

多视角认知诊断:学生评估方法

陈智斌, 顾克, 高硕

(河海大学信息学部计算机与信息学院, 江苏南京 211100)

摘要: 认知诊断是学生评估的主流方法之一。但是传统的认知诊断缺失了相关的知识点信息与试题文本信息, 导致建立的只是学生序号与试题序号的关系。另外由于试题文本较为复杂, 将其用于学生成绩的估计可能会在提取文本信息的过程中产生信息丢失的问题。多视角认知诊断针对上述问题, 将试题文本引入分析; 以认知诊断为基础, 引入试题文本调整所得到的试题参数, 使其具有更高的可靠性。同时为避免信息在提取的过程中丢失, 将一个词向量模型定义为提取文本信息的一个视角, 使用多视角学习融合各视角的信息, 实现了从多个视角看待试题文本, 从而在从文本到提取试题信息的过程中尽可能得还原出试题文本包含信息的全貌。实验证明了多视角认知诊断对于传统的认知诊断模型, 在 f-score 上提高了 0.10, 在均方根误差上降低了 0.10, 而对于最新的认知诊断方法例如多任务认知诊断框架, 多视角认知诊断在 f-score 上提升了 0.11, 在均方根误差上降低了 0.01, 具有非常明显的提升效果。

关键词: 认知诊断; 学生评估; 智慧教育; 多视角学习; 自然语言处理

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2022)0013-05

Multi View Cognitive Diagnose Model: A Framework for Student Assessment

CHEN Zhi-bin, GU Ke, GAO Shuo

(Department of Information School of Computer Science and Information, Hohai University, Nanjing 211100, China)

Abstract: Cognitive diagnosis is one of the mainstream method in student assessment. But because it is lack of related knowledge information and exam text information, leading to the traditional cognitive diagnosis only establish the relationship between serial number of students and questions. In addition, due to the complexity of the exam text, its use for estimating student performance may cause problems of information loss in the process of extracting textual information. The multi-perspective cognitive diagnosis addresses these problems by introducing the exam text into the analysis. Using the cognitive diagnosis as a basis, the exam text is introduced to adjust the resulting test parameters to make them more reliable. In order to avoid information loss in the extraction process, a word vector model is defined as a perspective for extracting text information, and multi-perspective learning is used to fuse the information from each perspective, so that the exam text can be viewed from multiple perspectives. Thus restoring the full information contained in the exam text as much as possible in the process of extracting the test information from the text. The experiments demonstrate that multi-perspective cognitive diagnosis improves the f-score by 0.10 and reduces the root mean square error by 0.10 for traditional cognitive diagnostic models. And it improves the f-score by 0.11 and reduces the root mean square error by 0.01 for the latest cognitive diagnostic methods such as the multi-task cognitive diagnostic framework, which is a significant improvement.

Key words: cognitive diagnose; student assessment; intelligent education; multi view learning; natural language learning

0 引言

在线教育因其不受时间空间的限制受到广泛的关注, 在 Coursera、Udacity 的影响下, 国内也逐渐出现了诸如 MOOC、微课网等在线教育平台^[1], 此外, 智能辅导系统也备受关注^[2]。在线教育的发展促进了教育数据的大量积累, 为教育数据挖掘提供了充足的

条件^[3-5]。

认知诊断^[6-7]是教育数据挖掘中评估学生的一种应用较为广泛的方法。但是传统的认知诊断仅通过学生的考试成绩对学生的进行学习情况进行诊断, 也就意味着建立的学生试题间的关系模型只是试题序号与学生序号之间的关系, 分析的结果不具有可解释性的同

收稿日期: 2021-08-15

基金项目: 国家自然科学基金资助项目(61976118)

作者简介: 陈智斌(2000-), 男, CCF 会员(A0682G), 研究方向为计算机视觉、智慧教育。

时,也缺失了泛化效果。

引入文本分析需要恰当的方法,否则可能造成信息丢失,甚至对学生试题关系建模造成负面的影响。由于试题文本的异构性和复杂性,直接将其考虑进学生评估中,不一定能取得很好的效果,而该文通过文本的异步分析及多视角学习,合理地将文本信息在恰当的时候以恰当的方法辅助学生学习情况诊断。

