

Otto-von-Guericke-University Magdeburg
Faculty of Process and Systems Engineering

Master Thesis



Prediction of Infinite Dilution Activity Coefficients in Polymer Solutions Using Graph Neural Networks

Author: Sreekanth Kunchapu

Matriculation number: 229775

Supervisor: M.Sc. Edgar Ivan Sanchez Medina

Submitted: April 11, 2023

Abstract

The present work focuses on using data-driven models to predict the weight fraction activity coefficient at infinite dilution Ω_i^∞ in polymer solutions, which is crucial for describing phase equilibria and the effects of solvents in polymer solutions. Three different models are employed, namely Graph Neural Networks (GNNs), Random Forest (RF), and Feed-Forward Neural Networks (FFNNs). The study compares the performance of these models against mechanistic models, including Entropic-FV and UNIFAC-ZM. According to the findings, the GNN model performs better than the mechanistic, RF, and FFN models. This study provides valuable insights into the potential use of GNN for accurately predicting Ω_i^∞ in polymer solutions and highlights the potential for further research in this area.

Acknowledgements

I am deeply grateful for the unwavering support and guidance I received during my journey towards completing my master's thesis. I extend my sincere thanks to M.Sc. Edgar Ivan Sanchez Medina, my supervisor, for his invaluable guidance and suggestions and for addressing all my concerns throughout this endeavour. I am also thankful to M.Sc. Ann-Joelle Minor for her role as my second reviewer and to Prof. Dr.-Ing. Kai Sundmacher, for accepting my thesis and providing me with the resources and opportunities at the Max-Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany.

I also wish to acknowledge the efforts of everyone who provided resources and ideas that made this journey easier. I am particularly grateful for the support of my family and friends, who were my source of inspiration and encouragement. Their love and support made all the difference.

This journey has marked the end of a long-term association with Otto-Von-Guericke University, and I am forever grateful for the opportunity to have learned and grown here.

To sum up, I want to extend my heartfelt appreciation to everyone who aided me on this path, and I take pride in having accomplished my master's thesis with their assistance.

Major tasks of the thesis



VST

FAKULTÄT FÜR VERFAHRENS-
UND SYSTEMTECHNIK

INSTITUT VERFAHRENSTECHNIK
Lehrstuhl Systemverfahrenstechnik

Otto-von-Guericke-Universität Magdeburg, Postfach 4120, 39016 Magdeburg

Student: Sreekanth Kunchapu
Studiengang: Chemical and Energy Engineering (M.Sc.)
Matrikelnummer: 229775

Otto-von-Guericke-Universität Magdeburg
Universitätsplatz 2
39106 Magdeburg
Telefon: +49 391 67-58414
Telefax: +49 391 67-11245
kai.sundmacher@ovgu.de
www.ovgu.de/vt/vsv

Ihre Zeichen. Ihre Nachricht vom:

Unsere Zeichen

Durchwahl:

Datum: 15.08.2022

Aufgabenstellung für eine Masterarbeit / Scope of tasks for a master thesis

Prediction of infinite dilution activity coefficients in polymer solutions using Graph Neural Networks

Motivation

Activity coefficients at infinite dilution are an important thermophysical property of compounds in liquid mixtures. They show the degree of deviation from ideality of a chemical compound in a liquid mixture that is caused by intermolecular or intramolecular forces. Besides this, activity coefficients at infinite dilution are commonly used to pre-select the best entrainers or solvents in an extractive separation process [1]. Given the limitations in time and resources that come along with the experimental determination of infinite dilution activity coefficients, and the vast chemical space that needs to be explore, many predictive methods have been develop for this property. Mechanistic models such as COSMO-RS, UNIFAC and MOSCED are examples of popular models that have been used for decades both in academia and industry. However, severe mismatches are still present for several systems when comparing them to experimental values [2]. This prediction mismatch increases when looking at complex mixtures such as the ones involving polymers. These mixtures are of special industrial relevance given that most of the current life standard commodities include the usage of polymers. Recently, machine learning models based on Graph Neural Networks have shown a remarkable ability and flexibility in predicting activity coefficients [3]. However, GNN-based models that are able to predict activity coefficients for polymer mixtures have not yet been developed. In fact, mechanistic models for such complex mixtures are still very limited. The main challenge for applying GNNs to polymers is the fact that polymers themselves are not define by a fixed structure, rather a distribution of different arrangements is the one defining it. Very recently stochastic edges in a graph have been proposed to target this type of compounds in the context of GNNs [4].

Ziel / Goal of the thesis

The main goal of this thesis is to develop the first GNN-based model that is able to predict infinite dilution activity coefficients of polymer mixtures (i.e., a polymer as solute and a small-molecule as solvent). For this, the collection of experimental data needs to be performed taking as a basis the

collection gathered in Vol. XIV of the DECHEMA Chemistry Data Series. Different predictive models (both mechanistic-based and data-driven) are to be compared for this task.

Aufgaben / Tasks

- Task 1: Gather experimental data from the literature for infinite dilution activity coefficients of polymer mixtures
- Task 2: Construct SMILES representations for the solvents and the monomers present in the polymer mixture.
- Task 3: Train GNN models on the molecular graphs to predict activity coefficients at infinite dilution and benchmark them with other machine learning and mechanistic models.
- Task 4: Discuss the performance of the GNN-based models for this task.
- Task 5: Provide a user-friendly and open-source code to perform predictions.

Literature

- [1] Brouwer, T., Kersten, S.R., Bargeman, G. and Schuur, B., 2021. Solvent pre-selection for extractive distillation using infinite dilution activity coefficients and the three-component Margules equation. *Separation and purification technology*, 276, p.119230.
- [2] Brouwer, T. and Schuur, B., 2019. Model performances evaluated for infinite dilution activity coefficients prediction at 298.15 K. *Industrial & Engineering Chemistry Research*, 58(20), pp.8903-8914.
- [3] Medina, E.I.S., Linke, S., Stoll, M. and Sundmacher, K., 2022. Graph neural networks for the prediction of infinite dilution activity coefficients. *Digital Discovery*.
- [4] Aldeghi, M. and Coley, C.W., 2022. A graph representation of molecular ensembles for polymer property prediction. *arXiv preprint arXiv:2205.08619*.

| | |
|----------------------------|----------------------------------|
| Dauer der Arbeit: | 20 Weeks |
| Start der Arbeit: | 05.12.2022 |
| Gutachter I: | Prof. Kai Sundmacher |
| Gutachter II: | M.Sc. Ann-Joelle Minor |
| 1. Betreuer intern: | M. Sc. Edgar Ivan Sanchez Medina |
| 2. Betreuer intern: | |

Magdeburg, den 8.11.2022



Prof. Kai Sundmacher
Lehrstuhl für Systemverfahrenstechnik

Declaration by the candidate

I, Sreekanth Kunchapu, hereby declare that this thesis is an original work of my own and has not been submitted elsewhere for any academic award or recognition. Any sources of information used in the preparation of this thesis have been appropriately cited and referenced. I also affirm that I have adhered to all ethical and academic standards in conducting this research and preparing this thesis. This declaration is made with a full understanding of the consequences of any breach of academic integrity.

Place and Date

Signature

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 6 |
| 1.1 | Poly informatics / Polymer thermodynamics | 6 |
| 1.2 | Activity coefficients | 7 |
| 1.3 | Mechanistic models for polymer solutions | 8 |
| 1.4 | Data-driven models for polymer solutions | 13 |
| 1.5 | Motivation / Goal | 14 |
| 1.5.1 | Thesis structure..... | 15 |
| 2 | State of the Art | 16 |
| 2.1 | Data-driven models for limiting activity coefficient γ_i^∞ | 16 |
| 2.2 | Models for weight fraction activitiy coefficient at infinite dilution Ω_i^∞ .. | 17 |
| 2.3 | Machine learning (ML) models for polymer property predictions .. | 17 |
| 3 | Dataset | 19 |
| 3.1 | Data sources and retrieval | 19 |
| 3.2 | Data cleaning..... | 19 |
| 3.2.1 | Construction of SMILES | 20 |
| 3.3 | Data Splitting | 21 |
| 3.3.1 | Random split..... | 22 |
| 3.3.2 | Extrapolation to solvents and solutes data split | 23 |
| 4 | Methodology | 24 |
| 4.1 | SMILES to ECFP | 24 |
| 4.2 | Convergence plots for oligomer | 25 |
| 4.3 | SMILES to molecular graph | 27 |
| 4.4 | Mixture graph..... | 30 |
| 4.5 | Message-passing layers..... | 31 |
| 4.6 | Architecture of graph neural networks..... | 32 |
| 5 | Results and Discussions | 35 |
| 5.1 | Comparisons | 35 |
| 5.1.1 | Random split..... | 35 |
| 5.1.2 | Extrapolation to solvents | 36 |
| 5.1.3 | Extrapolation to polymers | 38 |
| 5.2 | Mechanistic models vs GNN model | 39 |

| | |
|--|-----------|
| 6 Conclusion | 41 |
| 6.1 Scope of future work..... | 41 |
| A Considered polymer-solvent systems | 43 |
| B unique polymer-solvent systems used for training ML models..... | 44 |
| Bibliography | 50 |

List of Acronyms

GNNs Graph Neural Networks

RF Random Forest

FFNNs Feed-Forward Neural Networks

UNIFAC UNIQUAC Functional-group Activity Coefficients

COSMO-RS COnductor like Screening MOdel for Real Solvents

SMILES Simplified Molecular Input Line Entry System

RNNs Recurrent Neural Networks

CNNs Convolutional Neural Networks

LR Lasso Regression

SVM Support Vector Machine

GPR Gaussian Process Regression

MCMs Matrix Completion Methods

NLP Natural Language Processing

ECFP Extended Connectivity Circular Fingerprints

MN Number average molecular weight

MW Weight average molecular weight

PDI Polydispersity index

DP Degree of Polymerization

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

List of Figures

| | | |
|-----|--|----|
| 3.1 | Scanned image on the left and Detected text image on the right | 19 |
| 3.2 | Monomer | 21 |
| 3.3 | RepeatUnit-WO | 21 |
| 3.4 | RepeatUnit-W | 21 |
| 3.5 | Oligomer with DP equals 5..... | 21 |
| 3.6 | Polyisoprene structure representations | 21 |
| 4.1 | The ECFP bit vectors of the solute and solvent and values of MN, MW, or PDI are combined into a singular vector, depending on the dataset being analyzed. This vector is utilized for training a RF model that can be applied to prediction tasks..... | 25 |
| 4.2 | The ECFP bit vectors of the solute and solvent and the values of MN, MW, or PDI are merged into a single vector based on the dataset being examined. This combined vector is then utilized for training a FFNN model that can be employed for prediction tasks. | 25 |
| 4.3 | The degree of polymerization convergence in relation to substruc- tures is assessed using Tanimoto and Dice similarity. | 26 |
| 4.4 | SMILES to graph procedure for benzene molecule..... | 29 |
| 4.5 | Mixture Graph | 30 |
| 4.6 | Messaging passing with aggregation and updation steps | 32 |
| 4.7 | GNN-Architecture | 34 |
| 5.1 | Each plot represents the performance of ML models using different polymer representations and datasets, with MAE used as the metric. Each row corresponds to a different dataset, and the three varia- tions shown are denoted by the colours green (for random split), orange (for extrapolation to solvents), and blue (for extrapolation to polymers)..... | 37 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Cross-validation using 10 folds | 22 |
| 4.1 | The nodes in the molecular graphs are defined using atom features | 28 |
| 4.2 | The edges in the molecular graphs are defined using bond features | 28 |
| 4.3 | Global features used in the molecular graphs | 29 |
| 5.1 | MAPE for associated systems for Entropic-FV(E-FV), UNIFAC-ZM(U-ZM) and GNN model | 40 |
| A.1 | Polymer-solvent systems considered by Georgia et al. [23] | 43 |
| B.1 | Polymer-solvent systems | 45 |
| B.2 | Polymer-solvent systems (Contd) | 46 |
| B.3 | Polymer-solvent systems (Contd) | 47 |
| B.4 | Polymer-solvent systems (Contd) | 48 |
| B.5 | Polymer-solvent systems (Contd) | 49 |

1 Introduction

1.1 Poly informatics / Polymer thermodynamics

Polymer informatics is a branch of chemical informatics that involves developing and applying computational methods to study polymers. It combines the principles of polymer science and engineering with computer science and mathematics to create models that aid in understanding the behaviour of polymers [1]. An important aspect of polymer informatics is polymer thermodynamics, which is concerned with the thermodynamic properties of polymers, such as enthalpy, entropy, and free energy, and how they are affected by temperature and pressure. Polymer thermodynamics is critical in studying phase equilibria in polymer solutions because it helps in comprehending how temperature and pressure affect the distribution of polymers in a solution.

The Flory-Huggins theory, a thermodynamic model that describes a polymer solution's phase behaviour, was developed by Paul Flory and J.H. Huggins in the 1940s and is widely used to predict phase equilibria in polymer solutions [2]. The Flory-Huggins theory is grounded on the concept that polymers within a solution interact through changes in entropy and enthalpy. The theory assumes that the overall free energy of a polymer solution equals the combination of the interaction energy between the polymer and solvent and the mixing entropy of the polymer and solvent. The equation for the Gibbs free energy of mixing in the Flory-Huggins theory is given by

$$\Delta G_{\text{mix}} = kT \left(\phi_1 \ln \phi_1 + \frac{\phi_2}{DP} \ln \phi_2 + \phi_1 \phi_2 \chi \right) \quad (1.1)$$

Where k is the Boltzmann constant, T is the system's temperature, ϕ_1 is the volume fraction of solvent, ϕ_2 is the volume fraction of polymer, DP is the degree of polymerization, and χ is the Flory-Huggin's parameter that describes binary interaction between molecules. Since entropy favours mixing, the χ parameter predicts whether two components will mix or not. If χ is positive, mixing is opposed. If χ is zero, it is said to be an ideal mixture, which means no energy is associated with interactions. If χ is negative, mixing is promoted. The equation (1.1) has been the cornerstone of polymer thermodynamics for more than five decades and is useful for predicting phase diagrams [3–7].

Studying polymer informatics and thermodynamics is crucial in understanding the behaviour of polymers and aids in developing more accurate models for predicting phase equilibria.

1.2 Activity coefficients

This section discusses the activity coefficient γ_i , the activity coefficient at infinite dilution γ_i^∞ , the weight fraction activity coefficient Ω_i , and the weight fraction activity coefficient at infinite dilution Ω_i^∞ .

The activity coefficient, represented by γ_i , is of great importance in characterizing the non-ideal behaviour of liquid mixtures. Its calculation is necessary for determining phase equilibria in such systems. The activity coefficients quantify the extent of deviation from ideal behaviour for individual components within a mixture and, as such, are extremely relevant to the thermodynamic properties of the system. They serve as the foundation for determining the chemical potential, which is necessary for describing both phase and chemical equilibria. Therefore, activity coefficients are important in numerous chemical engineering and chemistry fields [8].

For liquid mixtures, the activity coefficient γ_i is influenced by composition, temperature, and pressure, but the effect of pressure is typically minimal and can be ignored [9]. At infinite dilution, the composition of a solute in a solution can be considered to be constant and independent of pressure, making it easier to calculate the chemical potential. This property is essential, especially in the distillation and extraction of separation processes, where the accurate determination of γ_i^∞ is critical for process design and optimization [10].

The activity coefficient at infinite dilution γ_i^∞ is calculated using the equation 1.2 from inverse gas chromatography. Where R is gas constant, V_g^0 is specific retention volume corrected to 273.15k, M_2 molecular weight of polymer, T is temperature, P_1^{sat} vapor pressure of solvent, B_{11} is second virial coefficient of solvent, V_1 is molar volume of solvent.

$$\ln \gamma_i^\infty = \ln \left(\frac{273.15 \times R}{V_g^0 \times M_2 \times P_1^{\text{sat}}} \right) - \frac{P_1^{\text{sat}} \times (B_{11} - V_1)}{RT} \quad (1.2)$$

However, the activity coefficient at infinite dilution γ_i^∞ is unsuitable for describing phase equilibria in polymer solutions because determining the molar activity

coefficient requires an accurate measurement of the molar mass. This can be a challenge for polymer solutions because the molar mass is often unknown, and synthetic polymers can have varying molecular weights (polydispersity) [11]. To overcome this problem, Patterson et al. [12] introduced the weight fraction activity coefficient Ω_i , which is appropriate for polymer solutions. Equation 1.3 shows how to calculate Ω_i .

$$\Omega_i^\infty = \gamma_i^\infty \frac{M_2}{M_1} \quad (1.3)$$

Here, M_1 and M_2 represent the solvent's and polymer's molecular weights, respectively. From equation 1.2 and 1.3, we get the following equation 1.4

$$\ln \Omega_i^\infty = \ln \left(\frac{273.15 \times R}{V_g^0 \times M_1 \times P_1^{\text{sat}}} \right) - \frac{P_1^{\text{sat}} \times (B_{11} - V_1)}{RT} \quad (1.4)$$

This formulation circumvents the issue of accurately determining the molar mass of the polymer and allows for more accurate predictions of phase equilibria in polymer solutions.

The Ω_i^∞ describes the behaviour of polymer solutions when polymer concentration in the solution approaches zero. This value provides crucial information about the strength of interactions between the polymer and the solvent, which is essential for predicting the thermodynamics of polymer solutions [13]. The book delves into the various methods employed for determining Ω_i .

Several methods are used to determine the weight fraction activity coefficient Ω_i , including experimental techniques, mechanistic models, and data-driven models. Experimental methods, such as inverse gas chromatography [14], are used to determine Ω_i . Based on theoretical calculations, mechanistic models, such as UNIFAC-ZM and Entropic-FV, are commonly used to predict Ω_i . Data-driven models, such as GNNs, FFNNs, and RF, are also used to predict Ω_i based on training data.

