The 24nd ACM SIGKDD Conference on
Knowledge Discovery and Data Mining
2018/08/19-08/23 , London

Finding Similar Exercises in Online Education Systems

Qi Liu[1], Zai Huang[1], Zhenya Huang[1], Chuanren Liu[2], Enhong Chen[1,*] , Yu Su[3], Guoping Hu[4]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China
{qiliuql,cheneh}@ustc.edu.cn,{huangzai,huangzhy}@mail.ustc.edu.cn
[2]Decision Sciences & MIS Department, Drexel University, chuanren.liu@drexel.edu
[3]School of Computer Science and Technology, Anhui University, yusu@iflytek.com
[4]iFLYTEK Research, gphu@iflytek.com

Reporter: Zai Huang
Date: 2018.07.22

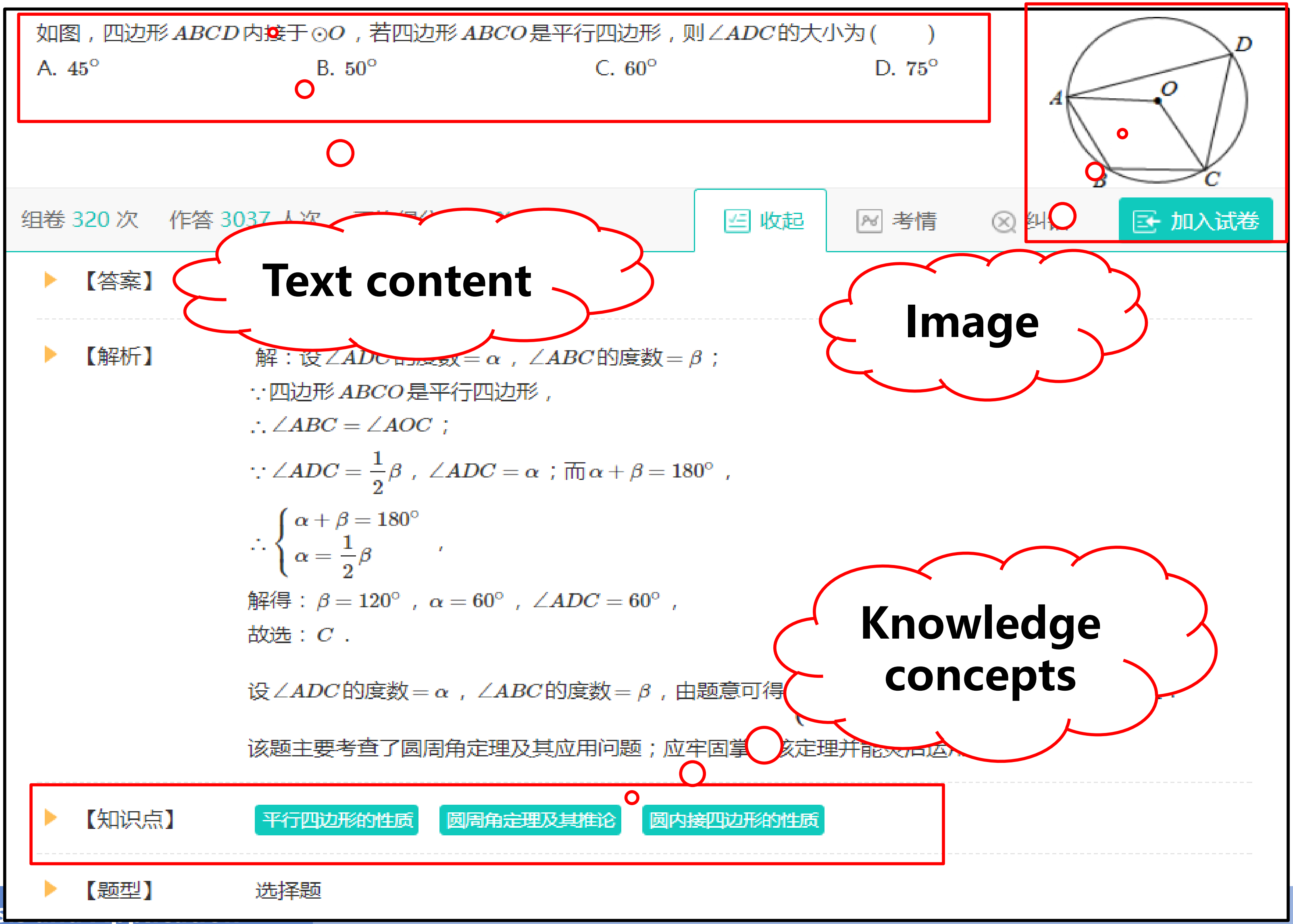
Outline

Background

- Online education systems
  - Such as KhanAcademy, Knewton, Zhixue
  - Exercise: collected millions of exercises
  - Applications: similar exercise retrieval and recommendation, personalized cognitive diagnosis based on exercise similarities
- Fundamental task
  - Finding Similar Exercises (FSE).
  - finding the similar ones of each given exercise

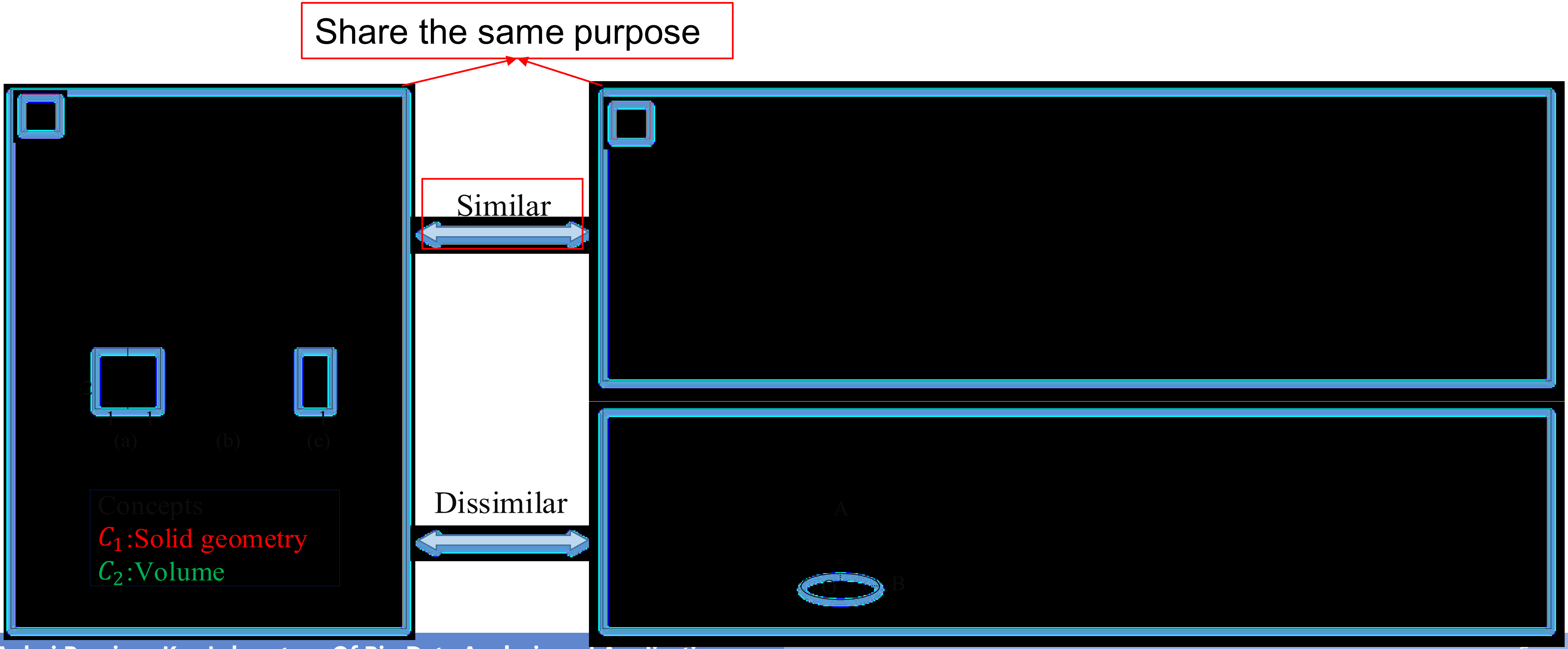
Exercise

- Exercise contains multiple heterogeneous data
  - Complex
  - Rich semantics



What are similar exercises?

