PolyRecommender: A Multimodal Recommendation System for Polymer Discovery

Xin Wang[*], **Yunhao Xiao, Rui Qiao**[*]
Department of Mechanical Engineering
Virginia Tech
Blacksburg, VA 24060
{xinwang, xyunhao, ruiqiao}@vt.edu

Abstract

We introduce PolyRecommender, a multimodal discovery framework that integrates chemical language representations from PolyBERT with molecular graph-based representations from a graph encoder. The system first retrieves candidate polymers using language-based similarity and then ranks them using fused multimodal embeddings according to multiple target properties. By leveraging the complementary knowledge encoded in both modalities, PolyRecommender enables efficient retrieval and robust ranking across related polymer properties. Our work establishes a generalizable multimodal paradigm, advancing AI-guided design for the discovery of next-generation polymers.

1 Introduction

The rational design of novel polymers is a grand challenge in materials science, with the potential to unlock breakthroughs in sustainable energy, advanced manufacturing, and medicine [1]. However, the chemical space of known polymers is astronomically large, making exhaustive experimental screening for specific applications intractable. This creates a critical need for AI-guided discovery frameworks that can intelligently navigate this landscape to recommend candidates with targeted property profiles. The primary bottleneck for such data-driven systems is the development of a polymer representation that is both computationally efficient and chemically informative.

Most prior work adopts unimodal molecular representations. Graph neural networks (GNNs) leverage explicit bond topology and perform well in smaller data regimes but can underrepresent higher-level chemical semantics [2, 3]. Conversely, transformer models trained on SMILES strings capture chemical "grammar" but may lose critical structural information [4–6]. Relying on any single modality provides an incomplete view of a material, limiting a model's ability to generalize and hindering the full potential of AI-guided material design.

To address these limitations and advance the paradigm of polymer design, we introduce PolyRecommender. Our framework operationalizes a two-stage "funnel" architecture [7, 8], a design crucial for practical discovery workflows that require both efficient exploration and precise ranking (Figure 1a). The first stage leverages a fine-tuned language model for rapid candidate retrieval from a space of 12,441 polymers, making the initial search computationally tractable. In the second, multimodal ranking stage, we fuse language and graph embeddings to perform a more holistic and accurate evaluation of the top candidates. After systematically investigating three fusion strategies [9], our results demonstrate that this multimodal approach consistently outperforms single-modality baselines (Figure 1b). This work establishes a powerful and scalable blueprint for next-generation AI-guided discovery systems, effectively integrating chemical language and structural data to accelerate the design of novel polymers.

[*]Corresponding authors.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI4Mat-NeurIPS-2025.

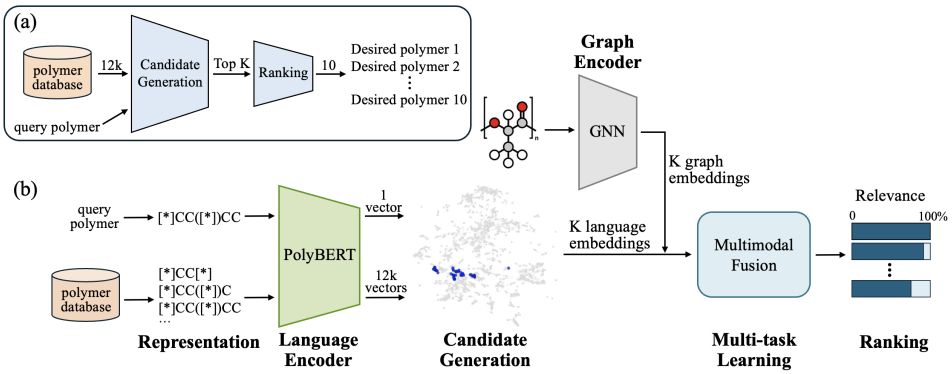

Figure 1: (a) The polymer recommender where polymers are recalled and ranked based on the similarity to the query polymer. (b) The detailed workflow to recommend candidates from the search space including material representation generation, candidate retrieval, fusion of multimodal representations, multi-task prediction for ranking.

