Enhancing Student Performance Prediction on Learnersourced Questions with SGNN-LLM Synergy

Lin Ni[1,2] , **Sijie Wang**[2], **Zeyu Zhang**[1*], **Xiaoxuan Li**[2], **Xianda Zheng**[2], **Paul Denny**[2], **Jiamou Liu**[2]

[1]Huazhong Agricultural University, Wuhan 430070, China
[2]School of Computer Science, The University of Auckland, Auckland 1010, New Zealand
l.ni@auckland.ac.nz, swan387@aucklanduni.ac.nz, zhangzeyu@mail.hzau.edu.cn,
{xli443, xzhe162}@aucklanduni.ac.nz, {p.denny, jiamou.liu}@auckland.ac.nz

Abstract

Learnersourcing offers great potential for scalable education through student content creation. However, predicting student performance on learnersourced questions, which is essential for personalizing the learning experience, is challenging due to the inherent noise in student-generated data. Moreover, while conventional graph-based methods can capture the complex network of student and question interactions, they often fall short under cold start conditions where limited student engagement with questions yields sparse data. To address both challenges, we introduce an innovative strategy that synergizes the potential of integrating Signed Graph Neural Networks (SGNNs) and Large Language Model (LLM) embeddings. Our methodology employs a signed bipartite graph to comprehensively model student answers, complemented by a contrastive learning framework that enhances noise resilience. Furthermore, LLM's contribution lies in generating foundational question embeddings, proving especially advantageous in addressing cold start scenarios characterized by limited graph data. Validation across five real-world datasets sourced from the PeerWise platform underscores our approach's effectiveness. Our method outperforms baselines, showcasing enhanced predictive accuracy and robustness.

Introduction

The rapid growth of online education in recent years has resulted in considerable interest in new approaches for delivering effective learning at scale (Xianghan and Stern 2022). A significant challenge for educators adopting new online learning platforms is the need to develop new content (Mulryan-Kyne 2010). Educators face the task of creating extensive repositories of items for supporting personalized learning and exploring innovative methods to immerse students in such material. Learnersourcing, a relatively new approach, has emerged as a promising technique to address the challenges of content development and student engagement in online learning (Khosravi et al. 2021; Denny et al. 2022). It involves leveraging the collective knowledge, skills, and contributions of learners to create educational resources. In general, accurate prediction of student performance (Li et al. 2020, 2022), such as future grades on assignments and exams, is immensely valuable for informing instructional

strategies and enabling personalized support. Within learnersourcing contexts, performance prediction on individual items can support the recommendation of new items from large repositories to match the current ability level of students (Paramythis and Loidl-Reisinger 2003).

Multiple-choice questions (MCQs) play a pivotal role in learnersourcing platforms like RiPPLE (Khosravi, Kitto, and Williams 2019), Quizzical (Riggs, Kang, and Rennie 2020), UpGrade (Wang et al. 2019), and PeerWise (Denny, Luxton-Reilly, and Hamer 2008). Predicting student performance on MCQs is a crucial research challenge (Abdi, Khosravi, and Sadiq 2021). MCQs, comprising a stem and multiple answer options, assess a wide range of cognitive skills, from factual recall to higher-order thinking (e.g., analysis, synthesis, and evaluation). They offer efficient grading and valuable insights into subject comprehension.

As learners engage in creating and responding to MCQs on platforms like PeerWise, a graph emerges, with learners and MCQs as nodes and their interactions as edges. GNNs provide a powerful means to capture these intricate learner-to-question, learner-to-learner, and question-to-question relationships (Zhou et al. 2020). GNNs excel at extracting nuanced features from graph topology, offering insights beyond traditional methods. However, while GNNs hold promise for predictive analytics in learnersourcing platforms, applying them to anticipate student performance on MCQs is relatively unexplored. A key challenge lies in modeling both students' responses and their correctness (i.e., correct or erroneous answers) within noisy data. Noise can stem from inadvertent errors during question creation or participation dynamics discouraging accurate responses. Furthermore, learnersourcing repositories constantly evolve, presenting a cold start dilemma. Newly created MCQs with limited responses lack substantial interaction data, hampering effective GNN utilization and predictive accuracy.

In summary, our study addresses two crucial challenges for predicting student performance on MCQs in learnersourcing platforms:

- Modeling complex relationships involving students and MCQs, accommodating both accurate and erroneous responses within noisy data.

- Tackling the cold start problem inherent to questions with low answer rates, where insufficient graph data undermines predictive accuracy.
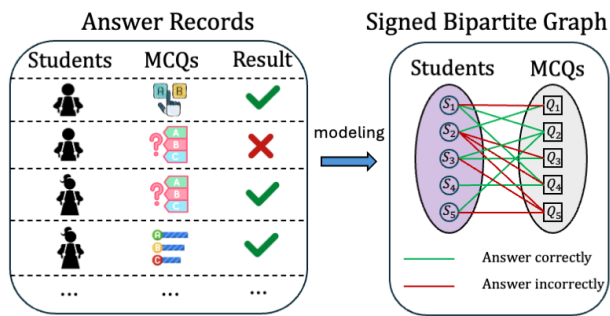

Figure 1: A scenario for the signed bipartite graph

To address these challenges, we propose an innovative technique that leverages the potential of GNNs for application to the unique dynamics present in learnersourcing repositories. Our approach begins by adopting a signed bipartite graph to model students' response accuracy, as illustrated in Figure 1. This graph structure segregates students and questions into separate sets of nodes. Positive edges denote correct student answers, while negative edges indicate incorrect responses. To counter the impact of noise, we introduce a novel contrastive learning framework tailored for signed bipartite graphs. This approach draws from the emerging paradigm of graph contrastive learning (Shuai et al. 2022; Zhu et al. 2021, 2020; You et al. 2020), renowned for its ability to resist noise by contrasting similar and dissimilar instances. Our choice of contrastive learning not only enhances model resilience but also addresses noise interference within the data.