- Following Educational Psychology, similar exercises are those having the same purpose embedded in exercise contents.



Background

- Existing solutions for Finding Similar Exercises (FSE) task
  - Manual Labeling
    - On a small quantity of exercises
    - requires strong expertise and takes much time
    - not suitable for large-scale online education systems containing millions of exercises
  - Methods based on text similarity
    - Use the same concepts or the similar words
    - cannot exploit rich semantics in the heterogeneous data

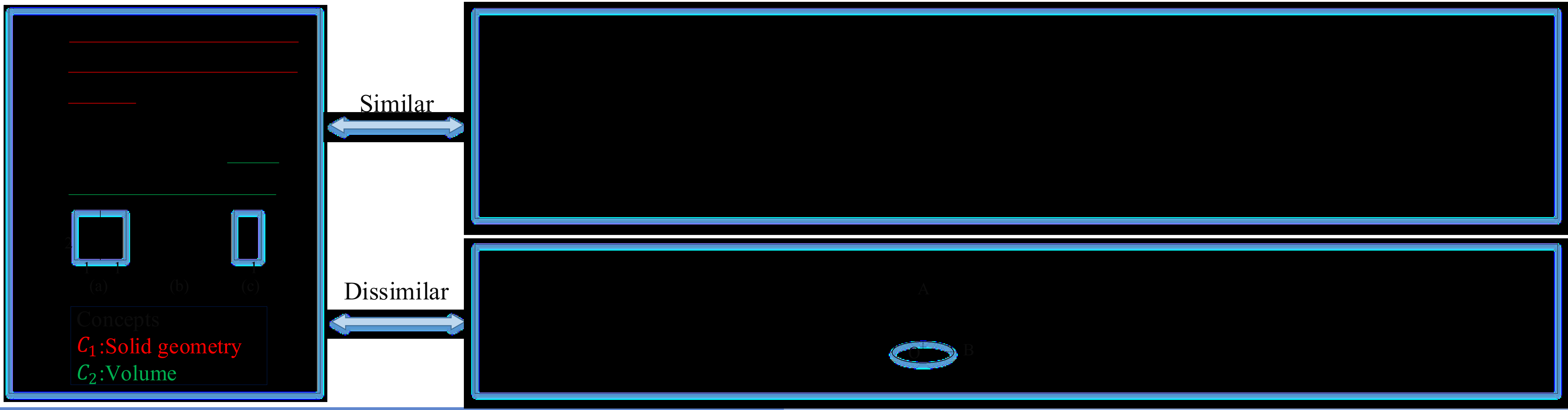
- Urgent Issue
  - Design an effective FSE solution for large-scale online education systems by exploit the heterogeneous data to understand exercise semantics and purposes.

Challenge 1 for FSE

- Exercises contain multiple heterogenous data.
  - texts
  - Images
  - knowledge concepts
- integrates multiple heterogeneous data to understand and represent exercise semantics and purposes.

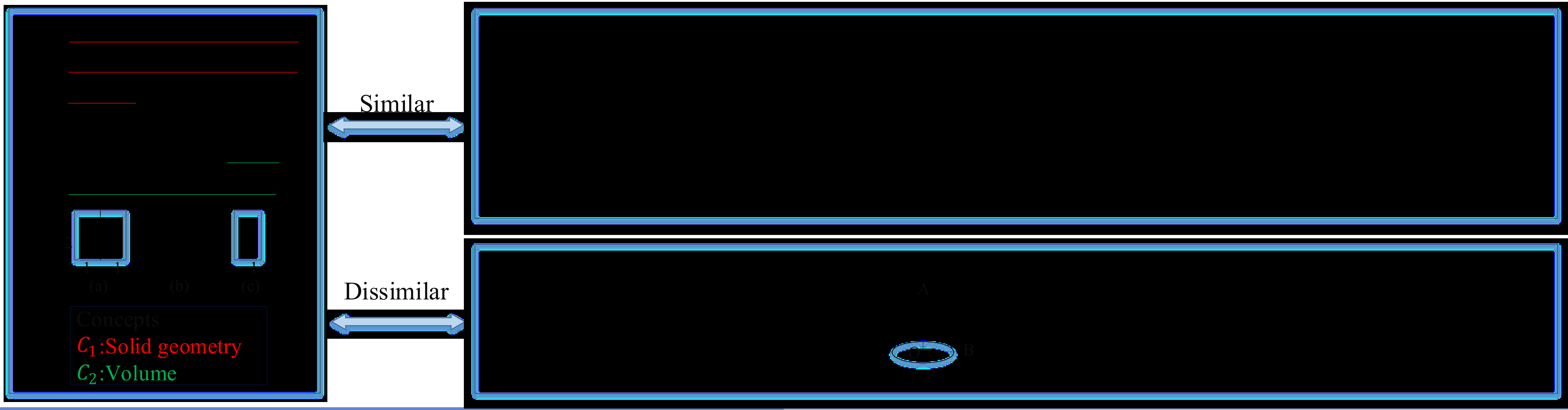
Challenge 2 for FSE

- In a single exercise, different parts/words of the text are associated with different concepts (text-concept) or images (text-image).
- For better understanding each exercise, it is necessary to capture these text-concept and text-image associations.



Challenge 3 for FSE

- A pair of similar exercises may consist of different texts, images and concepts.
- Finding similar exercises needs to measure the similar parts in each exercise pair by deeply interpreting their semantic relations.



Related Work

- Studies on FSE
  - Methods based on text similarity
    - Use the same concepts or the similar words
    - Vector Space Model (VSM)
  - Methods based on learners' performance data

Neglect semantics in heterogeneous materials of   exercises.

- Multimodal Learning
  - Powerful approach to handle heterogeneous data
  - Sound-video, video-text, image-text

Cannot understand exercise purposes or measure similar parts between two exercises

- Pair Modeling
  - Learn the relations between two instances in a pair
  - Sentence pair, image pair, video-sentence pair

Cannot handle instances having multiple heterogeneous data

Outline

Problem Definition

- Given: exercises with corresponding heterogeneous materials including texts, images and concepts

- Goal: learn a model $\mathcal{F}$ to measure the similarity scores of exercise pairs and find similar exercises for any exercise E by ranking the candidate ones $\mathcal{R}$ with similarity scores

