Concept Identification Visualizer (CIV) using Knowledge Tracing

Hui-Jun Huang, *I-Wei Lai
Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan
*iweilai@ntnu.edu.tw

***Abstract*—The COVID-19 pandemic has resulted in an escalation in the demand for online learning, leading to the need for experts, such as experienced math teachers, to classify exercises. However, achieving accurate and effective classification may present challenges due to differing expert opinions and complex exercise concepts. To address this challenge, we propose the use of the Concept Identification Visualizer (CIV). The CIV tool assists experts who lack *engineering programming* knowledge in comprehending the exercises and evaluating student responses. The tool leverages Knowledge Tracing to extract relevant information from students' answers and presents this data in a visual format. By providing a more comprehensive understanding of the exercises, the experts enable more informed exercise classification based on student feedback and improve the overall effectiveness of online learning.**

***Index Terms*—Knowledge Tracing, Embedding Visualization, Data Analysis**

I. Introduction

As we have experienced, in certain educational fields like math, the teachers generally split the information into various classifications termed as *concept*. Then, students receive instruction and assessment based on a concept basis. However, such concept classifications are usually manually defined by experienced experts and may contain flaws. For instance, a math question that actually belongs to the category of equations may be misclassified as a category of function.

To overcome the issue of misclassification, we proposed the Concept Identification Visualizer (CIV) that utilizes the technology of Knowledge Tracing (KT) [1]. Specifically, KT can automatically predict the students' mastery of a certain subject by analyzing the responses to exercises and assessments.

We discover that concept classification can leverage the information generated by KT. In particular, extracting the features of various concepts from embedding layers and sorting out the labels of concepts from the preprocessing layers respectively can utilize in subsequent visualization.

Since the extracted features of concepts have high dimensionality, the conventional visualization methods can not effectively process the features. To address this issue, we have employed T-SNE [2], a nonlinear dimensionality reduction method. T-SNE maintains local structure during dimensionality reduction and preserves more structural information than linear dimensionality reduction methods. The capability makes T-SNE well suited for visualizing the features of exercises in KT models.

CIV exhibits high flexibility and finds application in various KT models. Furthermore, extending the method to cover different dimensions of features offers scalability for meeting various needs in KT. The advantages make CIV a versatile solution that can be adapted to a range of different KT models and use cases.

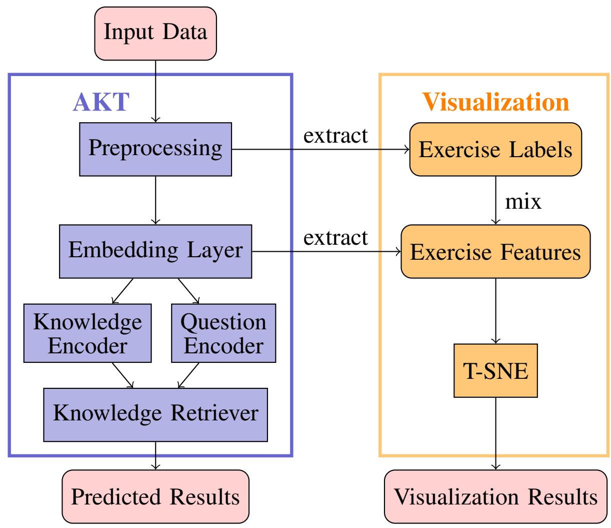

Fig. 1: The CIV framework is constituted by AKT and Visualization. In this figure, the boxes with pink, purple, and yellow colors represent data, AKT modules, and Visualization modules, respectively.

The rest of this paper is organized as follows. Section II elaborates on the design of the system framework and the dimensionality reduction method we used. Section III displays the results of the visualization graphs. The conclusions and future works are drawn in Section IV.

II. Concept Identification Visualizer (CIV)

In this section, we first introduce the system framework of CIV, followed by a description of the technologies, the KT model AKT [3], and the dimensionality reduction method T-SNE, respectively.

