Difficulty-Focused Contrastive Learning for Knowledge Tracing with a Large Language Model-Based Difficulty Prediction

Unggi Lee[1,5], Sungjun Yoon[2], Joon Seo Yun[3], Kyoungsoo Park[3], YoungHoon Jung[3], Damji Stratton[4], Hyeoncheol Kim[1†]

Korea University[1], Baedal Minjok[2], i-Scream Edu[3], The University of Missouri System[4], Enuma, Inc.[5]

LREC-COLING 2024

Introduction

LREC-COLING 2024

Knowledge Tracing and Difficulty

- Knowledge tracing (KT) is a field of research that aims to predict student learning progress by analyzing their past interactions with question items within an educational context (Abdelrahman et al., 2023; Corbett and Anderson, 1994).
- Difficulty estimation is crucial for understanding dynamic student learning progress
    - Item response theory (IRT)
    - Classical test theory (CTT)

Research Gap

- Few studies have focused on incorporating difficulty information to improve model performance in contrastive learning based KT
- The potential role of natural language in KT is not yet fully understood
  - Only use to improve performance, not considering difficulty

DCL4KT + LLM

- This study proposes a new model called Difficulty-Focused Contrastive Learning for Knowledge Tracing with a Large Language Model (DCL4KT+ LLM)
- Utilizes CTT to calculate concept difficulty and question difficulty
- Incorporates the contrastive learning framework
- Leverages textual features of questions using LLMs


Background
LREC-COLING 2024

Difficulty in KT

- Item Response Theory (IRT) (Edelen and Reeve, 2007)
  - Attentive Knowledge Tracing (AKT) used the Rasch embedding strategy inspired by IRT (Ghosh et al., 2020)
- Classical Test Theory (CTT) (Petrillo et al., 2015)
  - BEKT (Tiana et al., 2021), MonaCoBERT (Lee et al., 2022a), CL4KT (Lee et al., 2022b)
- Graph Neural Network (GNN)-based models in KT used difficulty representation based on question-student response relationships
  - Song et al., 2022 and Luo et al., 2022

Contrastive Learning in KT

- Contrastive learning
  - MOCO (He et al., 2020) improved performance in unsupervised visual representation tasks using a dynamic dictionary and moving-averaged encoder
  - SimCLR (Chen et al., 2020a, 2020b) achieved better performance using data augmentation for contrastive learning
- Few KT studies with contrastive learning framework
  - CL4KT (Lee et al., 2022b) used reversed answer data as negative samples and suggested data augmentation techniques
  - Some attempts to combine contrastive learning and GNN (Song et al., 2022; Wu and Ling, 2023; Dai et al., 2022)
- Limited exploration of the role of difficulty in enhancing contrastive learning-based KT model performance