2 Related Work

The machine learning-driven approach to polymer property prediction has evolved from traditional descriptor-based methods to sophisticated deep learning architectures. The current paradigm is dominated by two powerful, complementary axes of representation learning. First, GNNs directly leverage the molecular graph, enabling models to learn from the explicit topology of atoms and bonds. This approach has proven highly effective, with multitask frameworks demonstrating the capacity to generate robust, generalizable representations [2, 3]. In parallel, progress in natural language processing has inspired the use of Transformer-based models to treat SMILES strings as a chemical language. Models like PolyBERT [5] and TransPolymer [6] learn to encode rich chemical syntax and semantics into dense vector representations. Han et al.[10] established a multimodal transformer fusing semantic and structural embeddings, leading to superior performance in multitask property prediction. Our work operationalizes this validated multimodal synergy within a practical discovery framework, using language representations for efficient candidate retrieval and combined multimodal representations for high-precision ranking.

3 Methodology

3.1 Dataset

Our experiments are conducted on a dataset of 12,441 synthesized polymers curated from the PolyInfo database [11]. For each polymer, we use its SMILES representation and three experimental properties: glass-transition temperature (T_{g}), melting temperature (T_{m}), and band gap (E_{g}). The dataset was divided into training, validation, and test sets in a ratio of 80:10:10. We train all multitask models using a masked mean-squared-error (MSE) loss that is computed solely over available ground-truth labels.

3.2 Dual-modality polymer representations

To create a comprehensive polymer representation, we employ a dual-modality approach that fuses language embeddings, derived from SMILES strings, with graph embeddings that encode the molecular topology.

Language Embeddings. For our language-based modality, we leverage PolyBERT [5], a transformer pre-trained on a vast corpus of 100 million polymer SMILES. To adapt this powerful base model for our specific predictive tasks, we employ parameter-efficient fine-tuning using Low-Rank Adaptation (LoRA) [12], which modifies the network's attention layers. The resulting fine-tuned model processes polymer SMILES strings (truncated to 160 tokens) to generate 600-dimensional language embeddings.

Graph Embeddings. To capture structural and topological information, we generate graph embeddings using a Directed Message Passing Neural Network (D-MPNN) [13, 14]. Each polymer was represented as a molecular graph, where nodes correspond to atoms and edges represent covalent

bonds. Node features include atomic number, degree, formal charge, hybridization state, aromaticity, and hydrogen count, while edge features encode bond type and conjugation status. The D-MPNN, architected with 5 message-passing layers and a 512-dimensional hidden state, was trained in a multi-task regression framework to predict three key polymer properties. This pre-training step compels the network to learn chemically meaningful representations, from which the final 512-dimensional graph embeddings are extracted for downstream fusion.

3.3 Multi-modal fusion

We investigate three fusion architectures that combine the frozen, pre-computed language (z^{lang}) and graph (z^{graph}) embeddings, training each via multitask regression to predict our three target properties.

Early Fusion. Following the "shared-bottom" multitask paradigm [15], our early fusion model first concatenates the language (z^{lang}) and graph (z^{graph}) embeddings (Eq. 1), then processes the resulting vector through a shared 3-layer MLP to produce the 3-dimensional property prediction (Eq. 2):

$$x = [z^{\text{lang}}; z^{\text{graph}}] \tag{1}$$

$$y = h(x) \in \mathbb{R}^3 \tag{2}$$

Gated Late Fusion. To enable modality-specific processing, we implement a gated late fusion model. First, the language (z^{lang}) and graph (z^{graph}) embeddings are independently processed by two dedicated 3-layer MLP "experts" to produce per-task predictions:

$$y^{\text{lang}} = h^{\text{lang}}(z^{\text{lang}}), \qquad y^{\text{graph}} = h^{\text{graph}}(z^{\text{graph}}) \tag{3}$$

