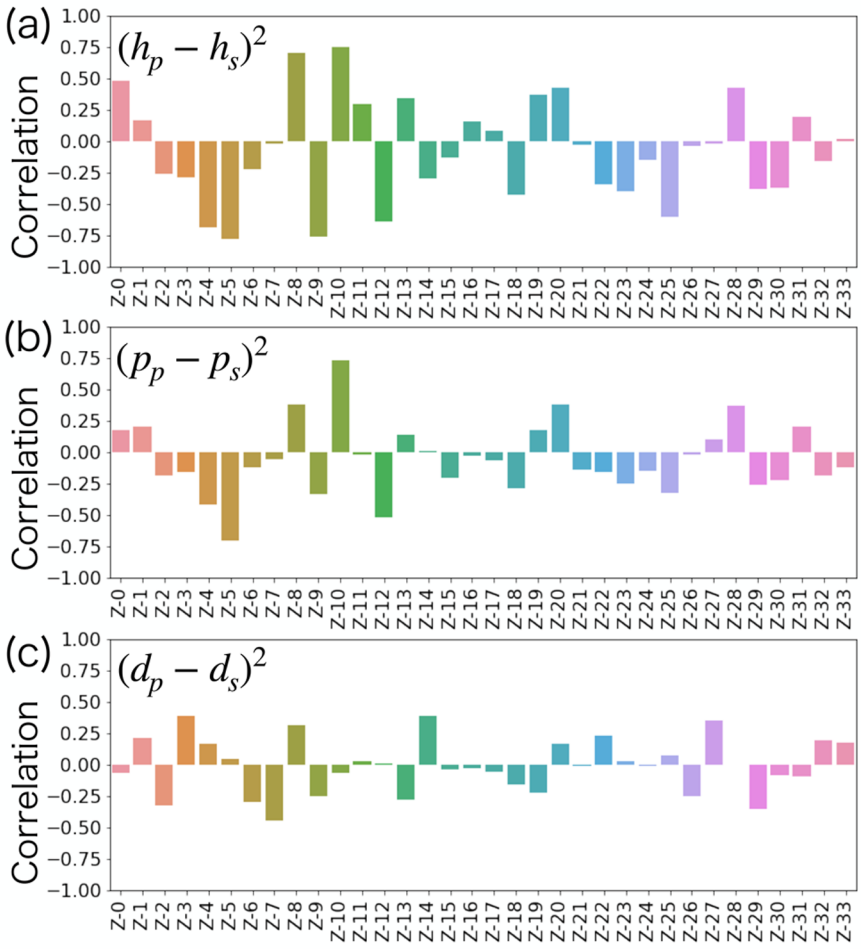


Figure 6. Pearson's correlations of the 34 machine-learned latent variables with respect to (a) the hydrogen-bonding energy term, (b) dipole interaction energy term, and (c) dispersion force term in the HSP solubility predictor.

tion learning of the latent parameters can be considered as replacing the three energy components in the HSP prediction model with a higher-dimensional vector. Figure 6 summarizes the Pearson's correlations between the machine-learned solubility parameters (k = 34) and the three HSP parameters calculated from all the polymer−solvent pairs in the test data set of the experimental χ parameter. As shown in Figure 6, several latent variables showed significant correlations with some of the HSP parameters. For example, the 4th, 5th, 8th, 9th, and 10th latent variables are clearly correlated with the hydrogen-bonding term $(h_p - h_s)^2$. The correlation pattern of the polarity term $(p_p - p_s)^2$ is similar to that of the hydrogen-bonding term. Therefore, these latent variables can be interpreted as those retaining the hydrogen-bonding and polarity terms in the HSP solubility representation. The dispersion term is highly correlated with the 3rd, 7th, 14th, 27th, and 29th variables, etc., which was encoded on a different coordinate axis than the other two terms. By contrast, the model learned latent parameters that were evidently uncorrelated with the three energy terms in the HSP prediction model. This observation indicates the existence of important unknown factors for χ parameter prediction that were ignored by the HSP representation. Understanding the physical meaning of these factors is an important issue to be addressed in the future. Methodologies developed in the context of explainable artificial intelligence (XAI) or interpretable machine learning may help address this problem. Nevertheless, these latent features, which were extracted autonomously from the given data, achieved a dramatic improvement in the prediction performance over the HSP predictor. In this regard, the present machine-learning framework successfully transformed the original concept of the HSP solubility sphere into a data-driven paradigm.

