The 28th ACM International Conference on Information and Knowledge Management

CIKM 2019 BEIJING CHINA November 03-07. 2019

DIRT: Deep Learning Enhanced Item Response Theory for Cognitive Diagnosis

Song Cheng[1], Qi Liu[1,*], Enhong Chen[1], Zai Huang[1], Zhenya Huang[1], Yuying Chen[1,2], Haiping Ma[3], Guoping Hu[3]

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, University of S&T of China [2]Ant Financial Services Group [3]IFLYTEK Co.,Ltd.

{chsong, huangzai, huangzhy, cyy3322}@mail.ustc.edu.cn; {qiliuql,cheneh}@ustc.edu.cn; {hpma, gphu}@iflytek.com

	1. In $\Delta ABC, AB = AC, \angle BAC = 108^{\circ}$. AD, AE and BC intersect at point D and E. And $\angle BAC$ is divided into three equal parts, what is wrong? Knowledges: Similar triangle properties, Similar triangle judgement, Proportional line segment	✓
	2. Calculate $4sin60^{\circ} + tan45^{\circ} - 2\sqrt{3}$ Knowledges: Quadratic root operation, Special trigonometric function	✗
	3. Two midlines AD, BE of ΔABC intersect at G, line $EF//BC$ through E and intersect with AD at F Knowledges: Parallel line segment proportion, centroid	✓

Figure 1: A toy example of student question records

Abstract

Cognitive diagnosis is the cornerstone of modern educational techniques. One of the most classic cognitive diagnosis methods is Item Response Theory (IRT), which provides interpretable parameters for analyzing student performance. However, traditional IRT only exploits student response results and has difficulties in fully utilizing the semantics of question texts, which significantly restricts its application. To this end, in this paper, we propose a simple yet surprisingly effective framework to enhance the semantic exploiting process, which we termed Deep Item Response Theory (DIRT). In DIRT, we first use a proficiency vector to represent student proficiency on knowledge concepts and represent question texts and knowledge concepts by dense embedding. Then, we use deep learning to enhance the process of diagnosing parameters of student and Question by exploiting question texts and the relationship between question texts and knowledge concepts. Finally, with the diagnosed parameters, we adopt the item response function to predict student performance. Extensive experimental results on real-world data clearly demonstrate the effectiveness and the interpretability of DIRT framework.

Problem Definition

Suppose there are L students, M questions and total P knowledge concepts. The history records that L students do M questions are represented by $R = \{1 \le i \le L, 1 \le j \le M\}$, where $R_{ij} = \langle S_j, Q_j, r_{ij} \rangle$ denotes the student S_j obtains score r_{ij} on question Q_j. $Q_j = \langle QT_j, QK_j \rangle$ is composed of question texts QT_j and knowledge concepts QK_j. Given students' responses r_{ij}, question texts QT_j and knowledge concepts QK_j, our goal is to build a model $\mathcal{M}$ to diagnose students' proficiency on each knowledge concept.

DIRT Framework

DIRT contains three modules, i.e., input, deep diagnosis and prediction modules. Input module initializes a proficiency vector for the student, and embeds question texts and knowledge concepts to vectors. Deep diagnosis module diagnoses latent trait, discrimination and difficulty with deep learning to enhance the model. Prediction module predicts the probability that the student answers the question correctly with item response function. In the section bellow, we give a specific implementation of DIRT which is shown in Figure 2.