Furthermore, we pioneer the integration of NLP technology to extract intrinsic knowledge essential for accurate MCQ responses, offering a solution to the cold start problem. Although previous student performance prediction research seldom leveraged NLP advancements, recent studies underscore the potential of integrating advanced NLP in education to unravel question intricacies (Ni et al. 2022). We posit that MCQ response accuracy hinges on students' comprehension of underlying knowledge within questions. Leveraging the gpt-3.5-turbo-0613 LLM[1], we capture nuances within MCQs, extracting and weighing inherent knowledge points. By fusing this semantic information with structural embeddings derived from the graph, our model gains a holistic view for enhanced student performance prediction, particularly in cold start scenarios. Our main contributions include:

1. Formulating the student performance prediction problem as a sign prediction on a signed bipartite graph.

2. Exploiting contrastive learning to enhance the robustness of representations for students and questions.

3. Employing an LLM to extract semantic embeddings of key knowledge points in questions, and integrating this information with structural embeddings learned from a graph to address the cold start problem.

---

[1]https://platform.openai.com/docs/models/continuous-model-upgrades

4. Validating the efficacy of our proposed model through experiments conducted on five real-world datasets sourced from a prominent learnersourcing platform, achieving the highest f1 value of 0.908.

Related Work

We investigate four domains: student academic performance prediction, the cold start problem, GNNs, and LLMs.

Performance Prediction: Significant research has been devoted to predicting student performance on online learning platforms, with the aim of personalizing and improving the learning experience (Wang et al. 2023). This body of work primarily encompasses static and sequential models (Li et al. 2020; Thaker, Carvalho, and Koedinger 2019). Static models, exemplified by Logistic Regression, leverage historical data such as student scores and learning activities to forecast future performance (Jiang et al. 2014; Wei et al. 2020; Daud et al. 2017). In contrast, sequential models like Knowledge Tracing and its variants explore the sequential relationships within learning materials, allowing dynamic knowledge updates (Piech et al. 2015; Nakagawa, Iwasawa, and Matsuo 2019). Knowledge Tracing, a subset of educational data mining and machine learning, aims to model and predict learners' knowledge or skill acquisition over time (Lyu et al. 2022). However, our task presents unique challenges. Unlike many previous approaches, we lack access to student profiles and additional contextual information. Furthermore, we do not possess information about the temporal sequence of learner responses. Consequently, our research is focused on exploring deep learning solutions that rely solely on question and learner-question interaction data to enhance student academic performance prediction.

Cold Start Problem: Addressing the cold start challenge is a fundamental concern in representation learning, particularly in machine learning scenarios in e-learning environments where minimal prior information is available about the target data (Liu et al. 2022). Researchers have explored diverse strategies to tackle this issue, tailoring their approaches to the specific nature of the problem, available resources, and data characteristics. One common strategy is transfer learning, which entails pre-training a model on a related task with a substantial dataset and then fine-tuning it for the target task with limited data (Pang et al. 2022). This method leverages prior knowledge gained during pre-training to mitigate cold start challenges. In cases where data is scarce, knowledge-based systems employ explicit rules or domain knowledge to provide recommendations or predictions, bypassing the need for extensive historical data (Du et al. 2022). Recommendation systems often turn to content-based filtering, relying on item and user attributes or features for making recommendations (Schein et al. 2002; Safarov et al. 2023). This approach proves effective, particularly in scenarios with little or no historical user-item interaction data, by suggesting items based on their inherent characteristics. In alignment with previous studies that have leveraged ontologies as supplementary data to mitigate the cold start problem (Sun et al. 2017; Jeevamol and Renumol 2021), our approach capitalizes on the inherent semantic richness of MCQs sourced from learnersourcing platforms. These

MCQs inherently contain vital knowledge points essential for predictive tasks. To effectively address the challenges posed by the cold start problem, should this be we incorporate embedding information from LLMs that encapsulate both common sense and domain-specific knowledge. This integration serves to enhance the representation of MCQs in our model.

Signed Graph Representation Learning Methods: Predicting student performance involves the unique task of sign prediction in a signed bipartite graph, where positive edges signify correct answers and negative edges denote incorrect ones (Derr, Ma, and Tang 2018; Zhang et al. 2023b). Traditional methods like SIDE (Kim et al. 2018) and SGDN (Jung, Yoo, and Kang 2020) leveraged random walks, while neural-based approaches like SGCN (Derr, Ma, and Tang 2018) extended Graph Convolutional Networks (GCN) (Kipf and Welling 2016) to balance theory-guided sign prediction. However, adapting these methods to signed bipartite graphs is non-trivial. Graph Contrastive Learning techniques, inspired by successes in image processing, aim to generate stable node representations under perturbations. Deep Graph Infomax (DGI) (Velickovic et al. 2019) maximizes the mutual information between global graph and local node representations. GraphCL (You et al. 2020) introduces varied graph augmentations, while Graph Contrastive Augmentation (GCA) employs diverse views for agreement of node representations. Most of these models are designed for unipartite graphs, with Signed Graph Contrastive Learning (SGCL) (Shu et al. 2021) adapting to signed unipartite graphs. However, Signed Bipartite graph Contrastive Learning (SBCL) (Zhang et al. 2023a) is the only model combining both signed and bipartite graph aspects, albeit lacking consideration for node features. Notably, these methods primarily rely on large-scale node interaction data to extract structural information for representation, which may limit their effectiveness in addressing the cold start problem.

LLM in NLP: Recent years have witnessed significant advancements in LLMs in natural language processing (NLP). Models like BERT (Devlin et al. 2018) and T5 (Raffel et al. 2020) excel in various NLP tasks by pre-training on extensive text data, followed by fine-tuning. However, fine-tuning often demands a substantial amount of task-specific data and introduces complexity. A groundbreaking development came with GPT-3 (Brown et al. 2020), a massive autoregressive language model with 175B parameters. It introduced In-Context Learning, allowing it to perform tasks without extensive fine-tuning or model updates. This innovation led to Instruction-Finetuned Language Models, exemplified by GPT-3.5 (Ye et al. 2023), which efficiently adapts to specific tasks using task-specific instructions. This approach overcomes previous limitations, making LLMs a cornerstone in NLP.

