The 37th AAAI Conference on Artificial Intelligence (AAAI-23)

Towards a Holistic Understanding of Mathematical Questions with Contrastive Pre-training

Yuting Ning[1,2], Zhenya Huang[1,2], Xin Lin[1,2], Enhong Chen[1,2], Shiwei Tong[1,2], Zheng Gong[1,2], Shijin Wang[2,3]

[1]Anhui Province Key Lab. of Big Data Analysis and Application, University of Science and Technology of China,
[2]State Key Laboratory of Cognitive Intelligence, [3]iFLYTEK AI Research (Central China), iFLYTEK Co., Ltd.

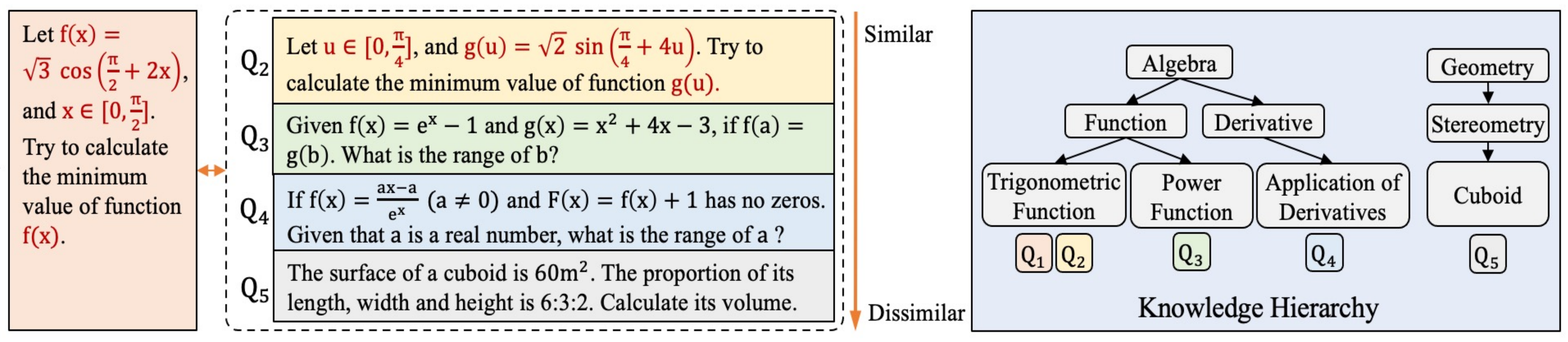
Introduction

Background

- Online learning systems collect massive educational questions.
- Mathematical question understanding is a crucial but challenging issue in intelligent education field.
  - more difficult to understand with special components (e.g., formulas) and complex mathematical logic
  - require more domain knowledge for comprehensive understanding

Challenges

- Mathematical questions are more complex with special components (e.g., formulas), and require more mathematical knowledge and logic.
- The holistic purposes of questions are more important than literal details.
- Related knowledge concepts play an important, since they reflect the purposes and mathematical domain of questions.



Problem Definition

Mathematical Questions

- consist of content and related knowledge concepts $\boldsymbol{q} = (\boldsymbol{x}, \boldsymbol{k})$
- Question content: $\boldsymbol{x} = \{x_1, x_2, \dots, x_T\}$
  a sequence of tokens, where each token is a word or symbol
- Related knowledge concepts: $\mathbf{k} = \{k_1, k_2, \dots, k_L\}$
  selected from a L-level knowledge hierarchy $KH = \{\mathcal{K}, \mathcal{E}\}$