1.3 Mechanistic models for polymer solutions

The UNIFAC-ZM and Entropic-FV models are modifications of the UNIFAC method. This section provides an overview of the UNIFAC method and compares the differences between the original UNIFAC and the UNIFAC-ZM and Entropic-FV models.

The UNIFAC model

The UNIFAC (UNIQUAC Functional-group Activity Coefficients) method is a group contribution approach commonly utilized for calculating the activity coefficient γ_i in solution. The method considers interactions between solutes and solvents and is based on contributions from various functional groups in a molecule. Fredenslund et al. introduced the UNIFAC method in their published works [15,16].

The UNIFAC method determines the activity coefficient γ_i of a substance by breaking down the calculation into the combinatorial and residual parts. Equation 1.5 shows the equation for the UNIFAC model. where $\ln \gamma_i^C$ is the combinatorial term and $\ln \gamma_i^R$ is the residual term.

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \quad (1.5)$$

The combinatorial aspect considers molecules' physical dimensions (size, shape), which measures their entropy contribution. The combinatorial term $\ln \gamma_i^C$ can be determined using equation 1.6.

$$\ln(\gamma_i^C) = \ln \phi_i - \left(\frac{\phi_i}{x_i} \right) + 1 - \left(\frac{zq_i}{2} \right) \left[\ln \left(\frac{\phi_i}{\theta_i} \right) + 1 - \frac{\phi_i}{\theta_i} \right] \quad (1.6)$$

The equation 1.6 involves two fractions, volume fraction and surface fraction, denoted by ϕ_i and θ_i , respectively. These fractions can be computed using known information about the composition of the molecules, their relative van der Waals volumes r_i using 1.7 and van der Waals surface areas q_i using 1.8 [17]. The coordination number, assumed to be 10, is denoted by z .

$$\phi_i = \frac{x_i r_i}{\sum_j x_j r_j} \quad (1.7)$$

$$\theta_i = \frac{x_i q_i}{\sum_j x_j q_j} \quad (1.8)$$

The UNIFAC model uses relative van der Waals properties r_i and q_i , which can be determined using the relative van der Waals group volumes R_k and surface areas

Q_k . These values can be obtained from tabulated values published by Hansen et al. [18] and Bondi [19], or from the DDB database [20].

$$r_i = \sum_k \nu_k^{(i)} R_k \quad (1.9)$$

$$q_i = \sum_k \nu_k^{(i)} Q_k \quad (1.10)$$

$\nu_k^{(i)}$, which represents the number of functional groups of type k present in the compound i.

The residual component of the activity coefficient calculation accounts for the enthalpic contributions between solute and solvent that are not included in the combinatorial calculation. The residual part is determined by applying the group concept to the solution and using Γ_k and $\Gamma_k^{(i)}$ using the equation 1.11

$$\ln \gamma_i^R = \sum_k \nu_k^{(i)} (\ln \Gamma_k - \ln \Gamma_k^{(i)}) \quad (1.11)$$

The UNIQUAC equation explains how the concentration of a particular group, represented by Γ_k , affects its activity coefficient in a mixture. The equation also takes into account how the group's activity coefficient in a pure compound, represented by $\Gamma_k^{(i)}$, can influence its behaviour in the mixture.

$$\ln \Gamma_k = Q_k \left[1 - \ln \left(\sum_m \Theta_m \Psi_{mk} \right) - \sum_m \frac{\Theta_m \Psi_{km}}{\sum_n \Theta_n \Psi_{nm}} \right] \quad (1.12)$$

Equations 1.13 and 1.14 can be used to calculate the surface area fractions, represented by Θ_m , and the group mole fractions, represented by X_m , for a specific group, denoted by m.

$$\Theta_m = \frac{\Theta_m X_m}{\sum_n Q_n X_n} \quad (1.13)$$

$$X_m = \frac{\sum_j \nu_m^{(i)} x_j}{\sum_j \sum_n \nu_n^{(i)} x_j} \quad (1.14)$$

The Parameter Ψ_{nm} is a combination of the group interaction parameter, denoted

by a_{nm} , which represents the way that functional groups n and m interact with each other [21].

$$\Psi_{nm} = \exp\left(-\frac{a_{nm}}{T}\right) \quad (1.15)$$

The UNIFAC method requires two temperature-independent interaction parameters (a_{nm}, a_{mn}) for every combination of sub-groups, which are mainly derived from experimental VLE data stored in the DDB [20]. Interaction parameters between sub-groups (represented by a_{nn} and a_{mm}) are considered to be zero if the groups are identical. This means that the parameter Ψ_{nn} and Ψ_{mm} will have one value. However, when adapting the UNIFAC model for use with polymer solutions, modifications were made to create the UNIFAC-ZM model. We will explore these modifications.

The UNIFAC-ZM model

The UNIFAC-ZM model was developed specifically to address the limitations of the original UNIFAC model in accurately predicting VLE in polymer solutions [22]. This model introduces a relation between the excluded volumes of the n -mer(a polymer composed of n monomer units) and the monomer, which is important in accurately modelling the behaviour of polymer solutions.

The excluded volume represents the space occupied by the polymer chain and affects the interaction between solute and solvent molecules. Since it is unavailable for solvent molecules due to the presence of polymer molecules in a solution, the excluded volume of the polymer chains needs to be considered in calculating the activity coefficients of solvent molecules. The magnitude of the excluded volume depends on the size of the polymer chain and the degree of overlap between the chains, making it a crucial factor in accurately modelling the behaviour of polymer solutions.

The excluded volumes of short chains can be determined through geometric calculations. A relationship can be used to estimate the excluded volumes of long chains, which suggests that the excluded volume of an n -mer is about 0.6583n times the excluded volume of the monomer [22]. After using the above-mentioned relationship to approximate the excluded volumes of long chains, the volume parameter $r(n)$ in the UNIFAC model is adjusted accordingly. Specifically, the modified volume parameter can be expressed as $r(n) = 0.6583nr(1)$, as proposed

by Zhong et al. [22].

The combinatorial part of the UNIFAC-ZM model uses the modification in volume fraction calculated using equation 1.16 rather than equation 1.7. The residual term is the same as that of the UNIFAC model.

$$\phi_1 = \frac{x_1 r_1}{x_1 r_1 + x_2 [0.6583nr(1)]} \quad (1.16)$$

Where x_1 is the mole fraction of solvent, x_2 is the mole fraction of polymer, and r_1 is the volume parameter of solvent.

The UNIFAC-ZM model has been altered in a certain way to enable it to make precise predictions of activity coefficients in solutions containing polymers. This modification is crucial for enhancing the process design and optimization in the polymer industries.

The Entropic-FV model

The Entropic-FV model includes a single equation that accounts for both the impact of combinatorial effects and free volume effects. Rather than using volume fractions, the model utilizes free-volume fractions. Free volume in polymer solutions refers to the amount of space or volume available for the solvent molecules to move or occupy within the polymer matrix. The combinatorial free volume equation is given by 1.17

$$\ln \gamma_i^{comb-fv} = \ln \left(\frac{\phi_i^{fv}}{x_i} \right) + 1 - \left(\frac{\phi_i^{fv}}{x_i} \right) \quad (1.17)$$

Here, ϕ_i^{fv} is free volume fraction given by:

$$\phi_i^{fv} = \frac{x_i V_{fi}}{\sum_i x_i V_{fi}} \quad (1.18)$$

The free volume V_{fi} is determined by:

$$V_{fi} = V_i - V_{wi} \quad (1.19)$$

Here, V_i is the molar volume of component i , and V_{wi} is the van der Waals volume

of component i , which is calculated using the Bondi method [19]. The Entropic-FV model uses a residual term of the UNIFAC model [23]. It's easy to get the weight fraction activity coefficient Ω_i using equation 1.3.

When both the volumes of the polymer and the solvent are available, it is advisable to apply the Entropic-FV model. Conversely, if the polymer and solvent volumes are unknown, it is more suitable to use the UNIFAC-ZM model [24]. In other words, depending on the availability of the necessary input data, one model may be more appropriate than the other for predicting activity coefficients in polymer solutions.

1.4 Data-driven models for polymer solutions

Experimental methods for determining Ω_i^∞ are subject to limitations such as high costs, long timelines, intensive manual labour, and narrow coverage of chemical space. Meanwhile, mechanistic models such as UNIFAC, UNIFAC-ZM, and Entropic-FV, though useful, depend on detailed information about the solute and solvent fragmentation and accurate volumes, leading to a limited exploration of chemical space. To broaden the examination of chemical space, data-driven models are used as a complementary approach. The performance of machine learning models is influenced by the quality and quantity of the training data, as the models rely on this data to generate predictions.

In this study, three data-driven models are employed to predict Ω_i^∞ : Random Forests (RF), Feed-forward Neural Networks (FFNNs), and Graph Neural Networks (GNNs). The subsequent sections provide a concise overview of these models.

Random Forest

Random Forest (RF) is a machine-learning algorithm that employs an ensemble of decision trees to make predictions. The algorithm involves training multiple decision trees on random subsets of the data and then aggregating their outputs to produce the final prediction. By using this approach, RF mitigates the risk of overfitting that can occur with a single decision tree and enhances the model's accuracy.

Feed Forward Neural Networks

FFNNs are a type of artificial neural network where information travels in a unidirectional manner, from the input layer to the output layer, without forming any loops. FFNNs employ multiple hidden layers, which enable them to learn and model complex nonlinear relationships between inputs and outputs. During training, the weights and biases of the neurons in each layer are adjusted using an optimization algorithm to optimize the network's predictions.

Graph Neural Networks

GNNs are gaining popularity in the chemical and pharmaceutical industries due to their ability to process graph data structures effectively. In these contexts, a molecule can be transformed into a graph. Graphs consist of nodes and edges, where nodes indicate atoms and edges indicate chemical bonds. GNNs aggregate information from neighbouring nodes in a graph and use it to make predictions about the nodes, incorporating chemical information such as bond types, molecular structures, and chemical properties. These algorithms have been applied to various problems in the chemical and pharmaceutical fields, including molecular property prediction, drug discovery, and virtual screening. The GNN algorithm uses message passing [25] between nodes in a molecular graph to propagate information and learn the representation of the molecular structure, leading to improved predictions and a better understanding of the underlying chemical processes.

1.5 Motivation / Goal

The weight fraction activity coefficient, Ω_i^∞ , plays a critical role in studying of polymer solutions and has numerous applications in industries such as petrochemicals, pharmaceuticals, and materials science. Despite its importance, to the best of our knowledge, there is currently no data-driven model for predicting Ω_i^∞ in polymer solutions. This presents a significant gap in the modelling of polymer solutions and limits the ability of industries to make accurate predictions and decisions.

The primary objective of this study is to create for the first time model that is based on data to predict Ω_i^∞ in polymer solutions using Graph Neural Networks (GNNs). By leveraging the ability of GNNs to handle graph data structures, the

molecular structure and other relevant chemical information can be incorporated into the model, leading to improved predictions of Ω_i^∞ . Developing a data-driven model for predicting Ω_i^∞ using GNNs is expected to fill the gap in the modelling of polymer solutions and provide valuable insights for related industries.

1.5.1 Thesis structure

In this subsection, the various components of the thesis are outlined and described, including a comprehensive explanation of each chapter involved.

Chapter 2 provides a comprehensive review of the current advancements and developments pertinent to the goal. Even though there are no existing models to predict Ω_i^∞ , the chapter delves into the examination of existing models for γ_i^∞ and their potential expansion to predict Ω_i^∞ in polymer solutions.

Chapter 3 outlines the process of obtaining and preprocessing data, including information about data retrieval and splitting data into appropriate subsets.

Chapter 4 outlines the process of obtaining SMILES for solvents and describes the methodology for constructing SMILES for long-chain molecules (polymers). The chapter also explains the conversion of SMILES to ECFP, which are utilized in the training of RF and FFNN. The chapter further discusses the conversion of SMILES to graphs for GNNs, implementing message-passing layers for GNNs, and capturing interactions between polymers and solvents through a mixture graph. All these implementations are comprehensively explained in the chapter.

Chapter 5 evaluates the performance of the GNN model and its ability to extrapolate solvents and solutes. The chapter compares the model's results with those from mechanistic models.

2 State of the Art

2.1 Data-driven models for limiting activity coefficient γ_i^∞

Recent advancements in machine learning have led to the gaining popularity and implementation of GNNs in chemical and process engineering. In a study by Sanchez Medina et al. [26], GNNs were used successfully to predict the limiting activity coefficient γ_i^∞ of organic compounds, where small molecules represent the solute and solvent. The graph structures were based on these compounds, and the temperature was assumed constant. The embeddings of the solute and solvent were then concatenated to predict γ_i^∞ . This model was found to be significantly better than other mechanistic predictive models, providing a promising approach for solving complex problems in the field of chemical and process engineering.

In a separate study, Sanchez Medina et al. [27] delve into the ability of GNNs to capture temperature dependency of the limiting activity coefficient γ_i^∞ by incorporating the Gibbs-Helmholtz equation. Their study expands on the previous work by including global-level features of the entire molecule and introducing a mixture graph to capture interactions between the solute and solvent. Rather than simply concatenating the embeddings of the solute and solvent, the authors concatenated these embeddings with a global-level vector to provide a more comprehensive representation of the system. With these modifications, the model demonstrated excellent performance predicting γ_i^∞ at various temperature ranges.

In addition to Graph Neural Networks (GNNs), other methods such as Matrix Completion Methods (MCMs) and Natural Language Processing (NLP) models have also been used to predict the limiting activity coefficient γ_i^∞ . In the MCM approach, Fabian Jirasek et al. [28] created a matrix of activity coefficients for a set of solutes and solvents, with some entries missing. They then used a matrix completion algorithm to fill in the missing entries, allowing for predicting activity coefficients for new solute-solvent combinations. However, this method did not predict γ_i^∞ at varying temperatures. In a further study on Matrix Completion Method (MCM), Julie Damay et al. [29] extended the model to predict the limiting activity coefficient γ_i^∞ at varying temperatures. The improved model demonstrated better performance than the UNIFAC (Dortmund) model.

The work of Benedikt Winter et al. [30] utilized a Natural Language Processing

(NLP) model that utilized the "Simplified Molecular Input Line Entry System (SMILES)" representation to predict the γ_i^∞ property. The SMILES representation of the molecules was used as the input for the NLP model. The authors then applied input encoding and multi-headed attention layers to extract valuable information from the SMILES representation. Finally, a regression head was used to predict the γ_i^∞ property. The model considers temperature variations and demonstrates superior performance compared to established models like COSMO-RS and UNIFAC (Dortmund).

2.2 Models for weight fraction activitiy coefficient at infinite dilution Ω_i^∞

Currently, there is no reported data-driven model for predicting Ω_i^∞ to the best of our knowledge. Nevertheless, mechanistic models such as UNIFAC-ZM and Entropic-FV are available for predicting Ω_i^∞ . For further details on these models, refer to section 1.3. In this thesis work, these mechanistic models were utilized for comparison with the GNN model.

In addition to UNIFAC-ZM and Entropic-FV, another COSMO-SAC model can predict Ω_i^∞ . According to Paula B. Staudt et al. [31], the calculation of the σ -profile of polymers using the COSMO-RS theory can be simplified by a method in which the σ -profile of a polymer is represented as the σ -profile of an average monomer, which is then scaled by the number of repeat units [32]. This approach aims to make the treatment of polymers with COSMO-RS less complex and time-consuming, although it does have limitations mentioned in the paper [31].

However, in our work, we found that the results published by Paula B. Staudt et al. [31] regarding the prediction of Ω_i^∞ are not reliable. This is because we considered the same dataset as they did the DECHEMA chemistry data series of polymer solutions [33]. According to their work, the natural logarithm of Ω_i^∞ ranges between [-8, 0], while in reality, it should range between [-1.2, 7.0].

2.3 Machine learning (ML) models for polymer property predictions

Several works have been cited for predicting polymer properties using machine learning models. In a paper by Lei Tao et al. [34], the authors investigated the use

of machine learning models for predicting polymers' glass transition temperature (T_g). They proposed structure representations of the polymer. Additionally, they evaluated different feature representations, including Morgan Fingerprint (MF), Morgan Fingerprint Frequencies (MFF), Molecular Embeddings (ME), and Molecular Graphs (MG).

Different types of machine learning algorithms were employed to predict T_g , including RF, FFNNs, GNNs, Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Lasso Regression (LR), Support Vector Machine (SVM), Gaussian Process Regression (GPR). The study showed that the highest-ranking feature representation was MFF, and the best-performing model was RF. The study's results offer significant insights into how machine learning algorithms and feature representations can be used to predict polymer properties. The authors' conclusions serve as a useful reference for future studies in this field and highlight the importance of using appropriate representations and machine learning models for accurate predictions.

In a study by Jaehong Park et al. [35], the authors utilized a Graph Neural Network (GNN) model to predict polymer properties such as glass transition temperature (T_g), melting temperature (T_m), density (ρ), and elastic modulus (E). The study compared the performance of the GNN model with that of the Extended Connectivity Circular Fingerprints (ECFP) representation. The findings indicated that the performance of the GNN model was marginally better than that of the neural network using ECFP representation. This study provides evidence for the GNN model's effectiveness in predicting polymer properties and highlights the advantages of using graph-based representations over traditional molecular fingerprints.

The above state-of-the-art provides a solid foundation for representing polymer structures and constructing machine learning models to predict the weight fraction activity coefficient at infinite dilution, Ω_i^∞ .

3 Dataset

3.1 Data sources and retrieval

The author of this study retrieved their data from the DECHEMA chemistry data series of polymer solutions [33], which is restricted by copyright and cannot be published elsewhere. To address this issue, the author utilized computer vision techniques such as OpenCV and optical character recognition software (Pytesseract) to digitize the data from the book. By doing so, the author could access the data without infringing on copyright laws.