KT with NLP

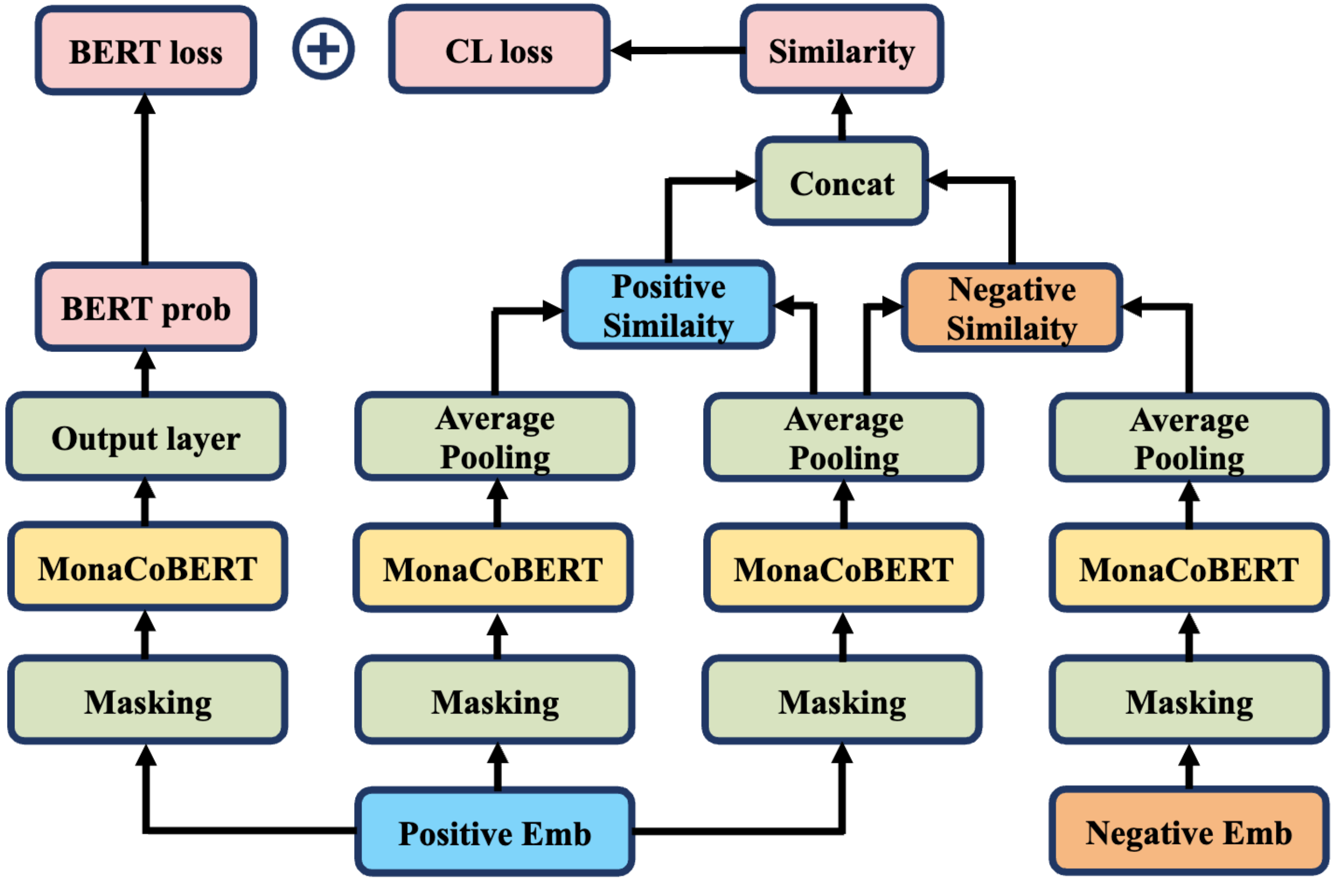
- Question text contains valuable information
- Several deep learning KT models have leveraged textual features to learn question representations and track students' knowledge states
  - RKT and HGKT extract features from question text to learn question representations
  - EERNN and EKT consider both exercising records and exercise texts for predicting student performance (Su et al., 2018; Liu et al., 2019)
  - AdaptKT and In Exercise Hierarchical Feature Enhanced Knowledge Tracing utilize BERT (Cheng et al., 2022; Tong et al., 2020)
  - QuesNet is an unsupervised learning method that leverages a large corpus of unlabelled questions (Yin et al., 2019)
- Previous research has not focused much on the latent difficulty representation of textual features in questions and concepts