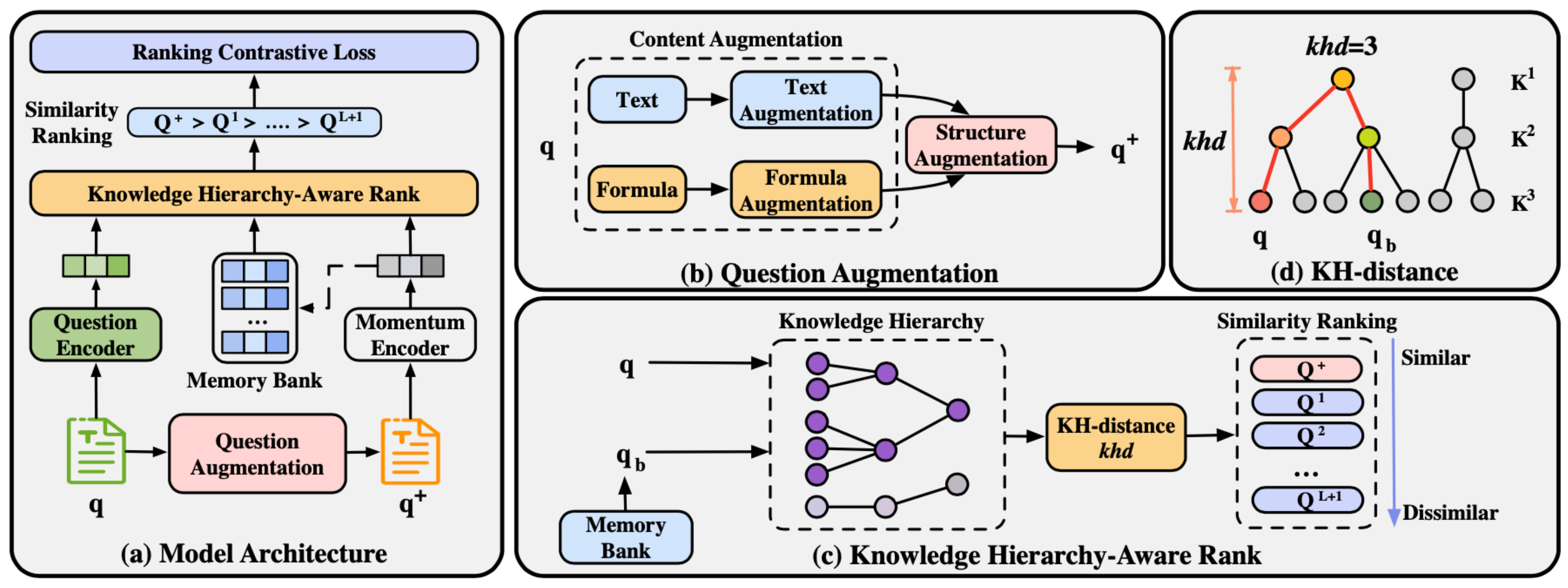
Question Representation Problem

- Given: mathematical question $\boldsymbol{q} = (\boldsymbol{x}, \boldsymbol{k})$
- Goal: a d-dimensional vector $\boldsymbol{v} \in \mathbb{R}^d$
  - be transferred to several downstream tasks and benefit their performances
  - capture latent purposes of questions
  - contain the rich information in question content and related knowledge concepts

QuesCo Framework

Main Idea

- learn comprehensive question representations by pulling questions with more similar purposes closer than those with less similar purposes



Question Augmentation

- To learn latent purposes of mathematical questions, generate questions with similar holistic purposes but diverse literal details
- Two-level augmentation
  - Content-level
    - Text Augmentation + Formula Augmentation
  - Structure-level
    - Structure Augmentation

Knowledge Hierarchy-Aware Rank

- Exploit fine-grained similarities between questions based on the relationship of mathematical knowledge concepts
- KH-distance

$$khd(q_i, q_j) = \begin{cases} L - u + 1 & \text{if } \exists u \in \{1, \cdots, L\}, \\ & k_i^u = k_j^u \text{ and } k_i^{u+1} \neq k_j^{u+1} \\ L + 1 & \text{if } \forall \ell \in \{1, \cdots, L\}, k_i^\ell \neq k_j^\ell \end{cases},$$

- Assign each question q_i in memory bank into one of $L + 1$ ranks

$$Q^u = \begin{cases} \{q^+\}, & u = 0 \\ \{p | khd(q, p) = u\}, & u \in \{1, \cdots, L+1\} \end{cases}.$$

- Similarity ranking $h(q, q^0) > h(q, q^1) > \cdots > h(q, q^{L+1}), \forall q^u \in Q^u,$