$$\mathcal{F}(E, \mathcal{R}, \Theta) \rightarrow \mathcal{R}^s,$$

Model

Candidates for E

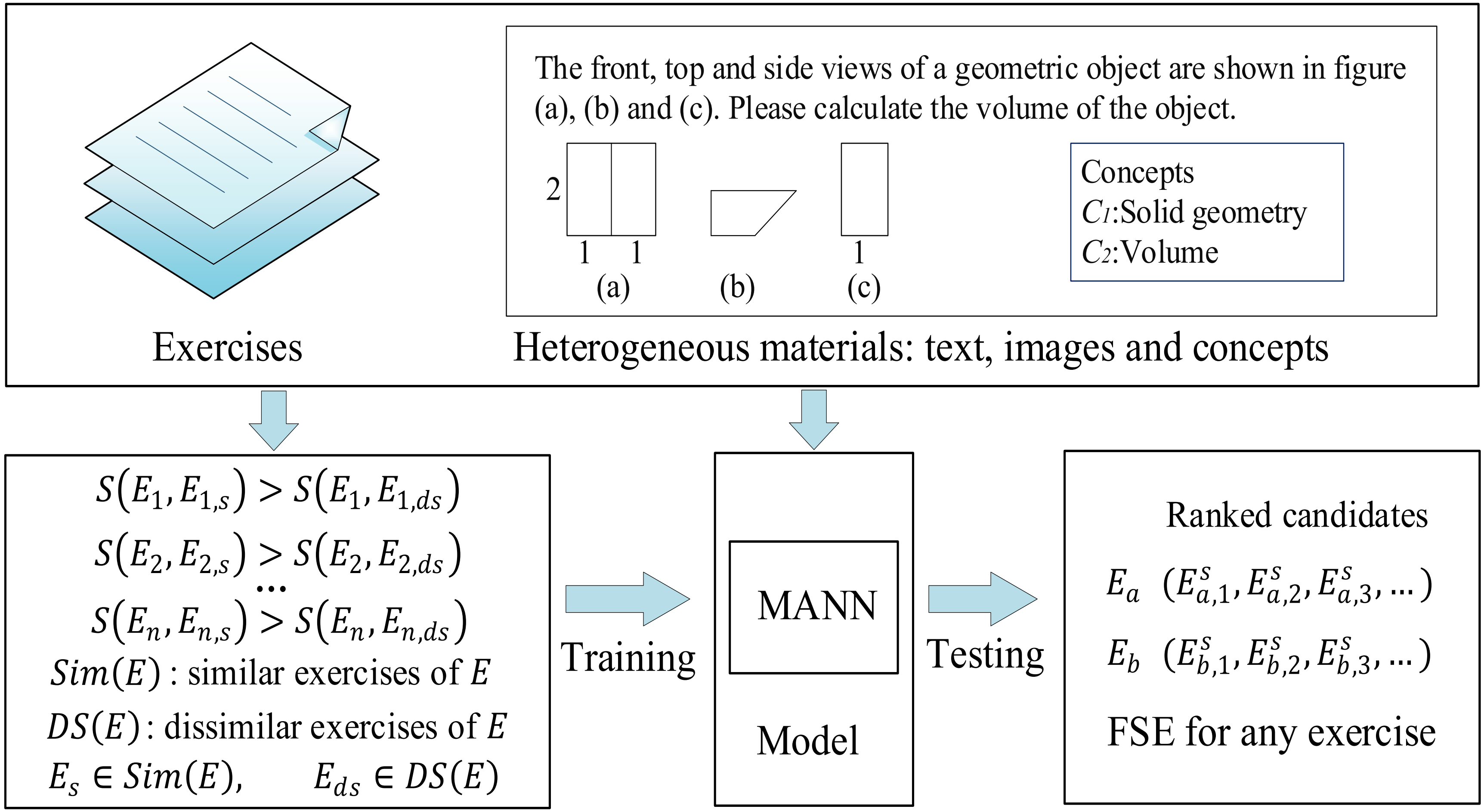
Parameters of $\mathcal{F}$

Similar exercises for E

Outline

Study Overview

- Two-stage solution
  - Training stage
    - MANN
    - Pairwise training
  - Testing stage
    - FSE for any exercise



Outline

MANN Framework

- Multimodal Attention-based Neural Network (MANN)

  - Learn a unified semantic representation of each exercise by handling its heterogeneous materials in a multimodal way

    Challenge 1: multimodal exercises understanding and representation

  - Propose two attention strategies to capture the text-image and text-concept associations in each single exercise

    Challenge 2: learning text-image, text-concept associations

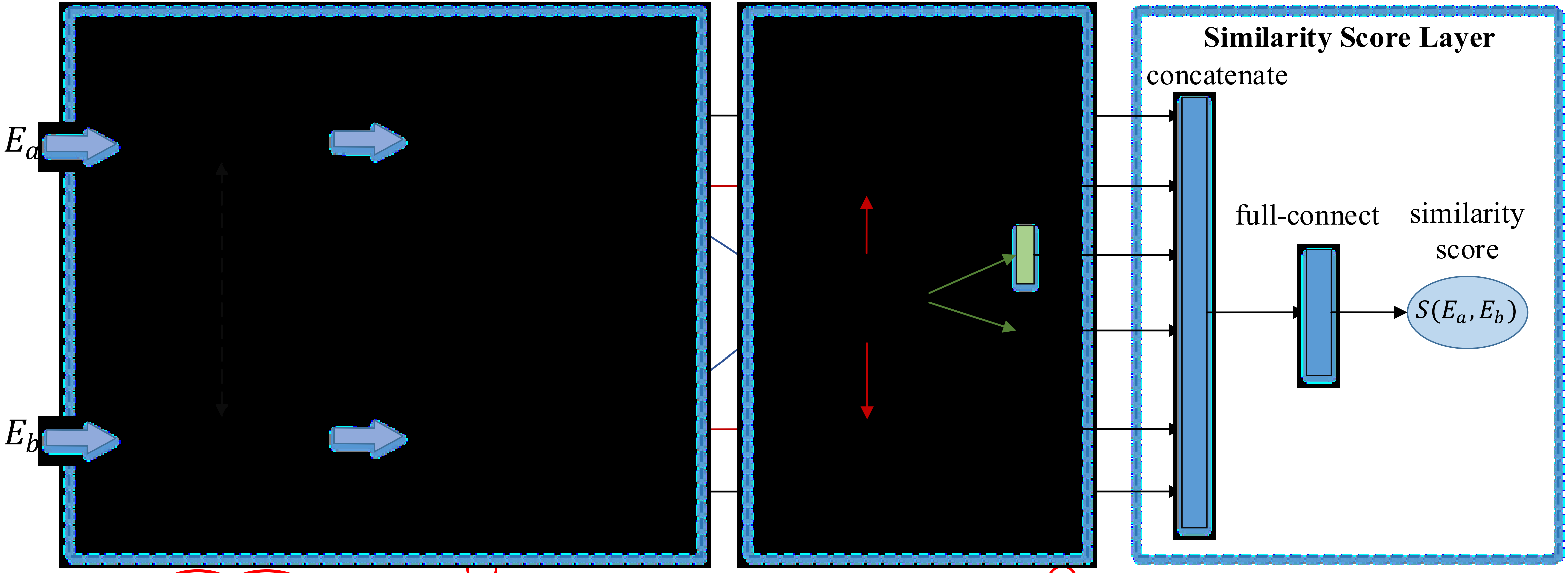
  - Design a Similarity Attention to measure the similar parts in each exercise pair with their semantic representations

    Challenge 3: learning similar parts

MANN Framework

Input： a pair of exercise ($\mathrm{E}_a, \mathrm{E}_b$)

