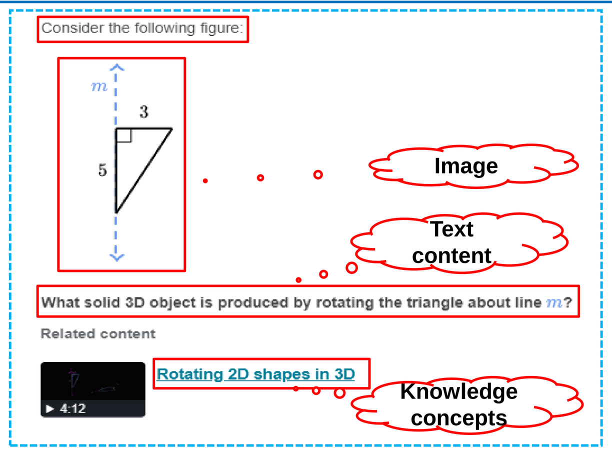
Finding Similar Exercises in Online Education Systems

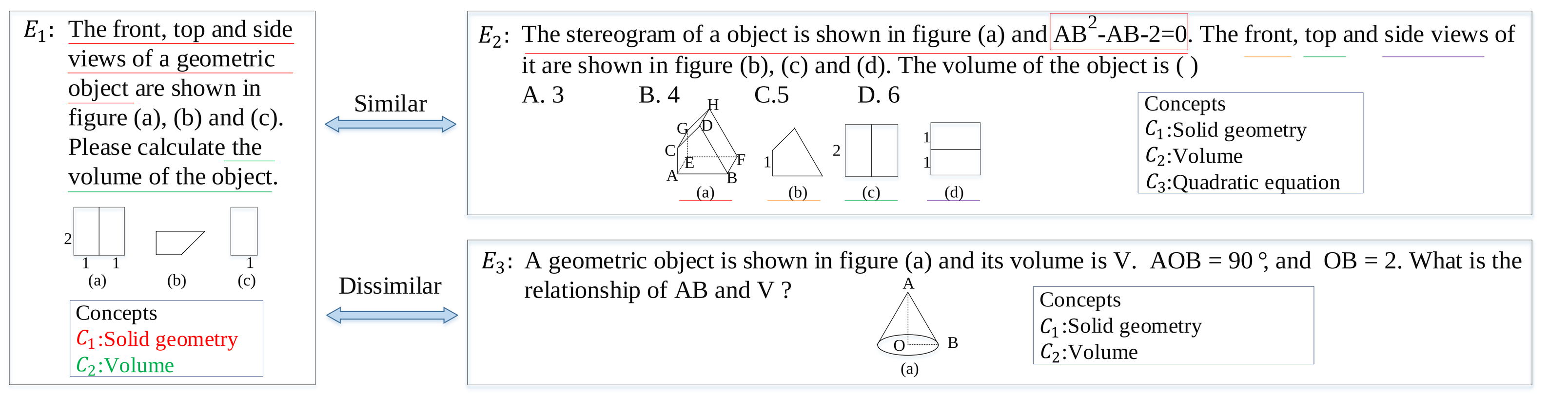
Qi Liu[1] Zai Huang[1] Zhenya Huang[1] Chuanren Liu[2] Enhong Chen[1*] Yu Su[3] Guoping Hu[4]

[1]Anhui Province Key Lab. 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

1. Introduction



- **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) for each exercise.
- **Similar Exercises:** having the same purpose embedded in exercise contents.
- **Challenges for FSE**
  - Exercises contain multiple heterogenous data, e.g., texts, images and concepts. It requires integrating multiple heterogeneous data to understand and represent exercise semantics and purposes in a multimodal way.
  - For understanding each exercise, it is necessary to capture text-concept and text-image associations in the exercise.
  - It is critical to to measure the similar parts in each exercise.