贡献在于以下三个方面:

(1)和直接将文本引入初始的分析阶段不同,考虑到现在的自然语言处理技术的局限和学生试题关系的复杂,直接端到端的引入分析会导致建模难度过大。因此,采用先使用认知诊断模型分析大致范围,再将文本引入分析的异步分析方式,达到在认知诊断的基础上小范围调整,最终提高准确性的目的。

(2)该文是第一个尝试将多视角学习引入学生评估的方法。考虑到单一的词向量提取的信息有所缺失,将不同的词向量视为观察文本的不同视角,从而使用多视角学习,综合多个词向量模型提取的信息,达到文本的分析合理、有效的目的。

(3)通过与传统、现有方法的对比,对学生的评估分析具有明显的提升效果,同时通过与单视角模型架构的对比,证明了多视角的引入确实能够进一步提升分析的准确性。

1 相关工作

1.1 认知诊断

认知诊断模型在不同的场景下有不同的应用,根据模型设计不同,可以分析学生的知识掌握状态,猜测率与失误率等。项目反应理论(item response theory, IRT)是认知诊断的主流方法之一^[8]。项目反应理论通过连续的向量来表征学生与试题。项目反应理论通常是由专家设计一个学生-试题反应函数,在假设其满足独立同分布的前提下^[9]通过统计学习方法得出学生与试题的参数。

1.2 多视角学习

通常描述一个问题具有多种不同的维度,这些不同的维度之间形式上不同单互有交集,也互有补充^[10]。从单个视角进行学习有可能造成信息的丢失,从而无法得到最合适的模型表征问题。多视角学习则尝试从多个视角出发,融合不同视角信息进行学习,使其最大化的利用信息。多视角主要分为三个方向:联合训练^[11],多核学习^[12]与子空间学习^[13]。

联合训练在每一个迭代过程中,一个视角的模型对验证集进行标注并将其加入另一视角的模型的下一迭代过程的训练集中。通过交替训练最终达到两个数据源的融合。在此之外还有诸多变体。例如将

EM^[14]、主动学习^[15-17]、贝叶斯无向图模型^[18]、高斯过程^[19]与联合训练结合。

2 多视角认知诊断

该文提出了 MV-CDM,一种学生评估方法。将文本信息引入认知诊断,可以做到直接通过试题文本分析试题参数。

该方法的合理性在于,认知诊断由于分析过程文本信息的缺失而导致结果有偏差,但大体上是可靠的,只需小范围的纠正。而只有引入了完整的文本信息,才能进行正确的纠正。而由于试题文本的复杂性,使用单一的词向量模型容易造成提取的信息不全面的问题。基于此,将多视角与认知诊断结合,具体为将每个词向量模型作为一个提取试题文本信息的视角,通过多视角学习,互相补充,以达到通过多个视角看待试题文本的目的,从而完整地提取文本信息。而多视角学习正是该方法奏效的关键所在。

本节首先在 2.1 给出问题定义,接着在 2.2 分小节依次详细描述 MV-CDM 各部分实现过程。

2.1 问题定义

给定学生集合 $S = \{s_1, s_2, \dots, s_u\}$, 试题集合 $J = \{j_1, j_2, \dots, j_v\}$, 试题文本集合 $T = \{t_1, t_2, \dots, t_v\}$, 以及学生得分矩阵 $R = [r_{uv}]_{U \times V}$, 其中 $r_{uv} = 1$ 表示学生 u 答对试题 v , $r_{uv} = 0$ 表示学生 u 答错试题 v 。该文主要的研究目标为如何通过上述信息准确预测学生的成绩, 以实现对学生的评估。

表 1 给出了问题所需要的符号和对应的描述。

表 1 涉及的一些重要符号及其描述

符号	描述
S	学生集合
s_u	学生 u
J	试题集合
j_v	试题 v
R	学生得分矩阵
T	试题文本集合
t_v	试题 v 的文本
θ_u	学生 u 参数
ξ_v	试题 v 参数
ε_k	模型 k 提取的试题文本的词向量
f_{kv}	模型 k 提取的试题文本 v 的信息