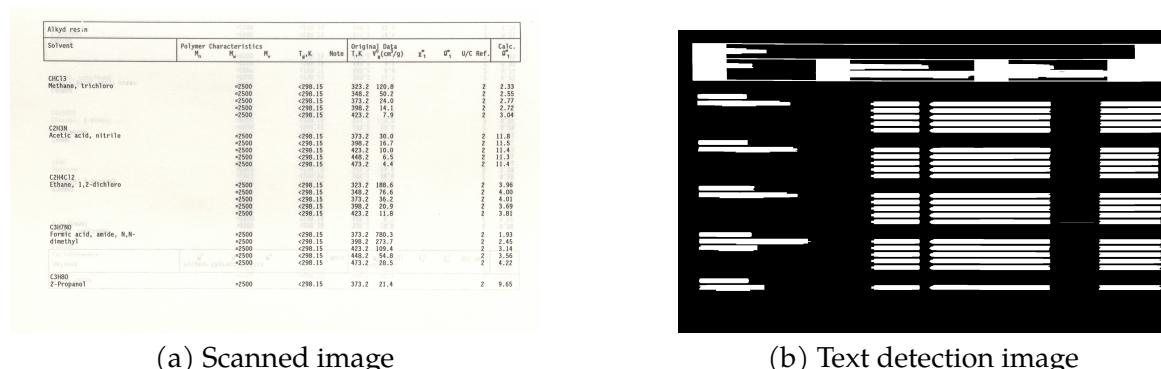


Figure 3.1: Scanned image on the left and Detected text image on the right

Authors used OpenCV, an open-source computer vision library, for image processing [36]. As shown in figure 3.1(b), text detection in images is one of the many applications of OpenCV. The Pytesseract library, a Python-based tool for optical character recognition, was then used to extract the detected text [37].

3.2 Data cleaning

Before proceeding to the data cleaning stage, it is important to discuss the features contained in the dataset and how SMILES are constructed. The dataset consists of solutes, which are polymers, and solvents, which are organic compounds. The following parameters are also included: the number average molecular weight (MN), the weight average molecular weight (MW), the polydispersity index (PDI), the temperature of the system (T), and the weight fraction activity coefficient at infinite dilution (Ω_i^∞).

3.2.1 Construction of SMILES

In this work, the SMILES notation, described in Weininger et al. [38], is used to represent the solvents and polymers in the dataset. Online resources such as PubChem [39] were utilized to obtain the SMILES representations for the solvents in the data.

The SMILES notation represents heavy atoms by their periodic table symbol. Carbon, for instance, is represented as "C", while hydrogen atoms are implicitly included and need not be explicitly stated. Single bonds between atoms are represented by simply connecting them, whereas double or triple bonds are denoted using the symbols "=" and "#", respectively. When a molecule has branches, they are enclosed in brackets, and if the molecule contains rings, the rings are represented using numbers indicating the connection points of the ring. The molecule Acetic acid, amide, N, N-dimethyl, for example, can be represented by its SMILES as "CN(C)C(C)=O".

Polymers are large, long-chain molecules that can be challenging to represent completely. However, as noted in the paper [34], they can be represented using their monomers, repeat units, and oligomers. Monomers are small organic molecules that undergo addition or condensation polymerization to form the repeat unit. The repeat unit of a polymer provides details about the bonding connections present along its chain, whereas oligomers are lengthy chains that comprise more substructures than just a single repeat unit. When the degree of polymerization (DP) is specified as five, the repeat unit is repeated five times to form the oligomer. In addition to the repeat unit, a repeat unit with an extra edge was also considered, as mentioned in the paper [40], with the connection between the start and end atoms explicitly indicated in the SMILES representation. The degree of polymerization is a hyperparameter that needs to be determined using similarity plots, as covered in the upcoming chapter 4.2.

For example, the monomer of Polyisoprene would be represented as "CC(=C)C=C", while the repeat unit without an edge (RepeatUnit-WO) would be represented as "C(C)CCC". The repeat unit with an extra edge (RepeatUnit-W) would be represented as "C*(C)CCC*" and the oligomer of this polymer, with a DP equal to five, would be represented as "C(C)CCCC(C)CCCC(C)CCCC(C)CCCC(C)CCC". The connections between the repeat units are indicated by the extra edge, represented by the asterisk (*). See figure 3.6

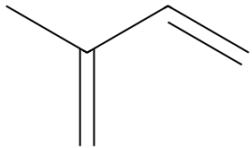


Figure 3.2: Monomer

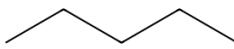


Figure 3.3: RepeatUnit-WO

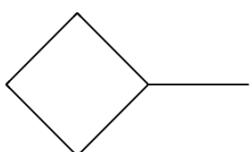


Figure 3.4: RepeatUnit- W

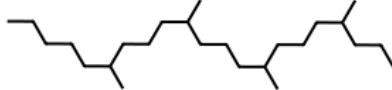


Figure 3.5: Oligomer with DP equals 5

Figure 3.6: Polyisoprene structure representations

The SMILES for the polymers were taken from two databases, namely, Poly-Info [41] and the polymer database [42]. When some of the SMILES were not present, new SMILES were constructed, and care was taken to ensure their accuracy by cross-referencing their structures using the PubChem sketcher tool [43].

To improve the quality of the data further, the MN, MW, PDI, and Ω_i^∞ values were transformed into their natural logarithmic form ($\ln(MN)$, $\ln(MW)$, $\ln(PDI)$, $\ln(\Omega_i^\infty)$). This was done to address the scaling issue and make the data more suitable for machine learning algorithms. To ensure data accuracy, certain polymers were excluded from the dataset due to uncertainties regarding their SMILES representation. In addition, certain data points were excluded from the analysis because the corresponding Ω_i^∞ values were missing in the dataset.

3.3 Data Splitting

This study has separated the data into three folders: Random-split, extrapolation to solvents, and extrapolation to solutes. The following subsections provide an overview of all three folders.

3.3.1 Random split

The Random-split folder contains the MN, MW, and PDI datasets, each with different features. The MN dataset includes solute, solvent, solute-SMILES, solvent-SMILES, $\ln(MN)$, temperature (T), and $\ln(\Omega_i^\infty)$. The MW and PDI datasets contain $\ln(MW)$ and $\ln(PDI)$ features, respectively, instead of the $\ln(MN)$ feature found in the MN dataset. The dataset for MN contains 2,438 data points, while the datasets for MW and PDI contain 2,525 and 1,588 data points, respectively. The MN dataset comprises 28 unique polymers and 136 unique solvents, while the MW and PDI datasets comprise 22 and 18 unique polymers and 117 and 102 unique solvents, respectively.

The authors utilized the sci-kit learn library [44] to execute random "k-fold cross-validation" on three datasets. They specified the value of k as 10. This approach entailed splitting each dataset into ten identical parts or folds, then training the model on nine folds and testing it on the remaining fold. This procedure was repeated ten times, with each fold being used as the validation set once.

| Fold | Training Data | Validation Data |
|------|----------------------------|-----------------|
| 1 | 1, 2, 3, 4, 5, 6, 7, 8, 9 | 10 |
| 2 | 1, 2, 3, 4, 5, 6, 7, 8, 10 | 9 |
| 3 | 1, 2, 3, 4, 5, 6, 7, 9, 10 | 8 |
| 4 | 1, 2, 3, 4, 5, 6, 8, 9, 10 | 7 |
| 5 | 1, 2, 3, 4, 5, 7, 8, 9, 10 | 6 |
| 6 | 1, 2, 3, 4, 6, 7, 8, 9, 10 | 5 |
| 7 | 1, 2, 3, 5, 6, 7, 8, 9, 10 | 4 |
| 8 | 1, 2, 4, 5, 6, 7, 8, 9, 10 | 3 |
| 9 | 1, 3, 4, 5, 6, 7, 8, 9, 10 | 2 |
| 10 | 2, 3, 4, 5, 6, 7, 8, 9, 10 | 1 |

Table 3.1: Cross-validation using 10 folds

The cross-validation using 10 folds is illustrated in Table 3.1. In this table, the folds are numbered from 1 to 10, and each fold's training and validation data are shown in separate columns. For example, in fold 1, the training data consists of folds 1 to 9, and the validation data consists of fold 10. In fold 2, the training data consists of folds 1 to 8 and fold 10, while the validation data consists of fold 9.

This process is repeated for each of the 10 folds.

"10-fold cross-validation" generates an average performance score by repeating the process ten times. This score provides an estimate of the model's generalization performance and helps identify overfitting, which is a common issue in machine learning where a model performs well on the training data but poorly on new data. By assessing the model on distinct portions of the data in each iteration, 10-fold cross-validation generates a more reliable estimate of the model's performance compared to training it on the entire dataset and assessing it on a separate test set.

3.3.2 Extrapolation to solvents and solutes data split

The authors utilized a Python function to divide the data into 5 groups to extrapolate solvents and solutes. The function was created to guarantee that each validation group would have different solvents or solutes that were not included in the corresponding training data. This method assists in assessing the model's performance on unseen solvents and solutes present in the test set.

4 Methodology

This chapter explains the conversion of SMILES to Extended Connectivity Fingerprints (ECFP) and molecular graphs, focusing on representing the solvent and different structures of polymers, as discussed in subsection 3.2.1. In addition, the chapter delves into the topic of using convergence plots to ascertain the most suitable degree of polymerization for oligomers, the characteristics taken into consideration for molecular graphs, and the execution of a Graph Neural Network (GNN) model.

4.1 SMILES to ECFP

ECFPs are a type of molecular fingerprint commonly used in computer-aided drug design and cheminformatics. ECFP represents the molecular structure of a compound that can be used to compare the similarity between different molecules.

In ECFP, the molecular structure of a compound is described by a set of circular fragments or "bits" of a specific size. These bits are centred on heavy atoms (atoms with an atomic number greater than 1) in the molecule, describing the atoms' connectivity within a defined radius. The ECFP representation is binary, meaning that each bit is either present or absent in the molecular structure. A specific bit indicates a particular pattern of atoms and bonds in the molecular structure [45].

For instance, SMILES for polystyrene with a degree of polymerization of 10 can be transformed into a molecule object using the RDKit [46]. RDKit is a toolkit used in cheminformatics to define molecules' chemical properties. The SMILES code is first converted into an RDKit molecule object in this process. Then, using the rdMolDescriptors module from RDKit, the molecule object is transformed into an Extended connectivity Fingerprint. The ECFP representation considers the molecular structure and creates a binary vector of length 1024, with the presence or absence of substructures encoded as 1s and 0s, respectively. The radius was specified as 4, and the number of bits was defined as 1024 in this particular scenario. A similar procedure was followed for the solvent as well.

The long bit vector of solute and solvent are concatenated with values of $\ln(MN)$

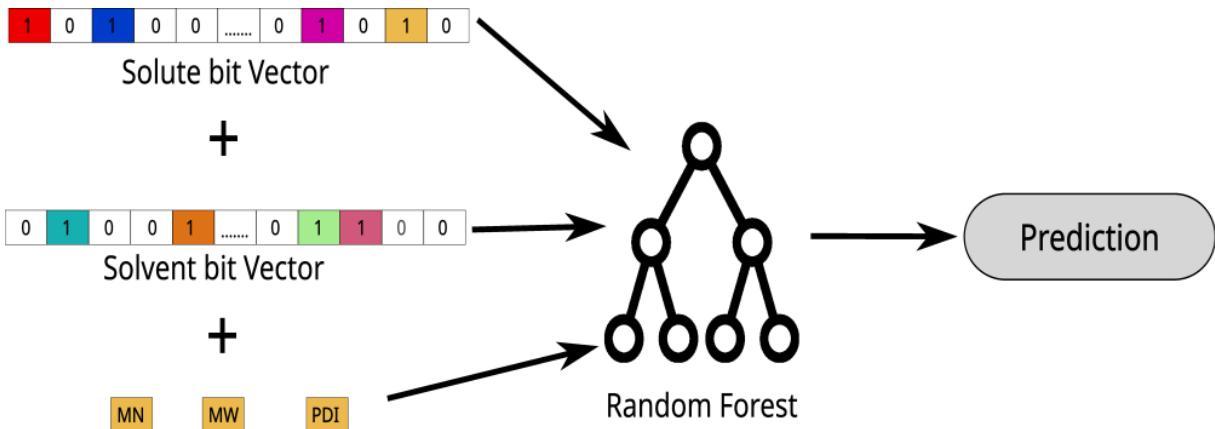


Figure 4.1: The ECFP bit vectors of the solute and solvent and values of MN, MW, or PDI are combined into a singular vector, depending on the dataset being analyzed. This vector is utilized for training a RF model that can be applied to prediction tasks.

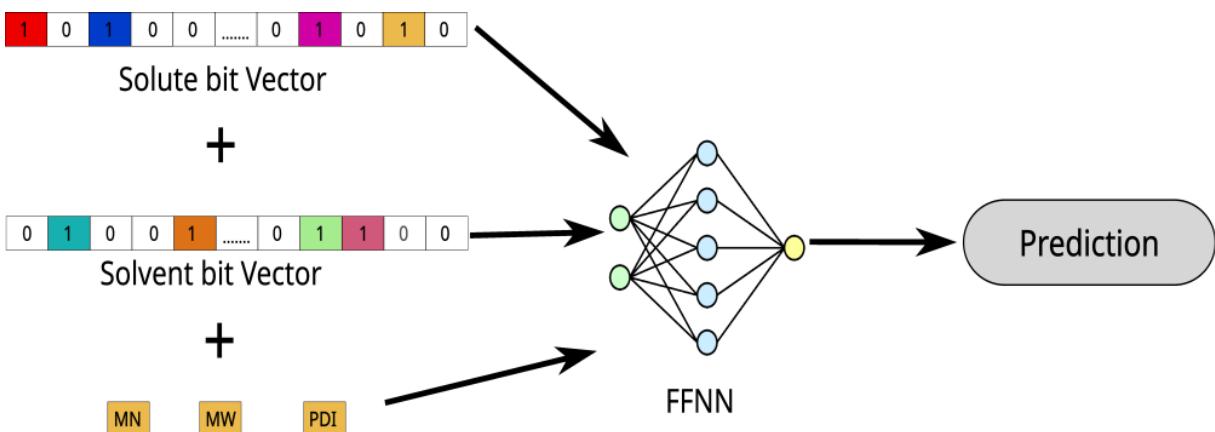


Figure 4.2: The ECFP bit vectors of the solute and solvent and the values of MN, MW, or PDI are merged into a single vector based on the dataset being examined. This combined vector is then utilized for training a FFNN model that can be employed for prediction tasks.

values when considering the MN dataset, values of $\ln(MW)$ when considering the MW dataset, and values of $\ln(PDI)$ when considering the PDI dataset, as well as values of T and $\ln(\Omega_i^\infty)$. This final molecular fingerprint is then utilized for training and predicting $\ln(\Omega_i^\infty)$ through RF and FFNN. The models utilize default parameters from the sci-kit learn library [44]. The RF model uses 100 estimators, while the FFNN model has a hidden layer dimension (100,100) and uses the ReLu activation function and adam optimizer. Refer to figures 4.1, 4.2.

4.2 Convergence plots for oligomer

To evaluate the similarities and differences between polymers, the Tanimoto coefficient [47] and the Dice coefficient are used to compare the structural features

of each polymer pair. This is done by transforming a specific molecular structure that is repeated in the polymer chain (repeat unit) into an ECFP and comparing the fingerprints pairwise. The Tanimoto coefficient is a measure that indicates the proportion of common substructures between two polymers relative to the overall number of substructures present in both. If the Tanimoto coefficient is 1, it signifies that the polymers are exactly the same, whereas a coefficient of 0 indicates that the polymers have no common substructures and are completely dissimilar.

$$TC = \frac{c}{a + b + c} \quad (4.1)$$

On the other hand, the dice coefficient measures the similarity of two sets based on the number of common elements relative to the average size of the sets. It computes the proportion of twice the number of shared elements between two sets relative to the total size of both sets. So, while both coefficients measure the similarity of two sets, they do in different ways. The Dice coefficient considers the frequency of repeated substructures.

$$DC = \frac{2c}{(a + c) + (b + c)} \quad (4.2)$$

Here, c is the number of common substructures between the two polymers, a is the number of unique substructures in the first polymer, and b is the number of unique substructures in the second polymer.

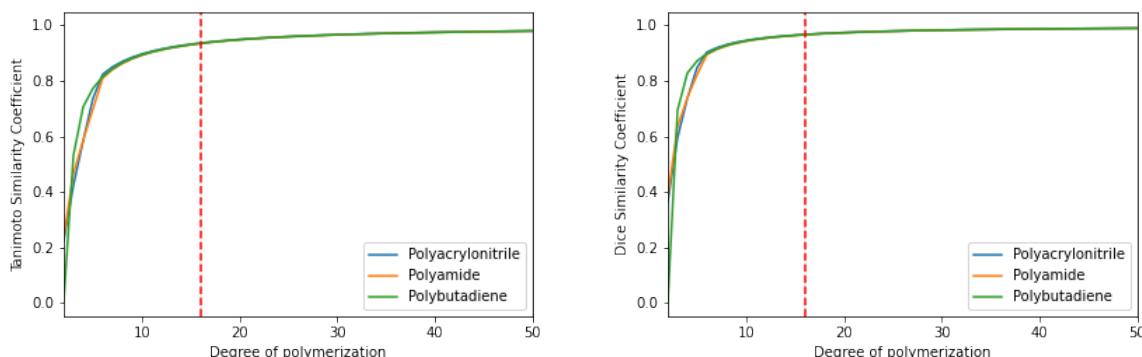


Figure 4.3: The degree of polymerization convergence in relation to substructures is assessed using Tanimoto and Dice similarity.

