

DKTwMF:一种融合多特征的知识追踪模型

王 斌,盛宇轩,冀星昀

(中南大学 计算机学院,湖南 长沙 410083)

摘要:知识追踪的目的是通过分析学习活动来量化学生掌握知识的能力,进而为每个学生提供更具有针对性的训练。伴随时间的推移,系统中记录的学习数据的数据量不断增加。如何充分利用这些数据,为学生提供个性化教育是目前教育数据挖掘领域的一个重要研究方向。现有的知识追踪模型大多只考虑学生练习相关的知识点和作答结果,并未充分利用数据集中其他数据。为解决上述问题,提出了一种融合多特征的知识追踪模型(deep knowledge tracing with multiple features,DKTwMF)。该模型首先利用邻域互信息和随机森林对特征进行选择,然后将多特征数据进行编码,最后使用深度学习对学生知识掌握状态进行建模。该模型可以自动提取出数据集中的重要特征,减少训练参数,保证模型能够收敛到全局最优,从而更准确地对学生技能状态进行评估。

关键词:数据挖掘;知识追踪;教育系统;特征融合;深度学习

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2021)07-0035-07

doi:10.3969/j.issn.1673-629X.2021.07.006

DKTwMF: A Deep Knowledge Tracing Model with Multiple Features

WANG Bin, SHENG Yu-xuan, JI Xing-yun

(School of Computer Science and Engineering, Central South University, Changsha 410083, China)

Abstract: The purpose of knowledge tracking is to quantify the students' ability of grasping knowledge by analyzing learning activities, and then provide more targeted training for each student. With the passage of time, the learning data recorded in this system is increasing. How to make full use of these data to provide students with personalized education is an important research direction in the field of educational data mining. Most of the existing models only consider knowledge points and answers related to student practice and do not make full use of other data in the data set. In order to solve the above problems, we present a multi-feature knowledge tracking model, DKTwMF. Firstly, the model uses the neighborhood mutual information and random forest to select the features, and then encodes the multi feature data. Finally, deep learning is used to model student knowledge. This model can automatically extract the important features of the data set, reduce the number of training parameters, and ensure the model can converge to the global optimal, so as to evaluate the students' skill state more accurately.

Key words: data mining; knowledge tracing; education system; feature integration; deep learning

0 引言

近年来,随着在线教育逐渐被人们认可,慕课网、网易云课堂、Coursera 和 Khan Academy 等大型在线开放式网络课程(massive open online courses, MOOC)平台由于其高质量的在线课程,吸引了全球范围内的大量用户注册使用^[1]。特别是在疫情防控期间,教育部要求在确保全体师生健康的基础上做好“停课不停学”工作,各地高校利用各大 MOOC 平台,开展了大规模的线上教学^[2]。利用从这些平台收集的数据,研究人员能够深入了解学生的学习过程,进而为教师、学生提供个性化服务,这引发了对教育数据挖掘

(educational data mining, EDM)的更多关注^[3-5]。

知识追踪是教育数据挖掘领域的一个重要研究课题^[6-7],其目标是随着时间推移不断对学生知识掌握状态进行建模,通过估计学生的知识掌握程度^[7],预测学生在后续练习中的表现状况。知识追踪模型被广泛应用于教育系统中,实现了学习评价的智能化,有助于推进个性化教育。知识追踪的方法主要分为三类^[8]。第一类是基于概率图模型的知识追踪,代表性模型是贝叶斯知识追踪模型(Bayes knowledge tracing, BKT)^[9-10]。该模型使用隐马尔可夫模型(hidden Markov model, HMM)学习学生知识状态,用一组二进

收稿日期:2020-07-25

修回日期:2020-11-27

基金项目:国家科技重大专项项目(2017ZX06002005)