2.2 MV-CDM 具体过程

提出了多视角认知诊断(multi view cognitive diagnose model, MV-CDM)的框架用于对学生成绩进行预测, MV-CDM 的框架如图 1 所示。

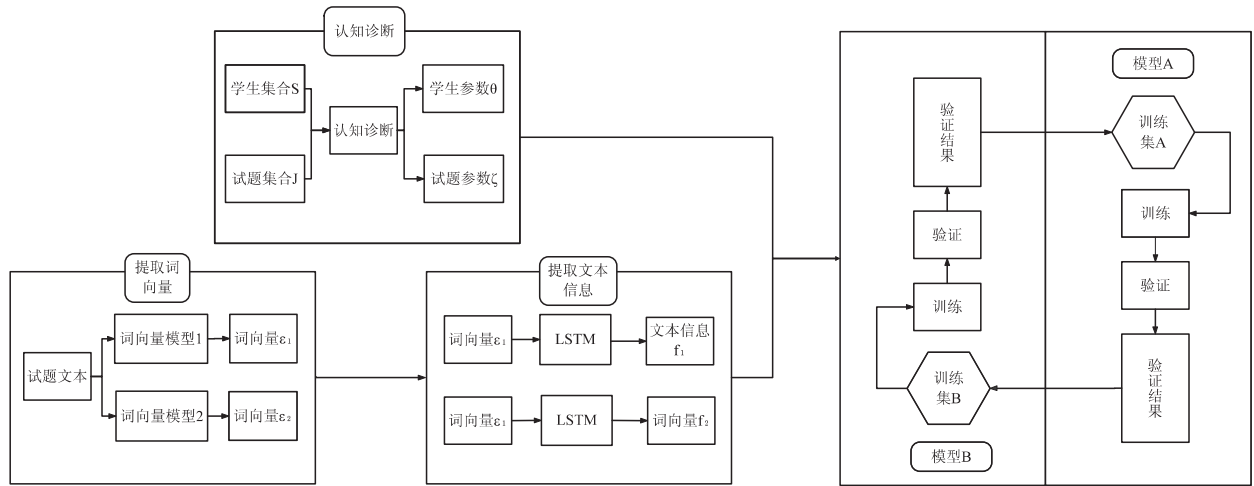


图1 MV-CDM 框架

2.2.1 认知诊断

据前所述,首先需要得到认知诊断的分析结果作为后续训练模型的基准,因此先选用任一满足公式(1)条件的认知诊断模型对给定的学生 S 、试题 J 进行分析得到学生参数 θ 、试题参数 ξ 。

$$P(R_{uv} = 1 | \theta_u, \xi_v) \equiv f(\theta_u, \xi_v) \quad (1)$$

其中, ξ 为 M 维的向量, θ 为 M 维的向量。

2.2.2 提取文本信息

首先选择两个不同的词向量模型。对于给定试题文本 T , 分别提取 T 的词向量集合 ϵ_1, ϵ_2 。将得到的 ϵ_1 按句子的顺序依次提取文本信息。为防止过拟合, 减少参数量, 先将每个词向量与矩阵 $W(D \times \frac{D}{2})$ 相乘进行降维。接下来采用双向 LSTM 得到每到试题的信息 f_{1v} , 模型 2 提取 f_{2v} 同理。

2.2.3 单视角的模型架构

单个视角的模型架构如图 2 所示。将 2.2.1 得到的学生隐藏属性 θ 求平均, 得到 $\bar{\theta}$, 作为模型 1 端的输入, 另一端的输入为 2.2.2 得到的 F 。经过图 2 所示的网络, 输出为预测的试题 v 的参数 $\bar{\xi}_v$ 。将 $\bar{\xi}_v$ 与真实的试题参数 ξ_v 进行比较, 得到损失函数:

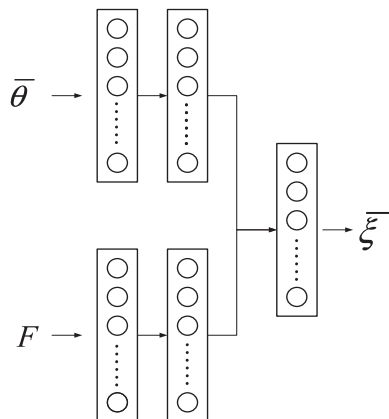


图2 模型架构

$$loss = \frac{1}{2} \sum_v \|\bar{\xi}_v - \xi_v\|_F^2 \quad (2)$$

根据得到的 θ_u 与 $\bar{\xi}_v$, 使用所选择的认知诊断模型的学生试题反应函数 $f(\theta_u, \bar{\xi}_v)$ 预测学生的答题情况 r'_{uv} , 从而得到预测的全体学生的得分矩阵 R' 。统计真实的得分矩阵 R 与预测的得分矩阵 R' ; 计算 R 的正确率 r 、 R' 的正确率 r' , 在 loss 中引入正则项 $\lambda \times |r - r'| \times |Y - thres|$, 则公式(2)中的 loss 变为:

$$loss = \frac{1}{2} \sum_v \|\bar{\xi}_v - \xi_v\|_F^2 + \lambda \times |r - r'| \times |Y - thres| \quad (3)$$

从而目标函数为:

$$\min_{\xi} \frac{1}{2} \sum_v \|\bar{\xi}_v - \xi_v\|_F^2 + \lambda \times |r - r'| \times |Y - thres| \quad (4)$$

其中, thres 为预设的目标阈值。引入正则项的目的是防止预测的结果倾向于全对或全错这样的极端化表现。

2.2.4 多视角学习

将词向量模型 1 作为一个视角的信息进行训练, 得到的模型记为模型 A, 将词向量模型 2 作为另一视角进行训练, 得到的模型记为模型 B。将数据划分为训练集 O 、验证集 V 与测试集 P , 模型 A、B 初始的训练集相同均为 O 。在第一遍训练结束后, 对得到的模型 A' 、 B' , 计算其在验证集 V 上的预测 Z_A 、 Z_B 。在第二遍训练时将 Z_A 加入初始训练集中作为模型 B 的训练集, 将 Z_B 加入初始训练集 O 中作为模型 A 的训练集, 即此时模型 A 的训练集为 $O_A = O \cup Z_B$, 模型 B 的训练集为 $O_B = O \cup Z_A$ 。之后的训练同理, 每一遍训练结束后都预测两个模型在 V 上的输出, 在下一训练阶段时, 分别与初始训练集 O 合并作为另一模型的训练集, 即 $O_A = O \cup Z_B$, $O_B = O \cup Z_A$ 。反复迭代, 直到两

个模型都收敛,得到的两个模型分别记作 \bar{A} 、 \bar{B} 。

2.2.5 学生得分预测

由于在训练时已采用多视角学习融合了两个视角的信息,因此在实际的预测阶段任选一个模型作为预测用模型即可,记为模型 X 。

对于给定的待考试试题文本 T ,先分别对每一道试题按 2.2.2 的方法提取每道试题文本的词向量 ε ,再对 ε 提取每道题的文本信息 F ,接下来采用模型 X 使用 2.2.3 的方法预测每一道试题的参数 ξ_v 。将试题参数 ξ_v 代入选用的认知诊断模型的学生-试题反应函数 $f(\theta_u, \xi_v)$,得到目标学生在待考试的试题文本上答对的概率 $P(Y_{uv} = 1 | \theta_u, \xi_v) \equiv f(\theta_u, \xi_v)$,若概率大于等于设定阈值,则判定学生能答对该题,若概率小于设定阈值,则判定学生答错该题。