Regarding the graph depicted in Figure 4.3, the DP is studied by examining a

list of structures, including the Monomer, Repeat Unit, and Polymerized-1 to Polymerized-50. The structural features of each structure are identified using substructure identifiers and quantification of their occurrence frequencies. Upon plotting the Tanimoto and Dice coefficients between neighbouring structures, a discernible trend of convergence is observed in the resulting graph. Based on the substructures, it has been determined that Polymerized-16 is a sufficiently long oligomer, which is in line with the result reported in the paper [34]. For the sake of clarity and ease of interpretation, only three polymers have been included in the plot.

4.3 SMILES to molecular graph

This section provides a procedure for converting solvent and polymer SMILES to a molecular graph. Firstly, the solvent and solute SMILES are converted to molecule objects with the help of the cheminformatics package RDKit (Landrum, 2013). Next, these generated molecule objects are transformed into molecular graphs using the Pytorch Geometric library (Fey et al., 2019). Molecular graphs are composed of two fundamental elements, nodes and edges. Nodes correspond to individual atoms, while edges signify their chemical bonds.

In molecular graph representation, three matrices are typically used: Atoms' features matrix (**A**), Bonds' features matrix (**B**), and Connectivity matrix (**C**). For example, the SMILES representation of benzene is "c1ccccc1". This SMILES string can be converted into a molecule object using **RDkit** and then transformed into a molecular graph using **Pytorch Geometric**. Benzene contains six atoms (n_a) and six bonds (n_b), so the molecular graph for benzene will have six nodes and six edges (see figure 4.4).

The features of each node and edge are specified to define the molecular graph. Tables 4.1 and 4.2 enumerate the characteristics of nodes and edges, respectively. Table 4.1 presents the node features pertaining to the properties of individual atoms in the molecular graph. Table 4.2, on the other hand, lists the edge features, which describe the nature of the chemical bonds between pairs of atoms in the molecular graph. The tables are encoded using a **one-hot encoded** scheme to create the feature vectors for nodes and edges. The resulting feature vectors comprise binary values 0's and 1's, indicating a specific feature's presence or absence.

Table 4.1: The nodes in the molecular graphs are defined using atom features

| Atom features | Description | Dimension |
|---------------|--|-----------|
| Atom Type | (C, N, O, Cl, S, F, Br, Si) | 8 |
| Ring | Does the atom form part of a ring? | 1 |
| Aromatic | Does the atom constitute an aromatic system? | 1 |
| Hybridization | Hybridization (sp, sp ² , sp ³) | 3 |
| Bonds | The number of bonds in which the atom participates (0, 1, 2, 3, 4) | 4 |
| Charge | Atom's formal charge (0, 1, -1) | 3 |
| Attached Hs | Count of hydrogen-bonded atoms (0, 1, 2, 3, 4) | 4 |
| Chirality | Unspecified | 1 |

Table 4.2: The edges in the molecular graphs are defined using bond features

| Bond features | Description | Dimension |
|-----------------|------------------------------------|-----------|
| Bond Type | (Single, double, triple, aromatic) | 4 |
| Conjugated | Does the bond exhibit conjugation? | 1 |
| Ring | Does the bond belong to a ring? | 1 |
| Stereochemistry | None | 1 |

With all the node feature vectors and edge feature vectors defined, they are stacked to create the Atoms' feature matrix A and the Bonds' feature matrix B, where the matrices have dimensions of $n_a \times 25$ and $n_b \times 7$, respectively. Additionally, a Connectivity matrix C is specified to store information about the connections between nodes in the molecular graph. C has dimensions of $2 \times 2n_b$, with the first row containing information about the source nodes and the second row containing information about the receiver nodes. As molecular graphs are undirected, the source nodes can also act as receiver nodes. These three matrices are implemented as Pytorch tensors using the Pytorch Geometric library [27].

In addition to the three matrices described earlier, the work includes global-level features for the solvent and polymer molecules, as listed in Table 4.3. The authors obtained Atomic polarizability, Bond polarizability, and Topological polar surface area from previous works by Brouwer and Schuur [48] and Edgar Medina et al. [26]. These features are used only for the solvent molecules and are represented as a global-level vector $\mathbf{U} \in \mathbb{R}^3$. For the polymers, the authors included four features: Atomic polarizability, Bond polarizability, Topological

Table 4.3: Global features used in the molecular graphs

| Global features | Description | Dimension |
|--------------------------------|--|-----------|
| Atomic polarizability | polarizabilities for each atom | 1 |
| Bond polarizability | Differences in atomic polarizabilities | 1 |
| Topological polar surface area | 2D approximation of the polar surface area | 1 |
| PDI | Polydispersity index | 1 |
| MN | Number average molecular weight | 1 |
| MW | weight average molecular weight | 1 |

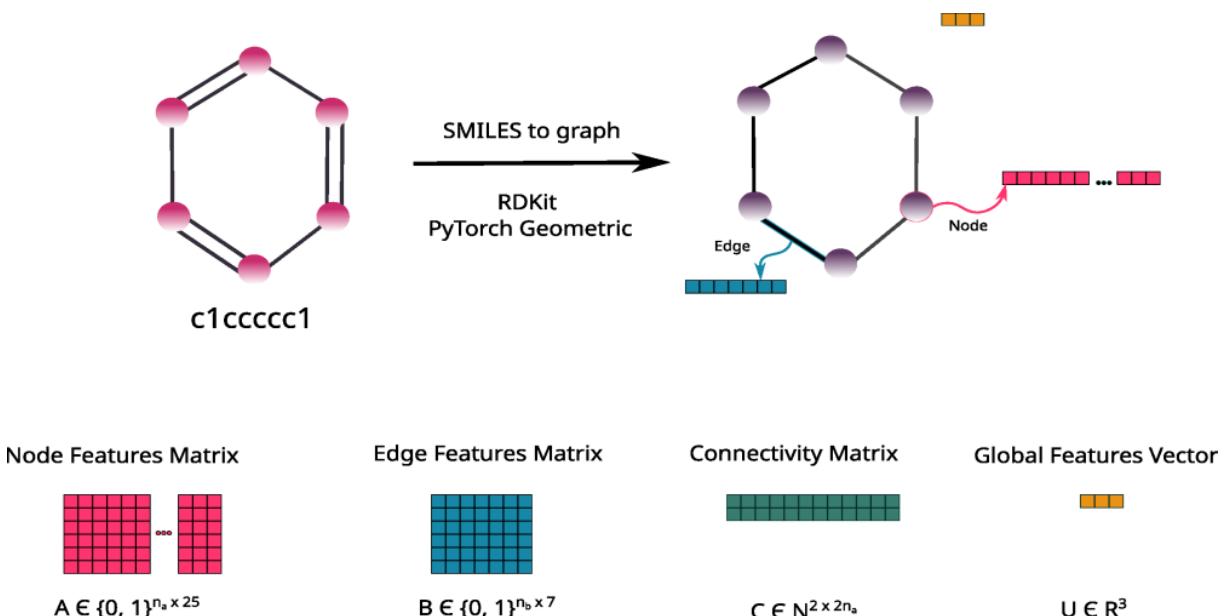


Figure 4.4: SMILES to graph procedure for benzene molecule.

polar surface area, and either $\ln(MN)$ for the MN dataset, $\ln(MW)$ for the MW dataset, or $\ln(PDI)$ for the PDI dataset. These features are represented as a global-level vector $U \in \mathbb{R}^4$.

Atomic polarizability, Bond polarizability, and Topological polar surface area are important for capturing inter-molecular interactions between the molecules. In order to establish a graph $G = (A, B, C, U)$, it is necessary to obtain the matrices A, B, and C, along with the global-level vector U. The procedure for SMILES to graph outlined above is followed for both the solvents and all the structural representations of the polymers, such as monomer, repeat unit without an edge, repeat unit with an extra edge, and Oligomer-16, as discussed in Section 3.2.1.

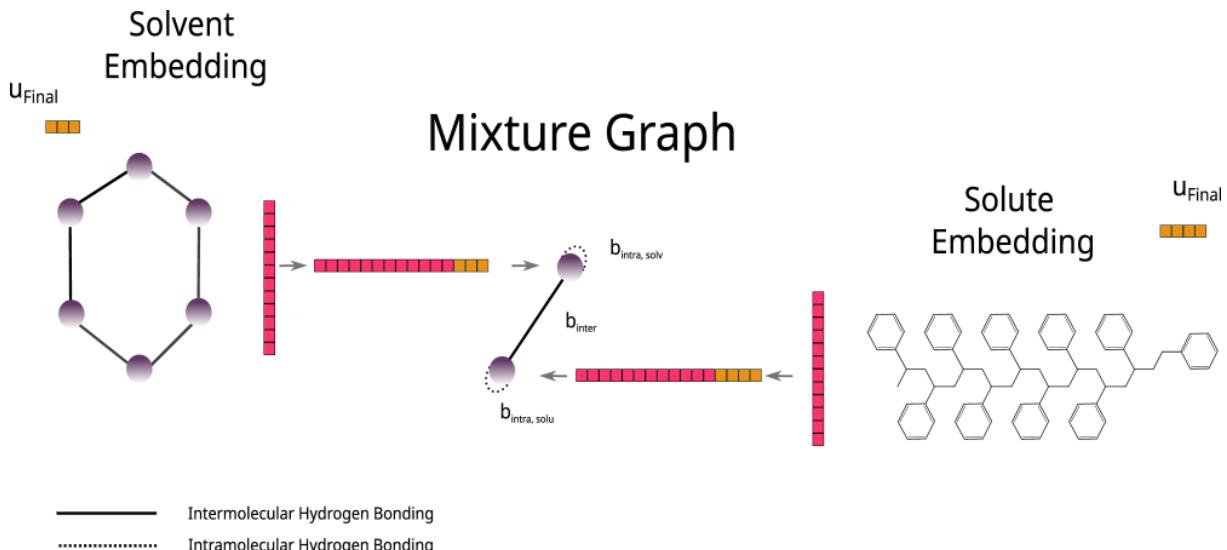


Figure 4.5: Mixture Graph

4.4 Mixture graph

When a solute is added to a solvent, inter-molecular and intra-molecular interactions occur. To capture these interactions, a mixture graph was used in the work of Sanchez Medina et al. [27], and this approach was also adapted in the current study.

This work constructed a mixture graph using the learned embeddings of the solvent and solute. These embedded vectors were concatenated with the respective global vectors of the solvent and solute. Additionally, hydrogen-bonding information was included as edge features in the mixture graph, the approach proposed by Qin et al. [49].

The inter-molecular hydrogen bonding between these molecules is stored in b_{inter} using equation 4.3. The count of hydrogen bond acceptors (HBA) in the solvent molecule is represented by $N_{\text{solv}}^{\text{HBA}}$, while the count of hydrogen bond donors (HBD) in the solute molecule is represented by $N_{\text{solu}}^{\text{HBD}}$. Atoms or functional groups in a molecule that are capable of accepting a hydrogen bond are referred to as hydrogen bond acceptors, while those that can donate a hydrogen bond are known as hydrogen bond donors. Equation 4.3 captures the maximum number of hydrogen-bonding sites possible.

$$b_{\text{inter}} = \min(N_{\text{solv}}^{\text{HBA}}, N_{\text{solu}}^{\text{HBD}}) + \min(N_{\text{solu}}^{\text{HBA}}, N_{\text{solv}}^{\text{HBD}}) \quad (4.3)$$

Similarly, the intra-molecular hydrogen bonding, which occurs within the molecules as self-loops, is stored in $b_{\text{intra},k}$ and captured using equation 4.4, where k indicates either the solvent or solute. The minimum number of HBA and HBD in each molecule is used to calculate $b_{\text{intra},k}$ (see figure 4.5).

$$b_{\text{intra},k} = \min(N_k^{\text{HBA}}, N_k^{\text{HBD}}) \quad (4.4)$$

4.5 Message-passing layers

Graph Neural Networks (GNNs) utilize message-passing layers (1) to process graph-structured data. In GNNs, these message-passing layers pass information between nodes in the graph through a series of transformations. These layers are designed to perform graph-to-graph transformations, with their parameters being optimized for specific tasks, such as regression.

After each l -level message-passing layer, each node in the graph acquires information about all other nodes. This comprehensive representation of the graph is due to the iterative process of information passing between nodes.

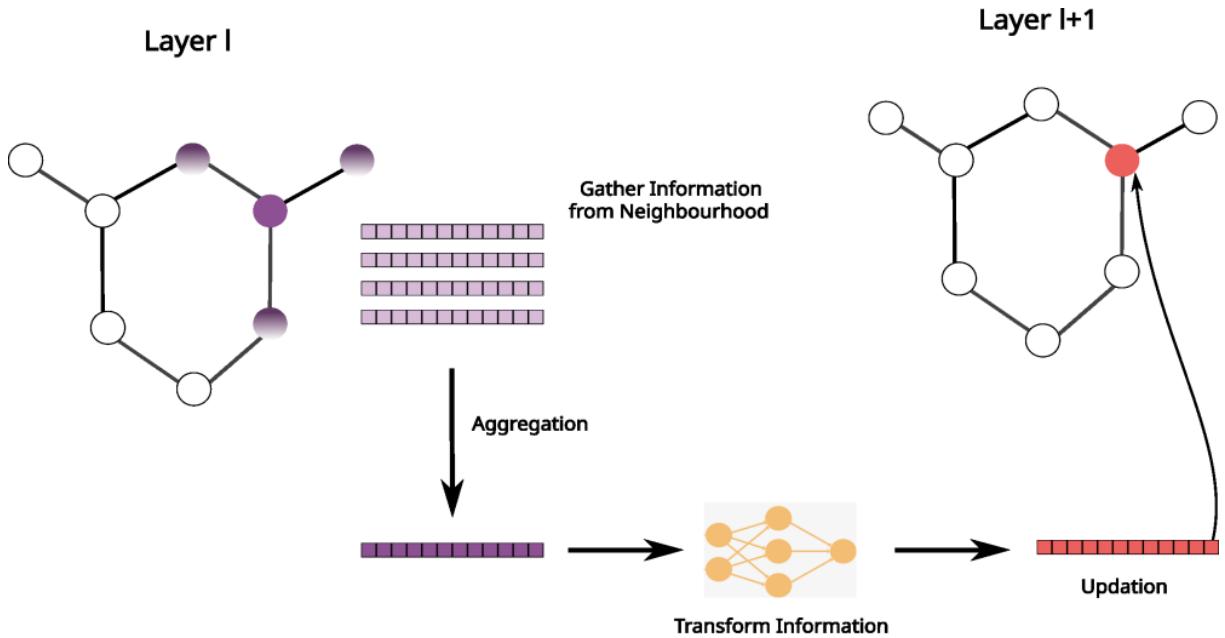
The number of nodes, edges, and the connectivity of the graph remains the same from layer to layer in a GNN, thus preserving the graph's structure, which is important for capturing relationships between nodes. The message-passing layers in GNNs are essential for processing graph-structured data as they enable the transformation and integration of information from different nodes in the graph (see figure 4.6).

The general message-passing scheme [50] in graph neural networks can be represented as:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N_i} W^{(l)} h_j^{(l)} + b^{(l)} \right) \quad (4.5)$$

where

- $h_i^{(l)}$, is hidden representation of node i in layer l , is a vector that summarizes the information about node i and its neighbours up to layer l .
- the weight matrix $W^{(l)}$ and bias vector $b^{(l)}$ for layer l are learned during training to determine how information should be propagated between nodes.



Message Passing Scheme

Figure 4.6: Messaging passing with aggregation and updation steps

- N_i , is number of neighbours for node i , represents nodes connected to node i in the graph.
- $\sigma(\cdot)$, the activation function is applied element-wise to the result of the linear combination of $h_j^{(l)}$ values and the weight and bias terms to produce the updated hidden representation of node i .
- $\sum_{j \in N_i}$, the sum over the neighbors of node i , aggregates the information from the neighbours of node i and combines it with node i 's information to producing the updated hidden representation of node i .

In this message-passing scheme, information is propagated from the neighbours of a node to the node itself in each layer of the network. The updated hidden representation of each node is used as input for the next layer of the network until the final layer produces the final graph representation.

4.6 Architecture of graph neural networks

This work employs an architecture that has been adapted with certain modifications from the one previously explained by Sanchez Medina et al. [27] in their research. In this architecture, graph convolution layers are used to process

the graphs representing the solvent and solute, with updates taking place for the graph's nodes, edges, and global representations. The input graph $G^{(l)} = (A^{(l)}, B^{(l)}, C, U^{(l)})$ is transformed into the updated graph $G^{(l+1)} = (A^{(l+1)}, b^{(l+1)}, C, U^{(l+1)})$. The embeddings of the solvent and solute, obtained by applying global pooling with the mean method, are then concatenated with their corresponding global-level vectors. This allows for capturing inter-molecular and intra-molecular interactions, as discussed in section 4.4.

Once the mixture graph undergoes processing through graph convolution layers and global pooling is applied to extract the mixture fingerprint, this fingerprint is utilized as input to a feed-forward neural network for predicting $\ln \Omega_i^\infty$. The temperature dependency is captured by predicting the parameters k_1 , k_2 , and k_3 using the Gibbs-Helmholtz derived equation 4.6. The constants for this equation are obtained by passing the mixture fingerprint through three separate multi-layer perceptrons, one for each constant. The neural network used in this work has two hidden layers with the ReLu activation function. In the case of computing k_3 , the mixture fingerprint is multiplied by the temperature to capture the temperature dependence of the parameter (see figure 4.7).

$$\ln \Omega_i^\infty(T) = k_1 + \frac{k_2}{T} + k_3(T) \quad (4.6)$$

This architecture provides a complete overview for training the graphs from the beginning to the desired property using backpropagation.