Output: similarity score

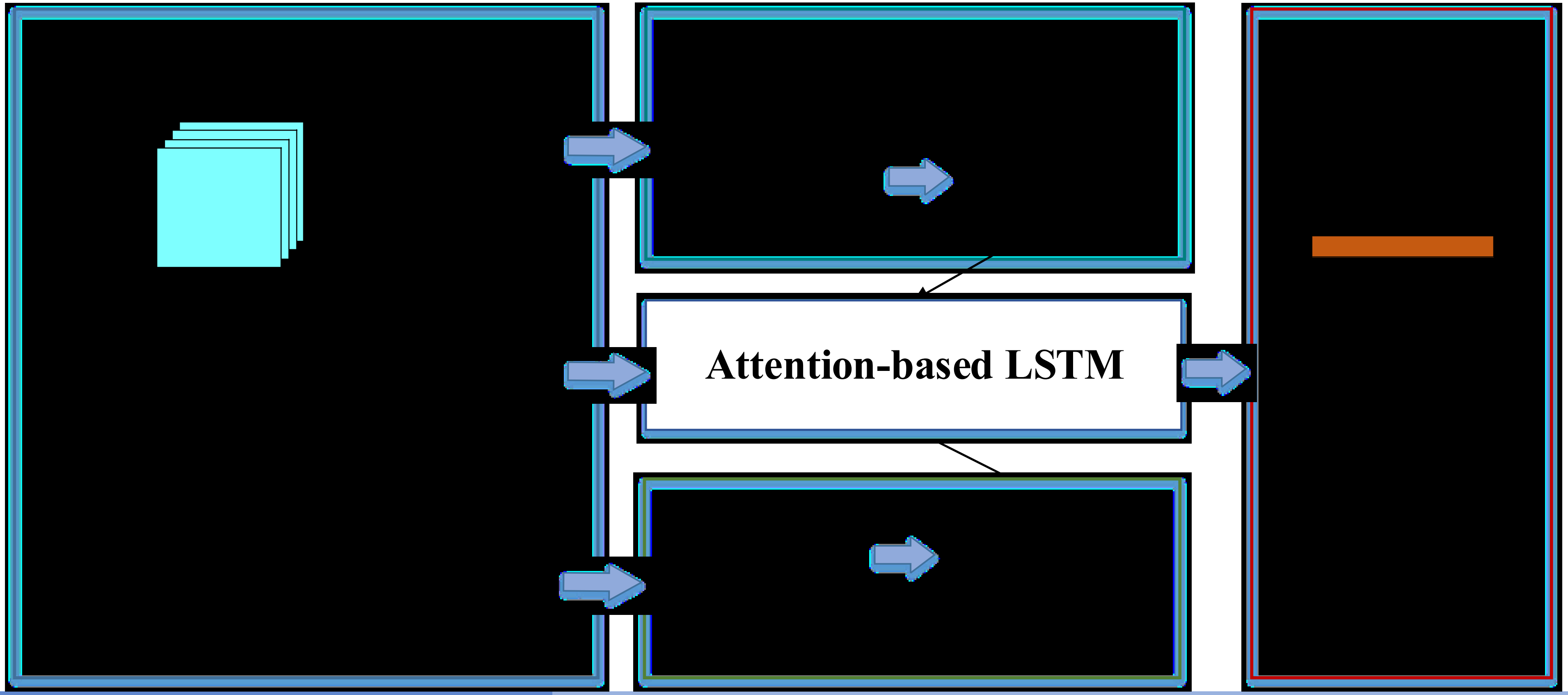


Challenge 1: multimodal exercises understanding and representation

Challenge 2: learning text-image, text-concept associations

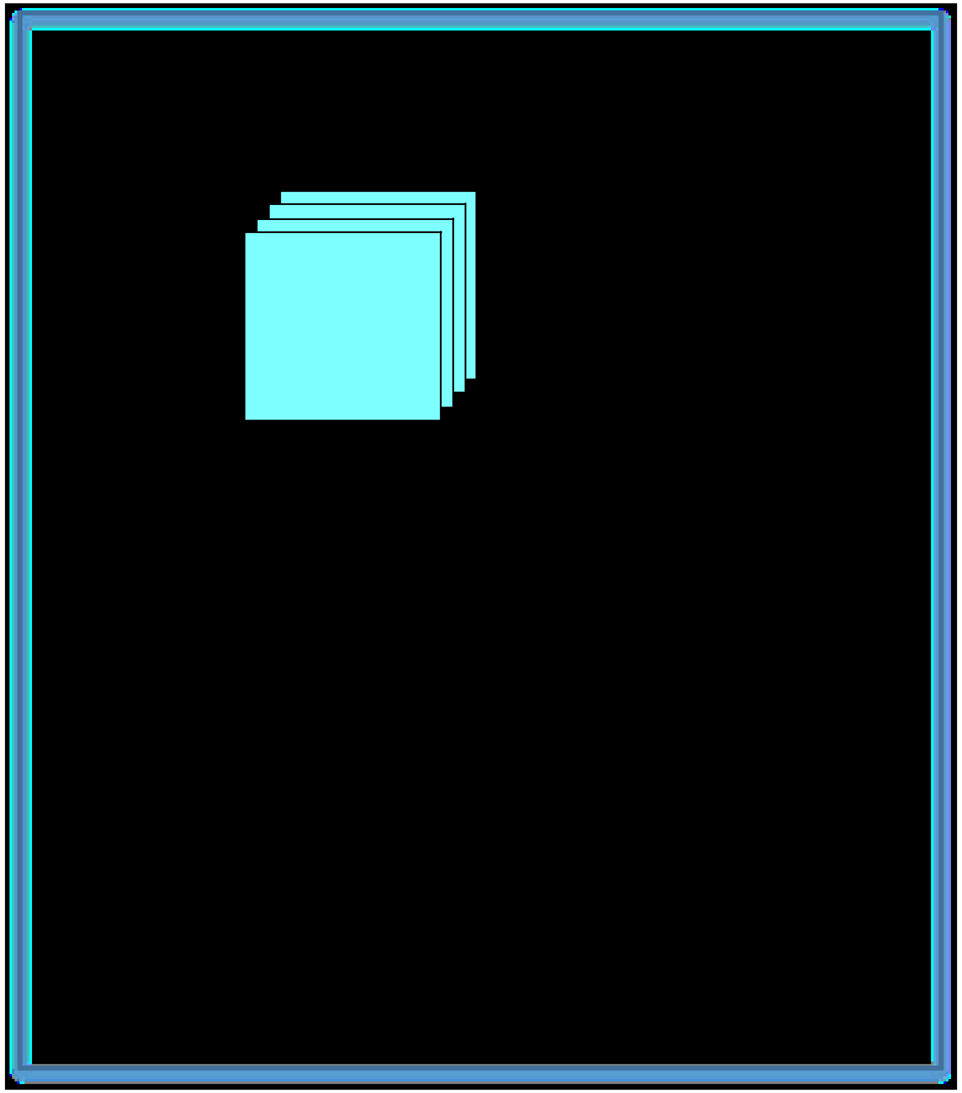
Challenge 3: learning similar parts

- Multimodal Exercise Representing Layer (MERL)
- Goal:
  - learn a unified semantic representation for each exercise by integrating its heterogeneous materials in a multimodal way



MANN Framework - Exercise Input

- Exercise Input: for each exercise
  - Text (*ET*):
    - Sequence words: $ET = (w_1, w_2, \ldots, w_N)$
    - Each word: d_0-dimensional word2vec
  - Images (*EI*):
    - A tensor : $EI = (p_1, p_2, \ldots, p_M) \in \mathbb{R}^{M \times 64 \times 64}$
    - Each image: a 64 x 64 matrix
  - Concepts (*EC*):
    - A matrix : $EC = (k_1, k_2, \ldots, k_L) \in \{0, 1\}^{L \times L_{all}}$
    - Each concept: one-hot vector



MANN Framework - Image CNN

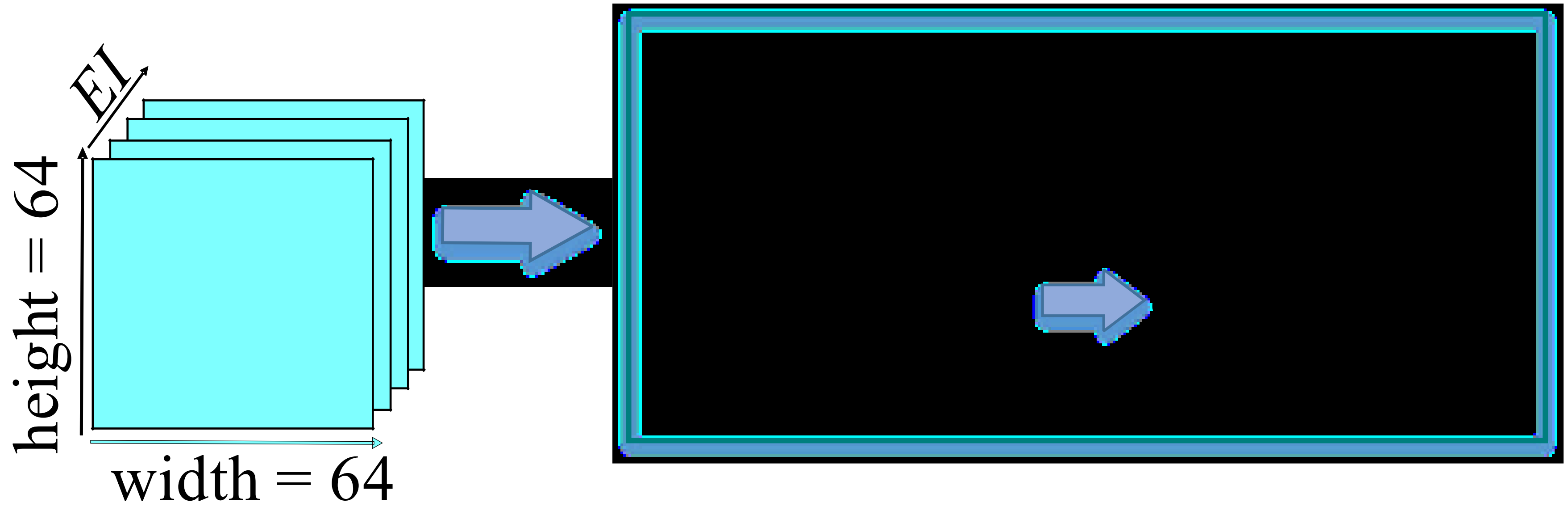
- Image CNN
- Goal:
  - gets the feature vector for each image.