Methodology
LREC-COLING 2024

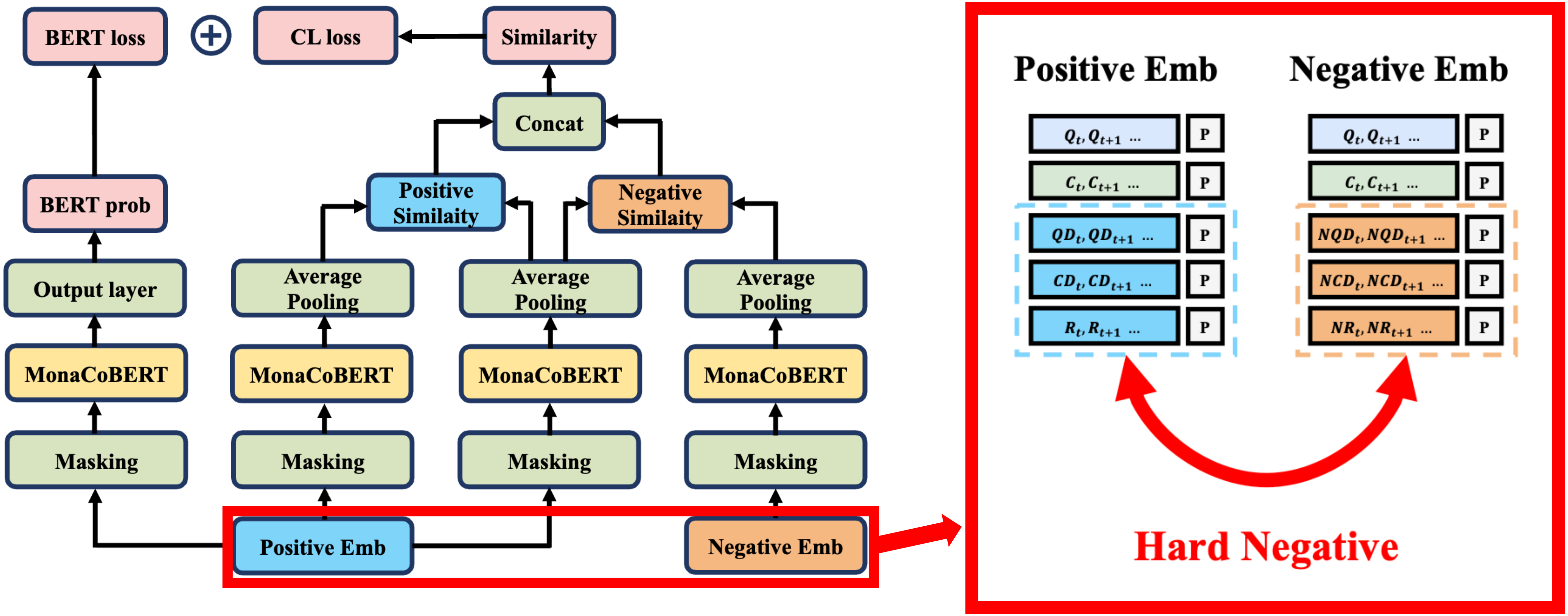
Problem Statement

- KT predicts the likelihood of a student answering accurately by analyzing interaction data from LMS or ITS
- Student interactions can be represented as $x_1, \ldots, x_t$
- Each interaction (x_t) consists of:
  - c_t : educational concept associated with the t-th inquiry
  - q_t: question's identifier
  - r_t: student's response (0 for incorrect, 1 for correct)
- Difficulties can be divided into:
  - cd_t: concept difficulties
  - qd_t: question difficulties
- Difficulty is set to an integer value ranging from 0 to 100
- Based on classical test theory (CTT), difficulty is calculated as:
  - Number of students who got the question (concept) correct / Total number of questions (concepts)