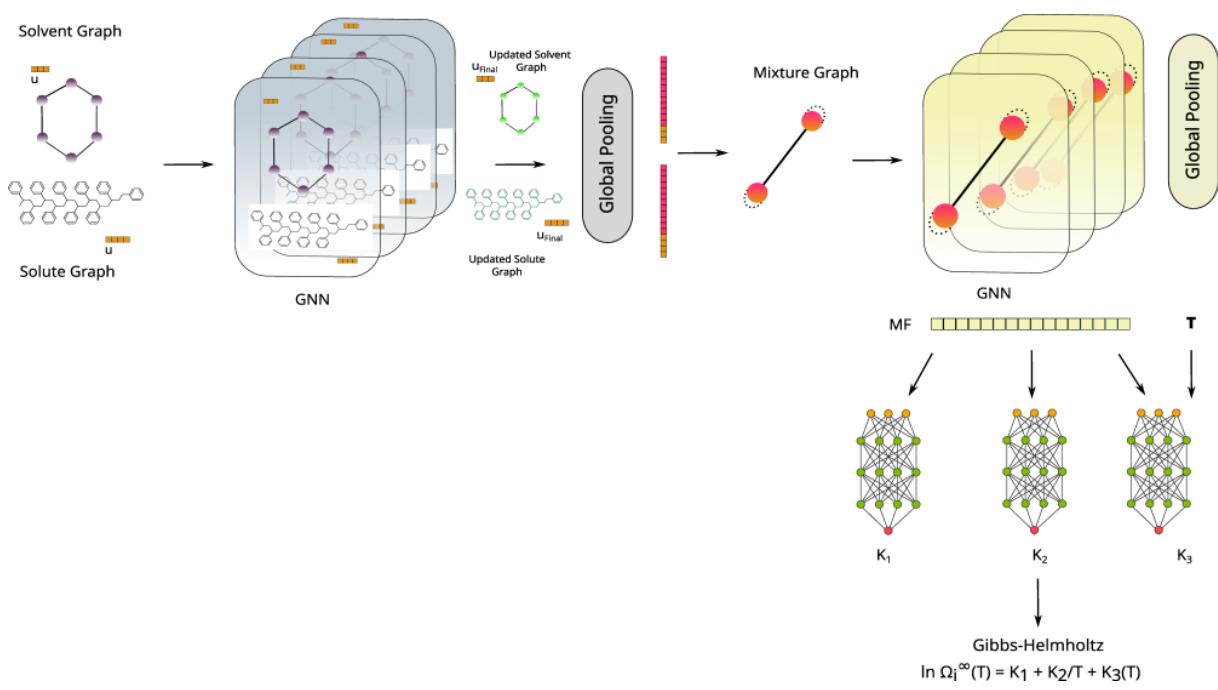


Figure 4.7: GNN-Architecture

5 Results and Discussions

5.1 Comparisons

The author compares the performance of GNN, FFNN, and RF models using structural representations of polymers: Monomer, RepeatUnit without an extra edge (RepeatUnit-WO), RepeatUnit with an extra edge (RepeatUnit-W), and Oligomer-16. The models are evaluated on three datasets corresponding to the $\ln(MN)$, $\ln(MW)$, and $\ln(PDI)$ features. Three different types of data splitting are used for each dataset, as explained in Section 3.3.

The metric used to evaluate these ML models is Mean Absolute Error (MAE). The mathematical equation for MAE is:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\ln(\Omega_{ipred}^\infty) - \ln(\Omega_{iact}^\infty)| \quad (5.1)$$

where $\ln(\Omega_{iact}^\infty)$ is the actual value, $\ln(\Omega_{ipred}^\infty)$ is the predicted value and N is the total number of data points. The equation 5.1 calculates the absolute value of the differences between the predicted and actual values for each sample and then averages to get the final MAE value. A lower MAE value indicates better performance of the model.

5.1.1 Random split

In Figure 5.1, the green colour represents a random split, which can result in interpolation between known polymers and solvents, indicating that these known entities may be present in both the training and test sets. The author suggests analyzing the type of dataset to be used, and the type of structure representations of the polymers, as the GNN model outperforms other models such as RF and FNN models.

Regarding the PDI dataset, the GNN model performed better than the RF and FFNN models. The monomer structure had an MAE of 0.38, while repeatUnit-W and repeatUnit-WO had an MAE of 0.31, and the oligomer-16 had an MAE of 0.36. The RF model had an MAE of 0.45 for the monomer structure, while repeatUnit-W and repeatUnit-WO had MAE values of 0.49 and 0.48, respectively. The oligomer-16 had an MAE of 0.46. The FFNN model had an MAE greater than

0.54 for all polymer structure representations.

The GNN model also performed better on the MN dataset compared to the RF and FFNN models. The monomer representation had an MAE of 0.27, while repeatUnit-W and repeatUnit-WO had MAE values of 0.30, and the oligomer-16 had an MAE of 0.35. The RF model had MAE values ranging from 0.39 to 0.44 for different polymer structure representations, while the FFNN model's MAE values ranged from 0.48 to 0.55.

The GNN model's performance for the MW dataset was better than the RF and FFNN models. The monomer representation had an MAE of 0.35, while repeatUnit-W and repeatUnit-WO had MAE values of 0.40 and 0.36, respectively. The oligomer-16 representation had an MAE of 0.35. The RF model had MAE values ranging from 0.41 to 0.48 for different polymer structure representations, while the FFNN model's MAE values ranged from 0.44 to 0.52.

In conclusion, the GNN model demonstrated varying levels of accuracy for different polymer structure representations on the MN, PDI, and MW datasets, with the monomer representation being the most accurately predicted and oligomer-16 being the least accurately predicted. The performance of the GNN model reduces for larger and more complex structures, indicating that the message passing along the longer polymer chain is less effective. The RF and FFNN models did not perform better than the GNN model for any of the datasets. Overall, it is important to consider the type of dataset and structure representations used when analyzing and choosing a model for polymer property prediction.

5.1.2 Extrapolation to solvents

In terms of solvent extrapolation, the GNN model demonstrates promising results across all three datasets and their respective structure representations. The orange colour in Figure 5.1 indicates that the polymer is present in both the training and test sets, but the solvent used in combination with the polymer differs between the two sets. Despite this challenging scenario, the GNN model outperforms both the RF and FFNN models in all cases, with MAE ranging from 0.23 to 0.29 for the different polymer representations.

Specifically, for the PDI dataset, the GNN model achieves MAE of 0.24, 0.25, and 0.23 for the monomer, repeatUnit-W, and repeatUnit-WO, respectively, and

5 Results and Discussions

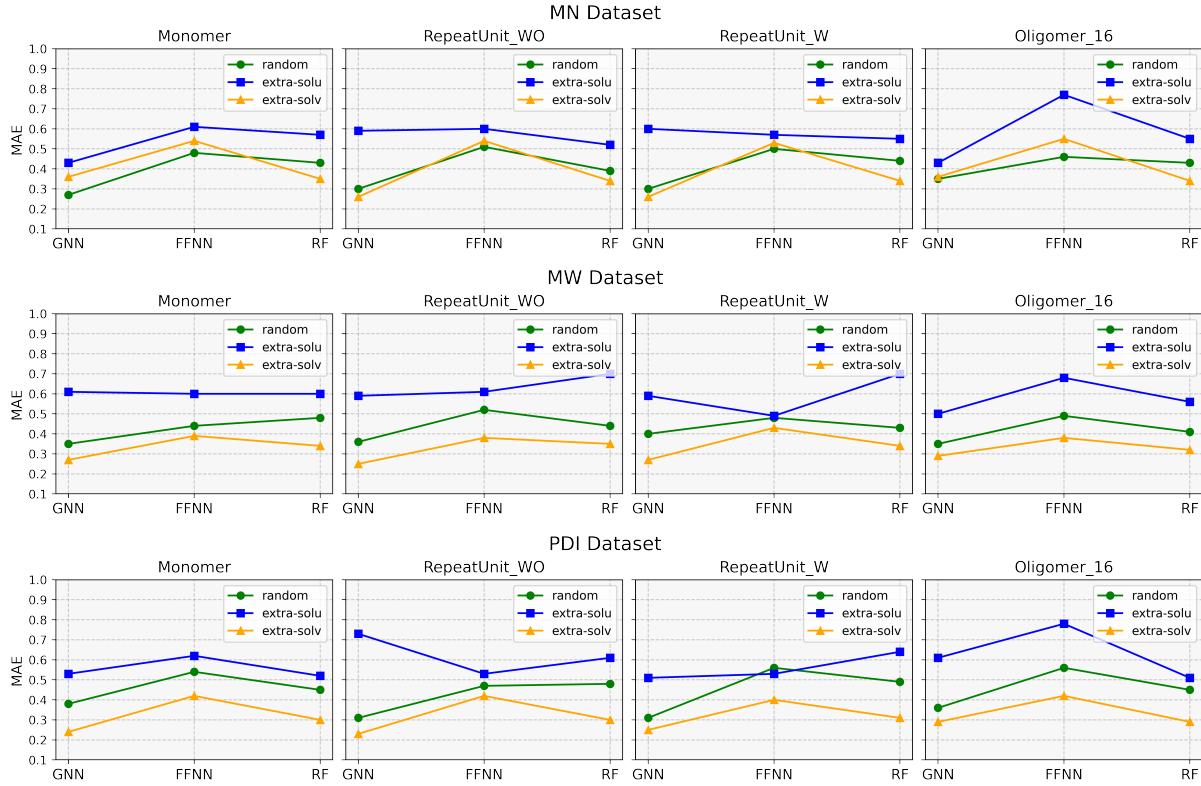


Figure 5.1: Each plot represents the performance of ML models using different polymer representations and datasets, with MAE used as the metric. Each row corresponds to a different dataset, and the three variations shown are denoted by the colours green (for random split), orange (for extrapolation to solvents), and blue (for extrapolation to polymers)

0.29 for oligomer-16. In contrast, the RF and FFNN models perform worse, with MAE ranging from 0.30 to 0.31 and 0.40 to 0.42, respectively, across the different polymer representations.

For the MN dataset, the GNN model achieves MAE of 0.36, 0.26, and 0.26 for the monomer, repeatUnit-W, and repeatUnit-WO, respectively, and 0.29 for oligomer-16. The RF model achieves MAE of 0.35, 0.34, and 0.34 for the same structures and 0.36 for oligomer-16. The FFNN model performs the worst among the three models, with MAE ranging from 0.53 to 0.54 for the different polymer representations.

Regarding the MW dataset, the GNN model again demonstrates good performance, achieving MAE of 0.27, 0.27, and 0.25 for the monomer, repeatUnit-W, and repeatUnit-WO, respectively 0.29 for oligomer-16. The RF model performs worse, with MAE of 0.34, 0.35, and 0.35 for the same structures and 0.32 for oligomer-16. The FFNN model also performs poorly, with MAE ranging from 0.38 to 0.43 for the different polymer representations.

In conclusion, the GNN model shows promising results for extrapolation to solvents, outperforming the RF and FFNN models for all three datasets and their respective polymer representations. These findings suggest that the GNN model could be useful for predicting polymer properties in different solvent environments.

5.1.3 Extrapolation to polymers

The blue colour in Figure 5.1 represents extrapolation to different polymers. The plot displays the MAE values for GNN, FFNN, and RF models when extrapolated to different types of polymers.

Overall, the performance of machine learning models in predicting $\ln \Omega_i^\infty$ is less consistent across the different structures of polymers compared to different datasets. When considering the PDI dataset, the GNN model's MAE for the monomer structure is 0.53, while the MAE for repeatUnit-W and repeatUnit-WO are 0.51 and 0.73, respectively. Furthermore, the MAE for oligomer-16 is 0.61. For the RF model, the MAE for different structures ranges from 0.52 to 0.64, while for the FFNN model, the MAE ranges from 0.53 to 0.78.

When considering the MN dataset, the GNN model's performance for the monomer structure is 0.43, while the MAE for repeatUnit-W and repeatUnit-WO are 0.60 and 0.59, respectively. The MAE for oligomer-16 is 0.43. For the RF model, the MAE for different structures ranges from 0.52 to 0.57, while for the FFNN model, the MAE ranges from 0.57 to 0.77.

When considering the MW dataset, the GNN model's performance for the monomer structure is 0.61, while the MAE for repeatUnit-W and repeatUnit-WO are 0.59 and 0.59, respectively. The MAE for oligomer-16 is 0.50. For the RF model, the MAE for different structures ranges from 0.56 to 0.50, while for the FFNN model, the MAE ranges from 0.49 to 0.68.

The MAE values for GNN are generally higher than those for extrapolation to solvents. This is because unseen polymers in the test set are sensitive to the model and may be less effective at predicting $\ln \Omega_i^\infty$ for more complex polymer representations.

The FFNN model generally performs the worst, with the highest MAE values across many of the polymer types. The GNN and RF models perform similarly,

with some variations in performance across different polymer types.

The study's findings suggest that machine learning models' accuracy in predicting $\ln \Omega_i^\infty$ depends on the type of polymer representation and the specific characteristic features employed, such as MN, MW, and PDI. In cases where polydispersity index values are available, it is recommended to use repeat unit structure representation. Monomer and repeat unit representations yielded promising results for the number average molecular weight. For weight average molecular weight, either monomer or oligomer-16 representations can be relied upon, as they demonstrated comparable performance. These results have important implications for the development of more effective and accurate machine learning models for predicting $\ln \Omega_i^\infty$.

5.2 Mechanistic models vs GNN model

The performance of GNN model was compared to mechanistic models, including UNIFAC-ZM and Entropic-FV, for associated systems. In a paper by Georgia D. Pappa [23], MAPE (Mean Absolute Percentage Error) was used to evaluate UNIFAC-ZM and Entropic-FV models. The MAPE is calculated using equation 5.2

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{\ln(\Omega_{iact}^\infty) - \ln(\Omega_{ipred}^\infty)}{\ln(\Omega_{iact}^\infty)} \right| \quad (5.2)$$

where N is amount of observations , $\ln(\Omega_{iact}^\infty)$ is the actual value and $\ln(\Omega_{ipred}^\infty)$ is the predicted value.

The PEO(10700)/ethanol system was not included in the analysis as it was unavailable in the databook. The appendix A provides information on the associated systems considered in the paper [23]. The MAPE was calculated for the remaining existing systems, and the results showed that the Entropic-FV and UNIFAC-ZM had MAPE of 29.2% and 25.7%, refer to Table 5.1.

To calculate the MAPE for the GNN, predicted values for $\ln(\Omega_i^\infty)$ were obtained from the test folds for each considered system, and the average of the predicted values for each system was taken. This procedure was performed for all systems, and the average MAPE for the GNN model was found to be 13.7% (see Table 5.1)

For these associated systems, the GNN model was used to calculate the MAPE, which was found to be 14.14% and using the RF model, it was 30.86%. It is worth

5 Results and Discussions

Table 5.1: MAPE for associated systems for Entropic-FV(E-FV), UNIFAC-ZM(U-ZM) and GNN model

| System | T range (K) | NP | MAPE | | |
|------------------------|-------------|----|------|------|------|
| | | | E-FV | U-ZM | GNN |
| PEO(2000)/butanol | 348-392 | 3 | 15.1 | 39.1 | 4.9 |
| PEO(10700)/propanol | 343-393 | 3 | 4.3 | 6 | 10.5 |
| PVAc(84300)/2-propanol | 398-473 | 4 | 2.1 | 13.1 | 4.1 |
| PS(20000)/1-propanol | 436-503 | 4 | 40.7 | 35 | 20.3 |
| PBD(93000)/1-propanol | 339-369 | 4 | 17.4 | 3.8 | 14.3 |
| PMMA(6107)/methanol | 423-473 | 3 | 78.2 | 7.1 | 9 |
| PS(20000)/acetic acid | 436-503 | 4 | 47 | 76.2 | 33.2 |
| | | 25 | 29.2 | 25.7 | 13.7 |

noting that using the GNN model reduced the percentage error for Entropic-FV and UNIFAC-ZM by 50.6% and 43.9%, respectively.

6 Conclusion

In conclusion, this study highlights the importance of understanding the weight fraction activity coefficient $\ln \Omega_i^\infty$ in predicting phase equilibria in polymer solutions. Combining polymer structures, feature engineering, and machine learning algorithms is crucial for effectively performing poly informatics tasks. The study also suggests that a 16-repeat unit oligomer can effectively represent long-chain polymers due to its substructure convergence and bonding information of the polymer's substructures. However, the selection of the appropriate polymer representation depends on the type of features available, such as MN, MW, and PDI values.

When PDI values are available, using the repeat unit structure representation is recommended, while Monomer and repeat unit representations are suitable for MN values. For MW values, either monomer or oligomer-16 representations can be relied upon, as they demonstrated comparable performance. The GNN model outperforms other ML models and mechanistic models in predicting $\ln \Omega_i^\infty$.

Overall, the findings have important implications for the development of more effective and accurate machine learning models for predicting $\ln \Omega_i^\infty$ in polymer solutions. The study highlights the importance of considering the specific characteristics of the polymer structures and features when selecting appropriate machine learning models for poly informatics tasks.

Limitation of the model

Further refinement is necessary to improve the GNN model's performance with different polymers by including more polymers in training $\ln \Omega_i^\infty$.

6.1 Scope of future work

The current work has focused on predicting $\ln \Omega_i^\infty$ for homopolymers. However, the scope of future work can be expanded by considering copolymers and substitution polymers. Copolymers have two or more repeating units in their structure and can significantly impact the polymer solution's phase equilibria. Similarly, substitution polymers with functional groups that replace one or more hydrogen atoms in the repeating unit exhibit different chemical and physical properties

than homopolymers.