$$v_i = \sigma(ImCNN(p_i)),$$

feature vector

Image CNN

A 64 x 64 Image



MANN Framework – Concept Embedding

- Concept Embedding
- Goal:
  - convert one-hot vectors of concepts into low-dimensional ones with dense values.

$$u_i = k_i \mathbf{W_u},$$

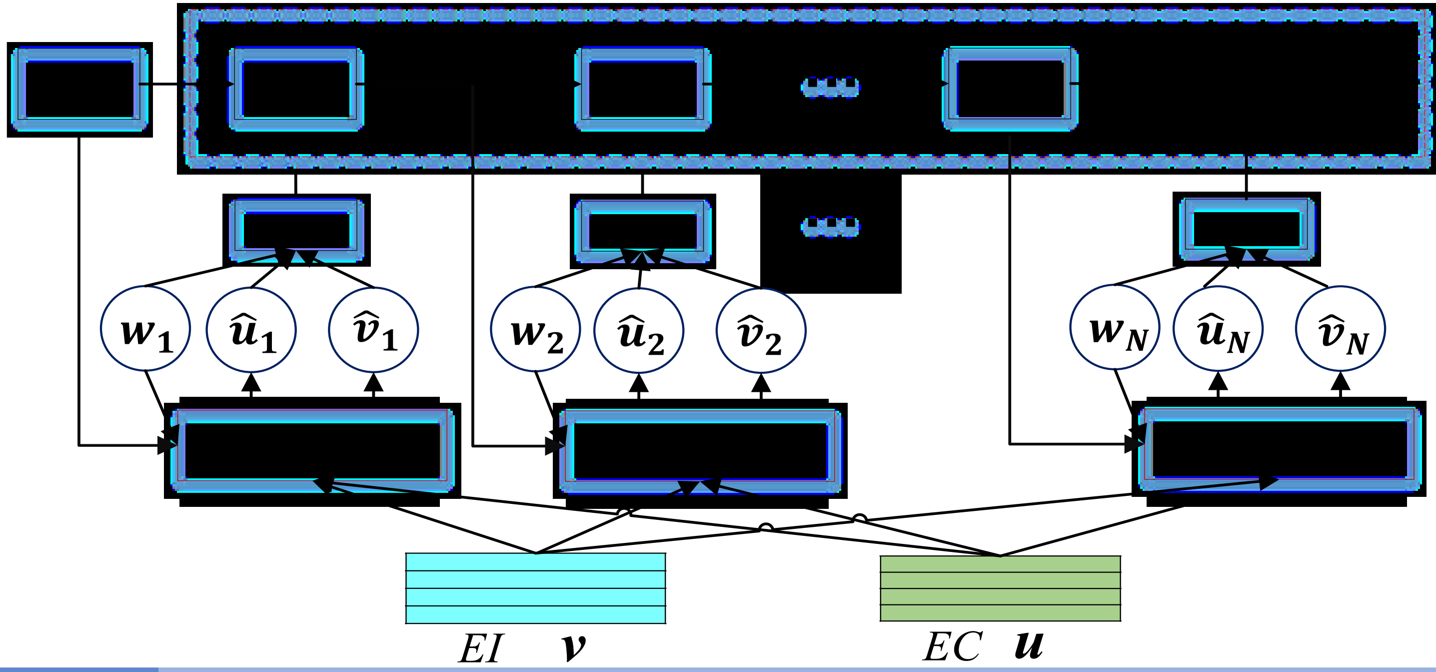
Dense vector of concept *i*

One-hot vector

Parameters of Concept Embedding

- Attention-based LSTM
- Goal:
  - learn a unified semantic representation for each exercise by integrating its all heterogeneous materials
  - capture text-concept and text-image associations with Text-Concept Attention (TCA) and Text-Image Attention (TIA), respectively.

semantic representations of different parts, $\boldsymbol{h}^{(E)}$



MANN Framework - Attention-based LSTM

- LSTM input: sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ combined with all materials of each exercise.

$$x_t = w_t \oplus \hat{u}_t \oplus \hat{v}_t,$$

t-th word representation in the text ET

Representation of the associated concepts learned by TCA

Representation of the associated images learned by TIA

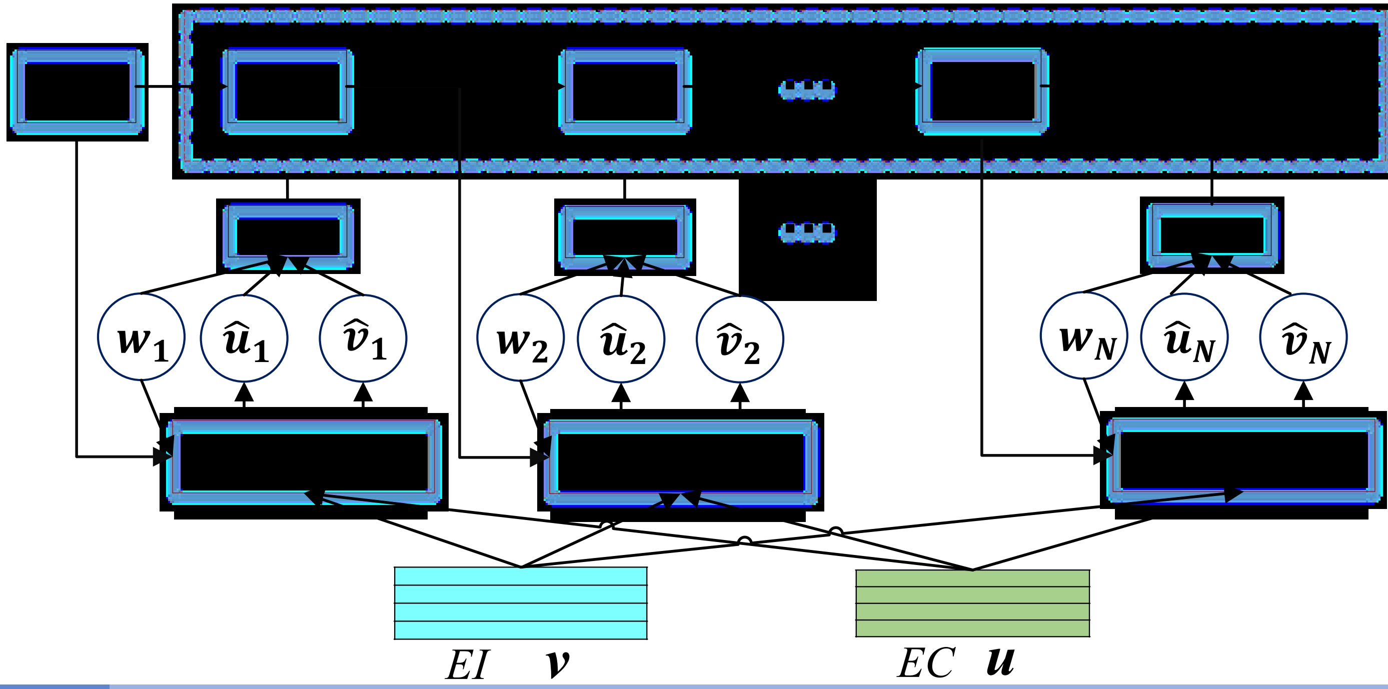
- Output: unified semantic representation for an input exercise *E*.

$$(r^{(E)}, \boldsymbol{h}^{(E)})$$

Whole semantic representation of *E*

Representations of different parts of *E*

semantic representations of different parts, $\boldsymbol{h}^{(E)}$



MANN Framework

- Text-Concept Attention (TCA): capture text-concept associations.

Representation of the associated concepts

Attention score after normalization

$$\hat{u}_t = \sum_{j=1}^{L} \alpha_j u_j, \quad \alpha_j = \frac{\varphi(u_j, w_t, h_{t-1})}{\sum_{i=1}^{L} \varphi(u_i, w_t, h_{t-1})},$$

$$\varphi(u_j, w_t, h_{t-1}) = \mathrm{V_{ac}} tanh(\mathrm{W_{ac}}[u_j \oplus w_t \oplus h_{t-1}]),$$

Association score between the j-th concept u_j and w_t in E

parameters of TCA

the $t-1^{th}$ hidden state

- Text-Image Attention (TCA): capture text-image associations.
  - modeled similarly as TCA.