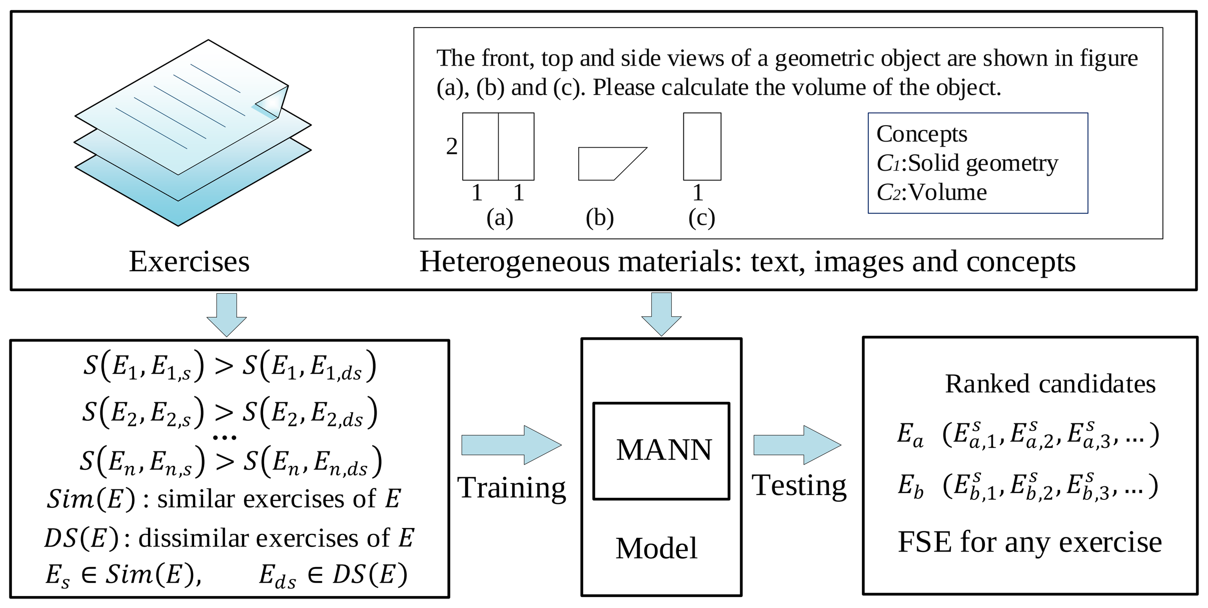
2. Problem Definition

Definition 1. *Given a set of exercises with corresponding heterogeneous materials including texts (ET), images (EI) and concepts (EC), our goal is to integrate these heterogeneous materials to learn a model $\mathcal{F}$, which can be used 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, i.e.*

$$\mathcal{F}(E, \mathcal{R}, \Theta) \to \mathcal{R}^s, \quad (1)$$

where Θ is the parameters of $\mathcal{F}$, $\mathcal{R} = (E_1, E_2, E_3, \dots)$ are the candidate exercises for E and $\mathcal{R}^s = (E_1^s, E_2^s, E_3^s, \dots)$ are the candidates ranked in descending order with their similarity scores $(S(E, E_1^s), S(E, E_2^s), S(E, E_3^s), \dots)$. The similar exercises for E are those candidates having the largest similarity score.