Incorporating copolymers and substitution polymers into the methodology will enhance the model's accuracy and predictability, making it more practical for a wider range of applications.

Overall, the scope of future work for this research is broad, and incorporating copolymers and substitution polymers into the methodology is an essential step towards a more comprehensive understanding of the behaviour of polymer solutions.

A Considered polymer-solvent systems

The table A.1 provides information related to the systems considered in the paper [23]. PEO refers to poly(ethylene oxide), PVAc to poly(vinyl acetate), PS to polystyrene, PBD to polybutadiene, and PMMA to poly(methyl methacrylate). The number inside the bracket represents either MN or MW. NP refers to the number of data points.

Table A.1: Polymer-solvent systems considered by Georgia et al. [23]

| System | T range (K) | NP |
|------------------------|-------------|----|
| PEO(2000)/butanol | 348-392 | 3 |
| PEO(10700)/propanol | 343-393 | 3 |
| PEO(10700)/ethanol | 348-393 | 3 |
| PVAc(84300)/2-propanol | 398-473 | 4 |
| PS(20000)/1-propanol | 436-503 | 4 |
| PBD(93000)/1-propanol | 339-369 | 4 |
| PMMA(6107)/methanol | 423-473 | 3 |
| PS(20000)/acetic acid | 436-503 | 4 |

B unique polymer-solvent systems used for training ML models

Table B.1 shows all the polymer-solvent systems employed to train the machine learning models. In total, 779 distinct polymer-solvent systems were utilized in this study.

Table B.1: Polymer-solvent systems

| Polymer-Solvent System | Polymer-Solvent System |
|---|--|
| Polyisoprene - Pentane,2-methyl | Polybutadiene - 2-Pentanone,4-methyl |
| Poly(n-butyl methacrylate) - Furane,tetrahydro | Polyethylene - Cyclohexane |
| Poly(vinyl acetate) - Ethanol,2-ethoxy | Polystyrene - Aniline |
| Poly(vinylidene fluoride) - Undecane | Polybutadiene - 2-Butanone |
| Poly(epi-chlorohydrin) - Furane,tetrahydro | Polysulfone - Cyclohexanone |
| Polystyrene - 2-Butanone | Poly(vinyl acetate) - Butanoicacid,nitrile |
| Poly(ethyl methacrylate) - Amino,butyl | Poly(methyl acrylate) - Dodecane |
| Polysulfone - Formicacid,amide,N,N-dimethyl | Poly(n-pentyl methacrylate) - Hexane |
| Polystyrene - Benzene,pentyl | Polybutadiene - Methane,tetrachloro |
| Poly(methyl acrylate) - Ethene,trichloro | Poly(n-hexyl methacrylate) - Aceticacid,nitrile |
| Poly(epi-chlorohydrin) - Cyclohexene | Polyethylene - Benzene,ethyl |
| Poly(ethylene oxide) - Ethanol,2,2,2-trifluoro | Poly(vinylidene fluoride) - Tetradecane |
| Poly(vinyl acetate) - Hexane,3,3,4,4-tetramethyl | Polybutadiene - 1-Butanol |
| Poly(vinyl acetate) - Methanol | Polystyrene - Pentane,2-methyl |
| Polyethylene - Dodecane | Poly(methyl methacrylate) - 2-Pentanone |
| Polystyrene - Methane,tetrachloro | Polysulfone - Methane,dichloro |
| Polyisobutylene - Cyclohexane | Poly(ethyl methacrylate) - Cyclohexane,ethyl |
| Polystyrene - 1-Butanol | Poly(n-hexyl methacrylate) - Hexane |
| Polyethylene - Sulfurdioxide | Polyethylene - Benzene |
| Polyisoprene - Furane,tetrahydro | Poly(2-hydroxyethyl methacrylate) - Tetradecane |
| Poly(ethyl methacrylate) - Methane,trichloro | Polyacrylonitrile - Dodecane |
| Poly(methyl acrylate) - Benzene | Polybutadiene - 2-Propanone |
| Poly(vinyl chloride) - Benzene | Poly(methyl acrylate) - Aceticacid,nitrile |
| Poly(vinyl acetate) - Aceticacid,butylester | Poly(vinyl chloride) - Aceticacid,nitrile |
| Polystyrene - Water | Poly(epi-chlorohydrin) - Dodecane |
| Polysulfone - Benzene,chloro | Poly(vinyl acetate) - Nonane |
| Polyamide - Methane,trichloro | Poly(n-propyl acrylate) - 1-Propanol |
| Polybutadiene - Furane,tetrahydro | Polystyrene - Furane,tetrahydro |
| Polyisoprene - Methane,trichloro | Poly(vinyl acetate) - Ethane,nitro |
| Poly(methyl methacrylate) - Benzene,1,4-dimethyl | Polystyrene - Heptane,2-methyl |
| Poly(n-propyl methacrylate) - Methane,trichloro | Poly(ethylene oxide) - Aceticacid,ethylester |
| Poly(methyl acrylate) - Decalin(trans) | Poly(methyl methacrylate) - 2-Propanol |
| Poly(vinyl acetate) - Decane | Poly(ethylene oxide) - 2-Butanone |
| Poly(n-butyl methacrylate) - 1,4-Dioxane | Poly(ethylene adipate) - 2-Butanone |
| Poly(DL-lactide) - Aceticacid,ethylester | Poly(methyl acrylate) - Hexane |
| Poly(ethyl methacrylate) - 1-Propanol | Poly(methyl methacrylate) - Aniline |
| Poly(methyl acrylate) - Toluene | Poly(vinyl chloride) - Cyclohexane |
| Polystyrene - 1-Octene | Poly(vinyl chloride) - Hexane |
| Poly(epi-chlorohydrin) - 1,4-Dioxane | Poly(butylene adipate) - Methane,trichloro |
| Poly(methyl methacrylate) - Aceticacid,ethylester | Poly(n-butyl methacrylate) - Butane,1-chloro |
| Poly(epi-chlorohydrin) - 2-Propanol,2-methyl | Poly(ethyl acrylate) - Methane,trichloro |
| Poly(n-butyl methacrylate) - Heptane | Poly(ethylene oxide) - Methane,tetrachloro |
| Poly(vinyl acetate) - Benzene,chloro | Polyacrylonitrile - Aceticacid,nitrile |
| Poly(dimethylsiloxane) - Pentane,2,2,4-trimethyl | Poly(ethylene adipate) - Methane,tetrachloro |
| Polyamide - 1,4-Dioxane | Poly(epi-chlorohydrin) - Butane,1-chloro |
| Poly(epi-chlorohydrin) - Heptane | Polystyrene - Styrene |
| Polystyrene - Toluene,alpha-hydroxy | Poly(ethylene oxide) - 1-Butanol |
| Polyisoprene - 1,4-Dioxane | Poly(methyl methacrylate) - Benzene,1,2-dichloro |
| Poly(n-propyl methacrylate) - 1-Propanol | Poly(vinyl acetate) - Decalin(cis) |
| Poly(vinyl acetate) - Hexadecane | Polyethylene - Hexane,2,4-dimethyl |
| Polystyrene - Dodecane | Poly(ethylene oxide) - Water |
| Polyisoprene - Heptane | Poly(n-butyl methacrylate) - Pentane |
| Poly(epi-chlorohydrin) - Toluene | Poly(epi-chlorohydrin) - Pentane,1-chloro |
| Polystyrene - 1-Propanol,2-methyl | Poly(vinylidene fluoride) - Cyclohexanol |
| Poly(DL-lactide) - 2-Propanone | Poly(vinylsobutyl ether) - Benzene,1,2-dimethyl |
| Poly(methyl acrylate) - Cyclohexane, butyl | Poly(epi-chlorohydrin) - Pentane |
| Poly(n-butyl methacrylate) - Octane | Poly(ethylene oxide) - 2-Propanone |
| Poly(methyl methacrylate) - 2-Propanone | Poly(vinyl acetate) - Propane,1-chloro |
| Polystyrene - Pyridine | Poly(ethylene adipate) - 2-Propanone |
| Poly(ethylene oxide) - Ethane,1,1,1-trichloro | Poly(ethyl acrylate) - 1-Propanol |
| Polybutadiene - Benzene | Polyethylene - Cyclohexanol |
| Poly(epi-chlorohydrin) - Octane | Poly(epi-chlorohydrin) - Hexane |
| Polyethylene, low-density - Heptane | Polysulfone - Aceticacid,methylester |
| Poly(vinylidene fluoride) - Octane,1-chloro | Polystyrene - Aceticacid |
| Poly(vinyl acetate) - Tetradecane | Poly(ethylene succinate) - 2-Butanone |
| Polystyrene - Formicacid,amide | Polystyrene - Benzene |
| Polystyrene - Pentane,3-methyl | Poly(butylene adipate) - Heptane |
| Poly(ethylene oxide) - Ethene,1,1-dichloro | Polyisoprene - Pentane |
| Poly(vinyl acetate) - Ethane,1,2-dichloro | Polystyrene - Heptane |
| Poly(vinylidene fluoride) - Benzene, butyl | Poly(methyl acrylate) - Benzene, butyl |
| Polyisulfone - Methane,trichloro | Poly(ethylene succinate) - Methane,tetrachloro |
| Polystyrene - Aceticacid,nitrile | Polybutadiene - Aceticacid,nitrile |
| Poly(dimethylsiloxane) - Heptane,2-methyl | Poly(vinyl acetate) - 2-Pentanol |
| Poly(n-butyl methacrylate) - Benzene,1,4-dimethyl | Poly(n-butyl methacrylate) - Ethanol,2-ethoxy |
| Poly(ethylene oxide) - Cyclohexane | Poly(vinyl acetate) - Aceticacid,methylester |
| Polyisoprene - Octane | Polystyrene - Decalin(trans) |
| Polyethylene - Benzene, butyl | Poly(n-propyl acrylate) - Aceticacid,ethylester |
| Polybutadiene - Toluene | Poly(n-butyl methacrylate) - 2-Propanol |
| Poly(butylene adipate) - Pentane | Polystyrene - Toluene |
| Polyethylene, low-density - Octane | Poly(n-butyl methacrylate) - Pentane,2,2,4-trimethyl |

Table B.2: Polymer-solvent systems (Contd)

| Polymer-Solvent System | Polymer-Solvent System |
|---|--|
| Polystyrene - Hexane | Poly(n-propyl acrylate) - 2-Butanone |
| Poly(ethyl methacrylate) - Aceticacid,ethylester | Polybutadiene - Hexane |
| Poly(dimethylsiloxane) - Cyclohexane | Poly(ethylene oxide) - Dodecane |
| Poly(methyl methacrylate) - Dodecane | Poly(n-butyl methacrylate) - Methanol |
| Poly(ethyl methacrylate) - 2-Butanone | Poly(methyl methacrylate) - Benzene,ethyl |
| Polysisoprene - Benzene,1,4-dimethyl | Polyamide - Ethanol,2-ethoxy |
| Polysulfone - 1-Propanol | Poly(epi-chlorohydrin) - Methanol |
| Poly(vinyl acetate) - Methane,trichloro | Poly(methyl acrylate) - Heptane,3,4,5-trimethyl |
| Polysulfone - 1,4-Dioxane | Polystyrene - Octane |
| Polyacrylonitrile - Cyclohexanone | Poly(vinyl methyl ether) - 2-Propanone |
| Polyamide - Aceticacid,ethylester | Poly(methyl methacrylate) - Benzaldehyde |
| Poly(hexamethylene sebacate) - Aceticacid,ethylester | Polyamide - 2-Propanol |
| Poly(dimethylsiloxane) - Benzene,ethyl | Polystyrene - Tetralin |
| Polysulfone - Ether,diethyl | Poly(vinyl acetate) - Cyclohexane,ethyl |
| Poly(hexamethylene sebacate) - 2-Butanone | Polyisoprene - 2-Propanol |
| Poly(ethylene oxide) - Benzene | Poly(DL-lactide) - Benzene |
| Polyethylene, low-density - Benzene,1,4-dimethyl | Polyacrylonitrile - Formicacid,amide,N,N-dimethyl |
| Poly(ethylene adipate) - Benzene | Polyisoprene - Pentane,2,2,4-trimethyl |
| Poly(ethyl methacrylate) - 1-Butanol | Poly(epi-chlorohydrin) - Aceticacid,butylester |
| Polysulfone - Heptane | Polyisoprene - Methanol |
| Poly(n-propyl methacrylate) - Aceticacid,ethylester | Poly(methyl methacrylate) - Benzene |
| Poly(hexamethylene sebacate) - Methane,tetrachloro | Poly(vinyl acetate) - 1-Hexene |
| Polystyrene, antishock - Cyclohexane | Poly(dimethylsiloxane) - Cyclohexane,methyl |
| Poly(epi-chlorohydrin) - Ethane,1,1-dichloro | Poly(ethylene oxide) - Aceticacid,nitrile |
| Poly(vinylidene fluoride) - Acetophenone | Poly(n-propyl acrylate) - 2-Propanone |
| Polysulfone - Ethanol | Poly(diethylsiloxane) - Benzene |
| Poly(n-butyl methacrylate) - Nonane | Poly(vinylidene fluoride) - Aceticacid,ethylester |
| Poly(vinyl chloride) - 2-Pentanone,4-methyl | Poly(n-butyl methacrylate) - Decane |
| Poly(epi-chlorohydrin) - Nonane | Poly(ethylene - 2-Pentanone,4-methyl |
| Poly(methyl acrylate) - Benzene,tert-butyl | Poly(vinylidene fluoride) - 2-Butanone |
| Poly(methyl acrylate) - Decalin(cis) | Poly(dimethylsiloxane) - Pentane,3-methyl |
| Poly(dimethylsiloxane) - Benzene | Poly(epi-chlorohydrin) - Decane |
| Poly(vinyl acetate) - 1-Propanol | Poly(butylene adipate) - Aceticacid,ethylester |
| Polyethylene - Hexane,3,4-dimethyl | Poly(ethyl acrylate) - Aceticacid,ethylester |
| Poly(ethyl methacrylate) - 2-Propanone | Polysulfone - Pentane |
| Poly(n-butyl methacrylate) - Methane,dichloro | Polyethylene - 2-Butanone |
| Polystyrene - Hexane,3,3,4,4-tetramethyl | Poly(vinylidene fluoride) - Methane,tetrachloro |
| Poly(ethylene oxide) - Toluene | Poly(n-pentyl methacrylate) - Methane,trichloro |
| Poly(epi-chlorohydrin) - Methane,dichloro | Poly(ethylene oxide) - Hexane |
| Polystyrene - Methanol | Poly(ethylene adipate) - Hexane |
| Polysulfone - Aceticacid,amide,N,N-dimethyl | Poly(vinyl chloride) - Ether,dipropyl |
| Poly(vinyl chloride) - Methane,tetrachloro | Polyethylene - Methane,tetrachloro |
| Poly(ethyl methacrylate) - 1-Octene | Poly(vinyl chloride) - Phenol |
| Poly(hexamethylene sebacate) - 2-Propanone | Poly(diethylsiloxane) - Toluene |
| Polyethylene - Phenol | Poly(methyl methacrylate) - Decane,1-chloro |
| Polybutadiene - Cyclohexanone | Poly(vinyl acetate) - Acetaldehyde |
| Polystyrene, antishock - Benzene | Poly(vinylidene fluoride) - Aceticacid,propylester |
| Polystyrene - Aceticacid,butylester | Polybutadiene - Formicacid,amide,N,N-dimethyl |
| Poly(vinyl acetate) - Ethanol | Poly(dimethylsiloxane) - Hexane |
| Polyisoprene - Methane,dichloro | Poly(ethyl methacrylate) - Cyclohexane |
| Poly(diethylsiloxane) - Hexane | Poly(vinylidene fluoride) - 2-Propanone |
| Polystyrene - Benzene,1,3-dimethyl | Poly(methyl methacrylate) - 1,2-Ethanediol |
| Polyisoprene - Hexane,2-methyl | Polystyrene - Nonane |
| Polysulfone - Benzene,dimethyl(isomernotspecified) | Poly(n-pentyl methacrylate) - 1-Propanol |
| Poly(ethylene succinate) - Heptane | Poly(n-butyl methacrylate) - Benzene,ethyl |
| Polyisobutylene - Methane,tetrachloro | Polystyrene - Ethane,nitro |
| Polystyrene - Decane | Poly(methyl acrylate) - Furane,tetrahydro |
| Poly(vinyl acetate) - 1-Nonene | Poly(ethyl methacrylate) - Benzene,ethyl |
| Polybutadiene - Methane,dichloro | Poly(vinyl chloride) - Furane,tetrahydro |
| Polyethylene - Butane | Poly(methyl methacrylate) - Hexadecane |
| Polysulfone - Diethyleneglycol,dimethylether | Poly(methyl methacrylate) - Cyclohexanol |
| Polyethylene - Furane,tetrahydro | Poly(vinyl acetate) - 2-Pentanone |
| Polyethylene - Heptane,2-methyl | Polystyrene - Methane,dichloro |
| Polysulfone - Aceticacid,ethylester | Polyisoprene - Butane,2-methyl |
| Polyethylene - Furane | Poly(ethylene succinate) - Pentane |
| Polystyrene - Benzene,chloro | Polystyrene, antishock - Hexane |
| Polybutadiene - Aceticacid,ethylester | Polysulfone - Methanol |
| Poly(vinyl acetate) - Methylchloride | Poly(n-butyl methacrylate) - Ethane,1,2-dichloro |
| Poly(vinyl methyl ether) - Octane | Poly(ethylene succinate) - Hexane |
| Poly(methyl methacrylate) - Benzene,ethyl | Polyisoprene - Benzene,ethyl |
| Poly(vinyl acetate) - Ethanol,2-methoxy | Polyethylene, low-density - Cyclohexane |
| Poly(epi-chlorohydrin) - Aceticacid,methylester | Poly(ethyl methacrylate) - Benzene |
| Poly(vinyl acetate) - Aceticacid,ethenylester | Poly(n-propyl acrylate) - Aceticacid,nitrile |
| Poly(ethyl methacrylate) - Aceticacid,nitrile | Poly(ethylene oxide) - Aceticacid,butylester |
| Poly(vinyl acetate) - Aceticacid,ethylester | Polyamide - Ethane,1,2-dichloro |
| Poly(ethylene oxide) - Nonane | Polystyrene - Decalin(cis) |
| Poly(methyl methacrylate) - Formicacid,amide,N,N-dimethyl | Poly(ethyl methacrylate) - Dodecane |
| Poly(vinyl acetate) - 2-Butanone | Poly(vinylidene fluoride) - Dodecane |
| Polystyrene - Aceticacid | Poly(vinyl acetate) - 2-Propanol |
| Poly(vinyl chloride) - Pyridine | Polyisoprene - Ethane,1,2-dichloro |