Regarding the temperature dependence of the χ parameter, the current model could not demonstrate predictive ability. As test instances to be demonstrated, we extracted χ parameters of polystyrene and *tert*-butyl acetate at 10 different temperatures from Schotsch et al.[70] The multitask model was trained using the procedure described above, in which this polymer−solvent pair was removed from the training set. As shown in Figure 7, all 10 independently trained models failed to capture the temperature dependency of this polymer−solvent pair. The reason behind the failure to predict the temperature-dependent curve is the lack of data representing the temperature dependence. In the experimental data set of χ parameters, only approximately 27% of the polymer−solvent pairs recorded multiple temperature measurements, whereas less than 10% included measurements at three or more temperatures. Even with multiple temperature observations, approximately one-third of the observations did not show clear, smooth monotonically increasing or decreasing temperature-dependent

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

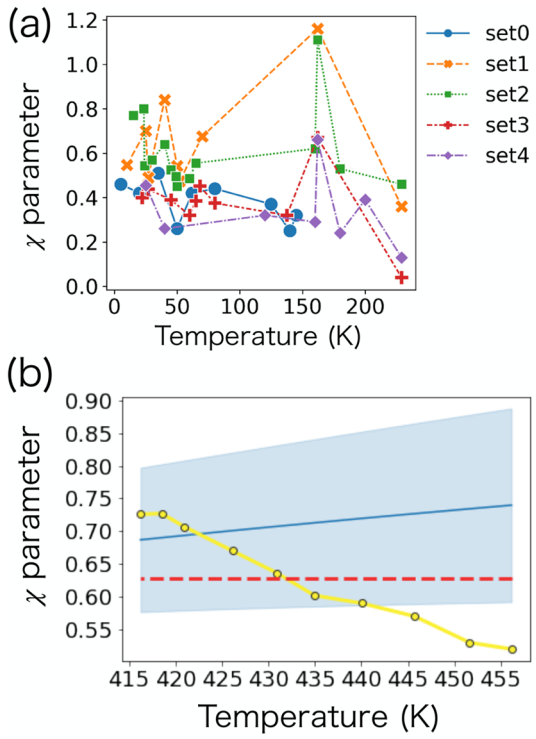


Figure 7. Prediction of temperature-dependent curves of the χ parameter. (a) Example of training instances for five polymer−solvent pairs. Set0−Set4 denote poly(vinyl acetate)/pyridine, polystyrene/2-butanone, polystyrene/cyclohexane, polystyrene/toluene, and polystyrene/benzene, respectively. (b) Temperature dependence of χ parameters for polystyrene dissolving in *tert*-butyl acetate. The yellow line denotes the experimental data. The dashed red line denotes the mean of the experimental data over the temperature range. The blue line denotes the mean predicted value of 10 different multitask models with the standard deviation shown as the shaded area.

patterns. As most samples were analyzed in different experimental environments or laboratories, it is not surprising that the nonsmooth temperature gradients were observed.

A possible strategy to overcome this limitation is to use the high-throughput computation of COSMO-RS to generate training data for temperature-dependent curves of χ parameters for multitask learning. There are several case studies in which COSMO-RS has been applied to analyze the temperature dependence of polymer−solution systems.[71] For example, Figure S5 in the Supporting Information shows a comparison of COSMO-RS temperature dependence curves for seven different polymer−solvent systems arbitrarily chosen with at least three data points that were measured at different temperatures. The use of such additional training data sets would be able to improve prediction accuracy. Note also that even if the temperature dependence of the χ parameter cannot be expressed appropriately, a phase diagram can be drawn according to eq 1; see the Supporting Information for details on the construction of phase diagrams based on Flory−Huggins theory. Figure S6 in the Supporting Information shows the phase diagrams created for the same seven polymer−solvent systems as those in Figure S5.