The Input:
$$x_t = \sum_{k_i \in \mathcal{K}_q} \mathrm{softmax}(\frac{\xi_j}{\sqrt{d_0}})k_i + w_t, \quad \xi_j = w_t^T k_i$$

Latent Trait:
$$\theta = \mathrm{DNN}_{\theta}(\Theta), \quad \Theta = \alpha \odot k = \sum_{k_i \in \mathcal{K}_q} \alpha_i k_i$$

Discrimination:
$$a = 8 \times \mathrm{sigmoid}(\mathrm{DNN}_a(A) - 0.5), \quad A = k \oplus k = \sum_{k_i \in \mathcal{K}_q} k_i$$

Difficulty:
$$b = 8 \times \mathrm{sigmoid}(\mathrm{averagePooling}(A) - 0.5)$$

Prediction:
$$P(\theta) = \frac{1}{1 + e^{-Da(\theta - b)}}$$

Objective Function:
$$\mathcal{L} = r_{ij}\log\widetilde{r_{ij}} + (1 - r_{ij})\log(1 - \widetilde{r_{ij}})$$

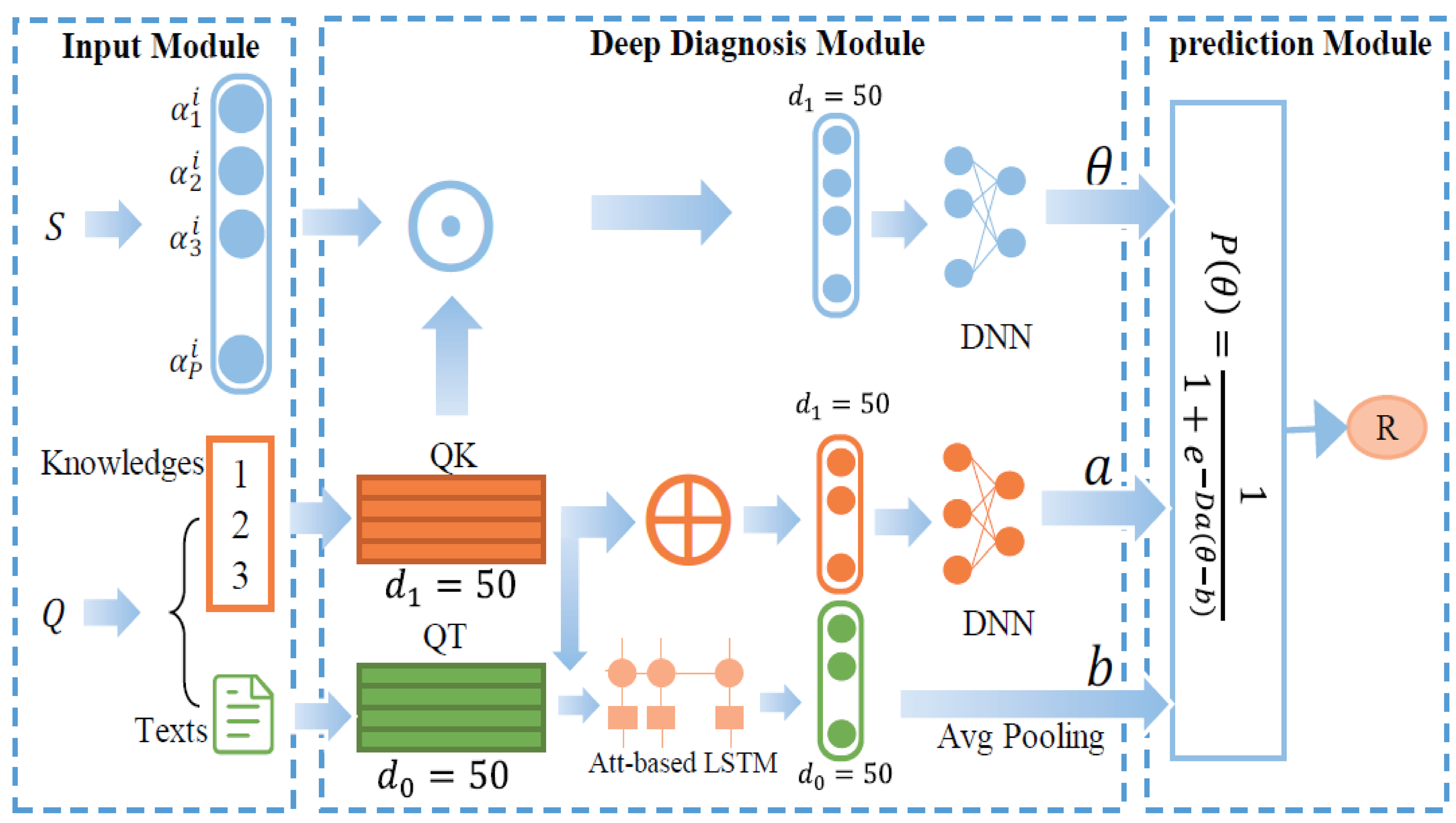