Pre-training

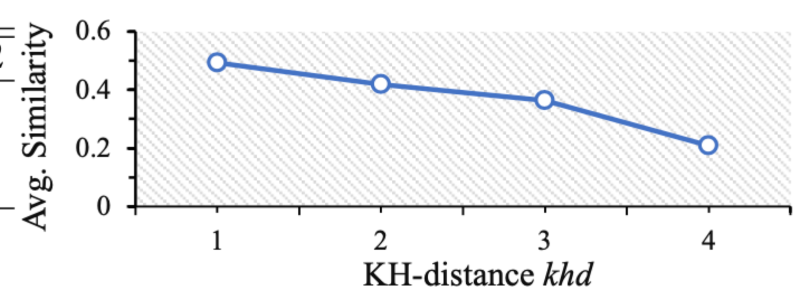
- Ranking Info Noise Contrastive Estimation loss (RINCE)
  - gradually decreasing similarity with increasing rank of samples

$$L_{rank} = \sum_0^L \ell_i \qquad \ell_i = -\log \frac{\sum_{p \in Q^i} \exp\left(\frac{h(q,p)}{\tau_i}\right)}{\sum_{p \in \bigcup_{j \geq i} Q^j} \exp\left(\frac{h(q,p)}{\tau_j}\right)}.$$

Experiments

Datasets

Statistics	SYSTEM1	SYSTEM2
# Questions	123,200	25,287
Avg. question length	80.08	130.91
Avg. formula length per question	48.02	84.48
# Hierarchical levels	3	3
# Knowledge in level-1	21	21
# Knowledge in level-2	81	54
# Knowledge in level-3	361	175
# Questions with difficulty label	7,056	/
# Questions with similarity label	/	6,873
Label sparsity	5.72%	27.18%



The relationship between *khd* and labeled similarity between questions in SYSTEM2.

Overall Performance

Tasks	Similarity Prediction		Concept Prediction								Difficulty Estimation			
Datasets	SYSTEM2		SYSTEM1				SYSTEM2				SYSTEM1			
Metrics	Pearson	Spearman	level-1 ACC	level-1 F1	level-2 ACC	level-2 F1	level-1 ACC	level-1 F1	level-2 ACC	level-2 F1	MAE	RMSE	PCC	DOA
BERT	0.2957	0.3655	0.7309	0.5213	0.4472	0.1833	0.4822	0.2374	0.2984	0.0945	0.1987	0.2463	0.3974	0.6318
DAPT-BERT	0.4856	0.5313	0.8032	0.6288	0.5597	0.2727	0.6522	0.3855	0.4960	0.1836	0.1880	0.2313	0.5087	0.6589
ConSERT	0.5060	0.4760	0.8064	0.6655	0.5933	0.3135	0.6987	0.4952	0.5020	0.2076	0.1873	0.2308	0.5115	0.6621
SCL	0.6901	0.7101	0.8985	0.8011	0.7492	0.4683	0.8083	0.6498	0.6225	0.3071	0.1996	0.2460	0.4002	0.6340
QuesNet	0.5370	0.5549	0.7881	0.6930	0.5693	0.3604	0.7194	0.6213	0.5810	0.3118	0.1865	0.2305	0.3959	0.6539
DisenQNet	0.6922	0.6955	0.8210	0.7064	0.6404	0.4332	0.7945	0.6805	0.2431	0.1023	0.1970	0.2424	0.4293	0.6338
QuesCo	**0.7385**	**0.7245**	**0.9176**	**0.8938**	**0.7857**	**0.5550**	**0.8340**	**0.7018**	**0.6719**	**0.3756**	**0.1778**	**0.2219**	**0.5623**	**0.6765**

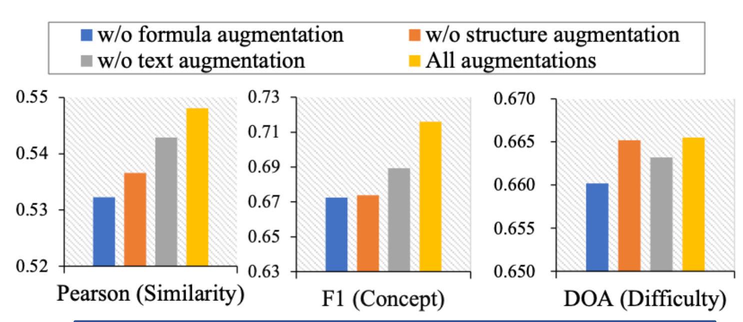
Performance comparisons on three typical downstream tasks.