A separate gating network (a 2-layer MLP) then takes the concatenated embeddings as input and outputs a task-specific *gating vector* $\mathbf{g}$. The final prediction of task k is a weighted combination of the expert outputs, dynamically controlled by g_k:

$$\mathbf{g} = \sigma\big(W_g[z^{\text{lang}}; z^{\text{graph}}]\big) \in \mathbb{R}^3 \tag{4}$$

$$y_k = g_k\, y_k^{\text{lang}} \;+\; (1 - g_k)\, y_k^{\text{graph}} \tag{5}$$

Multi-gate Mixture-of-Experts (MMoE). The MMoE model [16] processes the concatenated input $x = [z^{\text{lang}}; z^{\text{graph}}]$ using n shared experts $\{f_i\}_{i=1}^{n}$ where $n = 4$. For each task k, a dedicated gating network $g^{(k)}(x)$ produces a softmax distribution over experts. A task-specific tower $h^{(k)}$ then maps the gated expert mixture to the final prediction y_k:

$$f^{(k)}(x) = \sum_{i=1}^{n} g_i^{(k)}(x)\, f_i(x) \tag{6}$$

$$y_k = h^{(k)}\big(f^{(k)}(x)\big) \tag{7}$$

$$\text{where} \quad g^{(k)}(x) = \mathrm{softmax}(W_g^{(k)} x) \in \mathbb{R}^n \tag{8}$$

4 Results

We developed PolyRecommender, a two-stage multimodal recommendation system designed to efficiently search large chemical spaces for polymers with desired properties. The system employs a "funnel" architecture consisting of two sequential stages: candidate retrieval and multimodal ranking. In the retrieval stage, we use embeddings from a fine-tuned PolyBERT model [5] to represent each polymer. Based on the cosine similarity between these language embeddings, the system retrieves the top 100 candidates most relevant to a given query polymer. In the ranking stage, we refine this list using a multimodal approach that fuses the language embeddings with structural graph embeddings from a GNN. After systematically evaluating three fusion strategies, we selected the MMoE for its superior overall performance in predicting and ranking candidates based on their target properties. The final ranking score is defined in the Appendix.

To validate the quality of our multimodal representations before the final ranking, we visualized the concatenated language and graph embeddings for all 12,441 polymers using a two-dimensional UMAP projection (Figure 2). The resulting map reveals distinct clusters corresponding to key polymer properties, which demonstrates that our embeddings successfully capture chemically meaningful relationships and effectively structure the chemical space.

Table 1 presents a comprehensive ablation study to validate the effectiveness of each component within PolyRecommender, alongside a performance comparison to a state-of-the-art (SOTA) multimodal baseline [10]. Our results show that MMoE fusion achieves the best overall performance among the tested fusion strategies in predicting the glass transition temperature (T_g) and band gap (E_g). While the gated late fusion model showed a marginal advantage for melting temperature (T_m), MMoE provided the best-balanced performance across all tasks. Consistent with prior work [5, 10], we found that predicting T_m is markedly more challenging than predicting T_g or E_g. We attribute this difficulty to a potentially weaker correlation between a polymer's melting temperature and its molecular structure.

Our MMoE model demonstrates clear expert specialization (Figure 3a), successfully learning task-specific representations for predicting distinct polymer properties (T_g, T_m, and E_g) from a shared input. A case study using Polyethylene oxide (PEO) as a query highlights our system's practical utility. As shown in Figure 3b-c, the top 50 recommended candidates not only cluster near PEO in the chemical embedding space but also show a tight distribution of predicted melting temperatures close to the query's known value. This result validates our two-stage "retrieve and rank" framework, confirming it identifies candidates that are both structurally similar and functionally relevant to the user's query.

Table 1: Test R^2 scores for the multi-task prediction. The final model MMoE (Lang + Graph) is compared against a SOTA baseline and several ablation models. The best results are **bolded**.