作者简介:王 斌(1973-),男,博士,教授,博导,CCF 会员(E200006977M),研究方向为数据处理、智能软件。

制变量表示学生是否掌握相应的知识。BKT 可以通过严格的公式推导得到学生知识掌握情况、下次作答的预测结果,具有很强的解释性。但是由于该模型需要对各个知识单独建模(即只保留了练习的相对顺序),模型不能充分考虑知识之间的相互作用。第二类是基于矩阵分解的知识追踪,如概率矩阵分解(PMF)。该算法是推荐系统领域的经典算法之一,由于推荐领域与知识追踪的相似性,部分学者将这一算法改进以用于知识追踪领域。第三类是基于深度学习的知识追踪。

2015 年斯坦福大学的 Piech 等人^[11]在 NIPS 上提出深度知识追踪模型(deep knowledge tracing, DKT)。相比其他模型,DKT 能够利用学生的历史作答数据挖掘出各个练习之间的隐藏联系。与 BKT 相比,DKT 首次出现便达到了 25% 的 AUC 增长^[11]。尽管一些研究人员发现通过适当的扩展,BKT 可以达到和 DKT 相近的性能^[3],相信通过适当扩展 DKT 性能仍能够明显提升。传统的 DKT 模型只使用了知识点 ID 和作答结果作为模型的输入,它并没有考虑其他特征,如作答时间、作答次数、上次作答结果等。这些特征不仅表示了学生的作答行为,更为知识诊断提供了额外的信息,充分利用这些额外信息能够更加准确地评估学生知识掌握情况。

文献[12]提出一种基于手动选择特征、离散化连续特征的方法来提升 DKT 模型效果。文献[13]通过获取学生能力,并以固定时间间隔将学生动态分配至具有相似能力的不同组别中,再将这些信息送入 DKT 中。但这些方法受到两个方面的限制:(1)需要足够的专业知识来理解数据。当对数据理解不足时,模型的构建可能会产生较大偏差;(2)当数据巨大时,方法费时费力。与 DKT 模型类似,另一种 EERNN (exercise-enhanced recurrent neural network)模型^[14]也是通过 LSTM 网络来追踪学生知识掌握的情况。不同的是,EERNN 模型更加注重于充分利用题目当中的文本信息,以解决冷启动问题。同时,该模型还引入了注意力机制,通过考虑题目之间的相似性来提升在学生未来表现预测中的准确性。除了上述三种基本方法外,近年来,基于知识图谱的知识追踪也在不断发展^[15]。

为了解决上述 DKT 模型中的问题,该文提出一种融合多特征的知识追踪模型(deep knowledge tracing with multiple features, DKTwMF)。首先,提出基于邻域互信息和随机森林的混合特征选择算法(neighborhood mutual information and random forest, NMIRF),来提取巨大数据中的重要特征,然后将这些重要特征数据以及学生表现作为 LSTM 模型输入对

学生的知识水平进行建模。同时,当数据维度增大时,循环神经网络的训练代价相当昂贵,因此提出一种新的多特征编码方案(multi-feature encoding, MFE),通过使用交叉特征、one-hot 编码对多特征输入进行编码,兼顾了输入数据维度和模型效果。之后,再通过自动编码器(auto encoder, AE)对数据进行降维。

1 融合多特征的知识追踪模型

1.1 模型架构

该文提出的 DKTwMF 模型的整体架构如图 1 所示。模型首先通过 NMIRF 混合特征选择模块对输入数据进行处理,图中 a_i 表示模型的初始输入,包括数据集中可获得的全部特征数据,针对不同的数据集 a_i 代表的内容不尽相同。然后,对多个特征数据以及作答结果通过 MFE 进行编码。随后,使用 AE 对编码后的数据进行降维。最后,将编码结果 v_i 输入到神经网络进行训练,最终获得学生知识掌握状况的最终评估结果 y_i 。图中 h_0 表示神经网络的初始状态, y_i 的不同颜色代表不同的掌握程度,使用该模型能够同时对学生的所有知识点的掌握情况进行评估。