Ablation Study

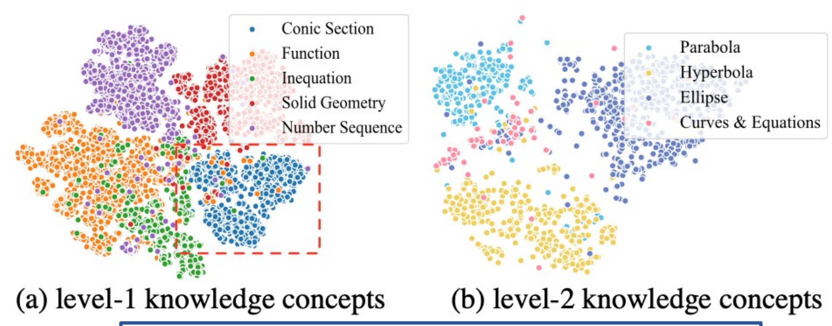
Tasks	Similarity Prediction		Concept Prediction								Difficulty Estimation			
Datasets	SYSTEM2		SYSTEM1				SYSTEM2				SYSTEM1			
Metrics	Pearson	Spearman	level-1 ACC	level-1 F1	level-2 ACC	level-2 F1	level-1 ACC	level-1 F1	level-2 ACC	level-2 F1	MAE	RMSE	PCC	DOA
QuesCo	**0.7385**	**0.7245**	**0.9176**	**0.8938**	**0.7857**	**0.5550**	**0.8340**	**0.7018**	**0.6719**	**0.3756**	**0.1778**	**0.2219**	**0.5623**	**0.6765**
w/o AUG	0.7028	0.7213	0.9079	0.8770	0.7305	0.4497	0.8320	0.6972	0.6443	0.3412	0.2007	0.2482	0.3797	0.6204
w/o KHAR	0.5481	0.5057	0.8202	0.7160	0.6181	0.3416	0.7332	0.5996	0.5613	0.2746	0.1810	0.2248	0.5475	0.6655

Effectiveness of each module.

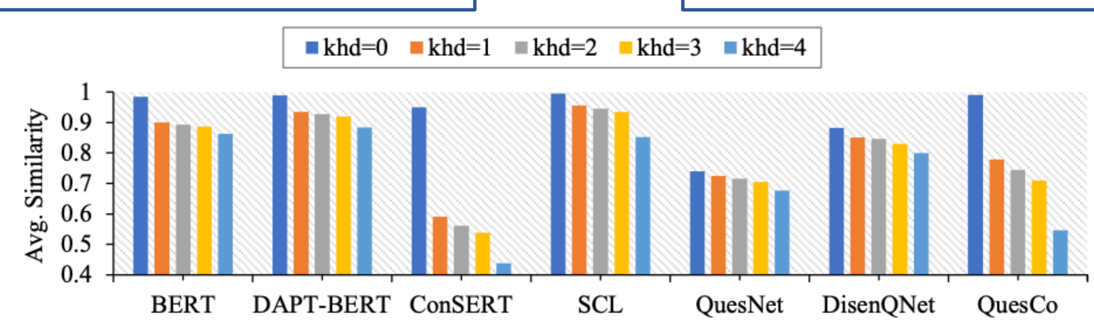
Model Analysis



Effectiveness of each augmentation.



Visualization.



Similarity Ranking Analysis: Relationship between *khd* and similarity.

Question	Knowledge Concept	*khd*	Similarity
$\boldsymbol{Q_1}$: Given that $x = 1.5^{-0.2}, y = 1.3^{0.7}, z = \left(\frac{2}{3}\right)^{\frac{1}{3}}$. What is the relationship between the magnitude of x, y, z?	Exponential Function	/	/
$\boldsymbol{Q_2}$: Given that $u = 1.5^{-0.2}, z = \left(\frac{2}{3}\right)^{\frac{1}{3}}, y = 2.3^{0.7}$. What is the relationship between the magnitude of u, y, z?		0	0.96
$\boldsymbol{Q_3}$: If $f(x) = e^x - 1, g(x) = -x^2 + 4x - 3$, and $f(a) = g(b)$. What the range of value of b?		1	0.70
$\boldsymbol{Q_4}$: What is the domain of definition of the function $f(x) = \ln(x^2 - x)$?	Logarithmic Function	2	0.64
$\boldsymbol{Q_5}$: $f(x) = (m^2 - m - 1) \cdot x^m$ $(m \in R)$ is a power function and is increasing when $x \in (0, +\infty)$. What is the value of m?	Power Function	3	0.54
$\boldsymbol{Q_6}$: If 2 cards are randomly selected from a mixed deck (52 cards in total), what is the probability of "both are Hearts"?	Independence of Events	4	0.42

Case Study: A case of questions with different similarities.