Figure 2: Deep Learning Enhanced Item Response Theory Framework

Experimental Results

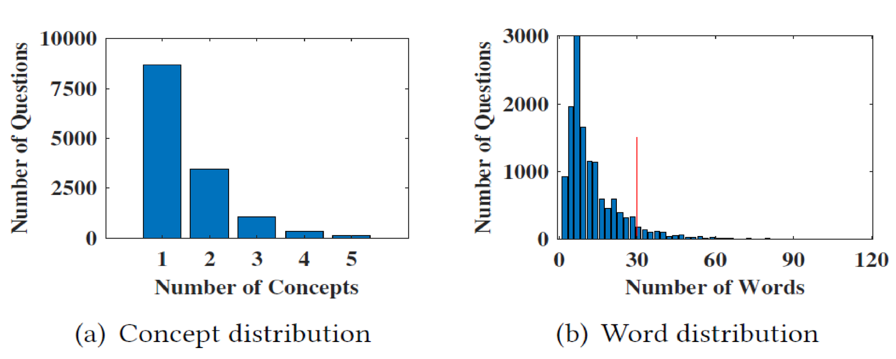


Figure 3: Distribution of words and knowledge concepts

Statistics	Original	Pruned
# of history records	65,368,739	5,068,039
# of students	1,016,235	81,624
# of questions	1,735,635	13,635
# of knowledge concepts	1,412	621
Avg. questions per student	/	62.09
Avg. concepts per question	/	1.49

Table 1: The statistics of the dataset

We conduct extensive experiments to demonstrate the effectiveness of our approach. First, we compare the performance between DIRT and baseline approaches for performance prediction. Then, we conduct a case study to visualize the explanatory of the DIRT.