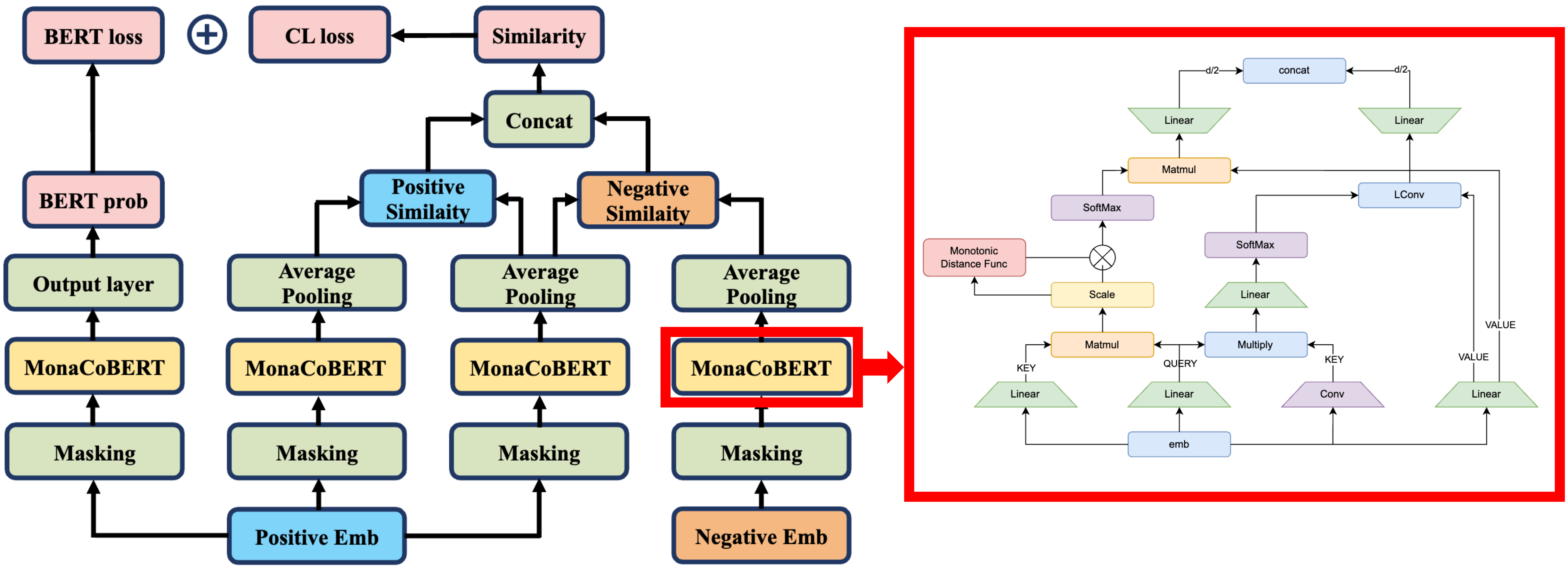
Proposed Model Architecture



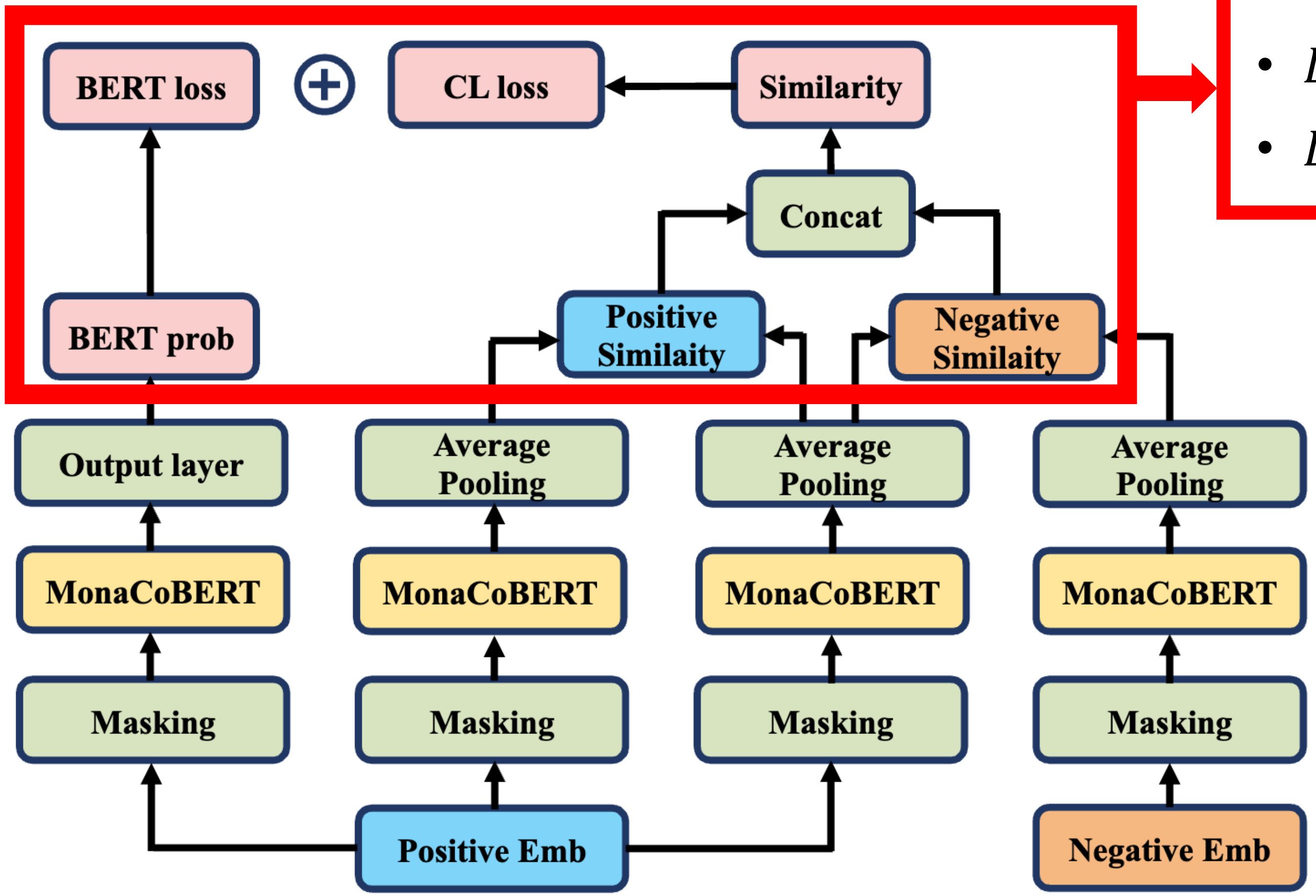
Embedding Layers with Hard Negative



MonaCoBERT based Encoders

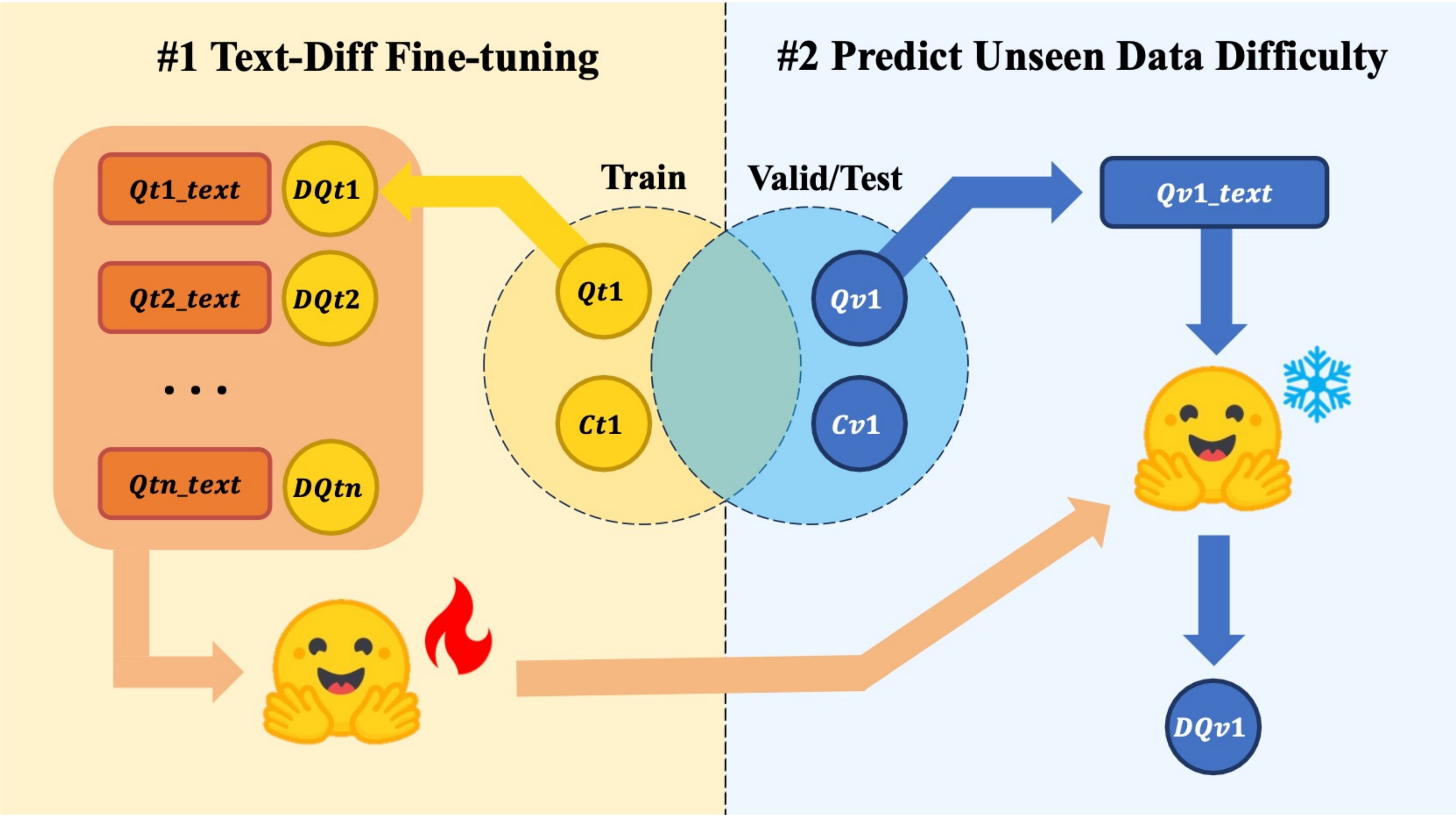


Loss Function



- $L = (1-\lambda_c) \times L_{bce} + \lambda_c \times L_{cl}$
- $L_{bce} = \sum_t -(r_t log\, \hat{r}_t + (1 - r_t) \log(1 - \hat{r}_t))$
- $L_{cl} = concat(sim_c, sim_q)$