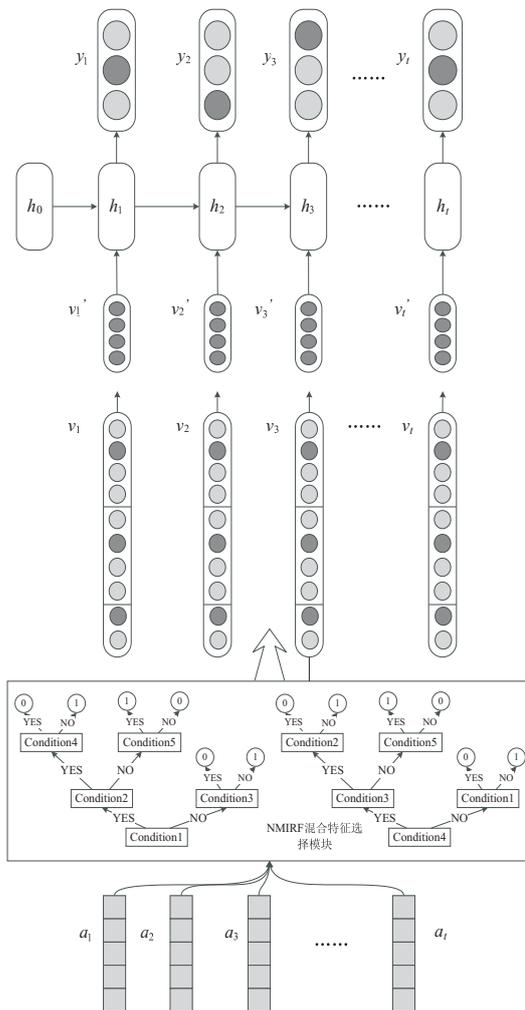


图 1 DKTwMF 知识追踪模型架构

1.2 NMIRF 混合特征选择算法

单一的特征选择算法都有各自的局限性,因此该文提出一种 NMIRF 混合特征选择算法。如图 2 所示,该算法首先使用方差选择过滤法对所有变量进行初步筛选,计算各个变量之间的方差,然后根据阈值(一般使用 0,即不存在变动的数据),选择方差大于阈值的特征,得到 k 个候选特征集合。然后,使用邻域互信息算法对计算各个特征之间的相关性。由于教育数据集中存在分类数据(如知识点 ID、题目 ID 等)和数值数据(如响应时间),因此使用邻域互信息能够很好地计算 k 个特征之间的相关性。最后,结合特征相关性使用随机森林对特征进行精选,得到该模块的最终结果。

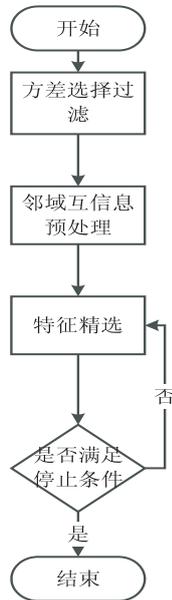


图 2 NMIRF 混合特征选择算法

评估变量特征重量性(variable importance measures, VIM)是随机森林的一个重要特征,常规计算方法有两种^[16-17]:(1)基于基尼指数的评估方法;(2)基于袋外数据(out of bag sets, OOB)分类准确率的评估方法。当数据中同时包含连续变量和分类变量时,基于基尼指数的评估方法比基于 OOB 分类准确率的评估方法效果差,因此,该文使用基于袋外数据分类准确率的变量重要性度量,即袋外数据变量特征值发生扰动后的分类正确率与未发生扰动的分类正确率的平均减少量。变量特征 x 重要性计算公式如下:

$$\text{vim}_x = \frac{1}{M} \sum_{i=1}^M (\text{ER}_{x_i}^{\text{ob}} - \text{ER}_{x_i}^{\text{ob}}) \quad (1)$$

其中, M 表示随机森林包含决策树的个数, x_i 表示第 i 个决策树中预评估特征变量 x 的取值, x_i' 表示对第 i 个决策树的预评估变量 x 进行扰动后对应的取值, $\text{ER}_{x_i}^{\text{ob}}$ 表示第 i 个分类树对应的袋外数据未发生扰动前在该树上的分类失误率, $\text{ER}_{x_i'}^{\text{ob}}$ 表示第 i 个分类树对应的袋外数据发生扰动后在该树上的分类失误率。