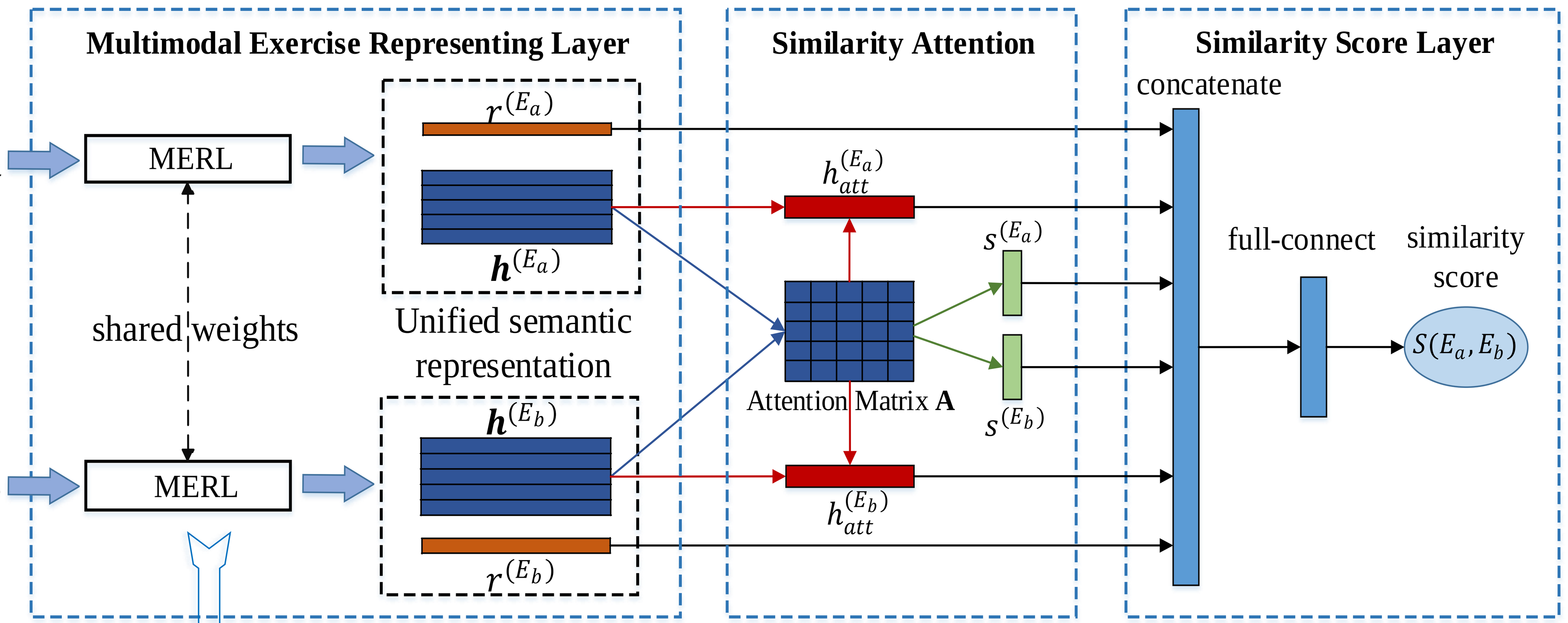
3. Study Overview



4. MANN Framework

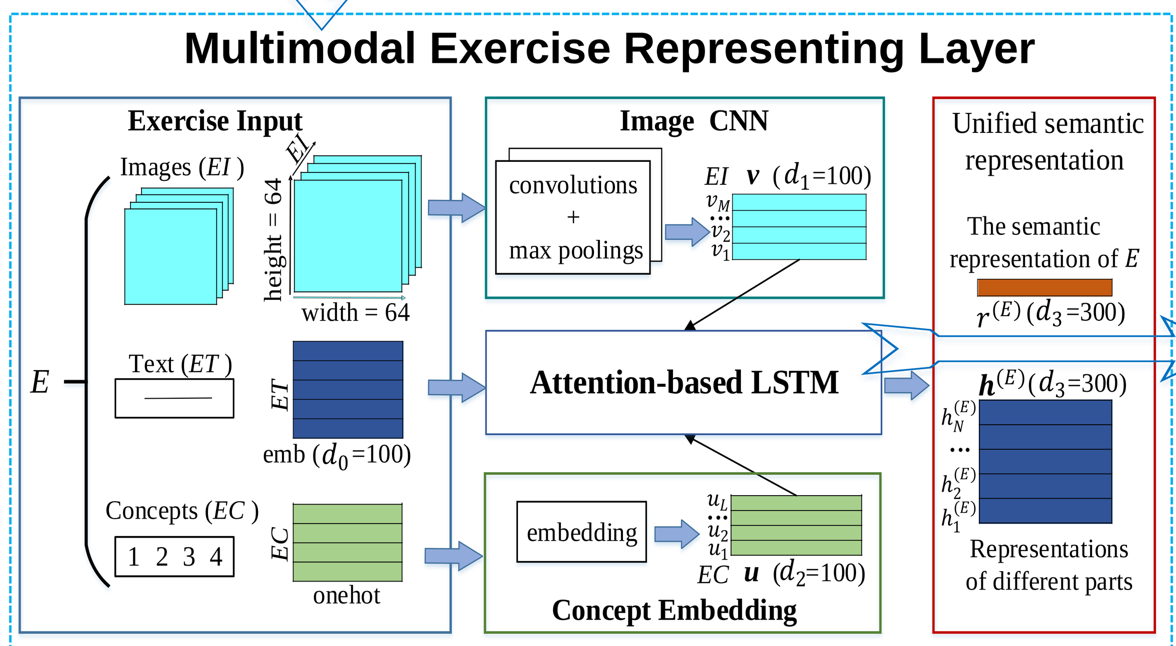
4.1 Multimodal Attention-based Neural Network (MANN)