A. Overall Architecture

Initially, we input the data that we want to analyze into the AKT model for training. Subsequently, the preprocessing layer and the embedding layer extract exercise-related information from the trained data. The resulting features are then conveyed to the T-SNE module to facilitate data dimensionality reduction and computation of results.

B. AKT

The AKT model consists of five components, but we primarily focus on the preprocessing layer and the embedding layer. The preprocessing layer employs the encoding function factorize to convert exercise title names into numerical representations, commonly known as exercise labels. Subsequently, sequential training of the embedding layer enables the projection of these labels onto multi-dimensional features. By incorporating both the exercise labels and embedding layer parameters, T-SNE can calculate visualizations.

C. T-SNE

T-SNE utilizes conditional probability to construct the Gaussian distribution of high-dimensional data, which represents the similarity between sample points. However, for low-dimensional data, T-SNE utilizes t-distribution instead of Gaussian distribution due to the congestion problem caused by the curse of dimensionality.

The CIV employs the KL distance as the cost function to approximate the similarity between high-dimensional data and low-dimensional data. By minimizing the cost function, we can obtain the desired low-dimensional data distribution, which represents the outcome of CIV.

III. Experimental Results

Here we present the configuration settings utilized for the CIV and demonstrate the visualization results of the extracted features.

A. Settings

The settings of our experiments are as follows:

a) Dataset: In our experiment, we employed the Junyi Academy Math Practicing Log as the dataset. This dataset comprises the practicing logs of 4851 students from Oct. 2012 to Jan. 2015, containing a total of 740110 answer records. The comprehensive exercise records including exercise names, problem concepts, and student feedback facilitated the training of the AKT model.

b) Feature Dimension: The extracted features correspond to the hidden information of the problems. In common sense, we generally think that problems with the same concept have the same hidden information. Therefore, we can infer that questions clustered together in visualization results belong to the same concept, which assists experts in identifying misclassified problems.

To achieve the goal mentioned above, we should select an appropriate feature dimension for the embedding layer. Using a sufficient number of feature dimensions to represent the high-dimensional features of the exercises in order to preserve the structure of the data. In our experiment, we choose the baseline model with the feature dimensions of 256, which yielded the highest accuracy during the training of the AKT model.

B. Visualization Results

When adopting T-SNE to reduce dimension, the visualization results reflect the concept differences. For instance, in Fig. 2(a), dots with blue and pink respectively represent addition-subtraction and basic-geometry with a larger gap. Conversely, in Fig. 2(b), dots with gray and yellow respectively represent fractions and rates-and-ratios with a smaller gap.

As can be seen, problems belong to different concepts in Fig. 2(a) are clustered to several groups, e.g., pink dots in the upper parts and blue dots in the right lower parts. Meanwhile, in Fig 2(b), problems belong to similar concept are mixed together. By such visualization, the math teachers can examine each problem to verify if the manual classification should be modified.

(a) The concepts with larger difference. (b) The concepts with small difference.

Fig. 2: The visualizations of features of exercises in the AKT model with different settings of concepts. Each point represents a math problem, where the same color means that points belong to the same concept.

IV. Conclusion

In this paper, we propose an effective approach CIV to visualize exercise features by combining the AKT model and T-SNE. With this approach, any answer records of students can be analyzed and visualized by experts without *engineering programming* knowledge, resulting in more accurate and efficient problem classification. The E-learning platforms and students both benefit from CIV. In future work, we aim to design a knowledge map that utilizes the CIV results to help students identify their weak subjects and enhance their learning experience.

References

[1] C. Piech, J. Bassen, J. Huang, *et al.*, “Deep knowledge tracing,” *Advances in neural information processing systems*, vol. 28, 2015.

[2] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.,” *Journal of machine learning research*, vol. 9, no. 11, 2008.

[3] A. Ghosh, N. Heffernan, and A. S. Lan, “Context-aware attentive knowledge tracing,” *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2330–2339, 2020.