Table B.3: Polymer-solvent systems (Contd)

| Polymer-Solvent System | Polymer-Solvent System |
|--|--|
| Polystyrene - Tetradecone | Poly(methyl methacrylate) - Cyclohexanone |
| Poly(vinylidene fluoride) - Benzene | Poly(ethyl methacrylate) - Propenoicacid,nitrile |
| Poly(ethyl methacrylate) - Hexane | Poly(ethyl methacrylate) - Propanoicacid,nitrile |
| Poly(ethylene adipate) - Methane,dichloro | Poly(n-propyl acrylate) - Hexane |
| Poly(ethylene oxide) - 2-Heptene | Polysulfone - Decane |
| Polyethylene - 2-Propanol,2-methyl | Poly(n-butyl methacrylate) - Methane,trichloro |
| Poly(vinyl acetate) - 1-Butanol | Poly(methyl acrylate) - 2-Propanol,2-methyl |
| Polystyrene - Ethane,1,1,2-trichloro-1,2,2-trifluoro | Poly(vinyl chloride) - 1,4-Dioxane |
| Poly(hexamethylene sebacate) - Hexane | Poly(epi-chlorohydrin) - Methane,trichloro |
| Poly(ethylene - Heptane | Poly(vinyl chloride) - 2-Propanol,2-methyl |
| Polyisoprene - Benzyllalcohol | Poly(methyl acrylate) - Heptane |
| Poly(vinyl acetate) - 2-Propanone | Poly(ethylene oxide) - Benzene,1,2-dimethyl |
| Poly(methyl methacrylate) - Acetophenone | Poly(vinyl chloride) - Heptane |
| Poly(vinylidene fluoride) - Toluene | Poly(ethylene oxide) - Benzene,chloro |
| Poly(methyl acrylate) - Pentane | Poly(methyl acrylate) - Ethanol |
| Poly(vinyl chloride) - Pentane | Polyethylene - Butane,1-chloro |
| Polyethylene - Decalin(trans) | Polysulfone - Cyclohexane |
| Polyethylene - Toluene | Poly(n-butyl methacrylate) - 1-Propanol |
| Poly(ethylene oxide) - 2-Butanol | Poly(n-butyl methacrylate) - Cyclohexanol |
| Poly(vinyl acetate) - 1-Octene | Poly(n-pentyl methacrylate) - Aceticacid,ethylester |
| Polystyrene - Methane,trichloro | Poly(dimethylsiloxane) - Benzene,1,2-dimethyl |
| Poly(dimethylsiloxane) - Benzene,1,3,5-trimethyl | Poly(ethyl methacrylate) - Butane,1-bromo |
| Polyethylene - Octane | Poly(methyl methacrylate) - 2-Pantanone,4-methyl |
| Poly(dimethylsiloxane) - Cyclooctane | Polyisobutylene - Benzene |
| Poly(ethylene succinate) - Methane,dichloro | Poly(DL-lactide) - 2-Butanone |
| Polysulfone - 2-Pyrrolidone,1-methyl | Poly(vinyl chloride) - Toluene |
| Poly(vinylisobutyl ether) - Decane | Poly(vinylidene fluoride) - Hexane |
| Polyethylene - Tetralin | Poly(methyl methacrylate) - 2-Butanone |
| Poly(dimethylsiloxane) - Methane,tetrachloro | Polybutadiene - Methane,trichloro |
| Polyisobutylene - Hexane | Poly(methyl acrylate) - Octane |
| Poly(vinyl acetate) - Undecane | Poly(DL-lactide) - Methane,tetrachloro |
| Poly(ethyl methacrylate) - Propane,1-nitro | Polyethylene - Hexane |
| Poly(vinyl acetate) - Dodecane | Polystyrene - Ether,diisopropyl |
| Polystyrene - 1-Propanol | Poly(vinyl chloride) - Octane |
| Poly(ethylene oxide) - Aceticacid,methylester | Poly(vinylisobutyl ether) - Nonane |
| Poly(n-butyl methacrylate) - 2-Pantanone | Poly(2-hydroxyethyl methacrylate) - Diethyleneglycol,monomethylether |
| Poly(methyl acrylate) - 2-Propanol | Poly(methyl methacrylate) - Methane,tetrachloro |
| Poly(methyl acrylate) - Pentane,2,2,4-trimethyl | Poly(methyl acrylate) - Tetralin |
| Poly(methyl acrylate) - 1,4-Dioxane | Poly(epi-chlorohydrin) - Ethanol |
| Polystyrene - Ether,diethyl | Poly(methyl methacrylate) - 1-Butanol |
| Poly(methyl acrylate) - Hexane,3,3,4,4-tetramethyl | Polysulfone - Ethane,1,2-dichloro |
| Poly(methyl methacrylate) - Furane,tetrahydro | Poly(vinyl acetate) - Cyclohexane |
| Poly(methyl acrylate) - Methanol | Polyisobutylene - Toluene |
| Poly(vinyl acetate) - Sulfurdioxide | Polyacrylonitrile - Aceticacid,amide,N,N-dimethyl |
| Poly(ethyl methacrylate) - 1-Pentanol | Poly(dimethylsiloxane) - Pentane,2-methyl |
| Poly(vinyl chloride) - Methanol | Polybutadiene - 1-Propanol |
| Polystyrene - Ethanol | Poly(vinyl acetate) - Benzene,ethyl |
| Poly(vinylidene fluoride) - Cyclohexanone | Polybutadiene - 1,4-Dioxane |
| Poly(vinyl chloride) - Aceticacid,butylester | Polysulfone - Formicacid,ethylester |
| Poly(vinyl acetate) - Aceticacid,nitrile | Poly(ethylene oxide) - Furane,tetrahydro |
| Polybutadiene - Pentane | Poly(methyl acrylate) - Aceticacid,tert-butylester |
| Polyacrylonitrile - Methanol | Polybutadiene - Heptane |
| Poly(vinylidene fluoride) - Nonane | Polystyrene - Naphthalene |
| Poly(n-butyl methacrylate) - Aniline | Poly(vinyl acetate) - Benzene |
| Poly(epi-chlorohydrin) - 2-Propanol | Poly(ethylene oxide) - Methane,trichloro |
| Poly(methyl acrylate) - Decane | Poly(ethylene adipate) - Methane,trichloro |
| Polyethylene - Nonane | Poly(vinylidene fluoride) - Aceticacid,butylester |
| Poly(vinyl chloride) - Decane | Poly(hexamethylene sebacate) - Methane,dichloro |
| Poly(vinyl acetate) - Hexane | Poly(n-butyl methacrylate) - Ethanol,2-methoxy |
| Poly(ethylene oxide) - Pyridine | Poly(vinylidene fluoride) - Formicacid,amide,N,N-dimethyl |
| Poly(ethyl methacrylate) - Ether,dipropyl | Polyethylene - Aceticacid,butylester |
| Polyisoprene - Aniline | Polyethylene - Benzene,1,3-dimethyl |
| Poly(vinyl chloride) - Benzene,chlro | Polystyrene - Pentane |
| Poly(DL-lactide) - Heptane | Poly(vinyl acetate) - Propanoicacid,nitrile |
| Poly(ethyl methacrylate) - Butanal | Poly(n-butyl methacrylate) - Aceticacid,ethylester |
| Polystyrene - Benzene,1,4-dimethyl | Poly(vinyl chloride) - Nonane |
| Polysulfone - Acetaldehyde | Poly(n-butyl methacrylate) - 2-Butanone |
| Polybutadiene - 2-Propanol | Poly(epi-chlorohydrin) - Aceticacid,ethylester |
| Poly(diethylsiloxane) - Heptane | Poly(vinyl chloride) - Ethane,nitro |
| Polyethylene - Benzene,1,3,5-trimethyl | Poly(epi-chlorohydrin) - Aceticacid,tert-butylester |
| Polybutadiene - Methanol | Polyamide - Ethanol,2-methoxy |
| Polyethylene - Decalin(cis) | Polyamide - 2-Pantanone,4-methyl |
| Polystyrene - Aceticacid,ethylester | Poly(methyl acrylate) - Methane,dichloro |
| Poly(butylene adipate) - 2-Butanone | Poly(n-butyl methacrylate) - Methane,tetrachloro |
| Poly(n-butyl methacrylate) - Heptane,2-methyl | Polyethylene - Decane |
| Poly(ethyl acrylate) - 2-Butanone | Poly(ethylene oxide) - 1,4-Dioxane |
| Poly(ethylene oxide) - Pentane | Poly(vinyl chloride) - Methane,dichloro |
| Poly(methyl acrylate) - Benzene,ethyl | Poly(ethyl methacrylate) - Methane,tetrachloro |
| Poly(epi-chlorohydrin) - Ethane,1,1,1-trichloro | Poly(n-butyl methacrylate) - 1-Butanol |
| Poly(ethylene adipate) - Pentane | Polyamide - 2-Butanone |

Table B.4: Polymer-solvent systems (Contd)

| Polymer-Solvent System | Polymer-Solvent System |
|--|---|
| Polystyrene - Aceticacid,tert-butylester | Poly(ethylene oxide) - Heptane |
| Poly(methyl methacrylate) - Toluene | Polybutadiene - Benzene,1,4-dimethyl |
| Poly(butylene adipate) - Methane,tetrachloro | Polyisoprene - 2-Butanone |
| Poly(methyl acrylate) - Tetradecane | Poly(ethylene adipate) - Heptane |
| Poly(methyl methacrylate) - Octane | Polystyrene - Cyclopentane |
| Poly(vinyl acetate) - Benzene, butyl | Poly(ethylene succinate) - Methane,trichloro |
| Poly(dimethylsiloxane) - Pentane | Poly(ethylene - Benzene, chloro |
| Poly(n-butyl methacrylate) - Cyclohexane | Poly(n-propyl methacrylate) - 2-Butanone |
| Poly(vinyl chloride) - Ethane,1,2-dichloro | Poly(vinyl acetate) - 1-Decene |
| Polyisoprene - Heptane,2-methyl | Poly(ethyl methacrylate) - Aceticacid,propylester |
| Poly(epi-chlorohydrin) - Cyclohexane | Poly(dimethylsiloxane) - 1,4-Dioxane |
| Poly(methyl methacrylate) - Tetralin | Poly(ethylene oxide) - Ethanol |
| Poly(n-hexyl methacrylate) - Methane,trichloro | Poly(ethylene oxide) - Diethyleneglycol,dieithylether |
| Poly(hexamethylene sebacate) - Methane,trichloro | Polyisoprene - Methane,tetrachloro |
| Poly(vinylidene fluoride) - Aceticacid,methylester | Polyamide - 1-Butanol |
| Poly(butylene adipate) - 2-Propanone | Poly(dimethylsiloxane) - Cycloheptane |
| Polyacrylonitrile - Benzaldehyde | Poly(n-butyl methacrylate) - 2-Propanone |
| Poly(ethyl acrylate) - 2-Propanone | Polyethylene - Hexane,2,5-dimethyl |
| Polystyrene - 2-Propanone | Polyisobutylene - Methane,dichloro |
| Poly(n-butyl methacrylate) - Ethene,trichloro | Poly(methyl methacrylate) - Aceticacid,nitrile |
| Polyacrylonitrile - Tetradeccane | Poly(styrene - Benzene,propyl |
| Poly(n-butyl methacrylate) - Benzaldehyde | Poly(styrene - 2-Propanone |
| Poly(epi-chlorohydrin) - Ethene,trichloro | Polyisoprene - 1-Butanol |
| Polyisoprene - Cyclohexane | Poly(styrene - Pentane,2,2,4-trimethyl |
| Poly(ethylene oxide) - 2-Propanol | Poly(vinyl acetate) - Cyclohexanol |
| Poly(dimethylsiloxane) - Cyclopentane | Poly(dimethylsiloxane) - Heptane |
| Poly(ethylene oxide) - Methanol | Poly(methyl acrylate) - Undecane |
| Poly(ethylene oxide) - Ethane,1,1,2-trichloro | Poly(vinyl acetate) - Cyclohexane, butyl |
| Poly(n-butyl methacrylate) - Aceticacid | Poly(DL-lactide) - Pentane |
| Polyulfone - Sulfone,tetramethylene | Poly(DL-lactide) - Hexane |
| Poly(methyl acrylate) - Methane,trichloro | Poly(styrene - Trichlorofluoromethane |
| Poly(n-butyl methacrylate) - Benzene | Poly(ethylene oxide) - Octane |
| Poly(epi-chlorohydrin) - Ethane,1,2-dichloro | Poly(n-propyl acrylate) - Methane,trichloro |
| Poly(n-hexyl methacrylate) - 1-Propanol | Polyisoprene - 2-Propanone |
| Poly(vinyl chloride) - Methane,trichloro | Poly(styrene - Benzene,1,2-dichloro |
| Poly(epi-chlorohydrin) - Benzene | Poly(n-propyl methacrylate) - 2-Propanone |
| Polybutadiene - Cyclohexane | Poly(dimethylsiloxane) - Toluene |
| Polysulfone - Phenol | Poly(dimethylsiloxane) - Pentane,2,4-dimethyl |
| Polyisoprene - Benzaldehyde | Poly(vinyl acetate) - Cyclohexanone |
| Poly(ethyl methacrylate) - Heptane | Poly(dimethylsiloxane) - Octane |
| Polyethylene - Benzylalcohol | Poly(vinylidene fluoride) - Furane,tetrahydro |
| Poly(ethylene oxide) - Methane,nitro | Poly(ethylene oxide) - Benzene,1,4-dimethyl |
| Polyamide - Benzene | Poly(epi-chlorohydrin) - Undecane |
| Poly(ethylene oxide) - Benzene,1,3-dimethyl | Polyethylene - Hexane,3-methyl |
| Polybutadiene - Benzene,ethyl | Poly(ethylene oxide) - Diethyleneglycol,dimethylether |
| Polyisoprene - Aceticacid | Polysulfone - 2-Butanone |
| Poly(hexamethylene sebacate) - Heptane | Poly(ethylene adipate) - Aceticacid,ethylester |
| Polyisoprene - Benzene | Poly(ethyl methacrylate) - Cyclohexane,methyl |
| Poly(n-butyl methacrylate) - Toluene | Poly(dimethylsiloxane) - Benzene,1,4-dimethyl |
| Polybutadiene - Benzaldehyde | Poly(methyl methacrylate) - Methanol |
| Polyisobutylene - Methane,trichloro | Polysulfone - Methane,tetrachloro |
| Poly(methyl acrylate) - 1-Propanol | Polyethylene - Methane,trichloro |
| Poly(vinyl chloride) - 1-Propanol | Poly(vinyl chloride) - Benzylalcohol |
| Poly(ethyl methacrylate) - Toluene | Poly(styrene - Cyclohexane |
| Poly(ethylene oxide) - Decane | Polysulfone - 1-Butanol |
| Poly(methyl methacrylate) - Nonane | Poly(n-butyl methacrylate) - Aceticacid,nitrile |
| Polystyrene - Benzene,hexyl | Poly(vinyl acetate) - 2-Pentanone,4-methyl |
| Poly(vinylisobutyl ether) - Benzene,1,4-dimethyl | Poly(ethyl methacrylate) - Benzene,bromo |
| Poly(dimethylsiloxane) - Benzene,1,3-dimethyl | Poly(styrene - Benzene,ethyl |
| Polyethylene, low-density - Benzene | Poly(styrene - Ethene,trichloro |
| Polysulfone - Furane,tetrahydro | Polyamide - Aceticacid,nitrile |
| Polybutadiene - Ethane,1,2-dichloro | Polysulfone - 2-Propanone |
| Poly(DL-lactide) - Methane,dichloro | Poly(styrene - Benzaldehyde |
| Poly(ethyl methacrylate) - Octane | Poly(vinyl acetate) - 2-Butanol |
| Polyamide - Toluene | Poly(n-butyl methacrylate) - Hexane |
| Poly(2-hydroxyethyl methacrylate) - Formicacid, amide,N,N-dimethyl | Poly(vinyl chloride) - Butylacrylate |
| Polyacrylonitrile - 1,2-Ethanediol | Poly(vinylidene fluoride) - 1,4-Dioxane |
| Poly(methyl methacrylate) - Methane,dichloro | Poly(vinyl acetate) - Methane,tetrachloro |
| Polysulfone - Benzene,fluoro | Polyisoprene - Aceticacid,nitrile |
| Poly(n-butyl methacrylate) - 1,2-Ethanediol | Poly(ethylene oxide) - Methane,dichloro |
| Polyisoprene - Toluene | Poly(methyl methacrylate) - Decane |
| Poly(vinyl chloride) - Ethanol | Poly(n-propyl methacrylate) - Aceticacid,nitrile |
| Polyacrylonitrile - Hexadecane | Poly(ethylene succinate) - Aceticacid,ethylester |
| Polybutadiene - Formicacid, ethylester | Poly(styrene - Ethane,1,2-dichloro |
| Poly(ethyl acrylate) - Aceticacid,nitrile | Poly(butylene adipate) - Benzene |
| Polyethylene, low-density - Decalin(trans) | Poly(vinylidene fluoride) - Heptane |
| Polyethylene, low-density - Toluene | Poly(hexamethylene sebacate) - Pentane |
| Poly(n-pentyl methacrylate) - 2-Butanone | Poly(dimethylsiloxane) - Nonane |
| Poly(vinylidene fluoride) - Pentane | Poly(methyl acrylate) - Naphthalene |
| Poly(vinyl acetate) - Furane,tetrahydro | |