Problem Definition

In this study, we model the student performance prediction problem on a learnersourcing platform as the edge sign prediction problem on a signed bipartite graph, as illustrated in Figure 1. We define our problem within the context of a bipartite graph $\mathcal{G} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, where $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$

and $\mathcal{V} = \{v_1, v_2, ..., v_{|\mathcal{V}|}\}$ represent the nodes for students and questions, respectively. $|\mathcal{U}|$ and $|\mathcal{V}|$ denote the total number of students and questions, where $\mathcal{U} \cap \mathcal{V} = \emptyset$. The edge set $\mathcal{E} \subset \mathcal{U} \times \mathcal{V}$ signifies the links between students and questions and is further partitioned into positive and negative edge sets, $\mathcal{E}^+$ and $\mathcal{E}^-$, respectively, with $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ and $\mathcal{E}^+ \cap \mathcal{E}^- = \emptyset$. Given this bipartite graph and any pair of nodes $u_i \in \mathcal{U}$, $v_j \in \mathcal{V}$ and the sign of the edge $e_{ij} \in \mathcal{E}$ is unknown, our goal is to leverage GNNs to learn the embeddings $z_{u_i} \in \mathbb{R}^d$ and $z_{v_j} \in \mathbb{R}^d$ of node u_i and v_j, respectively. Subsequently, we seek to find a mapping $f: (z_{u_i}, z_{v_j}) \mapsto \{-1, +1\}$ to determine the sign of the edge, thus producing our desired result.

One significant challenge in this context is the cold start problem for newly generated questions. In early-stage situations, a limited number of learners may have attempted the question, and consequently, the bipartite graph is sparsely connected. The edges connecting students to these questions are uncertain or unknown, making the prediction of student performance particularly challenging.

Proposed Method

In this section, we introduce the <u>L</u>arge <u>L</u>anguage <u>M</u>odel enhanced <u>S</u>igned <u>B</u>ipartite graph <u>C</u>ontrastive <u>L</u>earning (**LLM-SBCL**) model, a framework designed to learn representations for users and questions while facilitated by LLM.

Our framework, as depicted in Figure 2, capitalizes on student-MCQs relationships in learnersourcing platforms using SBCL. It employs separate encoders for students and MCQs, with a graph contrastive learning model to handle dataset noise. To address cold start challenges, we enhance question embeddings with semantic knowledge points extracted by LLM. This supplementary data augments the question structure embeddings to compute cross-entropy loss. In the following sections, we provide a detailed breakdown of our model's architecture.

Signed Bipartite Graph Contrastive Learning

Fig. 3 illustrates SBCL, which includes graph augmentation and graph encoding processes. During the graph augmentation phase, we use stochastic graph augmentation to generate two augmented graphs. Each graph is then bifurcated according to edge signs, resulting in a positive and a negative graph. These are further processed using GNNs to learn the structural embeddings for node representations.
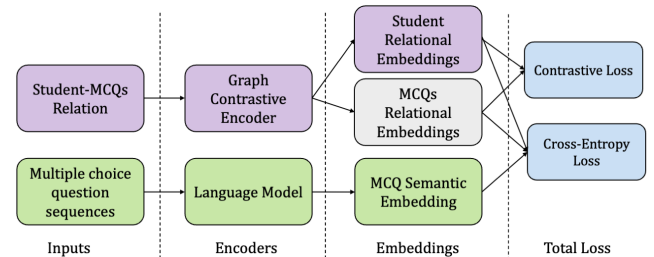

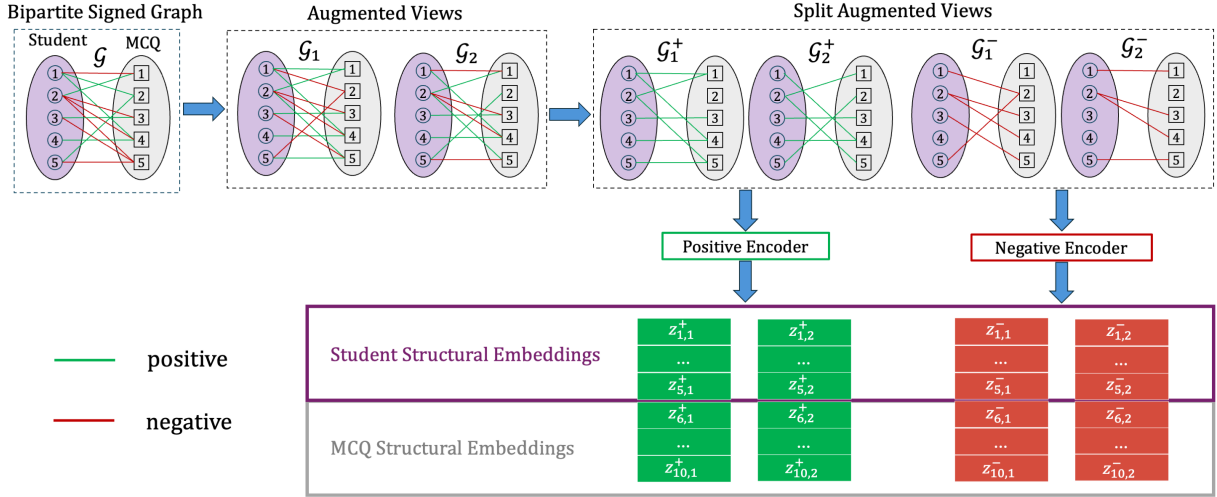Figure 2: The Framework of LLM-SBCL model

Figure 3: The framework of SBCL

Graph Augmentation: To enhance the robustness and generalization of learned node representations in the context of learnersourcing platforms with noisy data, we utilize stochastic perturbation for graph augmentation. We generate two distinct graphs from the original one by randomly flipping edge signs, transforming a positive edge into a negative one, and vice versa, with a probability p.