It is noteworthy that the χ parameter, which is generally known to depend on the molecular weight, could be predicted without using molecular weight as the input descriptor. This observation suggests that the polymer−solvent species is the dominant factor in the determination of the χ parameter. However, since the phase diagram prediction based on the Flory−Huggins theory is sensitive to the value of the χ parameter, the prediction accuracy of the χ parameter should be further improved by explicitly modeling the dependence on molecular weight and composition. Note that the calculation of the free energy of mixing based on the lattice model of the original Flory−Huggins theory is known to have limitations due to its simplicity, such as introducing the mean-field approximation and ignoring volume changes by mixing.[72,73] One approach to circumvent this limitation is to treat the χ parameter as a nonconstant quantity that depends on molecular weight and other latent factors. However, there is also a possible way to predict the phase diagram without going through the Flory−Huggins theory. For example, Ethier et al.[56] proposed to predict the phase diagram directly by applying machine learning without relying on the Flory−Huggins theory and reported that molecular weight is one of the most important descriptors in predicting the composition-dependent curve of polymer−solvent cloud points.

Our primary interest is in the use of this model for the exhaustive screening of miscible or immiscible polymer−solvent pairs. In such cases, the model must have predictive capability over a wide range of material spaces, which is typically outside the coverage of the training data set. To examine the extrapolative capability, we retrained the multitask model with a data set in which a particular class of polymers to be tested was excluded from the training data set. We manually labeled the polymers in the data set of the experimental χ parameters into seven classes: celluloses, polyacrylates, polychloro-olefins, polyesters, polyethers, polystyrenes, and polyvinyls. For each polymer class, we retrained the model by removing all polymers belonging to the same class from the training set of the experimental χ parameters. We also excluded the training polymers from the training sets of the calculated χ parameters and solubility−insolubility class labels. As shown in Figure 8, for all the polymer classes, two- or three-task joint learning generally demonstrated better or comparable prediction performance than single-task learning (Figure 8). In particular, multitasking performed overwhelmingly well in

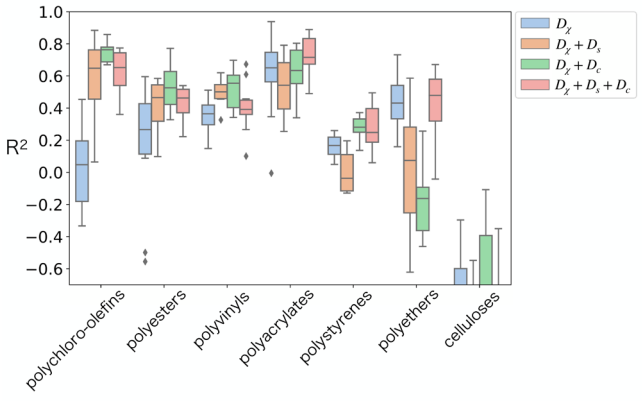


Figure 8. Comparison of extrapolative prediction performance of χ parameters for single-task and two- or three-task joint learning (the same notation as in Figure 4). The boxplot represents the distribution of R^2 for the 10 different independently trained models in which a particular type of polymer to be tested was removed from the training set of experimental χ parameters. Note that all methods failed to predict χ parameters of celluloses owing to lack of related data leading to negligibly low R^2.

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

the case of polychloro-olefins, polyesters, and polyvinyls. In cases where the performance did not improve significantly, the data sets for the two related subtasks contained few similar polymers. Overall, these results suggest that the joint use of D_c, D_s, and D_χ in the learning process contributes to expanding the applicability domain of the resulting model.