3 实验

3.1 实验设置

实验数据采自一所高中生物考试的测验成绩(数据集地址: https://drive.google.com/drive/folders/17CRXp45j6WC6NIYPa5iZjptZ_QX4FO-t?usp=sharing),共包括 2 次考试,每次考试 456 人,40 道题,题目皆为客观题,每题的得分根据选择答案的正确或错误分别为 0 分或 1.5 分。数据集提供了测验的试题文本,1.txt ~ 5.txt 为试题文本,每行为一道试题的文本。data1.txt ~ data5.txt 为考试答题情况记录,每行为一名学生的答题情况记录,每列为一道题的答题情况记录。将其平均分为 5 份,每份 16 题,依次标记为

1-5 组。其中 1-3 组为公用的训练集,4 组为验证集,5 组为测试集。

3.2 实验指标

预测学生成绩是学生评估的一个主要途径,也是验证学生评估方法是否有效的一个重要手段,本部分先分别对各方法进行训练,之后在公共的测试集上进行测试,比较预测的学生成绩和真实成绩的误差,以此证明本文方法的有效性。

实验从回归和分类两方面对各方法进行评估。从回归角度,实验结果用均方根差(root mean square error, rmse)、平均绝对误差(mean absolute error, mae)进行度量,均方根差与平均绝对误差越小方法预测的结果与真实情况越接近。从分类角度用召回率(recall)、F-score、准确度(accuracy rate, acc)、f-score 进行衡量,设置 0.5 为阈值,大于 0.5 的预测值认为是 1,反之认为是 0,结果的召回率、f-score、准确度越高方法越优,也就意味着能更好的对学生进行评估。

3.3 对比实验

本部分使用 M2PL^[20]作为基础认知诊断模型进行实验:Multi View M2PL(MV-M2PL);同时为证明多视角的有效性进行了单视角下的实验:Single View M2PL(SV-M2PL)、Single View M2PNO(SV-M2PNO);并与将多任务学习与认知诊断结合的 MT-MCD^[21]和基础的认知诊断模型 M2PL 进行比较。实验结果如表 2 所示。

表 2 对比实验

方法	acc	recall	f-score	rmse	mae
M2PL	0.673 961	0.700 414	0.686 933	0.526 361	0.360 122
MT-M2PL	0.712 527	0.841 058	0.771 476	0.536 134	0.287 512
SV-M2PL(ours)	0.75	0.791 897	0.770 379	0.443 231	0.291 341
MV-M2PL(ours)	0.762 856	0.805 963	0.783 817	0.428 111	0.278 184

实验结果证明,MV-CDM 应用于不同的基础认知诊断模型均能显著提升效果:单视角模型能明显提升认知诊断的效果,证明了文本的异步分析思路的正确性和可靠性;而多视角架构的模型与单视角模型比较也证明了引入多视角学习的有效性和对预测学生成绩,准确评估学生的巨大辅助作用。

4 结束语

提出了一种将文本信息引入认知诊断的框架,同时由于现阶段自然语言处理较难提取文本中包含的与学生学习情况的有关信息,结合了多视角学习方法补足了只从单一视角提取信息可能导致的信息缺失,并

通过实验证明了该方法的有效性。本文只是在该思路上的一个尝试,因此对于如何对于模型的设计以及多视角的应用都较为简单,如何设计更好的模型对文本、答题数据进行分析建模以及如何更好的应用多视角学习方法来更好的契合该任务,也是今后的研究方向。

参考文献:

- [1] 陈池,王宇鹏,李超,等.面向在线教育领域的大数据研究及应用[J].计算机研究与发展,2014,51(S1):67-74.
- [2] POLSON M C, RICHARDSON J J. Foundations of intelligent tutoring systems[M]. Psychology Press, 2013.
- [3] 刘淇,陈恩红,黄振亚,等.面向个性化学习的学生认知