Graph Encoder: Following data augmentation, we acquire two augmented graphs, denoted as $\mathcal{G}_1$ and $\mathcal{G}_2$, as illustrated in Fig 3. Considering the distinct semantic attributes of positive and negative edges—positive edges representing correct answers by users and negative edges indicating incorrect answers—we employ separate GNN encoders for each edge type. This dual-GNN approach aligns with our contrastive objective, which we will delve into in the next subsection. Consequently, we partition each augmented graph into two sub-graphs exclusively featuring positive and negative edges, denoted as positive graph $\mathcal{G}_m^+$ and negative graph $\mathcal{G}_m^-$, where $m \in \{1, 2\}$. To facilitate this encoding, we adopt the Positive and Negative Graph Attention Network (GAT) (Veličković et al. 2017), consistent with the design in (Zhang et al. 2023a). Formally, the representation for the i-th node, denoted as z_i, is computed as follows:

$$z_{i,m}^{(l+1),\sigma} = \text{PReLU}(\text{GNN}_m^\sigma(z_{i,m}^{(l),\sigma}, \mathcal{G}_m^\sigma)) \qquad (1)$$

$$z_i = \left[z_{i,1}^{(L),+} \| z_{i,1}^{(L),-} \| z_{i,2}^{(L),+} \| z_{i,2}^{(L),-} \right] W \qquad (2)$$

where $\sigma \in \{+, -\}$, m refers to the m-th augmented graph, L denotes the number of GNN layers and l denotes the l-th layer. $z_{i,m}^{(0),\sigma}$ denotes the input feature vector of the i-th node. W is a learn-able transformation matrix.

Contrastive Objective: Noise is commonly expected in a learnersourcing platform where the majority of content is student-generated. Variations in course rules for question creation and answering, along with student behaviors, can introduce considerable noise. For instance, when the motivation for answering a question is low (e.g., performance on
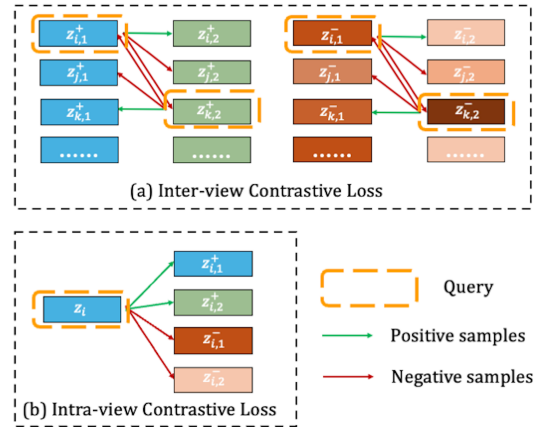

Figure 4: Inter-view and Intra-view contrastive loss

such platform does not impact final scores), students might guess answers randomly. Likewise, students may unintentionally create incorrect questions. We categorize these instances as noise within our dataset. To address this, we employ a contrastive objective for robust signed bipartite graph representation learning, encompassing two losses: (a) *Inter-view contrastive loss* and (b) *Intra-view contrastive loss*, where (a) is for nodes from the same encoder while (b) is for those from different encoders as depicted in Fig 4.

We utilize InfoNCE loss (Sohn 2016; Oord, Li, and Vinyals 2018) to define our inter-view contrastive loss for positive augmented graphs as follows:

$$\mathcal{L}_{\text{inter}}^+ = -\frac{1}{I} \sum_{i=1}^{I} \log \frac{\exp\left(\text{sim}\left(z_{i,m}^+, z_{i,m'}^+\right)/\tau\right)}{\sum_{j=1, j\neq i}^{I} \exp\left(\text{sim}\left(z_{i,m}^+, z_{j,m'}^+\right)/\tau\right)} \qquad (3)$$

where I is the number of nodes in a mini-batch, $z_{i,m}^+$ represents the representation of node i in the m-th augmented
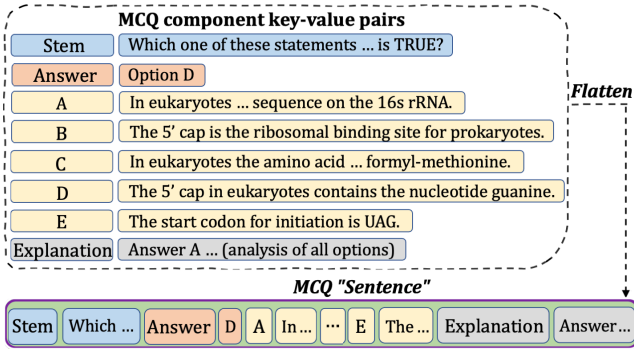
Figure 5: Flatten MCQ component key-value pairs into a sequence to form an MCQ "sentence"

positive graph, $sim(\cdot, \cdot)$ represents the similarity function between the two representations (e.g., cosine similarity) and τ denotes the temperature parameter. The perspective-specific contrastive loss for negative graphs is similar.

The intra-view contrastive loss is defined as:

$$\mathcal{L}_{\text{intra}} = -\frac{1}{I} \sum_{i=1}^{I} \log \frac{\sum_{m=1}^{M} \exp\left(\text{sim}\left(z_i, z_{i,m}^{+}\right)/\tau\right)}{\sum_{m=1}^{M} \exp\left(\text{sim}\left(z_i, z_{i,m}^{-}\right)/\tau\right)} \quad (4)$$

where M denotes the number of augmented graphs, which is equal to 4 in our paper. The Combined contrastive loss is

$$\mathcal{L}_{CL} = (1-\alpha) \cdot (\mathcal{L}_{\text{inter}}^{+} + \mathcal{L}_{\text{inter}}^{-}) + \alpha \cdot \mathcal{L}_{\text{intra}} \quad (5)$$

where α is the weight coefficient that controls the significance between two losses.