MANN Framework - Similarity Attention

- Similarity Attention
- Goal:
  - measure similar parts between two exercises with their unified semantic representations, and learn attention representations for them.
- Attention Matrix **A**
  - measure similar parts between E_a and E_b

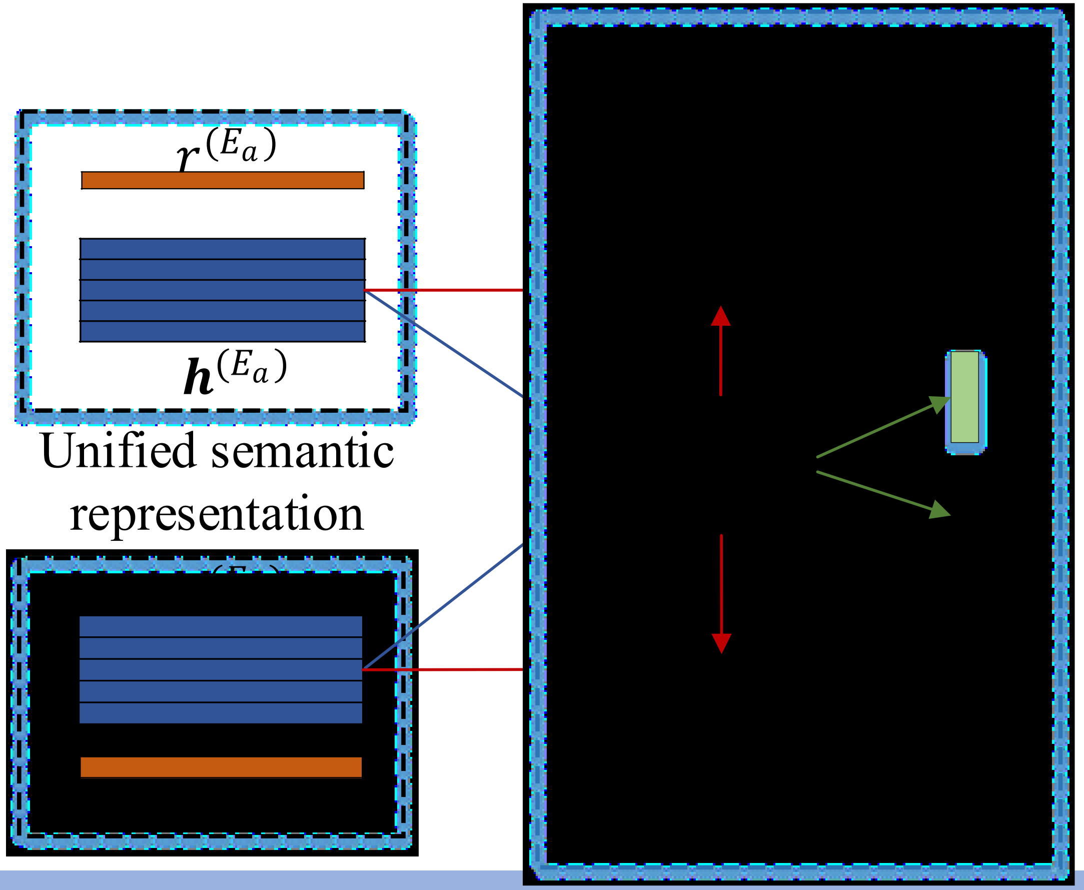
$$A_{i,j} = \cos(h_i^{(E_a)}, h_j^{(E_b)}) \text{ , } 1 \le i \le N_{E_a}, 1 \le j \le N_{E_b}$$

- Similarity attention representations $s^{(E_a)}$ and $s^{(E_b)}$

$$s_i^{(E_a)} = \sum_{k=1}^{N_{E_b}} A_{i,k} \qquad s_j^{(E_b)} = \sum_{k=1}^{N_{E_a}} A_{k,j}$$

- Semantic attention representations $h_{att}^{(E_a)}$ and $h_{att}^{(E_b)}$

$$h_{att}^{(E_a)} = \sum_{i=1}^{N_{E_a}} A_{i,N_{E_b}} h_i^{(E_a)}, \qquad h_{att}^{(E_b)} = \sum_{j=1}^{N_{E_b}} A_{N_{E_a},j} h_j^{(E_b)}.$$



MANN Framework - Similarity Score Layer

- Similarity Score Layer
- Goal:
  - calculating the similarity score of each exercise pair to rank candidate exercises to find similar ones for any exercise.

Whole semantic representation

similarity attention representation

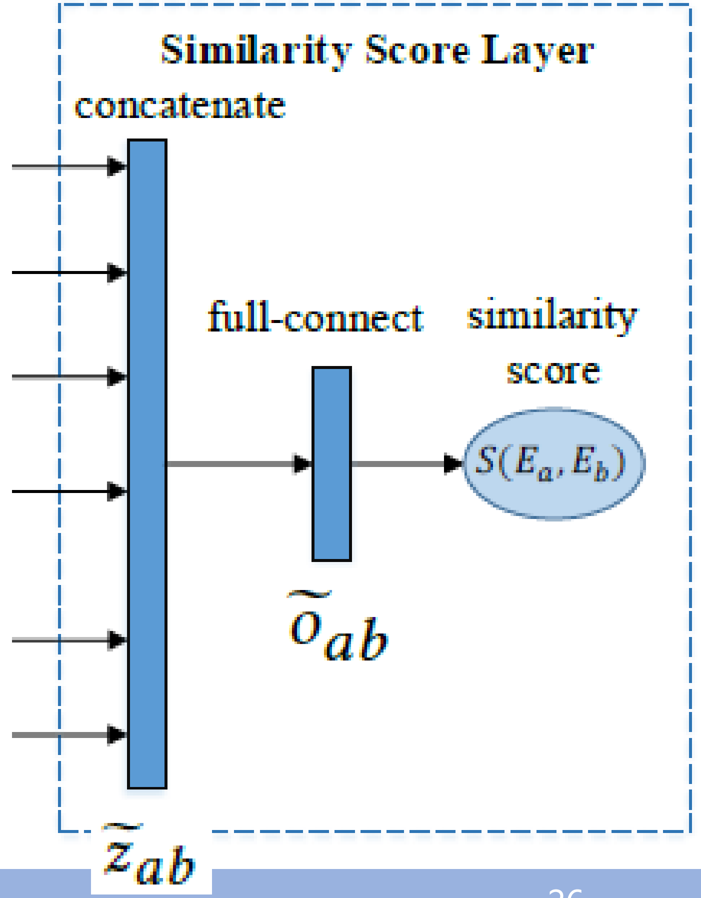
semantic attention representation

$$\widetilde{z}_{ab} = r^{(E_a)} \oplus r^{(E_b)} \oplus s^{(E_a)} \oplus s^{(E_b)} \oplus h_{att}^{(E_a)} \oplus h_{att}^{(E_b)}$$

$$\widetilde{o}_{ab} = ReLU(\mathbf{W}_1 \widetilde{z}_{ab} + \mathbf{b}_1)$$

$$S(E_a, E_b) = \sigma(\mathbf{W}_2 \widetilde{o}_{ab} + \mathbf{b}_2)$$

similarity score



MANN Learning

- Pairwise loss function

a margin

similarity score

regularization hyperparameter

$$\mathcal{L}(\Theta) = \sum_{E, E_s, E_{ds}} max(0, \mu - (S(E, E_s) - S(E, E_{ds}))) + \lambda_{\Theta} ||\Theta||^2$$

a labeled similar exercise of E

a dissimilar exercise of E

parameters of MANN

- Similar exercises (e.g. E_s) are labeled by education experts (e.g. teachers).
- Dissimilar exercises (e.g. E_{ds}) are sampled in the training process:
  - **Sampling Randomly (Random):** At each iteration, we randomly select a number of dissimilar exercises from all the dissimilar ones of *E*.
  - **Sampling by Concepts (Concept):** At each iteration, we randomly select a number of dissimilar exercises from those having at least one common concept with *E*.