LLM-based Difficulty Prediction



Data Augmentation

- Token cutoff, span cutoff (Shen et al., 2020)
- Concept and question mask (Lee et al., 2022b)
- Crop (Lee et al., 2022b)
- Summarize
- Reverse
- Permute (Lee et al., 2022b; Yang et al., 2019)
- Segment permute
- Replace higher and lower difficulty (Lee et al., 2022b)
- Concatenate sequence

Experiment Setting

- Datasets
  - ASISSTment09
  - Algebra05, 06
  - EdNet
  - Homerun20: Text contain (not published)
- Evaluatoin Metrics and Validation
  - AUC, RMSE
  - Five-fold cross validation

- Hyperparameters
  - Batch size: 512
  - Early stop: 10
  - Train, validation, test ratio
    - Train: 0.8
    - Test: 0.2
    - Valid: 10% of Train
  - LR: 0.001
  - Optimizer: Adam
  - Embedding size: 512

Result and Discussion

LREC-COLING 2024

Overall Performance

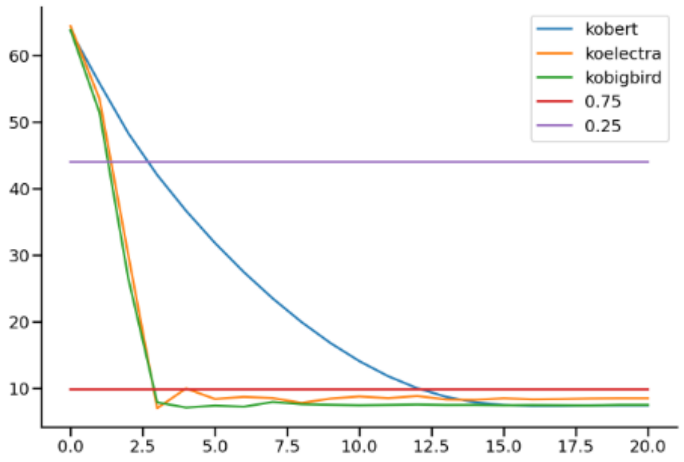
Dataset	Metrics	DKT	DKVMN	AKT	CL4KT	MCB-C	DCL4KT	DCL4KT-A
ASSISTments09	AUC	0.7285	0.7271	0.7449	0.7600	0.8059	<u>0.8111</u>	**0.8153**
	RMSE	0.4328	0.4348	0.4413	0.4337	<u>0.4063</u>	0.4068	**0.4034**
Algebra05	AUC	0.8088	0.8146	0.7673	0.7871	0.8201	<u>0.8288</u>	**0.8295**
	RMSE	0.3703	0.3687	0.3918	0.3824	**0.3584**	0.3657	<u>0.3644</u>
Algebra06	AUC	0.7939	0.7961	0.7505	0.7789	0.8064	<u>0.8258</u>	**0.8278**
	RMSE	0.3666	0.3661	0.3986	0.3863	0.3672	<u>0.3522</u>	**0.3504**
EdNet	AUC	0.6609	0.6602	0.6687	0.6651	0.7336	<u>0.7392</u>	**0.7403**
	RMSE	0.4598	0.4597	0.4783	0.4750	0.4516	<u>0.4505</u>	**0.4500**
Homerun20	AUC	0.7619	0.7543	0.5903	0.6014	0.7659	<u>0.7766</u>	**0.7808**
	RMSE	0.4092	0.4212	0.4745	0.4631	0.4880	<u>0.4042</u>	**0.4014**

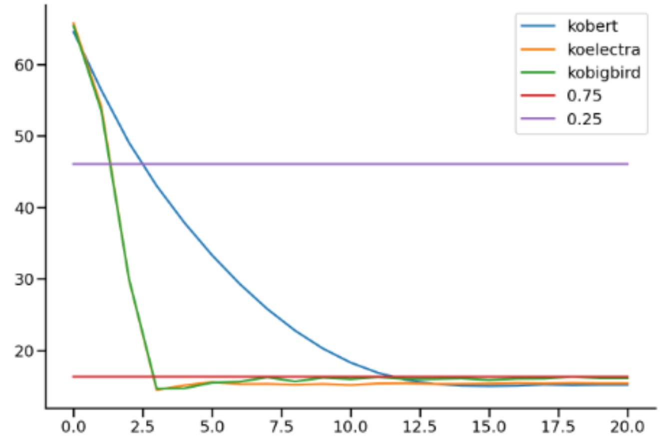
Effect of Difficulty-focused CL

- Non-Diff-CL.
  - Difficulty level of 0.75 is applied to all unseen data in both positive and negative embeddings.
- Diff-CL
  - Difficulty level is at 0.75 for positive embedding and 0.25 for negative embedding.
- Diff-CL achieved higher performance on all of the benchmark datasets..