在特征精选模块,该文提出一种基于序列后向选择法的混合特征选择算法,即不断从特征子集中删除不重要的特征,直至达到截止条件。为更准确地评估变量的重要性,需要逐步剔除最不重要的变量,然后重新构建随机森林,并在新的随机森林上重新对待筛选特征子集中特征的重要性进行排序。数据集中可能存在相关性很高的特征,即使在其他变量(除待评估变量之外)特征不发生扰动,其他特征也可能对待评估变量特征重要性评估产生影响,因此该文结合相关性矩阵与随机森林决定是否剔除该特征。具体特征精选算法描述如下:

(1)初始化待选择特征子集 U ,使用邻域互信息计算各个特征之间的相关性矩阵 S 。

(2)根据数据集构建包含 M 棵决策树的随机森林,初始化决策树 $i=1$ 。

(3)计算特征子集中每个特征在第 i 棵决策树上的分类失误率 $\text{ER}_{x_i}^{\text{ob}}$ 。

(4)将袋外数据中特征 x 的取值打乱顺序,计算发生扰动后分类失误率 $\text{ER}_{x_i'}^{\text{ob}}$ 。

(5)对于 $i=2,3,\dots,M$,重复步骤(3)~(4),计算特征 x 的重要性 vim_x 。

(6)对待选择特征子集 U 中的每个特征重复(3)~(5),并对特征重要性进行排序。

(7)选择特征重要性排序中最不重要的特征 y ,相关性矩阵中与特征 y 相关性高于阈值 δ 的特征集合 W 。如果集合 W 为空,更新待选择特征子集 $U=U-y$,跳转至第(8)步;否则,对于 W 中任一特征,分别计算在使用 $U-W+w$ 集合时随机森林的袋外分数,最大袋外分数对应的特征记为 w ,更新待选择特征子集 $U=U-W+w$ 。

(8)根据新的特征子集重新构建随机森林,重复步骤(2)~(7),直到剩余两个特征,确定最终特征子集。

1.3 MFE 多特征编码

由 Piech 等人提出的 DKT 模型只考虑了知识点 ID 和作答结果,该文考虑加入更多的特征作为神经网络输入,为此提出多特征编码方案 MFE(multi feature encoding)。通过对多个单独特征的编码进行拼接,从而构建模型的输入。MFE 方案中的多特征编码构建公式如下:

$$C(s_i, c_i) = s_i + [\max(s) + 1] * c_i \quad (2)$$

$$v_i = O(C(s_i, c_i)) \oplus O(C(f_i, c_i)) \oplus O(f_i) \quad (3)$$

其中, $C()$ 表示使用交叉特征, $O()$ 表示 one-hot 编码, \oplus 表示将编码结果进行拼接。公式(2)、公式(3)中 s_i 代表知识点 ID, c_i 代表对应练习的作答结果(1 表示作答正确,0 表示作答错误), f_i 代表通过特征工程

模块选取的重要特征对应的取值。

交叉特征是将两个或多个特征通过一定规则编码成一个特征的方法。例如,两个单独特征的交叉特征编码方式如下:

$$C(f, g) = f + [\max(F) + 1] * g \quad (4)$$

$$C(g, f) = g + [\max(G) + 1] * f \quad (5)$$

式中, f 、 g 分别表示需要编码的两个特征, F 、 G 分别表示特征 f 、 g 的取值集合。文献[18]中指出使用交叉特征相比使用单独特征可以明显提高模型性能。如

