# 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]
codingchild@korea.ac.kr, yseong555@gmail.com, {jsyun0501, kspark0818, yhjung}@i-screamedu.co.kr, dhsdfn@umsystem.edu, harrykim@korea.ac.kr

## Intro

### Knowledge Tracing

· Knowledge tracing (KT) is a field of research that aims to predict student learning pro gress by analyzing their past interactions with question items within an educational co ntext (Abdelrahman et al., 2023; Corbett and Anderson, 1994)

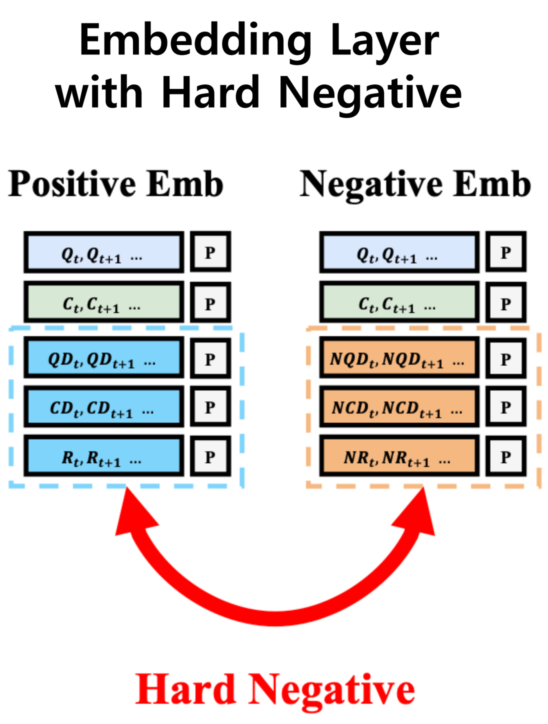
· Difficulty estimation is crucial for understanding dynamic student learning progress: Item response theory (IRT), Classical test theory (CTT)

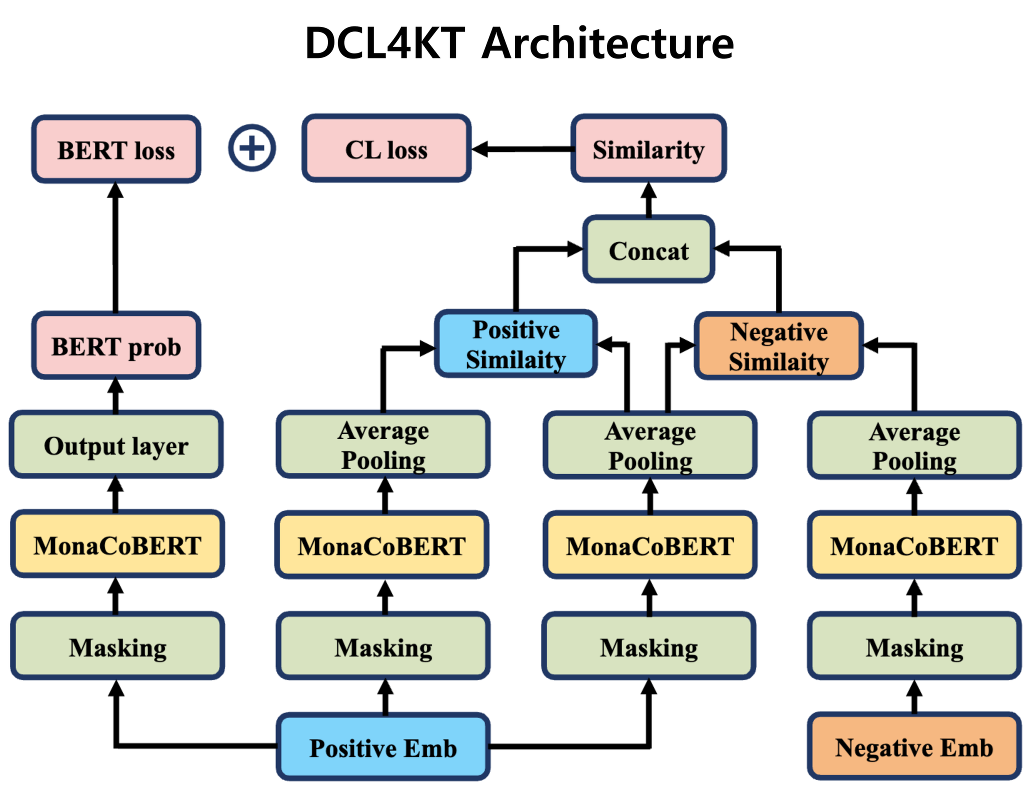
### Research Gap

· Few studies have focused on incorporating difficulty information to improve model p erformance 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.