CONCLUSIONS

This study addressed the problem of predicting the miscibility of polymers in solvents—a critical problem in polymer chemistry. Up until now, various quantum chemical methods or empirical models using solubility parameters have been proposed to predict the χ parameter, which is a dimensionless quantity of great importance in the prediction and understanding of polymer miscibility. However, as shown in this study, their predictive capabilities are still far from being practically applicable. The significance of this study is that the machine-learning-based prediction of χ parameters can outperform traditional computational methods in terms of accuracy.

The objective of applying machine learning was to estimate the mapping from the chemical structures of polymer−solvent molecules to χ parameters. While the workflow itself was a conventional setup of machine learning, the challenge lay in handling the technical difficulty arising from the availability of biased and limited data. To overcome this barrier, we created a large number of the simulated χ parameters and extracted soluble/insoluble polymer−solvent pairs from PoLyInfo; we then employed the multitask-learning strategy to train the neural network. From a technical point of view, the most significant contribution of this study is that it presents this novel idea and demonstrates its effectiveness.

In addition to these contributions, our study has a few limitations. Clarifying physicochemical implications of the learned latent features and the range of chemical space that can be predicted by the model is an important issue in conducting prediction and decision making based on uncertain machine-learning models. In addition, the current model cannot represent the dependence of polymer miscibility on molecular weight, temperature, or compositional features. These issues can be addressed by expanding the training data set and by further multitasking and multifunctionalizing the prediction model. For example, as already discussed, we can create an arbitrarily large data set by quantum-chemically computing the χ parameter at arbitrary temperatures and volume fractions. The polymer−solvent cloud point data set of Ethier et al.[56] can also be used to learn the dependence on molecular weight and composition. To facilitate subsequent research, the Python source code and other related materials have been uploaded on GitHub.[68] It is expected that these results will contribute to overcoming important unsolved problems in the field of polymer science.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.macromol.2c02600.

Detailed descriptions of data preparation and processing, neural network modeling and performance evaluations, prediction of χ parameters using empirical equations or computational simulations, and potential extensions of the proposed model (PDF)

AUTHOR INFORMATION

Corresponding Authors

Kazuya Shiratori − *Mitsubishi Chemical Corporation, Science & Innovation Center, Yokohama 227-8502, Japan*; Email: kazuya.shiratori.ms@mcgc.com

Ryo Yoshida − *Research Organization of Information and Systems, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan; The Graduate University for Advanced Studies, Department of Statistical Science, Tachikawa 190-8562, Japan; National Institute for Materials Science, Research and Service Division of Materials Data and Integrated System, Tsukuba 305-0047, Japan*; orcid.org/0000-0001-8092-0162; Email: yoshidar@ism.ac.jp

Authors

Yuta Aoki − *Research Organization of Information and Systems, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan*

Stephen Wu − *Research Organization of Information and Systems, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan; The Graduate University for Advanced Studies, Department of Statistical Science, Tachikawa 190-8562, Japan*; orcid.org/0000-0002-7847-8106

Teruki Tsurimoto − *Mitsubishi Chemical Corporation, Science & Innovation Center, Yokohama 227-8502, Japan*

Yoshihiro Hayashi − *Research Organization of Information and Systems, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan; The Graduate University for Advanced Studies, Department of Statistical Science, Tachikawa 190-8562, Japan*; orcid.org/0000-0002-7650-4083

Shunya Minami − *Research Organization of Information and Systems, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan; The Graduate University for Advanced Studies, Department of Statistical Science, Tachikawa 190-8562, Japan*; orcid.org/0000-0002-3566-817X

Okubo Tadamichi − *Mitsubishi Chemical Corporation, Science & Innovation Center, Yokohama 227-8502, Japan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.2c02600