Model Configuration	T_{g}	T_{m}	E_{g}
Multimodal Transformer [10]	0.880	0.720	–
PolyBERT (Pretrained)	0.801	0.498	0.667
GNN (Graph only)	0.895	0.829	0.908
PolyBERT (Finetuned)	0.888	0.726	0.904
MMoE (Lang only)	0.898	0.761	0.915
Early Fusion (Lang + Graph)	0.912	0.835	0.926
Gated Late Fusion (Lang + Graph)	0.915	**0.840**	0.926
MMoE (Lang + Graph)	**0.923**	0.838	**0.933**

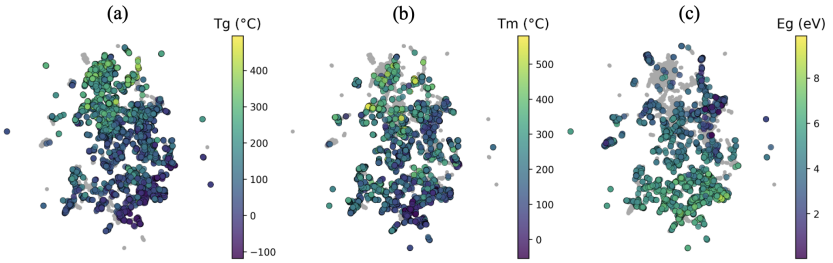

Figure 2: Two-dimensional UMAP projection of the multimodal polymer embeddings. The distribution is colored by three distinct properties: (a) T_g, (b) T_m, and (c) E_g.

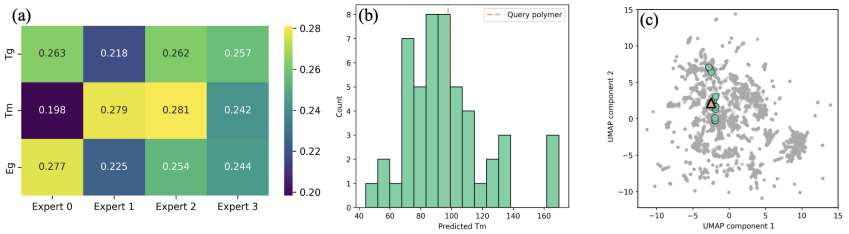

Figure 3: MMoE model analysis: (a) heatmap of task-specific expert utilization; (b) predicted T_m distribution and (c) UMAP projection in space for top 50 candidates from a PEO query.

References

[1] Mohammad Harun-Ur-Rashid and Abu Bin Imran. Emerging trends in engineering polymers: a paradigm shift in material engineering. *Recent Progress in Materials*, 6(3):1–37, 2024.

[2] Rishi Gurnani, Christopher Kuenneth, Aubrey Toland, and Rampi Ramprasad. Polymer informatics at scale with multitask graph neural networks. *Chemistry of Materials*, 35(4):1560–1567, 2023.

[3] Owen Queen, Gavin A McCarver, Saitheeraj Thatigotla, Brendan P Abolins, Cameron L Brown, Vasileios Maroulas, and Konstantinos D Vogiatzis. Polymer graph neural networks for multitask property learning. *npj Computational Materials*, 9(1):90, 2023.

[4] 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.

[5] Christopher Kuenneth and Rampi Ramprasad. polybert: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nature communications*, 14(1):4099, 2023.

[6] Changwen Xu, Yuyang Wang, and Amir Barati Farimani. Transpolymer: a transformer-based language model for polymer property predictions. *npj Computational Materials*, 9(1):64, 2023.

[7] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198, 2016.

[8] Jiaxing Qu, Yuxuan Richard Xie, and Elif Ertekin. A language-based recommendation system for material discovery. In *1st Workshop on the Synergy of Scientific and Machine Learning Modeling@ ICML2023*, 2023.

[9] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37:98–125, 2017.