### This Research

· In this research, we proposed Difficulty-Focused Contrastive Learning for Knowledge Tracing with a Large Language Model-Based Difficulty Prediction (DCL4KT+LLM).



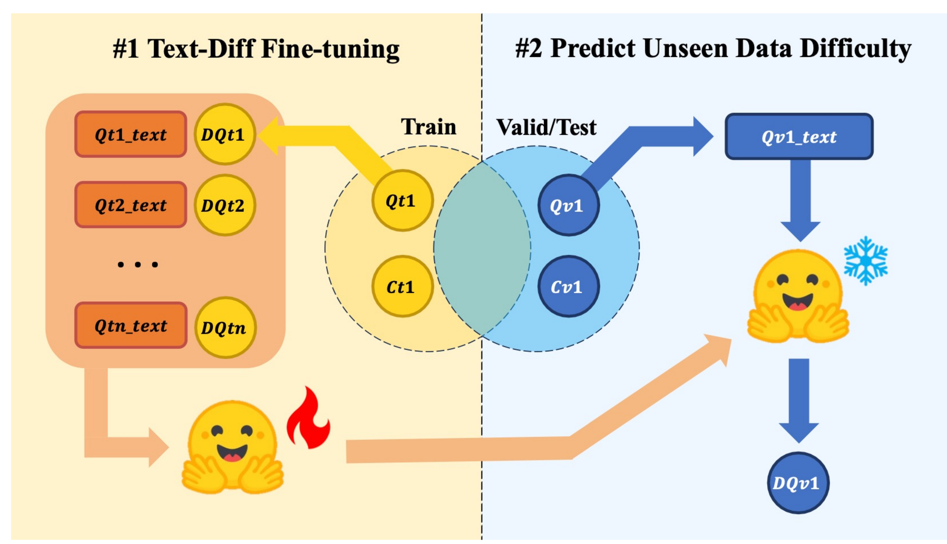




## Method

### DCL4KT+LLM Architecture

- **Embedding Layers with Hard Negative**: We implement hard negative strategies to response as well as difficulty from questions and concept.
- **Encoder Architecture**: We use MonaCoBERT as Transformer encoder blocks.
- **Contrastive Learning Framework**: By using the positive similarity and negative similarity, we calculate contrastive loss.
- **Loss Function**: We use combination of BERT loss and contrastive loss.
  - $L = (1-\lambda_c) \times L_{bce} + \lambda_c \times L_{cl}$
  - $L_{bce} = \sum_t -(r_t log\ \hat{r}_t + (1 - r_t)\log(1 - \hat{r}_t))$
  - $L_{cl} = concat(sim_c, sim_q)$
- **LLM-based Difficulty Prediction Framework**
  - Encoder based-LMs are trained pairs of the textual features and difficulties in train data, to predict difficulties of unseen questions and concepts in valid, test data.
- **Data Augmentation**
  - Token cutoff, span cutoff, Concept and question mask, Crop, Summarize, Reverse, Permute, Segment permute, Replace higher and lower difficulty, Concatenate sequence

LLM-based Difficulty Prediction




### Experiment Setting

- **Dataset**: Assist09, Algebra05,06, EdNet, Homerun20 (unpublished)
- **Hyperparameters**
  - Batch size: 512
  - Early stop: 10
  - Train, Valid, Test
    - Train: 0.8
    - Valid: 0.1 of Train
    - Test: 0.2
  - LR and Opt: 0.001, Adam
  - CL ratio: 0.1