Dataset	Metric	*Non-Diff-CL*	*Diff-CL*
ASSISTments09	AUC	0.8080	0.8111
	RMSE	0.4070	0.4068
Algebra05	AUC	0.8223	0.8288
	RMSE	0.3721	0.3657
Algebra06	AUC	0.8254	0.8258
	RMSE	0.3525	0.3522
EdNet	AUC	0.7357	0.7392
	RMSE	0.4598	0.4505

Difficulty Prediction with LLMs

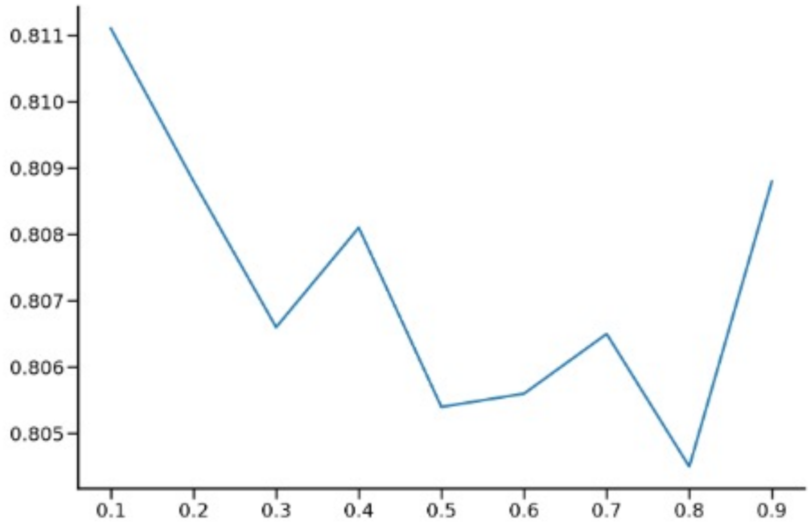




- Left: Concept difficulty prediction.
- Right: Question difficulty prediction between LLMs.
- The x-axis is training step and y-axis means RMSE score.
- The RMSE score of LLMs are lower than hyperparameter 0.75.
  - That means LLMs can predict difficulty by using text data of questions and concepts.