Incorporating Semantic Embedding into SBCL

In our LLM-SBCL framework (Fig 2), we employ a multimodal approach, utilizing GNNs and NLP Models, to process User-Question Relationship and MCQ Textual Information respectively. We recognize that infusing semantic context into MCQs has great potential for uncovering the underlying knowledge, which are crucial for the prediction task.

Extracting Semantic Embeddings via LLM: To include semantic information, we utilize a LLM to extract knowledge from MCQs. We combine these semantic representations with relational embeddings from the Graph Encoder, training them with cross-entropy loss $\mathcal{L}_{\text{CE}}$ for a more comprehensive representation.

Our approach involves transforming MCQs into sentence format by flattening key-value pairs (Fig 5). This results in input data, denoted as mcq_j, comprising content from Stem, Answer, Options, and Explanation: $mcq_j = \{S_k, S_v, A_k, A_v, O_{A,k}, O_{A,v}, \ldots, O_{E,k}, O_{E,v}, E_k, E_v\}$.

Next, we employ LLM to extract n knowledge point terms t_i and their corresponding weights h_i for each MCQ. GloVe word embeddings (Pennington, Socher, and Manning 2014) provide word embeddings $emb_{i,k}$ for words in t_i. The average of these embeddings represents a single knowledge point kp_i. We compute the weighted average of all knowledge points to obtain the Question Semantic Embedding w_j.

$$w_j = \frac{\sum_{i=1}^{n} h_i \cdot kp_i}{\sum_{i=1}^{n} h_i}, \quad kp_i = \frac{\sum_{k=1}^{m} emb_{i,k}}{m} \quad (6)$$

NLP Large Language Model: We employ gpt-3.5-turbo-0613 [2] as our LLM engine, capable of retrieving n pairs of t_i and h_i with a single prompt API request.

- **Task Objective**: Your task is to extract key knowledge points from an MCQ created by a first-year student who studies biology in college. The question consists of a stem, up to five options, an answer, and an explanation provided by the student author.

- **Special Requirements**: Please provide your response in JSON format with the following keys and format:

  ```
  {"Keywords": [{"keyword": "keyword_name", "percentage": 0}]}
  ```

  1. Each `"keyword_name"` should be a word or a short term with less than five words.

  2. The percentages of all keywords should add up to 1.

  3. Only include the top five keywords to avoid excessive keyword extraction.

- **In Context Learning**: Here's an example response format:

  ```
  {"Keywords": [
    {"keyword": "Menstrual cycle", "percentage": 0.8},
    {"keyword": "Luteal phase", "percentage": 0.2 }
  ]}
  ```

- **Context**: Here's the MCQ from the student author:

  Question stem: Jamie has a Atherosclerosis, what does this imply?

  [A]: She constantly needs to visit the ladies room to pass urine.

  [B]: She has impaired vision.

  [C]: She has lost a lot of weight recently due to no apparent reason. (She hasn't suddenly turned paleo).

  [D]: Her arterial walls have thickened making her more likely to have a heart attack.

  Question answer: D

  Explanation: Atherosclerosis is the thickening of the artery walls (generally due to poor diet and lack of exercise). This can lead to heat attacks and strokes, as when pressure builds up in the arteries, arteries are more likely to burst. In a normal person the walls of arteries are more likely to stretch and allow for the fluctuations of pressure in the blood.

- **LLM Response**:

  ```
  [{"keyword": "Atherosclerosis", "percentage": 0.44},
  {"keyword": "Arterial walls", "percentage": 0.22},
  {"keyword": "Thickening", "percentage": 0.11},
  {"keyword": "Heart attack", "percentage": 0.11},
  {"keyword": "Strokes", "percentage": 0.11}]
  ```

The provided prompt encompasses the task objective, requirements, in-context learning format, and context, ensuring consistent and aligned responses. This well-crafted prompt guided the LLM to analyse the specific knowledge topics related to the MCQs. As a result, the Question Semantic Embedding derived from this prompt becomes more meaningful, thereby enhancing the accuracy of student performance prediction.

Semantic Embedding Integration: LLM-SBCL enhances SBCL with semantic question embeddings. Instead of directly predicting the edge sign between student u_i and

---
[2]https://platform.openai.com/docs/models/continuous-model-upgrades

question v_j, we also include semantic embeddings w_j for question representations.

This results in d-dimensional embeddings for u_i and v_j and an additional d-dimensional embedding for w_j. To maintain uniform dimensions, we transform v_j and w_j into a unified d-dimensional representation q_{v_j} using a learnable matrix W_q. The transformed question embeddings, q_{v_j}, are concatenated with u_i and processed through a single-layer MLP for edge sign prediction:

$$q_{v_j} = (v_j \| w_j) W_q, \quad y_{\text{pred}} = \text{MLP}\left(u_i \| q_{v_j}\right) \qquad (7)$$

Use the predicted y_{pred} in the cross-entropy loss function:

$$\mathcal{L}_{\text{CE}} = -y \cdot \log y_{\text{pred}} + (1 - y) \cdot \log\left(1 - y_{\text{pred}}\right) \qquad (8)$$

where y is the ground truth from $\{-1, 1\}$ to $\{0, 1\}$.
The final joint loss of cross-entropy and contrastive loss:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \beta \cdot \mathcal{L}_{CL} \qquad (9)$$

Here, β controls the contribution of the contrastive loss.

Experiments

In this section, we evaluate the LLM-SBCL model's performance on five real-world datasets: biology, law, cardiff20102, sydney19351, and sydney23146. We introduce the datasets and compare our model against state-of-the-art baselines. We provide experimental details and present results from both baseline and our models.

Dataset Background and Introduction

PeerWise is a platform that allows students to share their knowledge with each other. Teachers can create courses and manage student access to them. MCQs are a common type of assessment format that presents a question or statement followed by a set of predetermined options or choices. Within these courses, students can create and explain their own MCQs related to the course material. They can also answer and discuss questions posed by their peers. This allows for a collaborative learning experience where students can learn from each other and deepen their understanding of the course material.