Author Contributions

[‖]Y.A. and S.W. contributed equally to this work. K.S. and R.Y. devised the project, main conceptual ideas, and proof outline. Y.A. and S.W. implemented the machine-learning algorithms and conducted the experiments with the support of S.M. and Y.H. T.T. carried out the quantum chemistry calculations. O.T. complied the data set. S.W., Y.A., T.T., K.S., and R.Y. wrote the manuscript. R.Y. supervised this project.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank all members of the ISM−MCC Frontier Materials Design Laboratory, a joint laboratory of Mitsubishi Chemical Corporation (MCC) and the Institute of Statistical Mathematics (ISM), for useful discussions. Some of the calculations were performed on the MCC high-performance computer system. This research was supported in part by JST CREST Grant Number JPMJCR19I3, MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Project ID: hp210264) JPMXP1020210314, MEXT KAKEN-

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

HI Grant-in-Aid for Scientific Research on Innovative Areas 19H05820, Grant-in-Aid for Scientific Research (A) 19H01132, and Grant-in-Aid for Young Scientists 21K14675 from the Japan Society for the Promotion of Science. Part of the computations for data analysis was conducted on the supercomputer at the Research Center for Computational Science, Okazaki, Japan (Project: 22-IMS-C125).

■ REFERENCES

(1) Kampouris, E. M.; Diakoulaki, D. C.; Papaspyrides, C. D. Solvent Recycling of Rigid Poly(Vinyl Chloride) Bottles. *J. Vinyl Technol.* **1986**, *8*, 79.
(2) Pappa, G. D.; Kontogeorgis, G. M.; Tassios, D. P. Prediction of Ternary Liquid-Liquid Equilibria in Polymer-Solvent-Solvent Systems. *Ind. Eng. Chem. Res.* **1997**, *36*, 5461−5466.
(3) Kricheldorf, H. R., Ed. *Handbook of Polymer Synthesis*; Plastics Engineering 24; Marcel Dekker: New York, 1992.
(4) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular Thermodynamics of Fluid Phase Equilibria*, 3rd ed.; Prentice-Hall International: Upper Saddle River, NJ, 1999.
(5) Levinson, H. J. *Principles of Lithography*, 2nd ed.; SPIE Press: Bellingham, WA, 2005.
(6) Strathmann, H.; Kock, K. The Formation Mechanism of Phase Inversion Membranes. *Desalination* **1977**, *21*, 241−255.
(7) Wienk, I.; Boom, R.; Beerlage, M.; Bulte, A.; Smolders, C.; Strathmann, H. Recent Advances in the Formation of Phase Inversion Membranes Made from Amorphous or Semi-Crystalline Polymers. *J. Membr. Sci.* **1996**, *113*, 361−371.
(8) Barson, C.; Bevington, J.; Hunt, B. Efficacy of Precipitation in Non-Solvents for the Recovery of Pure Polymers from Solution. *Eur. Polym. J.* **1996**, *32*, 1055−1059.
(9) Moad, G.; Solomon, D. H.; Johns, S. R.; Willing, R. I. Fate of the Initiator in the Azobisisobutyronitrile-Initiated Polymerization of Styrene. *Macromolecules* **1984**, *17*, 1094−1099.