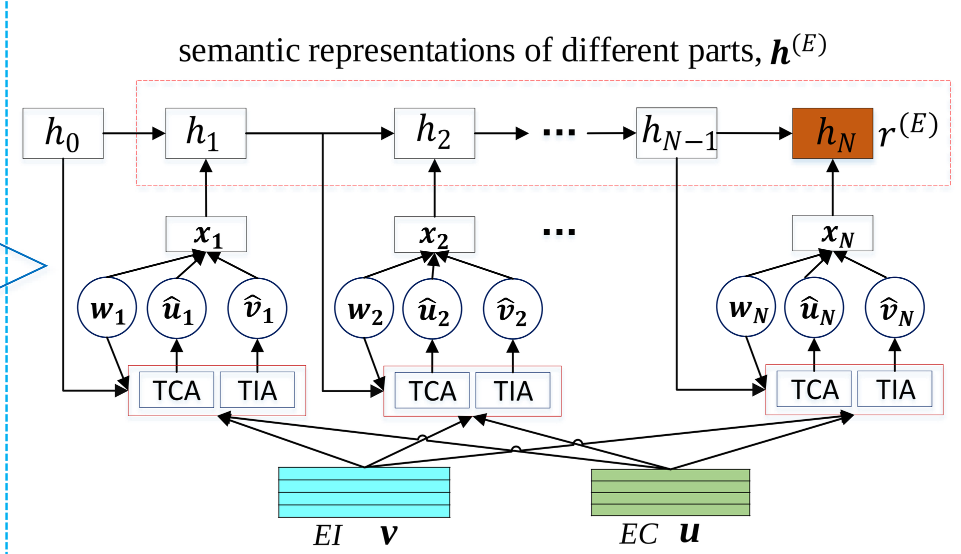
- **Multimodal Exercise Representing Layer (MERL):** outputs a unified semantic representation of each exercise in a multimodal way.
- **Similarity Attention:** measures similar parts between two exercises.
- **Similarity Score Layer:** calculates the similarity scores of exercise pairs.



4.2 Multimodal Exercise Representing Layer (MERL)

- **Exercise Input:** materials of each exercise E, Text (ET), Images (EI), Concepts(EC).
- **Image CNN:** gets the feature vector for each image.
- **Concept Embedding:** converts one-hot vectors of concepts into low-dimensional ones with dense values.
- **Attention-based LSTM:** learns a unified semantic representation $(r^{(E)}, \boldsymbol{h}^{(E)})$ for an input exercise E, and capture the text-concept and text-image associations with Text-Concept Attention (TCA) and Text-Image Attention (TIA), respectively.





- **LSTM Input:** sequence $x = (x_1, x_2, \dots, x_N)$, $x_t = w_t \oplus \hat{u}_t \oplus \hat{v}_t$, w_t is the t-th word representation in the text ET, $\hat{u}_t$ and $\hat{v}_t$ are the representations of the associated concepts and images of this word, which are learned by TCA and TIA respectively.
- **Text-Concept Attention (TCA):**

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

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

- **Text-Image Attention (TIA):** TIA is modeled similarly as TCA.

4.3 Similarity Attention

- **Similarity attention matrix *A*:** $A_{i,j} = cos(h_i^{(E_a)}, h_j^{(E_b)})$ measuring similar parts of the input pair (E_a, 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 $h_{att}^{(E_b)}$:

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

4.4 Similarity Score Layer

$$\tilde{z}_{ab} = r^{(E_a)} \oplus r^{(E_b)} \oplus s^{(E_a)} \oplus s^{(E_b)} \oplus h_{att}^{(E_a)} \oplus h_{att}^{(E_b)} \quad \tilde{o}_{ab} = ReLU(\mathrm{W}_1 \tilde{z}_{ab} + \mathrm{b}_1) \quad S(E_a, E_b) = \sigma(\mathrm{W}_2 \tilde{o}_{ab} + \mathrm{b}_2)$$

4.5 MANN Learning

- **Pairwise loss function**

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

μ is a margin, E_s is a labeled similar exercise of E, E_{ds} is a dissimilar exercise of E.

- **Sampling dissimilar exercises**
  - **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.

5. Experiments

5.1 Dataset

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

Table 2: The statistics of the dataset.