## Result

### Overall Performance

- Overall performance of KT models on datasets. The best performance is denoted in bold, and the second is underlined. DCK4KT-A means DCL4KT with augmentation strategies.

| 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** |

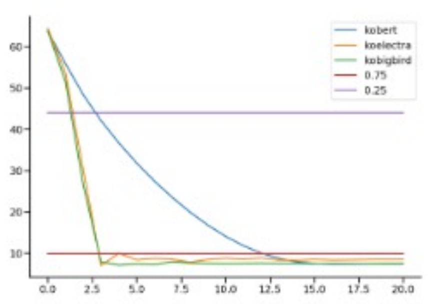
### Effect of Difficulty-focused Contrastive Learning

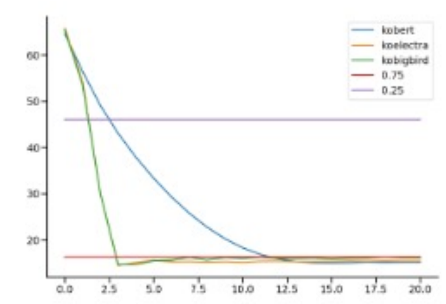
- Non-Diff-CL: Difficulty level is 0.75 is ap plied to all unseen data in both positive a nd negative embeddings.
- Diff-CL: Difficulty level is 0.75 for positiv e embedding and 0.25 for negative embe dding.
- Diff-CL achieved higher performance on all of the benchmark datasets.

| 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 using LLM

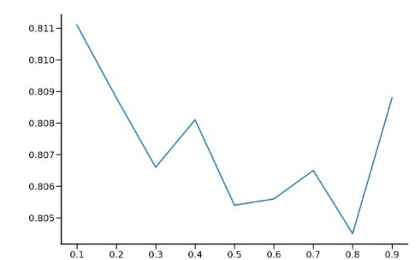
- Concept difficulty (left) and question difficulty (right) prediction between LLMs.
- X-axis is training step and Y-axis is RMSE score. LLMs can predict difficulty by using text data of questions and concepts.





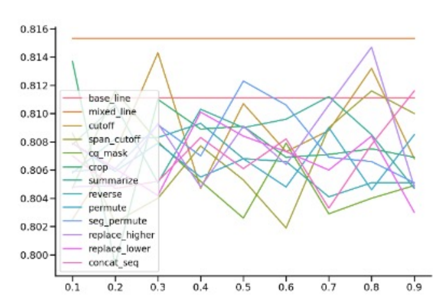
### Contrastive Learning Loss Ratio

- Relationship between contrastive learning ratio (x-axis) and model's AUC score (y-axis)
- Contrastive learning ratio 0.1 is best, 0.8 is worst performance.



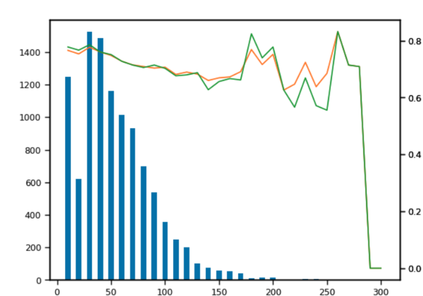
### Effect of Data Augmentation

- Baseline is non-augmented DCL4KT (0.8111)
- Mixed augmentation is best (0.8153).
- Independent augmentation strategies are not good to mixed augmentation strategies.



### Relationship between language and difficulty

- Relationship between character length and difficulty. X-axis is the character count of questions. Blue histogram mean the number of character length. Orange line is mean of correctness. Green line is median of correctness.
- When the character count is less than 120, student's correctness decreases as the character length increases.



## Conclusion

- We proposed a Difficulty-Focused Contrastive Learning for Knowledge Tracing with Large Language Model-based Difficulty Prediction Framework.
- As a result, most of the benchmark, DCL4KT with augmentation strategies achieve best performance in our settings.
- We investigate each part of model, we found that the effect of difficulty-focused contrastive learning , difficulty prediction using LLMs, contrastive learning loss ratio, eff ect of data augmentation and relationship between language and difficulty
- However, the relationship between language and difficulty requires additional study. In subsequent research, we intend to identify the linguistic characteristics that pos sibly indicate difficulty level.