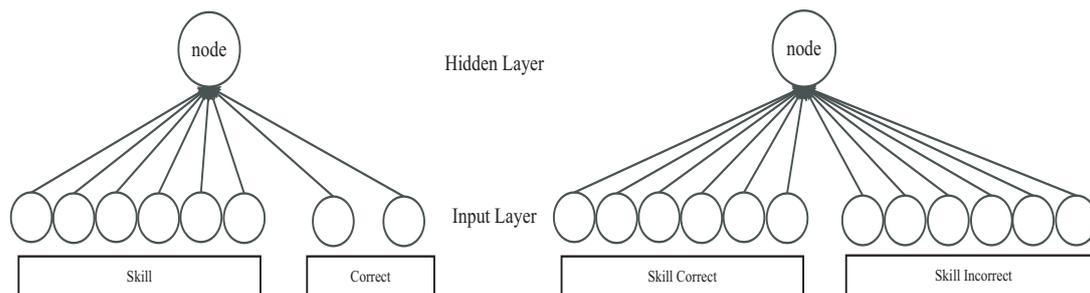


图 3 交叉特征

然而,使用交叉特征的缺点是输入向量的维数极速增加。随着维数的增加,模型需要训练的参数随之增加,这不仅会消耗大量的计算资源和时间,模型也很难收敛到全局最优。该文使用自动编码器将数据维度减少了四分之一。公式(6)中 v_i 表示原始编码 v_i 使用 AE 处理后得到的编码结果, W_{ae} 、 b_{ae} 分别表示通过 AE 训练得到的权重矩阵和偏差矩阵。

$$v_i' = \tanh(W_{ae} * v_i + b_{ae}) \quad (6)$$

2 实验与分析

2.1 数据集

为验证 DKTwMF 模型的有效性,该文使用三个教育相关的公开数据集进行了对比实验。这三个数据集分别来自不同的智能教育系统,每个系统都对学生练习的交互信息进行了记录。

Assistments 2009-2010: Assistments 是一个基于计算机的学习系统,可以同时评估学生的掌握水平和教师的授课水平。该文使用的数据集是从 Assistments skill builder 集合中收集的,这些集合是学生在类似的问题(同一个知识点)上不断进行练习,直到能正确地回答 n 个连续的问题(其中 n 通常是 3)。当系统认为学生掌握相应的技能后,学生通常不会重复练习同样的技能。与其他数据集不同,该数据集中学生的作答记录可能不连续,在实验中,通过预处理,将所有属于一个学生的记录连接在一起。

KDD 2010: 该数据集是 ACM 的知识发现和数据挖掘小组(knowledge discovery and data mining)在 2010 年发布的竞赛题目使用的数据集。该次竞赛以

图 3 所示,假设知识点数为 6 种,作答结果为 2 种(正确/错误)。如果不使用交叉特征,神经网络中每个隐藏节点与 6+2 个节点进行连接,对应 6+2 的连接权重。使用交叉特征神经网络中每个隐藏节点与 6*2 个节点进行连接,此时对于每个知识都有单独作答正确权重、作答错误权重与之对应,而不是所有知识点共用正确权重与错误权重。使用交叉特征而不是单独特征的关键原因是:神经网络中的权重不平衡,从而导致模型效果下降。

基于学生在线做题记录,预测学生认知能力(学习成绩)为目的。该数据集来自 2005 年至 2009 年卡内基学习认知诊断系统。数据集包含 574 名学生,437 个知识点,809 694 条学生作答记录。与 Assistments 平台不同的是,该平台记录了学生的详细操作数据,学生的每一个行为都可能是评估学生知识的依据。使用数据集中的 KC(knowledge component)作为知识点,对于多个 KC 组合的情况视为新的知识点。

OLI F2011: 开放学习计划(open learning initiative, OLI)是卡内基梅隆大学的一个计算机学习系统,该系统用于为在校学生提供在线课程。OLI 系统在学生许可的情况下收集了学生课堂学习记录和网络学习数据。该文使用 2011 年秋季学期大学工程力学的数据作为实验数据,包含 335 名学生,361 092 条学生练习记录。该系统详细记录了学生的操作情况,即学生在练习过程中的鼠标点击、翻页、保存等信息。实验中对数据记录进行了删减,只保留了学生的最终作答记录。