Outline

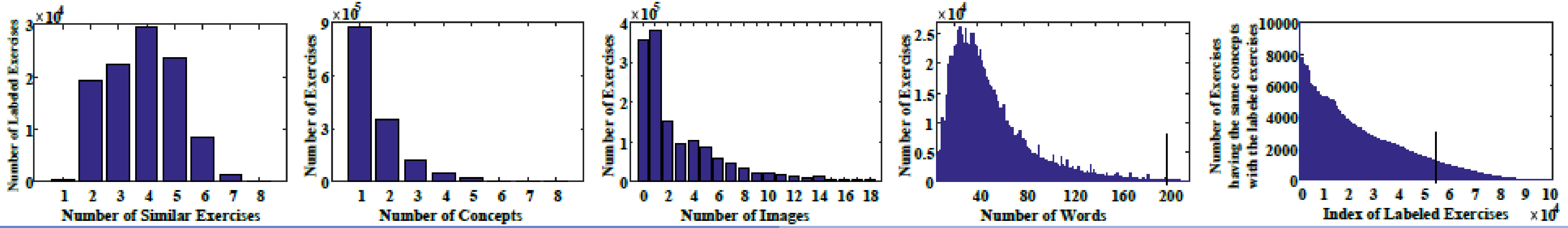
Experiments

- Experiments dataset
  - supplied by iFLYTEK, collected from Zhixue.
  - contains 1,420,727 math exercises.

Table 2: The statistics of the dataset.

Statistics	Values
number of exercises	1,420,727
number of exercises having images	1,064,964
number of labeled exercises	104,515
number of similar pairs	401,476
number of similar pairs having the same concepts	174,672
Average similar pairs per labeled exercise	3.84
Average concepts per exercise	1.61
Average images per exercise	3.04

- Observations in dataset
  - On average 3.84 similar exercises are labeled for the given one.
  - Each exercise consists of about 1.61 concepts and 3.04 images.
  - About 75% exercises have at least one image.
  - 99% exercises contain less than 200 words in the text.
  - More than 55% labeled exercises have the same concepts with at least 1,000 exercises.

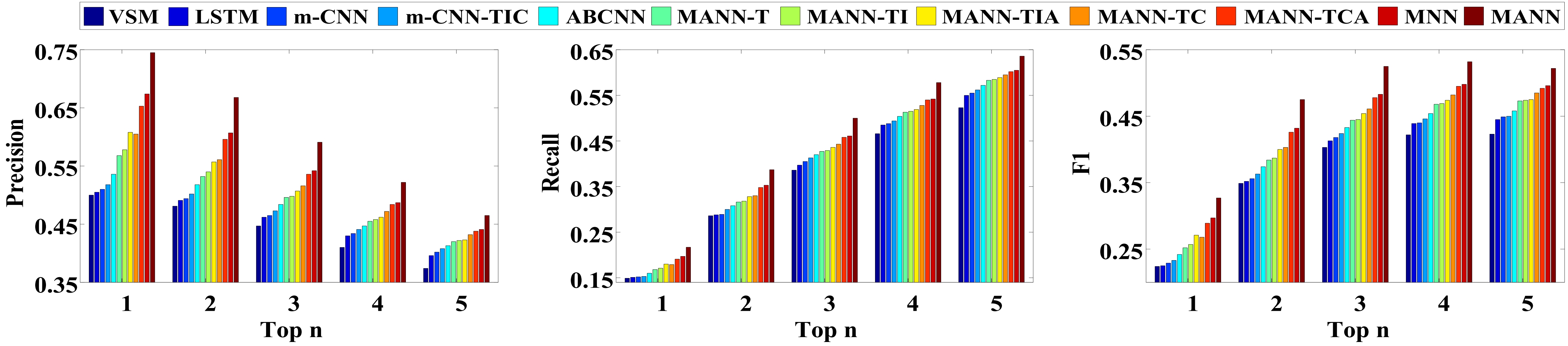


Experiments

- Baseline Approaches
  - Variants of MANN: **MANN-T** (only Text), **MANN-TI** (Text and Images), **MANN-TIA** (with TIA), **MANN-TC** (Text and Concepts), **MANN-TCA** (with TCA), **MNN** (Using Text, Images and Concepts, but without TIA and TCA).
  - **VSM**: Vector space model (VSM) is applied for the FSE task based on texts of exercises.
  - **LSTM**: learn the semantic similarity between sentences based on the texts.
  - **ABCNN**: a network architecture based on texts for modeling sentence pairs.
  - **m-CNN**: integrating texts and images into a vectorial representation.
  - **m-CNN-TIC**: a variant of m-CNN integrating texts, images and concepts.
- Evaluation Metrics
  - **Precision**, **Recall**, and **F1** at top n = 1, 2, 3, 4, 5.
  - As on average 3.84 similar exercises are labeled for the given one.

Experiments

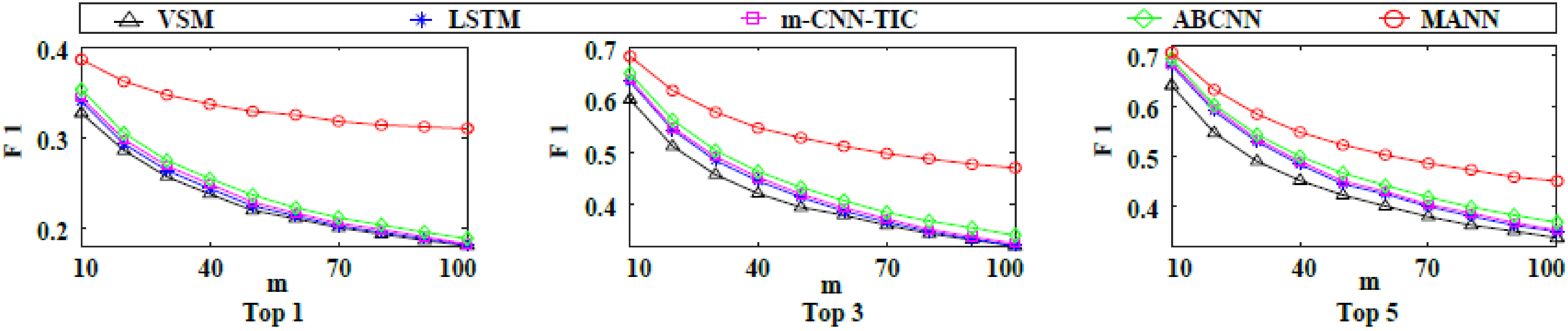
Performance Comparison



- MANN achieves the best performance, and the variants of MANN also perform better than other baselines.
- MANN-T performs better than ABCNN, indicating the effectiveness of Similarity Attention to measure similar parts of an exercise pair.
- MANN-TIA beats MANN-TI ,and MANN-TCA performs better than MANN-TC, demonstrating the effectiveness of TIA and TCA.
- MANN performs best and MNN ranks the second, suggesting that it is more effective for the FSE task by integrating texts, images and concepts, and further demonstrating the effectiveness of TIA and TCA.