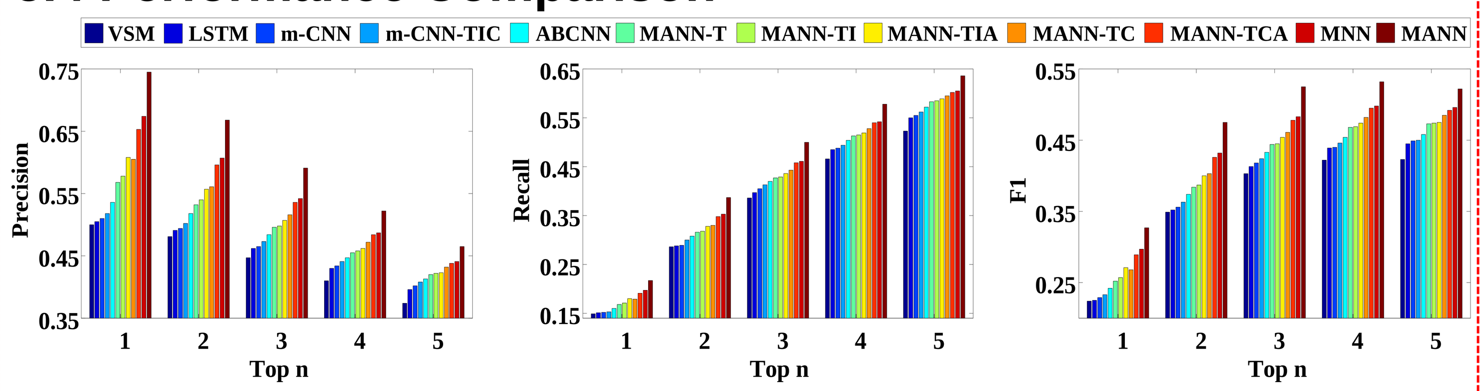
Statistics	Values
number of exercises	1,420,727
number of exercises having images	1,064,964
number of labeled exercises	104,515
number of similar pairs	401,476
number of similar pairs having the same concepts	174,672
Average similar pairs per labeled exercise	3.84
Average concepts per exercise	1.61
Average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

5.2 Baselines

- **VSM**: Vector space model (VSM) is widely applied for the FSE task based on the 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.
- **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).

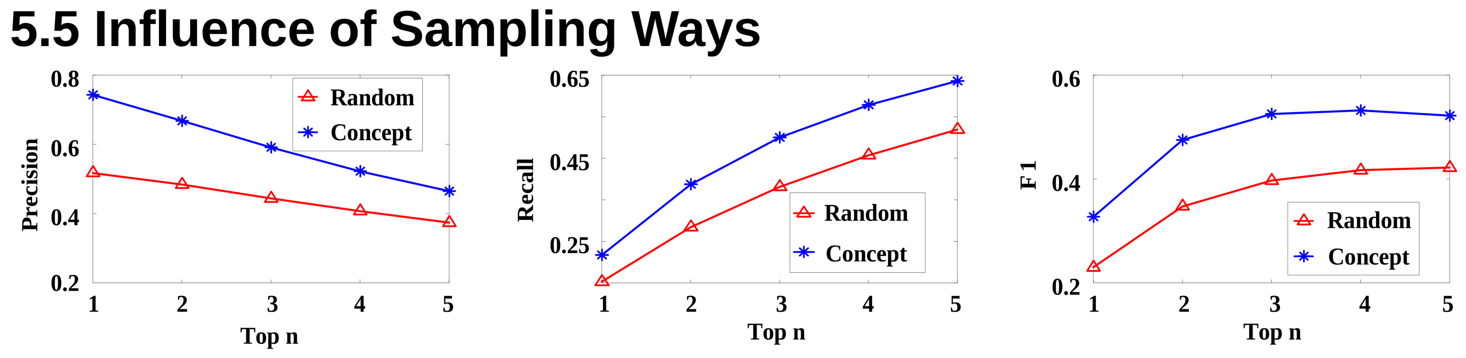
5.3 Evaluation Metrics: Precision, Recall, and F1 at top n = 1, 2, 3, 4, 5.

5.4 Performance Comparison



- MANN achieves the best performance, and the variants of MANN also have better performance 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.

5.5 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**.

5.6 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.