Finding Similar Exercises in Online Education Systems

Qi Liu¹, Zai Huang¹, Zhenya Huang¹, Chuanren Liu², Enhong Chen¹*, Yu Su³, Guoping Hu⁴

¹Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China
{qiliuq1, cheneh}@ustc.edu.cn, {huangzai, huangzhy}@mail.ustc.edu.cn
²Decision Sciences & MIS Department, Drexel University, chuanren.liu@drexel.edu
³School of Computer Science and Technology, Anhui University, yusu@iflytek.com
⁴iFLYTEK Research, gphu@iflytek.com

Reporter: Zai Huang
Date: 2018.07.22
## Outline

<table>
<thead>
<tr>
<th></th>
<th>Background and Related Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Problem Definition</td>
</tr>
<tr>
<td>3</td>
<td>Study Overview</td>
</tr>
<tr>
<td>4</td>
<td>MANN Framework</td>
</tr>
<tr>
<td>5</td>
<td>Experiments</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion and Future Work</td>
</tr>
</tbody>
</table>
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 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.

Concepts:
- $C_1$: Solid geometry
- $C_2$: Volume

$E_1$: 

$E_2$: 

$E_3$: 

Share the same purpose

Similar

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

The stereogram of an object is shown in figure (a) and $AB^2 - AB - 2 = 0$. The front, top and side views of it are shown in figure (b), (c) and (d). The volume of the object is $\frac{1}{3}$. 

Concepts:
- Solid geometry
- Volume
- Quadratic equation

$E_2$: 
$C_1$ 
$C_2$ 
$C_3$
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.

The volume of the object is $A$.

Concepts:
- Solid geometry
- Volume

$E_2$:

$E_3$: Similar

$E_3$: Dissimilar A

B

O
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.

The front, top and side views of a geometric object are shown in figure (a), (b) and (c). Please calculate the volume of the object.

Concepts:
Solid geometry: Volume

The stereogram of a object is shown in figure (a) and AB - AB - AB = 0. The front, top and side views of it are shown in figure (b), (c) and (d). The volume of the object is A, B, C or D.

Concepts:
Solid geometry: Volume: Quadratic equation

A geometric object is shown in figure (a) and its volume is $V$. $\angle AOB = 90^\circ$, and $OB = 2$. What is the relationship of AB and $V$ ?

Concepts:
Solid geometry: Volume
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

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

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

Neglect semantics in heterogeneous materials of exercises.

Cannot understand exercise purposes or measure similar parts between two exercises.

Cannot handle instances having multiple heterogeneous data.
Outline

1 Background and Related Work
2 Problem Definition
3 Study Overview
4 MANN Framework
5 Experiments
6 Conclusion and Future Work
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 $E$
Study Overview

- Two-stage solution

  - Training stage
    - MANN
    - Pairwise training

  - Testing stage
    - FSE for any exercise

---

The front, top and side views of a geometric object are shown in figure (a), (b) and (c). Please calculate the volume of the object.

Concepts

<table>
<thead>
<tr>
<th>C1: Solid geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2: Volume</td>
</tr>
</tbody>
</table>

Exercises

Heterogeneous materials: text, images and concepts

\[
S(E_1, E_{1,s}) > S(E_1, E_{1,ds}) \\
S(E_2, E_{2,s}) > S(E_2, E_{2,ds} \\
\quad \ldots \\
S(E_n, E_{n,s}) > S(E_n, E_{n,ds})
\]

Sim(E) : similar exercises of E
DS(E) : dissimilar exercises of E
E_s \in \text{Sim}(E), \quad E_{ds} \in \text{DS}(E)

\[
E_a (E_{a,1}^s, E_{a,2}^s, E_{a,3}^s, \ldots) \\
E_b (E_{b,1}^s, E_{b,2}^s, E_{b,3}^s, \ldots)
\]