We use five real-world datasets from PeerWise: biology and law from the University of Auckland, cardiff20102 from Cardiff University School of Medicine with course id 20102, sydney19351 and sydney23146 are from two University of Sydney biochemistry courses with the highest number of answering records (i.e. links). Detailed summaries for these datasets are provided in Table 1.

	Biology	Law	Cardiff	Sydney1	Sydney2
Students	761	528	383	382	198
MCQs	380	5600	1171	457	748
Links	76613	88563	64524	24032	24050
$Link_{P\%}$	0.665	0.931	0.600	0.531	0.706
$Link_{N\%}$	0.335	0.069	0.400	0.469	0.294

Table 1: Statistics on Signed Bipartite Networks.

Since each course is independent of the others, we construct a signed bipartite graph for each dataset using the answering records. For a student $u_i \in \mathcal{U}$ answering a question $v_j \in \mathcal{V}$, we establish an edge between u_i and v_j with its sign being +1 if the answer of u_i matches the correct answers for v_j's and -1 otherwise. We observe that the overall correctness for the questions exceeds 50% for each course, with most courses ranging from 60% to 70%. However, there are exceptions such as Law, where the overall correctness reaches 93%. The inherent nature of such online peer-practice platforms suggests an innate data imbalance issue, which is handled by class weights in the experiment.

Experimental Settings

We compared our model against state-of-the-art methods, including Random Embedding, Graph Convolutional Network (GCN), Graph Attention Networks (GAT), Signed Graph Convolutional Network (SGCN), Signed Bipartite Graph Neural Networks (SBGNN) (Huang et al. 2021), and Signed Bipartite Graph Contrastive Learning (SBCL).

- **Random Embedding:** Concatenates random 64-dimensional student ($z_{u_i} \in \mathbb{R}^{64}$) and question ($z_{v_j} \in \mathbb{R}^{64}$) embeddings, followed by logistic regression to predict.

- **GCN:** base on an efficient variant of convolutional neural networks which operate directly on graphs. The choice of convolutional architecture is motivated via a localized first-order approximation of spectral graph convolutions.

- **GAT:** Implemented using PyTorch Geometric (PyG) [3] with two convolutional layers and 300 training epochs, initializing with standard normal distribution embeddings.

- **SGCN:** harnesses balance theory to correctly incorporate negative links during the aggregation process. SGCN addresses the unique challenge posed by negative links, which not only carry a distinct semantic meaning compared to positive links, but also form complex relations with them due to their inherently different principles.

- **SBGNN:** Utilizes innovative techniques for modeling signed bipartite graphs [4]. We used publicly available code with default settings and 300 training epochs.

- **SBCL:** Utilizes a single-headed GAT with 2 convolutional layers and 64-dimensional random embeddings [5]. We used publicly available code with default settings and 300 training epochs. The final LLM-SBCL variant incorporates semantic question embeddings.

Data was split into training (85%), validation (5%), and test (10%) sets. 64-dimensional embeddings were generated for students and questions. Question representations were enriched by concatenating stem, choices, and explanations, then processed through GPT-3.5 for knowledge points, reduced to 64 dimensions with PCA. Hyperparameters were set to $\alpha = 0.8$, $\beta = 5e-4$, and $r = 0.1$ (where r denotes the mask ratio, governing the percentage of links flipped during graph augmentation) after experimentation.

---

[3]https://pytorch-geometric.readthedocs.io/en/latest/index.html
[4]https://github.com/huangjunjie-cs/SBGNN
[5]https://github.com/Alex-Zeyu/SBGCL

Dataset	Random Embedding	Unsigned Network Embedding		Signed/Bipartite Network Embedding		Signed Bipartite graph Contrastive Learning models	
	Random	GCN	GAT	SGCN	SBGNN	SBCL	LLM-SBCL
Biology	0.35±0.01	0.682±0.058	0.618±0.013	0.768±0.04	0.753±0.014	0.772±0.016	**0.787±0.014**
Law	0.472±0.01	0.823±0.01	0.817±0.05	0.84±0.013	0.861±0.034	0.901±0.016	**0.908±0.018**
Cardiff20102	0.136±0.062	0.677±0.024	0.571±0.013	0.607±0.033	0.712±0.016	0.718±0.018	**0.734±0.023**
Sydney19351	0.29±0.014	0.642±0.021	0.564±0.022	0.635±0.044	0.673±0.016	0.674±0.021	**0.694±0.021**
Sydney23146	0.288±0.035	0.728±0.013	0.608±0.02	0.726±0.04	0.712±0.021	0.733±0.019	**0.76±0.022**

Table 2: The results of link sign prediction on five real-world educational datasets (average binary-F1 $\pm$ standard deviation).

Dataset	Metrics	SBCL	LLM-SBCL
Biology	Binary-F1	0.616±0.049	**0.629±0.061**
	Micro-F1	0.537±0.027	**0.553±0.043**
	Macro-F1	0.512±0.015	**0.528±0.036**
	AUC	0.524±0.007	**0.539±0.032**
Law	Binary-F1	0.885±0.039	**0.894±0.031**
	Micro-F1	0.803±0.059	**0.814±0.048**
	Macro-F1	**0.559±0.025**	0.548±0.017
	AUC	**0.628±0.022**	0.594±0.017
Cardiff	Binary-F1	0.595±0.08	**0.625±0.045**
	Micro-F1	0.532±0.037	**0.543±0.029**
	Macro-F1	0.506±0.017	**0.517±0.018**
	AUC	**0.521±0.019**	0.52±0.019
Sydney1	Binary-F1	0.573±0.057	**0.623±0.055**
	Micro-F1	0.539±0.035	**0.555±0.031**
	Macro-F1	0.533±0.027	**0.533±0.033**
	AUC	0.538±0.022	**0.544±0.026**
Sydney2	Binary-F1	0.779±0.039	**0.801±0.009**
	Micro-F1	0.66±0.036	**0.681±0.012**
	Macro-F1	**0.505±0.031**	0.495±0.018
	AUC	**0.518±0.017**	0.511±0.013

Table 3: The results for cold-start problem.