Experiments

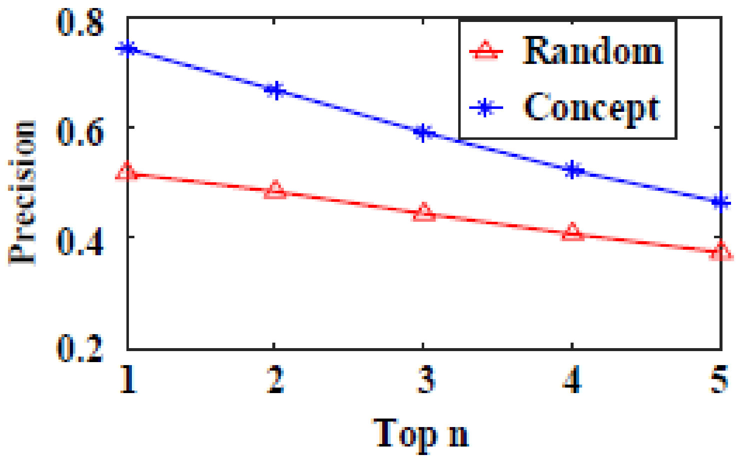
➤ Performance with Different Number of Sampled Dissimilar Exercises (m)

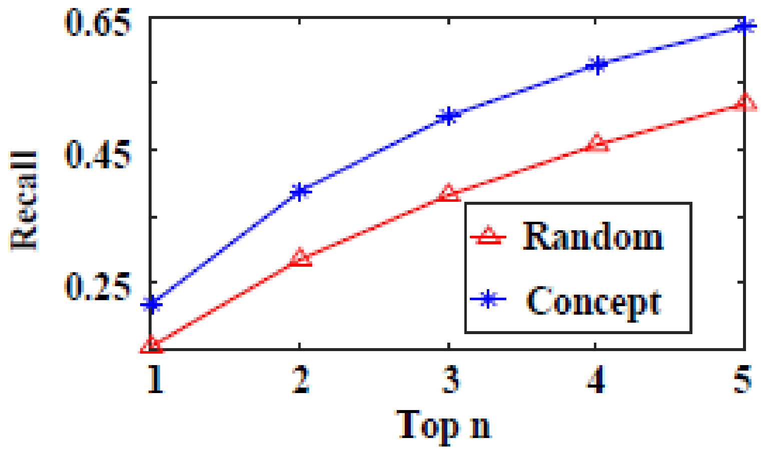


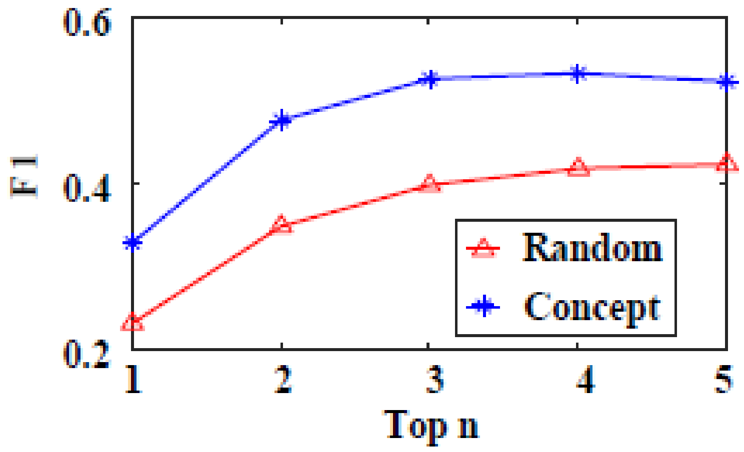
- MANN still outperforms baselines with different m.
- The F1 value of MANN degrades the most slowly while m increases.
- The more unlabeled exercises (i.e. negative samples) in the testing set, the more improvement of MANN compared with the baselines could be observed.

Experiments

➢ Influence of Sampling Ways







➢MANN trained in the sampling way of **Concept** performs much better than that in **Random.**

➢MANN can focus on the subtle differences between its similar pairs and dissimilar ones in **Concept**, because for each given exercise, its similar exercises are close to the dissimilar ones in **Concept**, while they are very different from most sampled dissimilar ones in **Random**

Experiments

- Case Study
  - MANN explanatory power

- The parts in the green box (or blue, red box) in E_a and E_b are the similar parts that express the same meaning.
- This implies that MANN provides a good way to capture the similarity information between exercises by Similarity Attention.

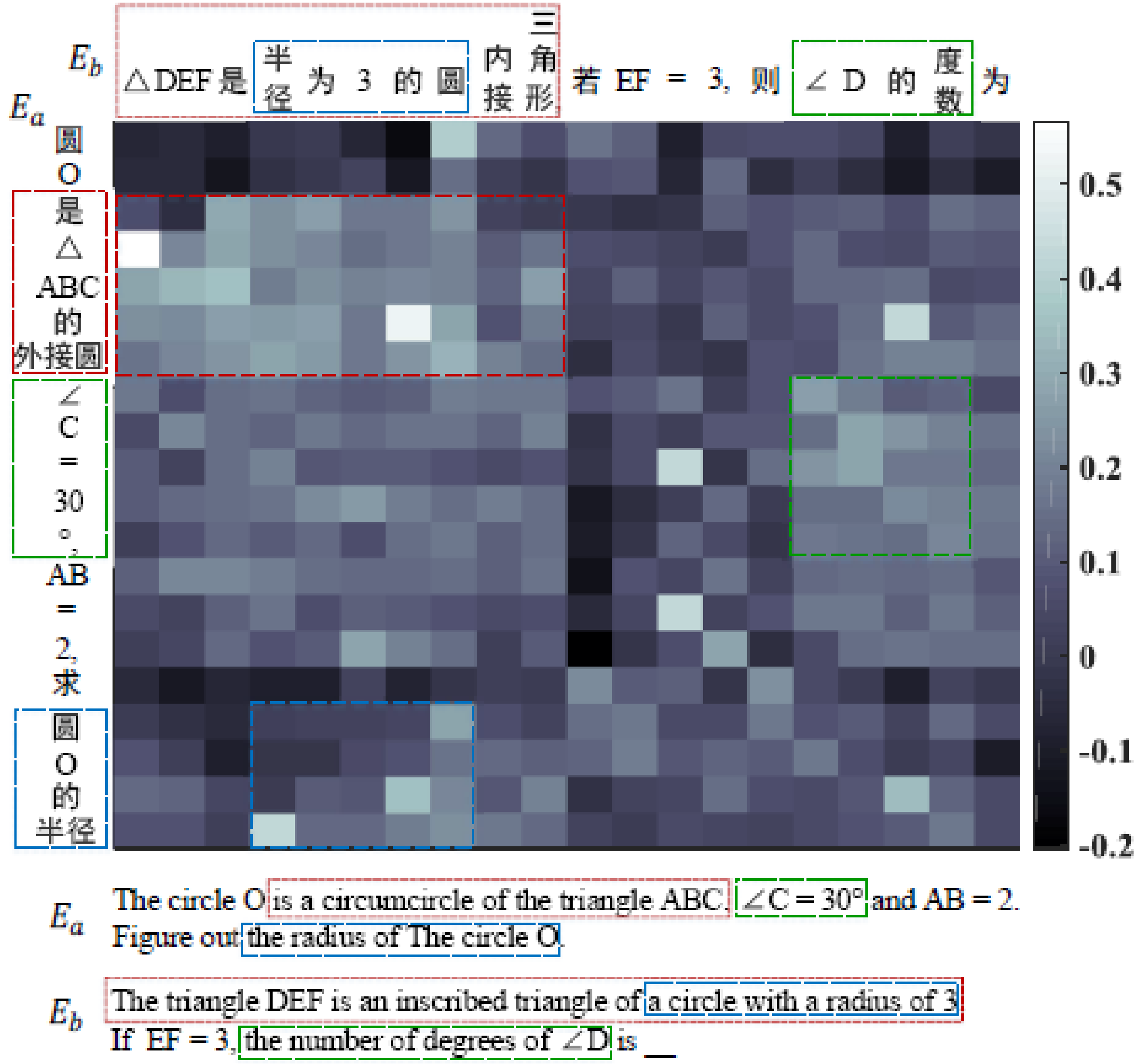


Figure 10: Visualization of the similar parts between two example exercises E_a and E_b.

Outline

Conclusion

- Provided a focused study on finding similar exercises (FSE) in online education systems.
- Proposed a novel Multimodal Attention-based Neural Network (MANN) framework for the FSE task by modeling the heterogeneous materials of exercises semantically.
- Designed an Attention-based LSTM network to learn a unified semantic representation of each exercise, where two attention strategies were proposed to capture text-image and text-concept associations.
- Designed a Similarity Attention to measure similar parts in exercise pairs.
- Experiments on a large-scale real-world dataset clearly demonstrated both the effectiveness and explanatory power of MANN.

Future Work

- We would like to measure the relation of exercises in more aspects, e.g. by considering the difficulty of exercises.

- We will also try to develop the semi-supervised or unsupervised learning methods for the FSE task.

- As our MANN is a general framework, we will test its performance on other disciplines (e.g. Physics), and meanwhile, on the similar applications in other domains, such as the measurement of product similarities in e-commerce.

Q & A

Thanks!