(10) Coleman, M. M.; Graf, J. F.; Painter, P. C. *Specific Interactions and the Miscibility of Polymer Blends*; Technomic Publishing Co., Inc.: Lancaster, PA, 1995.
(11) Utracki, L. A. *Polymer Alloys and Blends: Thermodynamics and Rheology*; Hanser; distributed in the U.S. by Oxford University Press: Munich, Germany; New York, 1990.
(12) Holten-Andersen, J.; Hansen, C. M. Solvent and Water Evaporation from Coatings. *Prog. Org. Coat.* **1983**, *11*, 219−240.
(13) Holten-Andersen, J.; Eng, K. Activity Coefficients in Polymer Solutions. *Prog. Org. Coat.* **1988**, *16*, 77−97.
(14) Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **1942**, *10*, 51.
(15) Huggins, M. L. Some Properties of Solutions of Long-chain Compounds. *J. Phys. Chem.* **1942**, *46*, 151−158.
(16) Zhang, H.; Bhagwagar, D. E.; Graf, J. F.; Painter, P. C.; Coleman, M. M. The Effect of Hydrogen Bonding on the Phase Behaviour of Ternary Polymer Blends. *Polymer* **1994**, *35*, 5379−5397.
(17) Coleman, M. M.; Painter, P. C. Hydrogen Bonded Polymer Blends. *Prog. Polym. Sci.* **1995**, *20*, 1−59.
(18) Wolf, B. A. Making Flory−Huggins Practical: Thermodynamics of Polymer-Containing Mixtures. *Adv. Polym. Sci.* **2010**, *238*, 1−66.
(19) Orwoll, R. A.; Arnold, P. A. *Physical Properties of Polymers Handbook*, 2nd ed.; Springer: New York, 2007.
(20) Lee, J. K.; Yao, S. X.; Li, G.; Jun, M. B. G.; Lee, P. C. Measurement Methods for Solubility and Diffusivity of Gases and Supercritical Fluids in Polymers and Its Applications. *Polym. Rev.* **2017**, *57*, 695−747.
(21) Kontogeorgis, G.; von Solms, N. In *Handbook of Surface and Colloid Chemistry*, fourth ed.; Birdi, K., Ed.; CRC Press: Boca Raton, FL, 2015; pp 199−246.
(22) Hildebrand, J. H.; Scott, R. L. *The Solubility of Nonelectrolytes*, 3rd ed.; Reinhold Publishing Corporation: New York, 1950.
(23) Hansen, C. *The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient and Their Importance in Surface Coating Formulation*; Danish Technical Press: Copenhagen, Denmark, 1967.
(24) Hansen, C. *Hansen Solubility Parameters: A User's Handbook*, 2nd ed.; CRC Press: Boca Raton, FL, 2007.
(25) Lindvig, T.; Michelsen, M. L.; Kontogeorgis, G. M. A Flory−Huggins Model Based on the Hansen Solubility Parameters. *Fluid Phase Equilib.* **2002**, *203*, 247−260.
(26) Barton, A. *Handbook of Solubility Parameters and Other Cohesion Parameters*; CRC Press: Boca Raton, FL, 1983.
(27) Stefanis, E.; Panayiotou, C. Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *Int. J. Thermophys.* **2008**, *29*, 568−585.
(28) Abbott, S.; Hansen, C. M. *Hansen Solubility Parameters in Practice*; Hansen-Solubility, 2008.
(29) *Hansen Solubility Parameters*. https://www.hansen-solubility.com/.