Results

In this section, **SBCL** represents our proposed model with contrastive learning, excluding NLP information. Variations of SBCL, denoted as **LLM-SBCL**, incorporate semantic embeddings from LLM. We evaluate model performance using metrics such as Area Under the Receiver Operating Characteristic Curve (AUC), binary F1, micro-F1, and macro-F1 scores. Higher metric values indicate superior link prediction.

We conducted each experiment ten times on each dataset for each model. Table 2 presents mean and standard deviation results for baseline models (Random Embedding, GCN, GAT, SGCN, and SBGNN) alongside our proposed SBCL models. The best results are highlighted in bold, and runner-up results are underlined. Key insights from Table 2 include:

- SGCN and SBGNN, which consider the sign of the links in signed networks, outperform unsigned methods (GCN, GAT) on most datasets.

- SBCL consistently outperforms SGCN and SBGNN, highlighting the ability of contrastive learning.

- LLM-enhanced variants outperform SBCL across all datasets, showcasing the benefit of incorporating NLP semantic embeddings.

Cold Start Problem Results

Our study demonstrates a substantial enhancement in student performance prediction by incorporating semantic embeddings of MCQs, as shown in Table 3. To ensure a fair comparison, we maintained a consistent experimental setup. We randomly selected 10% of the questions to simulate a cold start scenario. Edges related to these questions were placed in the test set, while the rest formed the training set. Our primary evaluation metric was the binary F1 score.

As a baseline, we used a generic representation learning model without semantic embeddings, which performed well with abundant data but struggled in cold start scenarios. Our enhanced model consistently outperformed the baseline across various question segments, achieving an average increase of approximately 6.4% in binary F1 score. This improvement was particularly notable in subjects where semantic richness in MCQs was critical. The inclusion of MCQ-derived semantic embeddings holds promise for improving representation learning in e-learning environments, especially under cold start conditions. Future research can focus on optimizing these embeddings for diverse educational contexts and expanding their applicability.

Conclusion and Future Work

In this paper, we initiate the study of performance prediction for students using learnersourcing platforms. We modeled the relationship between students and questions using a signed bipartite graph. To mitigate the negative impact of noise in the datasets, which is a significant problem in learnersourcing contexts, we adopted contrastive learning as a novel learning paradigm. Additionally, we attempted to integrate both semantic and structural information related to the questions into our models. We evaluated the effectiveness of our approach on five real-world datasets from PeerWise, a widely-used learnersourcing platform. Our experimental results show that our approach achieves more accurate and stable predictions of student performance over several other baseline methods.

For future work, we aim to utilize our performance prediction model to enhance personalized learning experiences on learnersourcing platforms. Our plan involves developing a recommendation system for dynamically tailoring question suggestions based on individual student strengths and weaknesses. Exploring adaptive feedback mechanisms and real-time interventions based on predicted performance will contribute to a more effective and engaging learning environment.

Acknowledgements

References

Abdi, S.; Khosravi, H.; and Sadiq, S. 2021. Modelling Learners in Adaptive Educational Systems: A Multivariate Glicko-Based Approach. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, LAK21, 497–503. New York, NY, USA: Association for Computing Machinery. ISBN 9781450389358.

Brown, T.; Mann, B.; Ryder; et al. 2020. Language models are few-shot learners. *Neurips*, 33: 1877–1901.

Daud, A.; Aljohani, N. R.; Abbasi, R. A.; Lytras, M. D.; Abbas, F.; and Alowibdi, J. S. 2017. Predicting student performance using advanced learning analytics. In *Proceedings of the 26th international conference on world wide web companion*, 415–421.

Denny, P.; Luxton-Reilly, A.; and Hamer, J. 2008. The PeerWise system of student contributed assessment questions. In *Proceedings of the tenth conference on Australasian computing education-Volume 78*, 69–74. Citeseer.

Denny, P.; Sarsa, S.; Hellas, A.; and Leinonen, J. 2022. Robosourcing Educational Resources – Leveraging Large Language Models for Learnersourcing. arXiv:2211.04715.

Derr, T.; Ma, Y.; and Tang, J. 2018. Signed graph convolutional networks. In *2018 IEEE International Conference on Data Mining (ICDM)*, 929–934. IEEE.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Du, Y.; Zhu, X.; Chen, L.; Fang, Z.; and Gao, Y. 2022. Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Huang, J.; Shen, H.; Cao, Q.; Tao, S.; and Cheng, X. 2021. Signed Bipartite Graph Neural Networks. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 740–749.

Jeevamol, J.; and Renumol, V. 2021. An ontology-based hybrid e-learning content recommender system for alleviating the cold-start problem. *Education and Information Technologies*, 26: 4993–5022.

Jiang, S.; Williams, A.; Schenke, K.; Warschauer, M.; and O'dowd, D. 2014. Predicting MOOC performance with week 1 behavior. In *Educational data mining 2014*.

Jung, J.; Yoo, J.; and Kang, U. 2020. Signed Graph Diffusion Network. *arXiv preprint arXiv:2012.14191*.

Khosravi, H.; Demartini, G.; Sadiq, S.; and Gasevic, D. 2021. Charting the Design and Analytics Agenda of Learnersourcing Systems. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, LAK21, 32–42. New York, NY, USA: Association for Computing Machinery. ISBN 9781450389358.

Khosravi, H.; Kitto, K.; and Williams, J. J. 2019. RiPPLE: A crowdsourced adaptive platform for recommendation of learning activities. *arXiv preprint arXiv:1910.05522*.

Kim, J.; Park, H.; Lee, J.-E.; and Kang, U. 2018. Side: representation learning in signed directed networks. In *Proceedings of the 2018 World Wide Web Conference*, 509–518.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Li, H.; Wei, H.; Wang, Y.; Song, Y.; and Qu, H. 2020. Peer-inspired student performance prediction in interactive online question pools with graph neural network. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2589–2596.