Table B.5: Polymer-solvent systems (Contd)

| Polymer-Solvent System | Polymer-Solvent System |
|--|---|
| Poly(n-butyl methacrylate) - Cyclohexane,butyl | Poly(vinyl methyl ether) - Aceticacid,ethylester |
| Poly(vinyl acetate) - Heptane,2-methyl | Poly(vinyl acetate) - Aceticacid,propylester |
| Poly(ethylene oxide) - 1,2-Ethanediol,dimethylether | Polystyrene - 2-Octene(cis) |
| Polyisobutylene - Heptane | Poly(vinyl chloride) - Acetaldehyde |
| Poly(ethylene oxide) - Benzene,ethyl | Polyisoprene - Hexane |
| Polyisoprene - 1,2-Ethanediol | Poly(dimethylsiloxane) - Methane,dichloro |
| Polystyrene - Diethyleneglycol,monomethylether | Poly(n-propyl methacrylate) - Hexane |
| Poly(butylene adipate) - Hexane | Poly(dimethylsiloxane) - Hexane,2-methyl |
| Poly(ethyl acrylate) - Hexane | Polyethylene, low-density - Hexane |
| Poly(dimethylsiloxane) - Butane,2-methyl | Poly(ethylene oxide) - Undecane |
| Poly(vinyl chloride) - Propane,2-nitro | Polyethylene - Heptane,3-methyl |
| Polyethylene - 2-Pentanone | Poly(ethyl methacrylate) - Propanoicacid |
| Poly(n-hexyl methacrylate) - Aceticacid,ethylester | Poly(vinylidene fluoride) - Octane |
| Poly(n-hexyl methacrylate) - 2-Butanone | Polysulfone - Sulfoxide,dimethyl |
| Poly(methyl acrylate) - Hexane,2,2,5-trimethyl | Polybutadiene - Benzylalcohol |
| Poly(ethylene oxide) - Ethane,1,2-dichloro | Polyethylene - 1-Nonene |
| Poly sulfone - Benzene | Poly(ethylene succinate) - 2-Propanone |
| Poly(n-butyl methacrylate) - Cyclohexanone | Polyacrylonitrile - Benzene,butyl |
| Poly(n-pentyl methacrylate) - 2-Propanone | Poly(vinylidene fluoride) - Tetralin |
| Polyisobutylene - Octane | Poly(n-butyl methacrylate) - Benzene,butyl |
| Poly(methyl methacrylate) - Propenoicacid,2-methyl,methylester | Polysulfone - Cyclohexane,methyl |
| Poly sulfone - Propenoicacid,nitrile | Poly sulfone - Pyridine |
| Poly(ethyl methacrylate) - Nonane | Polyethylene - Propane,2-nitro |
| Poly(methyl acrylate) - Aceticacid,ethylester | Poly(vinylidene fluoride) - Methylchloride |
| Poly sulfone - Propenoicacid,nitrile | Poly(methyl methacrylate) - Tetradecone |
| Poly(vinyl methyl ether) - Cyclohexane | Polyisobutylene - Pentane |
| Poly(ethyl methacrylate) - Ethane,nitro | Poly(n-butyl methacrylate) - Aceticacid, butylester |
| Poly(methyl acrylate) - 2-Butanone | Polystyrene - 1,2-Ethanediol |
| Poly(vinyl chloride) - Aceticacid,ethylester | Polyethylene - Methylchloride |
| Polyethylene - 2-Propanol | Poly(methyl methacrylate) - Ethane,1,2-dichloro |
| Polyethylene - Pentane,2,2,4-trimethyl | Poly(ethyl methacrylate) - Methane,nitro |
| Poly(vinyl acetate) - 1,4-Dioxane | Poly(ethyl methacrylate) - Aceticacid, butylester |
| Polyisoprene - Cyclohexanone | Polysulfone - Butane,1-chloro |
| Poly(dimethylsiloxane) - Hexane,3-methyl | Poly(n-butyl methacrylate) - Formicacid, amide,N,N-dimethyl |
| Poly sulfone - Toluene | Polystyrene - Hexadecane |
| Poly(vinyl acetate) - 2-Propanol,2-methyl | Polyethylene - Benzene,1,4-dimethyl |
| Poly(vinyl chloride) - 2-Butanone | Polysulfone - Aceticacid,nitrile |
| Poly(n-hexyl methacrylate) - 2-Propanone | Polyamide - Aceticacid, butylester |
| Poly(methyl acrylate) - Methane,tetrachloro | Poly(vinyl chloride) - 2-Propanol |
| Poly(vinyl methyl ether) - Benzene,ethyl | Polyethylene - Aceticacid, ethenylester |
| Poly(vinylidene fluoride) - Benzene,1,2-dichloro | Polystyrene - Cyclohexane, butyl |
| Poly(vinyl acetate) - Heptane | Poly(vinyl acetate) - Cyclohexane, methyl |
| Poly styrene - Benzene, butyl | Poly(ethyl methacrylate) - Decane |
| Poly sulfone - Octane | Polyisoprene - Formicacid, amide,N,N-dimethyl |
| Poly(vinyl chloride) - 1-Butanol | Polysulfone - Hexane |
| Poly(n-butyl methacrylate) - 2-Pentanone,4-methyl | Polybutadiene - Benzene, butyl |
| Poly(methyl methacrylate) - Benzylalcohol | Poly(DL-lactide) - Methane,trichloro |
| Poly(vinyl acetate) - Hexane,1-chloro | Poly(n-butyl methacrylate) - Benzene, chloro |
| Poly(vinyl acetate) - Diethyleneglycol,diethylether | Poly(vinyl acetate) - Butane,1-chloro |
| Poly(ethylene succinate) - Benzene | Polystyrene - Benzene, methoxy |
| Poly(n-butyl methacrylate) - Benzene,tert-butyl | Poly(methyl methacrylate) - Methane, trichloro |
| Poly(dimethylsiloxane) - Methane,trichloro | Poly(ethyl methacrylate) - Benzene, chloro |
| Polystyrene - Formicacid, amide,N,N-dimethyl | Polyacrylonitrile - Aniline |
| Poly(vinylidene fluoride) - Decane | Polyethylene - 1-Butanol |
| Poly(methyl acrylate) - 2-Propanone | Polystyrene - Cyclohexanone |
| Poly(vinyl acetate) - Toluene | Poly(ethylene oxide) - 1-Hexene |
| Poly(epi-chlorohydrin) - 2-Butanone | Poly(vinyl acetate) - Pentane,1-chloro |
| Poly(vinyl chloride) - 2-Propanone | Poly(vinyl acetate) - Pentane |
| Poly(ethylene oxide) - 1-Propanol | Polyethylene - 2-Propanone |
| Polystyrene - 1-Pentanol | Poly(n-butyl methacrylate) - Phenol |
| Poly(vinyl acetate) - 1-Heptene | Poly(butylene adipate) - Methane, dichloro |
| Poly(epi-chlorohydrin) - Methane,tetrachloro | Poly(vinyl chloride) - Benzene, fluoro |
| Polyisoprene - 2-Pentanone,4-methyl | Polyethylene - 1-Octene |
| Poly(vinyl acetate) - Octane | Polysulfone - 3-Pentanone |
| Poly(ethylene oxide) - Ether,diethyl | Polystyrene - 1-Octanol |
| Poly(vinyl acetate) - Tetralin | Poly(dimethylsiloxane) - Disiloxane,hexamethyl |
| Polystyrene - Benzene,1,2-dimethyl | Poly(methyl acrylate) - Cyclohexane |
| Poly(n-pentyl methacrylate) - Aceticacid,nitrile | Polystyrene - Benzene, nitro |
| Poly(epi-chlorohydrin) - Aceticacid, propylester | Poly(methyl methacrylate) - Ethanol |
| Polybutadiene - Aniline | |

Bibliography

- [1] Jozef Bicerano. *Prediction of polymer properties*. cRc Press, 2002.
- [2] Wilson CK Poon and David Andelman. *Soft condensed matter physics in molecular and cell biology*. CRC Press, 2006.
- [3] Paul C Hiemenz and Timothy P Lodge. Polymer chemistry 2nd edition. *NICE (News & Information for Chemical Engineers)*, 25(4):409–409, 2007.
- [4] JW Kennedy, M Gordon, and R Koningsveld. Generalization of the flory-huggins treatment of polymer solutions. In *Journal of Polymer Science Part C: Polymer Symposia*, volume 39, pages 43–69. Wiley Online Library, 1972.
- [5] Ronald Koningsveld, Ronald Koningsveld, Walter H Stockmayer, and Erik Nies. *Polymer phase diagrams: a textbook*. Oxford University Press on Demand, 2001.
- [6] Alfred Rudin. *Elements of Polymer Science & Engineering: An Introductory Text and Reference for Engineers and Chemists*. Elsevier, 1998.
- [7] Nam Hyung Kim, Yong Sik Won, and Joong So Choi. Partial molar heat of mixing at infinite dilution in solvent/polymer (peg, pmma, p (et-va)) solutions. *Fluid phase equilibria*, 146(1-2):223–246, 1998.
- [8] Joseph Mauk Smith, Hendrick C Van Ness, Michael M Abbott, and Mark Thomas Swihart. *Introduction to chemical engineering thermodynamics*. McGraw-Hill Singapore, 1949.
- [9] Jürgen Gmehling, Michael Kleiber, Bärbel Kolbe, and Jürgen Rarey. *Chemical thermodynamics for process simulation*. John Wiley & Sons, 2019.
- [10] Clare McCabe and Amparo Galindo. Saft associating fluids and fluid mixtures. *Applied thermodynamics of fluids*, 8:215–279, 2010.
- [11] Gareth J Price, Simon J Hickling, and Ian M Shillcock. Applications of inverse gas chromatography in the study of liquid crystalline stationary phases. *Journal of Chromatography A*, 969(1-2):193–205, 2002.
- [12] D Patterson, YB Tewari, HP Schreiber, and JE Guillet. Application of gas-liquid chromatography to the thermodynamics of polymer solutions. *Macromolecules*, 4(3):356–359, 1971.

- [13] Paul J Flory. *Principles of polymer chemistry*. Cornell university press, 1953.
- [14] Agustín Etxeberria, Jokin Alfageme, Cristina Uriarte, and Juan J Iruin. Inverse gas chromatography in the characterization of polymeric materials. *Journal of Chromatography A*, 607(2):227–237, 1992.
- [15] Aage Fredenslund, Russell L Jones, and John M Prausnitz. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AICHE Journal*, 21(6):1086–1099, 1975.
- [16] Aage Fredenslund. *Vapor-liquid equilibria using UNIFAC: a group-contribution method*. Elsevier, 2012.
- [17] Roberta Ceriani and Antonio JA Meirelles. Predicting vapor–liquid equilibria of fatty systems. *Fluid Phase Equilibria*, 215(2):227–236, 2004.
- [18] Henrik K Hansen, Peter Rasmussen, Aage Fredenslund, Martin Schiller, and Jürgen Gmehling. Vapor-liquid equilibria by unifac group contribution. 5. revision and extension. *Industrial & Engineering Chemistry Research*, 30(10):2352–2355, 1991.
- [19] A_ Bondi. A correlation of the entropy of fusion of molecular crystals with molecular structure. *Chemical Reviews*, 67(5):565–580, 1967.
- [20] Dortmund data bank. www.dortmunddatabank.com, 2021.
- [21] Antje Jakob, Hans Grensemann, Jürgen Lohmann, and Jürgen Gmehling. Further development of modified unifac (dortmund): revision and extension 5. *Industrial & Engineering Chemistry Research*, 45(23):7924–7933, 2006.
- [22] Chongli Zhong, Yoshiyuki Sato, Hirokatsu Masuoka, and Xiaoning Chen. Improvement of predictive accuracy of the unifac model for vapor-liquid equilibria of polymer solutions. *Fluid Phase Equilibria*, 123(1-2):97–106, 1996.
- [23] Georgia D Pappa, Epaminondas C Voutsas, and Dimitrios P Tassios. Prediction of activity coefficients in polymer and copolymer solutions using simple activity coefficient models. *Industrial & engineering chemistry research*, 38(12):4975–4984, 1999.
- [24] Hamid Reza Radfarnia, Georgios M Kontogeorgis, Cyrus Ghotbi, and Vahid Taghikhani. Classical and recent free-volume models for polymer solutions: A comparative evaluation. *Fluid phase equilibria*, 257(1):63–69, 2007.

- [25] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.
- [26] Edgar Ivan Sanchez Medina, Steffen Linke, Martin Stoll, and Kai Sundmacher. Graph neural networks for the prediction of infinite dilution activity coefficients. *Digital Discovery*, 1(3):216–225, 2022.
- [27] Edgar Ivan Sanchez Medina, Steffen Linke, Martin Stoll, and Kai Sundmacher. Gibbs-helmholtz graph neural network: capturing the temperature dependency of activity coefficients at infinite dilution. *arXiv preprint arXiv:2212.01199*, 2022.
- [28] Fabian Jirasek, Rodrigo AS Alves, Julie Damay, Robert A Vandermeulen, Robert Bamler, Michael Bortz, Stephan Mandt, Marius Kloft, and Hans Hasse. Machine learning in thermodynamics: Prediction of activity coefficients by matrix completion. *The journal of physical chemistry letters*, 11(3):981–985, 2020.
- [29] Julie Damay, Fabian Jirasek, Marius Kloft, Michael Bortz, and Hans Hasse. Predicting activity coefficients at infinite dilution for varying temperatures by matrix completion. *Industrial & Engineering Chemistry Research*, 60(40):14564–14578, 2021.
- [30] Benedikt Winter, Clemens Winter, Johannes Schilling, and André Bardow. A smile is all you need: predicting limiting activity coefficients from smiles with natural language processing. *Digital Discovery*, 1(6):859–869, 2022.
- [31] Paula B Staudt, Renata L Simões, Leonardo Jacques, Nilo SM Cardozo, and Rafael de P Soares. Predicting phase equilibrium for polymer solutions using cosmo-sac. *Fluid Phase Equilibria*, 472:75–84, 2018.
- [32] Ntebogeng Mqoni, Sangeeta Singh, Indra Bahadur, Hamed Hashemi, and Deresh Ramjugernath. Ionic liquids, the mixture of ionic liquids and their co-solvent with n, n-dimethylformamide as solvents for cellulose using experimental and cosmo study. *Results in Engineering*, 15:100484, 2022.
- [33] W Hao, HS Elbro, and P Alessi. Polymer solution data collection: Chemistry data series. *DECHEMA, Frankfurt, Germany*, 14, 1992.

- [34] Lei Tao, Vikas Varshney, and Ying Li. Benchmarking machine learning models for polymer informatics: an example of glass transition temperature. *Journal of Chemical Information and Modeling*, 61(11):5395–5413, 2021.
- [35] Jaehong Park, Youngseon Shim, Franklin Lee, Aravind Rammohan, Sushmit Goyal, Munbo Shim, Changwook Jeong, and Dae Sin Kim. Prediction and interpretation of polymer properties using the graph convolutional network. *ACS Polymers Au*, 2(4):213–222, 2022.
- [36] Gary Bradski. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000.
- [37] Supriya Kurlekar, O Deshpande, A Kamble, A Omanna, and D Patil. Reading device for blind people using python, ocr and gtts. *International journal of Science and Engineering Applications*, 9(4):049–052, 2020.
- [38] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.
- [39] Pubchem. <https://pubchem.ncbi.nlm.nih.gov>, 2021.
- [40] Matteo Aldeghi and Connor W Coley. A graph representation of molecular ensembles for polymer property prediction. *Chemical Science*, 13(35):10486–10498, 2022.
- [41] Shingo Otsuka, Isao Kuwajima, Junko Hosoya, Yibin Xu, and Masayoshi Yamazaki. Polyinfo: Polymer database for polymeric materials design. In *2011 International Conference on Emerging Intelligent Data and Web Technologies*, pages 22–29. IEEE, 2011.
- [42] Polymer database. www.polymerdatabase.com, 2015-2022.
- [43] Wolf D Ihlenfeldt, Evan E Bolton, and Stephen H Bryant. The pubchem chemical structure sketcher. *Journal of cheminformatics*, 1:1–9, 2009.
- [44] Gaël Varoquaux, Lars Buitinck, Gilles Louppe, Olivier Grisel, Fabian Pedregosa, and Andreas Mueller. Scikit-learn: Machine learning without learning the machinery. *GetMobile: Mobile Computing and Communications*, 19(1):29–33, 2015.

- [45] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.
- [46] Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8, 2013.
- [47] Dávid Bajusz, Anita Rácz, and Károly Héberger. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics*, 7(1):1–13, 2015.
- [48] Thomas Brouwer and Boelo Schuur. Model performances evaluated for infinite dilution activity coefficients prediction at 298.15 k. *Industrial & Engineering Chemistry Research*, 58(20):8903–8914, 2019.
- [49] Shiyi Qin, Shengli Jiang, Jianping Li, Prasanna Balaprakash, Reid C Van Lehn, and Victor M Zavala. Capturing molecular interactions in graph neural networks: a case study in multi-component phase equilibrium. *Digital Discovery*, 2022.
- [50] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.