[10] Seunghee Han, Yeonghun Kang, Hyunsoo Park, Jeesung Yi, Geunyeong Park, and Jihan Kim. Multimodal transformer for property prediction in polymers. *ACS Applied Materials & Interfaces*, 16(13):16853–16860, 2024.

[11] 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.

[12] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1 (2):3, 2022.

[13] 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.

[14] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, et al. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling*, 59 (8):3370–3388, 2019.

[15] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

[16] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1930–1939, 2018.

A Experimental Settings

A.1 Hyperparameter Settings

In this section, we provide the detailed hyperparameter settings used for the models in PolyRecommender. The PolyBERT fine-tuning, GNN training, and the training of the three multimodal fusion models were all conducted on a single NVIDIA A10G GPU. Table 2 lists the common training hyperparameters and specific network architectures for each fusion model. The hyperparameters for the GNN and PolyBERT fine-tuning stages largely followed those outlined in their original works.

Table 2: Hyperparameters for the three multimodal fusion models.

Hyperparameter	**Value**
Batch Size	128
Learning Rate	1e-5
Weight Decay	1e-3
Total Epochs	100
Dropout	0.4
Optimizer	AdamW
Learning Rate Scheduler	ReduceLROnPlateau
Architectures	
MLP in Early Fusion	3-layer (hidden sizes [256, 128])
Expert network (Gated Late Fusion)	3-layer MLP (hidden sizes [256, 128])
Gate network (Gated Late Fusion)	2-layer MLP (hidden size 128)
Number of Experts (MMoE)	4
Expert network (MMoE)	3-layer MLP (hidden size 256)
Gate network (MMoE)	2-layer MLP (hidden size 256)
Tower network (MMoE)	2-layer MLP (hidden size 128)

A.2 Dataset Information

The dataset used in this work consists of 12,441 synthesized polymers curated from the PolyInfo database. The distributions of the available experimental data for the three target properties are shown in Figure 4. The dataset exhibits a significant label imbalance, with many more values available for glass transition temperature (T_{g}, 6900 labels) than for melting temperature (T_{m}, 3633 labels) and band gap (E_{g}, 3379 labels).

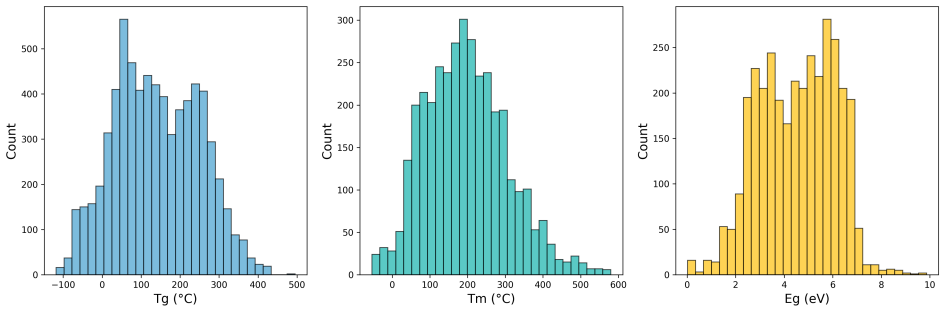

Figure 4: Distributions of available experimental data for the three target properties.

A.3 Relevance Score

To rank the final list of candidates for a given query, we defined a unitless relevance score, R (from 0 to 100), based on the Total Absolute Percentage Difference (TAPD) across all shared properties:

$$R = \frac{100}{\mathrm{TAPD} + 1}, \quad \mathrm{TAPD} = \sum_{i=1}^{n} \left| \frac{y_i^c - y_i^q}{y_i^q} \right| \tag{9}$$

where y_i^c and y_i^q are the predicted values of the i-th property for the candidate and query polymer, respectively, and n is the number of properties being compared. This metric normalizes the error across properties of different scales, and the $+1$ in the denominator ensures a perfect match (TAPD=0) yields a maximum score of 100.