(30) Oishi, T.; Prausnitz, J. M. Estimation of Solvent Activities in Polymer Solutions Using a Group-Contribution Method. *Ind. Eng. Chem. Process. Des. Dev.* **1978**, *17*, 333−339.
(31) Elbro, H. S.; Fredenslund, A.; Rasmussen, P. A New Simple Equation for the Prediction of Solvent Activities in Polymer Solutions. *Macromolecules* **1990**, *23*, 4707−4714.
(32) Kontogeorgis, G. M.; Fredenslund, A.; Tassios, D. P. Simple Activity Coefficient Model for the Prediction of Solvent Activities in Polymer Solutions. *Ind. Eng. Chem. Res.* **1993**, *32*, 362−372.
(33) Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. New Reference Equation of State for Associating Liquids. *Ind. Eng. Chem. Res.* **1990**, *29*, 1709−1721.
(34) Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001**, *40*, 1244−1260.
(35) Klamt, A., Ed. *COSMO-RS*; Elsevier: Amsterdam, 2005.
(36) Loschen, C.; Klamt, A. Prediction of Solubilities and Partition Coefficients in Polymers Using COSMO-RS. *Ind. Eng. Chem. Res.* **2014**, *53*, 11478.
(37) *BIOVIA COSMOtherm*; http://www.3ds.com.
(38) Okuwaki, K.; Mochizuki, Y.; Doi, H.; Ozawa, T. Fragment Molecular Orbital Based Parametrization Procedure for Mesoscopic Structure Prediction of Polymeric Materials. *J. Phys. Chem. B* **2018**, *122*, 338−347.
(39) Loschen, C.; Klamt, A. COSMO *quick*: A Novel Interface for Fast σ-Profile Composition and Its Application to COSMO-RS Solvent Screening Using Multiple Reference Solvents. *Ind. Eng. Chem. Res.* **2012**, *51*, 14303−14308.
(40) Kawakami, T.; Shigemoto, I.; Matubayasi, N. Free-Energy Analysis of Water Affinity in Polymer Studied by Atomistic Molecular Simulation Combined with the Theory of Solutions in the Energy Representation. *J. Chem. Phys.* **2012**, *137*, 234903.
(41) Krüger, K.-M.; Sadowski, G. Fickian and Non-Fickian Sorption Kinetics of Toluene in Glassy Polystyrene. *Macromolecules* **2005**, *38*, 8408−8417.
(42) Boom, R. M.; Reinders, H. W.; Rolevink, H. H. W.; van den Boomgaard, T.; Smolders, C. A. Equilibrium Thermodynamics of a Quaternary Membrane-Forming System with Two Polymers. 2. Experiments. *Macromolecules* **1994**, *27*, 2041−2044.
(43) Shiomi, T.; Izumi, Z.; Hamada, F.; Nakajima, A. Thermodynamics of Solutions of Poly(dimethylsiloxane). 1. Solutions of Poly(dimethylsiloxane) in Methyl Ethyl Ketone, Methyl Isobutyl Ketone, Ethyl n-Butyl Ketone, and Diisobutyl Ketone. *Macromolecules* **1980**, *13*, 1149−1154.
(44) Smidsrod, O.; Guillet, J. E. Study of Polymer-Solute Interactions by Gas Chromatography. *Macromolecules* **1969**, *2*, 272−277.
(45) Danner, R.; Tihminlioglu, F.; Surana, R.; Duda, J. Inverse Gas Chromatography Applications in Polymer-Solvent Systems. *Fluid Phase Equilib.* **1998**, *148*, 171−188.
(46) Chandrasekaran, A.; Kim, C.; Venkatram, S.; Ramprasad, R. A Deep Learning Solvent-Selection Paradigm Powered by a Massive