Li, M.; Zhang, Y.; Li, X.; Cai, L.; and Yin, B. 2022. Multi-view hypergraph neural networks for student academic performance prediction. *Engineering Applications of Artificial Intelligence*, 114: 105174.

Liu, T.; Wu, Q.; Chang, L.; and Gu, T. 2022. A review of deep learning-based recommender system in e-learning environments. *Artificial Intelligence Review*, 55(8): 5953–5980.

Lyu, L.; Wang, Z.; Yun, H.; Yang, Z.; and Li, Y. 2022. Deep knowledge tracing based on spatial and temporal representation learning for learning performance prediction. *Applied Sciences*, 12(14): 7188.

Mulryan-Kyne, C. 2010. Teaching large classes at college and university level: Challenges and opportunities. *Teaching in higher Education*, 15(2): 175–185.

Nakagawa, H.; Iwasawa, Y.; and Matsuo, Y. 2019. Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *IEEE/WIC/ACM International Conference on Web Intelligence*, 156–163.

Ni, L.; Bao, Q.; Li, X.; Qi, Q.; Denny, P.; Warren, J.; Witbrock, M.; and Liu, J. 2022. Deepqr: Neural-based quality ratings for learnersourced multiple-choice questions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12826–12834.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Pang, H.; Giunchiglia, F.; Li, X.; Guan, R.; and Feng, X. 2022. PNMTA: A pretrained network modulation and task adaptation approach for user cold-start recommendation. In *Proceedings of the ACM Web Conference 2022*, 348–359.

Paramythis, A.; and Loidl-Reisinger, S. 2003. Adaptive learning environments and e-learning standards. In *Second european conference on e-learning*, volume 1, 369–379.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L. J.; and Sohl-Dickstein, J. 2015. Deep knowledge tracing. *Advances in neural information processing systems*, 28.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1): 5485–5551.

Riggs, C. D.; Kang, S.; and Rennie, O. 2020. Positive impact of multiple-choice question authoring and regular quiz participation on student learning. *CBE—Life Sciences Education*, 19(2): ar16.

Safarov, F.; Kutlimuratov, A.; Abdusalomov, A. B.; Nasimov, R.; and Cho, Y.-I. 2023. Deep Learning Recommendations of E-Education Based on Clustering and Sequence. *Electronics*, 12(4): 809.

Schein, A. I.; Popescul, A.; Ungar, L. H.; and Pennock, D. M. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 253–260.

Shu, L.; Du, E.; Chang, Y.; Chen, C.; Zheng, Z.; Xing, X.; and Shen, S. 2021. SGCL: Contrastive Representation Learning for Signed Graphs. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 1671–1680.

Shuai, J.; Zhang, K.; Wu, L.; Sun, P.; Hong, R.; Wang, M.; and Li, Y. 2022. A review-aware graph contrastive learning framework for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1283–1293.

Sohn, K. 2016. Improved deep metric learning with multiclass n-pair loss objective. *Advances in neural information processing systems*, 29.

Sun, G.; Cui, T.; Beydoun, G.; Chen, S.; Dong, F.; Xu, D.; and Shen, J. 2017. Towards massive data and sparse data in adaptive micro open educational resource recommendation: a study on semantic knowledge base construction and cold start problem. *Sustainability*, 9(6): 898.

Thaker, K.; Carvalho, P.; and Koedinger, K. 2019. Comprehension Factor Analysis: Modeling student's reading behaviour: Accounting for reading practice in predicting students' learning in MOOCs. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 111–115.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. *ICLR (Poster)*, 2(3): 4.

Wang, X.; Talluri, S. T.; Rose, C.; and Koedinger, K. 2019. UpGrade: Sourcing student open-ended solutions to create scalable learning opportunities. In *Proc. of the Sixth (2019) ACM Conf. on Learning@ Scale*, 1–10.

Wang, Z.; Yan, W.; Zeng, C.; Tian, Y.; Dong, S.; et al. 2023. A unified interpretable intelligent learning diagnosis framework for learning performance prediction in intelligent tutoring systems. *International Journal of Intelligent Systems*, 2023.

Wei, H.; Li, H.; Xia, M.; Wang, Y.; and Qu, H. 2020. Predicting student performance in interactive online question pools using mouse interaction features. In *Proceedings of the tenth international conference on learning analytics & knowledge*, 645–654.

Xianghan, O.; and Stern, J. 2022. Virtually the same?: Online higher education in the post Covid-19 era. *British Journal of educational technology*, 53(3): 437.

Ye, J.; Chen, X.; Xu, N.; Zu, C.; Shao, Z.; Liu, S.; Cui, Y.; Zhou, Z.; Gong, C.; Shen, Y.; et al. 2023. A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv preprint arXiv:2303.10420*.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33: 5812–5823.

Zhang, Z.; Liu, J.; Zhao, K.; Yang, S.; Zheng, X.; and Wang, Y. 2023a. Contrastive Learning for Signed Bipartite Graphs. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1629–1638.

Zhang, Z.; Liu, J.; Zheng, X.; Wang, Y.; Han, P.; Wang, Y.; Zhao, K.; and Zhang, Z. 2023b. RSGNN: A Model-agnostic Approach for Enhancing the Robustness of Signed Graph Neural Networks. In *Proceedings of the ACM Web Conference 2023*, 60–70.

Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph neural networks: A review of methods and applications. *AI open*, 1: 57–81.

Zhu, Y.; Xu, Y.; Yu, F.; Liu, Q.; Wu, S.; and Wang, L. 2020. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*.

Zhu, Y.; Xu, Y.; Yu, F.; Liu, Q.; Wu, S.; and Wang, L. 2021. Graph contrastive learning with adaptive augmentation. In *Proceedings of the Web Conference 2021*, 2069–2080.