- 能力分析[J]. 中国计算机学会通讯,2017,13(4):28-34.
- [4] 牟智佳,俞显,武法提. 国际教育数据挖掘研究现状的可视化分析:热点与趋势[J]. 电化教育研究,2017,38(4):108-114.
- [5] 柯清超. 大数据与智慧教育[J]. 中国教育信息化·基础教育,2013(12):8-11.
- [6] 刘声涛,戴海崎,周骏. 新一代测验理论——认知诊断理论的源起与特征[J]. 心理学探新,2006,26(4):73-77.
- [7] 陈秋梅,张敏强. 认知诊断模型发展及其应用方法述评[J]. 心理科学进展,2010(3):522-529.
- [8] RUPP A A, TEMPLIN J. The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model[J]. Educational and Psychological Measurement,2008,68(1):78-96.
- [9] 刘淇,陈恩红,朱天宇,等. 面向在线智慧学习的教育数据挖掘技术研究[J]. 模式识别与人工智能,2018,31(1):77-90.
- [10] XU C, TAO D, XU C. A survey on multi-view learning[J]. arXiv:1304.5634,2013.
- [11] BLUM A, MITCHELL T. Combining labeled and unlabeled data with co-training[C]//Proceedings of the eleventh annual conference on computational learning theory. New York:Association for Computing Machinery,1998:92-100.
- [12] GÖNEN M, ALPAYDIN E. Multiple kernel learning algorithms[J]. The Journal of Machine Learning Research,2011,12:2211-2268.
- [13] DE LA TORRE F, BLACK M J. A framework for robust subspace learning[J]. International Journal of Computer Vision,2003,54(1):117-142.
- [14] NIGAM K, GHANI R. Analyzing the effectiveness and applicability of co-training[C]//Proceedings of the ninth international conference on information and knowledge management. New York:Association for Computing Machinery,2000:86-93.
- [15] MUSLEA I, MINTON S, KNOBLOCK C A. Active+ semi-supervised learning= robust multi-view learning[C]//The nineteenth international conference on machine learning. Sydney:Morgan Kaufmann,2002:435-442.
- [16] MUSLEA I, MINTON S, KNOBLOCK C A. Active learning with strong and weak views;a case study on wrapper induction[C]//Proceedings of the eighteenth international joint conference on artificial intelligence. Acapulco:Morgan Kaufmann,2003:415-420.
- [17] MUSLEA I, MINTON S, KNOBLOCK C A. Active learning with multiple views[J]. Journal of Artificial Intelligence Research,2006,27:203-233.
- [18] YU S, KRISHNAPURAM B, STECK H, et al. Bayesian co-training[C]//Advances in neural information processing systems. Vancouver:Curran Associates, Inc,2008:1665-1672.
- [19] YU S, KRISHNAPURAM B, ROSALES R, et al. Bayesian co-training[J]. The Journal of Machine Learning Research,2011,12:2649-2680.
- [20] RECKASEM D. Multidimensional item response theory models[M]//Multidimensional item response theory. New York,NY:Springer,2009:79-112.
- [21] ZHU T, LIU Q, HUANG Z, et al. MT-MCD:a multi-task cognitive diagnosis framework for student assessment[C]//International conference on database systems for advanced applications. [s. l.]:Springer,2018:318-335.
- (上接第12页)
- gium:[s. n.],2018:4231-4242.
- [47] HUANG X, ZHANG J, LI D, et al. Knowledge graph embedding based question answering[C]//Proceedings of the twelfth ACM international conference on web search and data mining. Melbourne, VIC, Australia:ACM,2019:105-113.
- [48] SAXENA A, TRIPATHI A, TALUKDAR P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. [s. l.]:[s. n.],2020:4498-4507.
- [49] YASUNAGA M, REN H, BOSSELUT A, et al. QA-GNN: reasoning with language models and knowledge graphs for question answering[C]//Proceedings of the 2021 conference of the north American chapter of the association for computational linguistics; human language technologies. [s. l.]:[s. n.],2021:535-546.
- [50] BORDES A, CHOPRA S, WESTON J. Question answering with subgraph embeddings[C]//Proceedings of the 2014 conference on empirical methods in natural language processing. Doha, Qatar:[s. n.],2014:615-620.
- [51] ZHANG Y, LIU K, HE S, et al. Question answering over knowledge base with neural attention combining global knowledge information[J]. arXiv:1606.00979,2016.
- [52] BAHDANAU D, CHO K, BENGIO Y J A P A. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473,2014.