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX

Solvent/Nonsolvent Database for Polymers. *Macromolecules* **2020**, *53*, 4764−4769.
(47) Xu, J.; Liu, H.; Li, W.; Zou, H.; Xu, W. Application of QSPR to Binary Polymer/SolventMixtures: Prediction of Flory-HugginsParameters. *Macromol. Theory Simul.* **2008**, *17*, 470−477.
(48) Xu, Q.; Jiang, J. Machine Learning for Polymer Swelling in Liquids. *ACS Appl. Polym. Mater.* **2020**, *2*, 3576−3586.
(49) Kern, J.; Venkatram, S.; Banerjee, M.; Brettmann, B.; Ramprasad, R. Solvent Selection for Polymers Enabled by Generalized Chemical Fingerprinting and Machine Learning. *Phys. Chem. Chem. Phys.* **2022**, *24*, 26547−26555.
(50) Nistane, J.; Chen, L.; Lee, Y.; Lively, R.; Ramprasad, R. Estimation of the Flory-Huggins Interaction Parameter of Polymer-Solvent Mixtures Using Machine Learning. *MRS Commun.* **2022**, *12*, 1096−1102.
(51) Sanchez-Lengeling, B.; Roch, L. M.; Perea, J. D.; Langner, S.; Brabec, C. J.; Aspuru-Guzik, A. A Bayesian Approach to Predict Solubility Parameters. *Adv. Theory Simul.* **2019**, *2*, 1800069.
(52) Hagita, K.; Aoyagi, T.; Abe, Y.; Genda, S.; Honda, T. Deep Learning-based Estimation of Flory−Huggins Parameter of A−B Block Copolymers from Cross-Sectional Images of Phase-Separated Structures. *Sci. Rep.* **2021**, *11*, 1−16.
(53) Liu, T.-L.; Liu, L.-Y.; Ding, F.; Li, Y.-Q. A Machine Learning Study of Polymer-Solvent Interactions. *Chin. J. Polym. Sci.* **2022**, *40*, 834−842.
(54) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. Polymer Database for Polymeric Materials Design. *2011 International Conference on Emerging Intelligent Data and Web Technologies*; 2011; pp 22−29.
(55) *PoLyInfo*; https://polymer.nims.go.jp/.
(56) Ethier, J. G.; Casukhela, R. K.; Latimer, J. J.; Jacobsen, M. D.; Rasin, B.; Gupta, M. K.; Baldwin, L. A.; Vaia, R. A. Predicting Phase Behavior of Linear Polymers in Solution Using Machine Learning. *Macromolecules* **2022**, *55*, 2691−2702.
(57) Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41−75.
(58) Zhang, Y.; Yang, Q. An Overview of Multi-Task Learning. *Natl. Sci. Rev.* **2018**, *5*, 30−43.
(59) Yu, J.; Zhang, C.; Cheng, Y.; Yang, Y.-F.; She, Y.-B.; Liu, F.; Su, W.; Su, A. SolvBERT for Solvation Free Energy and Solubility Prediction: a Demonstration of an NLP Model for Predicting the Properties of Molecular Complexes. *Digital Discovery* **2023**, *2*, 409.
(60) Mahmoudabadi, S. Z.; Pazuki, G. Investigation of COSMO-SAC Model for Solubility and Cocrystal Formation of Pharmaceutical Compounds. *Sci. Rep.* **2020**, *10*, 19879.
(61) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(62) Hayashi, Y.; Shiomi, J.; Morikawa, J.; Yoshida, R. RadonPy: Automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *npj Comput. Mater.* **2022**, *8*, 222.
(63) *RDKit*; https://www.rdkit.org/.
(64) Han, J.; Morag, C. The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning. *From Natural to Artificial Neural Computation*; Lecture Notes in Computer Science, 1995; Vol. *930*, pp 195−201.
(65) Bronskill, J.; Gordon, J.; Requeima, J.; Nowozin, S.; Turner, R. TaskNorm: Rethinking Batch Normalization for Meta-Learning. *Proceedings of the 37th International Conference on Machine Learning*; 2020; Vol. *119*, pp 1153−1164.
(66) Young, N. P.; Balsara, N. P. *Encyclopedia of Polymeric Nanomaterials*; Springer: Berlin, Hidelberg, Germany, 2015.
(67) Knychala, P.; Timachova, K.; Banaszak, M.; Balsara, N. P. *50th Anniversary Perspective*: Phase Behavior of Polymer Solutions and Blends. *Macromolecules* **2017**, *50*, 3051−3065.
(68) Sample code for this paper. https://github.com/yoshida-lab/MTL_ChiParameter.
(69) Reddi, S. J.; Kale, S.; Kumar, S. On the Convergence of Adam and Beyond. *Proceedings of the Sixth International Conference on Learning Representations*; 2018.
(70) Schotsch, K.; Wolf, B. A.; Jeberien, H.-E.; Klein, J. Concentration Dependence of the Flory-Huggins Parameters at Different Thermodynamic Conditions. *Makromol. Chem.* **1984**, *185*, 2169−2181.
(71) Reinisch, J.; Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the Temperature Dependence of a Polyether−Water Mixture Using COSMOtherm. *Fluid Phase Equilib.* **2011**, *310*, 7−10.
(72) Dudowicz, J.; Freed, K. F. Effect of monomer structure and compressibility on the properties of multicomponent polymer blends and solutions: 1. Lattice cluster theory of compressible systems. *Macromolecules* **1991**, *24*, 5076−5095.
(73) Dudowicz, J.; Freed, M. S.; Freed, K. F. Effect of monomer structure and compressibility on the properties of multicomponent polymer blends and solutions. 2. Application to binary blends. *Macromolecules* **1991**, *24*, 5096−5111.

https://doi.org/10.1021/acs.macromol.2c02600
Macromolecules XXXX, XXX, XXX−XXX