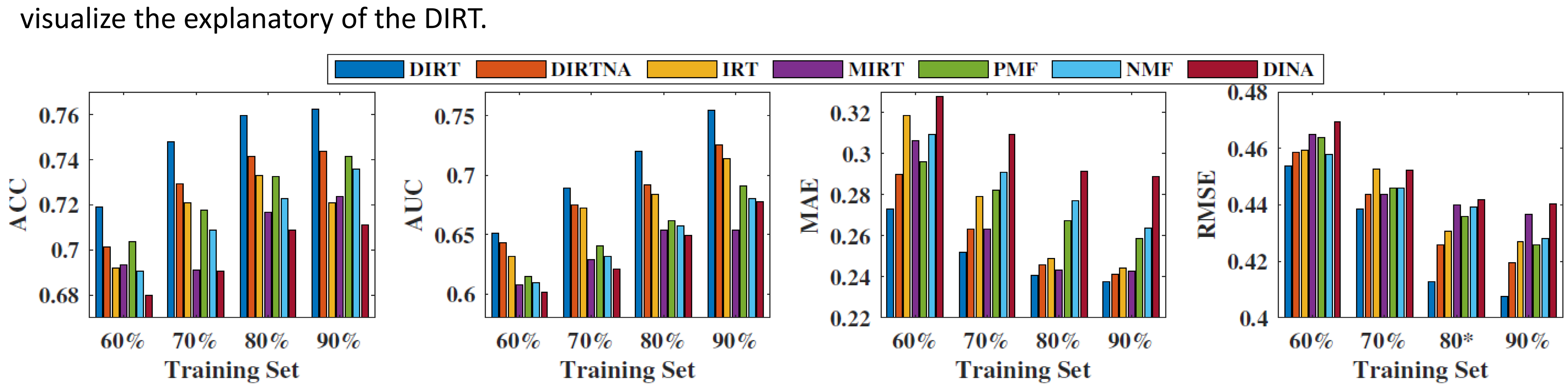
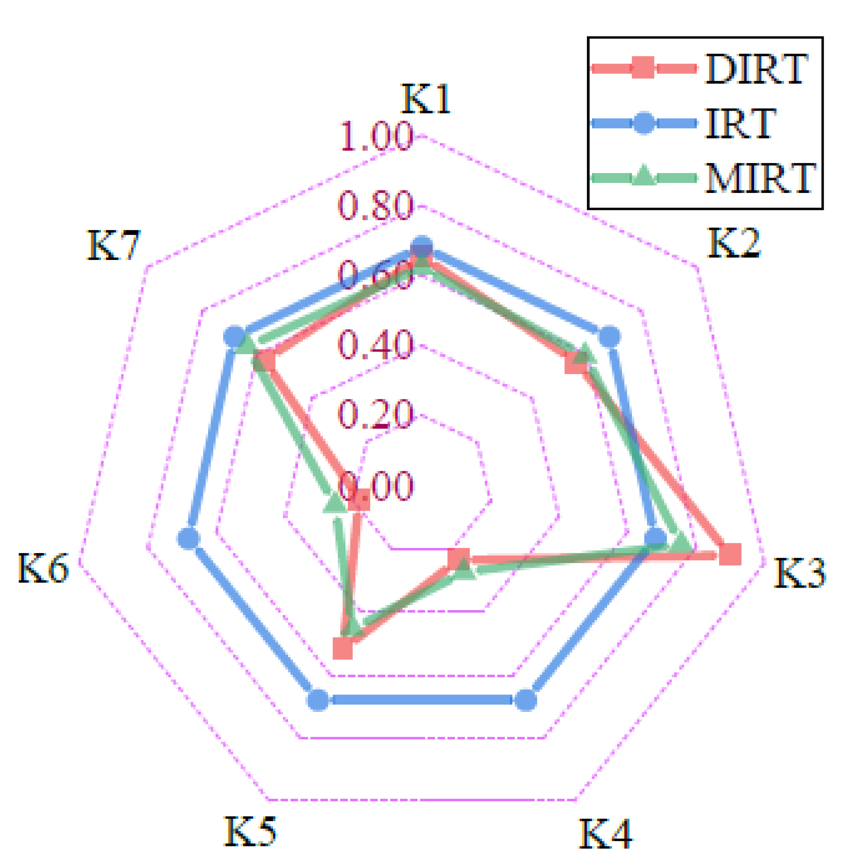


Figure 4: Overall results of student performance prediction on four metrics



	Q	real	prediction			Discrimination a			Difficulty b		
			IRT	MIRT	DIRT	IRT	MIRT	DIRT	IRT	MIRT	DIRT
1. In $\Delta ABC, AB = AC, \angle BAC = 108^{\circ}$. AD, AE and BC intersect at point D and E. And $\angle BAC$ is divided into three parts,what is wrong? *Concepts*: K1, K2, K3.	1	✓	✓	✓	✓	1.3545	1.4437	1.69	0.6352	0.5439	0.51
2. Calculate $4sin60^{\circ} + tan45^{\circ} - 2\sqrt{3}$ *Concepts*: K4, K5 .	2	✗	✓	✗	✗	0.6358	0.362	1.47	-0.171	0.1374	0.18
3. What is the minimal positive period of function y = 1 – cos(2x) ? *Concepts*: K6, K7	3	✗	✗	✓	✗	0.5102	0.3957	0.358	1.265	0.7352	0.6475

Figure 5: Visualization of a student's proficiency on knowledge concepts and the parameters of three questions.