FSE for any exercise
Outline

1. Background and Related Work
2. Problem Definition
3. Study Overview
4. MANN Framework
5. Experiments
6. Conclusion and Future Work
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
  - Propose **two attention strategies** to capture the text-image and text-concept associations in each single exercise
  - Design a **Similarity Attention** to measure the similar parts in each exercise pair with their semantic representations

**Challenge 1:** multimodal exercises understanding and representation

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

**Challenge 3:** learning similar parts
MANN Framework

Input: a pair of exercise \((E_a, E_b)\)

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
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: \( EC = (k_1, k_2, \ldots, k_L) \in \{0, 1\}^{L \times L_{alt}} \)
  - Each concept: one-hot vector
MANN Framework - Image CNN

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

\[ \nu_i = \sigma(ImCNN(p_i)), \]

feature vector

Image CNN

A 64 x 64 Image

A 64 x 64 Image

height = 64

width = 64
MANN Framework – Concept Embedding

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

\[ u_i = k_i W u_i, \]

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, $h^{(E)}$
MANN Framework - Attention-based LSTM

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

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

\[
x_t = w_t \oplus \hat{u}_t \oplus \hat{v}_t
\]

- **t-th word representation in the text** \( ET \)
- **Representation of the associated concepts learned by TCA**
- **Representation of the associated images learned by TIA**

Whole semantic representation of \( E \)  
Representations of different parts of \( E \)

Semantic representations of different parts, \( h^{(E)} \)
MANN Framework

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

Representation of the associated concepts

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

Attention score after normalization

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

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

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 $E_b$
  
  \[ A_{i,j} = \cos(h_i^{(E_a)}, h_j^{(E_b)}) , \quad 1 \leq i \leq N_{E_a}, 1 \leq j \leq N_{E_b} \]

- **Similarity attention representations $s_i^{(E_a)}$ and $s_j^{(E_b)}$**
  
  \[ s_i^{(E_a)} = \sum_{k=1}^{N_{E_b}} A_{i,k} \quad s_j^{(E_b)} = \sum_{k=1}^{N_{E_a}} A_{k,j} \]

- **Semantic attention representations $h_{att}^{(E_a)}$ and $h_{att}^{(E_b)}$**
  
  \[ h_{att}^{(E_a)} = \sum_{i=1}^{N_{E_a}} A_{i,N_{E_b}} h_i^{(E_a)} , \quad h_{att}^{(E_b)} = \sum_{j=1}^{N_{E_b}} A_{N_{E_a},j} h_j^{(E_b)} . \]
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.

\[ \tilde{z}_{ab} = r(E_a) \oplus r(E_b) \oplus s(E_a) \oplus s(E_b) \oplus h_{att}(E_a) \oplus h_{att}(E_b) \]

\[ \tilde{o}_{ab} = \text{ReLU}(W_1\tilde{z}_{ab} + b_1) \]

\[ S(E_a, E_b) = \sigma(W_2\tilde{o}_{ab} + b_2) \]
**MANN Learning**

- **Pairwise loss function**

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

  - a margin
  - similarity score
  - regularization hyperparameter
  - 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 \(E\).
Experiments

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

- 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.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of exercises</td>
<td>1,420,727</td>
</tr>
<tr>
<td>number of exercises having images</td>
<td>1,064,964</td>
</tr>
<tr>
<td>number of labeled exercises</td>
<td>104,515</td>
</tr>
<tr>
<td>number of similar pairs</td>
<td>401,476</td>
</tr>
<tr>
<td>number of similar pairs having the same concepts</td>
<td>174,672</td>
</tr>
<tr>
<td>Average similar pairs per labeled exercise</td>
<td>3.84</td>
</tr>
<tr>
<td>Average concepts per exercise</td>
<td>1.61</td>
</tr>
<tr>
<td>Average images per exercise</td>
<td>3.04</td>
</tr>
</tbody>
</table>
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.
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.
Thanks!