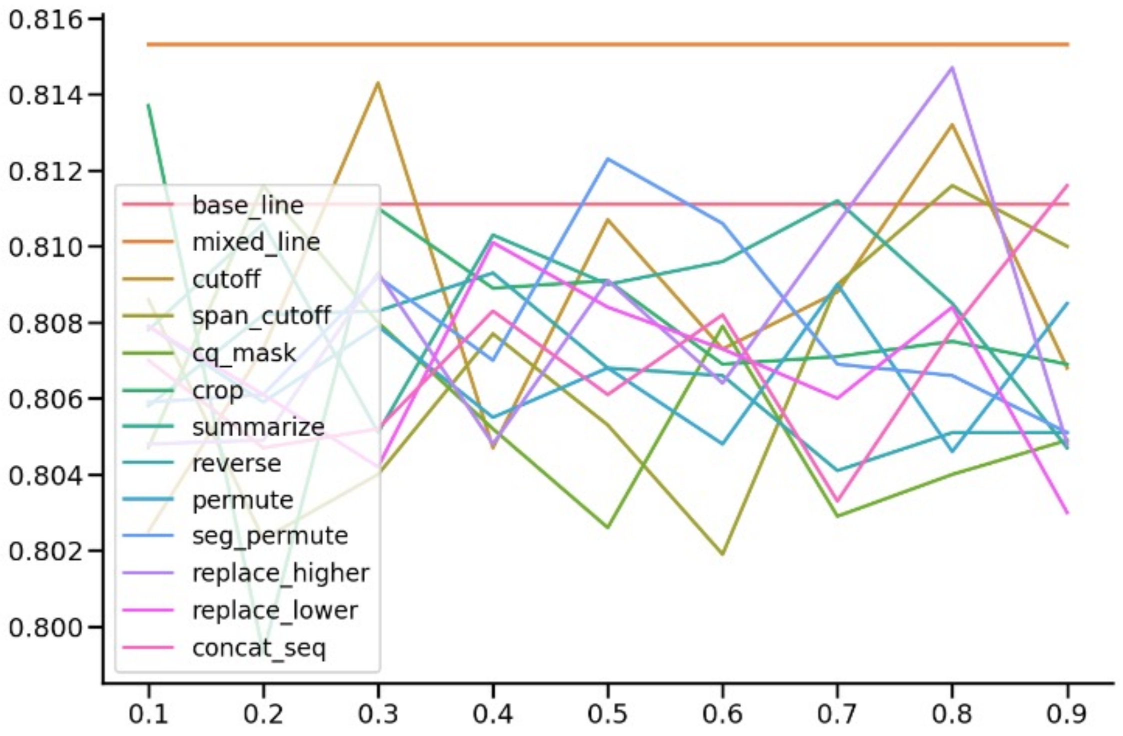
Contrastive Learning Loss Ratio

- Experiment how the contrastive learning framework affects the performance of the model (AUC)
  - Augmentation strategies have not been applied
- Contrastive learning loss ratio is 0.1, the performance is best (0.8111).
- Contrastive learning loss ratio is 0.8, the performance is worst (0.8045)



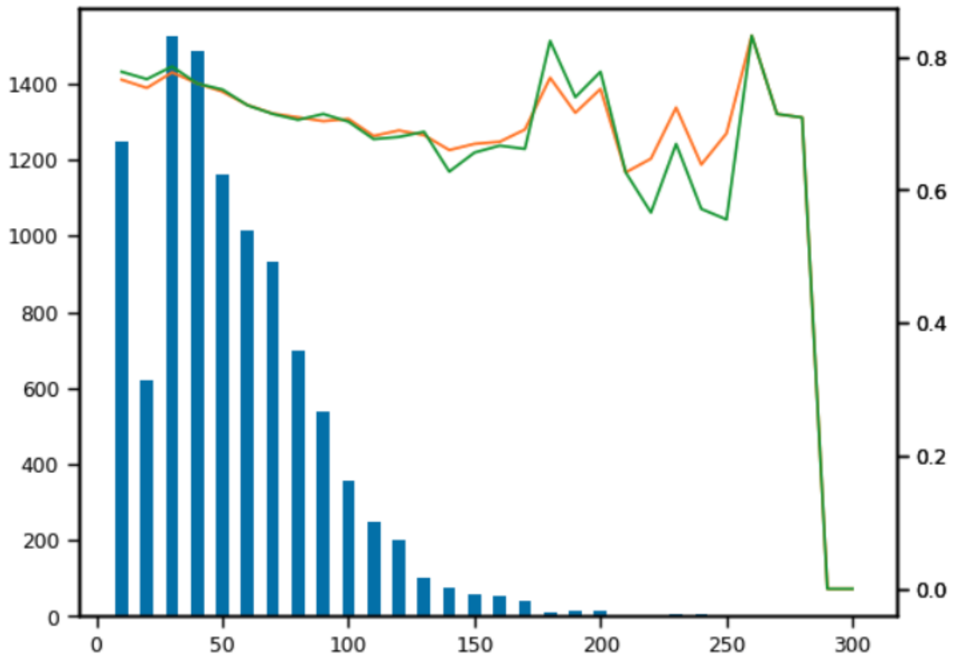
Effect of Data Augmentation

- The baseline is nonaugmented DCL4KT (0.8111)
- When we estimate performance of mixed augmentation, the probabilities are higher (0.8153) than performance of each augmentation independently



Relationship between language and difficulty

- x-axis: character length of query
- y-axis: character count
- Orange line: the mean level of correctness
- Green line: median level of correctness
- The graph depicts a decline in students' accuracy as character length increases, which holds for both the mean and median correctness



Conclusion

LREC-COLING 2024

Conclusion

- Difficulty level significantly impacts student learning habits and KT model efficacy
- Previous KT research has not fully exploited difficulty to improve performance and struggled to calculate difficulty in unseen data
- This study:
  - Developed a difficulty-centered contrastive learning technique for KT models
  - Proposed an LLM-based difficulty prediction framework
- These novel techniques can:
  - Optimize the performance of the KT model
  - Estimate the difficulty level of unknown data
  - Ablation investigation confirmed the efficacy of these new techniques for improving the KT model
- Future research:
  - Further study the relationship between language and difficulty
  - Identify linguistic characteristics that possibly indicate difficulty level

Reference

- The whole reference can be found in the original papers.