针对多知识点题目,不同的系统处理方法也不相同。其中,Assistments 中的数据对于多知识点题目,会生成多条记录。而 KDD 提供的数据只会生成一条记录,系统采用 KC(knowledge component)记录题目关联的知识点,知识点之间采用 ~ ~ 串联起来。例如,题目 Q_1 关联了知识点 k_1 、 k_2 , 题目 Q_2 关联了知识点 k_1 。假设学生 S_1 参与一次作答,其中 Q_1 作答错误, Q_2 作答正确。

Assistments 生成的记录如表 1 所示, KDD 生成的记录如表 2 所示。

表 1 Assistments 记录样例

Record ID	Student ID	Question ID	Skill ID	...	correct
1	S ₁	Q ₁	k ₁	...	0
2	S ₁	Q ₁	k ₂	...	0
3	S ₁	Q ₂	k ₁	...	1

表 2 KDD 记录样例

Record ID	Student ID	Question ID	KC	...	correct
1	S ₁	Q ₁	k ₁ ~ ~ k ₂	...	0
2	S ₁	Q ₁	k ₁	...	1

2.2 评价指标

模型的评价标准采用 AUC(area under the curve), 即 ROC 曲线线下面积。ROC 曲线每个点反映模型对同一信号刺激的感受性, 横轴为负正类率 (false positive rate, FPR), 纵轴为真正类率 (true positive rate, TPR)。对于二分类问题, 所有实例可以被划分为正类、负类。这里将作答结果为正确的归为正类, 作答结果为错误的归为负类, 对应的 FPR 取值为 $FP/(FP + TN)$, TPR 取值为 $TP/(TP + FN)$ 。

表 3 二分类 AUC 指标

实际	1	0
1	True Positive, TP	False Negative, FN
0	False Positive, FP	True Negative, TN

2.3 实验对比方法

该文将数据中的学生随机分成 5 组, 依次取其中一组作为测试数据集进行实验, 最终使用五次实验结果的平均值作为评价结果。通过五折交叉验证得到的实验数据更趋于稳定值, 结果能够更让人信服。实验使用 TensorFlow 中的 LSTM Cell 实现循环神经网络模型, 为消除其他因素的影响, 各模型均采用 Adam 梯度下降算法, 神经网络包含 200 个隐藏神经元, dropout 概率设置为 0.4。为保证得到的是稳定性结果, 实验采用五折交叉的方式进行验证, 模型评价标准采用 AUC^[11]。实验使用基本 DKT 作为基本模型, 分别通过 AE 降维、NMIRF 自动特征提取对模型进行扩充。一些研究人员发现通过适当的扩展, BKT 可以达到和 DKT 相近的性能^[3], 因此实验也与 BKT 及其扩展模型进行比较。更具体的对比模型介绍如下:

DKT: 该模型为基本模型, 如文献[11]所述, 即只对 skill ID 和 correct 进行 one-hot 编码, 然后使用编码数据对 LSTM 网络进行训练。

DKT+AE: 在基本 DKT 模型的基础上, 使用自动编码器对 DKT 模型的输入进行预处理, 降低输入维度。

DKT+NMIRF: 首先使用 NMIRF 算法对数据集进

行筛选, 然后将筛选后的多特征数据使用 MFE 编码, 最后使用编码结果作为循环神经网络输入, 并对网络进行训练。

DKTwMF: 该文提出的模型, 将经过 NMIRF 筛选的多特征数据使用 MFE 算法进行编码, 然后使用 AE 进行预处理, 最后使用 LSTM 神经网络进行训练。

BKT: 经典贝叶斯知识追踪模型, 该模型中知识一旦被掌握就不会遗忘。

BKT+F: 在经典 BKT 模型的基础上扩展了知识遗忘属性, 该模型中知识可以在掌握和未掌握之间相互转化, 该项扩展有助于对于同一知识的预测。

BKT+S: 在经典 BKT 的基础上扩展了技能发现。经典 BKT 对各知识点单独建模, 未对技能之间的交互进行建模, 该模型将独立的 BKT 相互关联起来。

BKT+A: 在经典 BKT 模型的基础上扩展了学生潜在能力属性, 针对不同的学生设置个性化的猜测、失误概率, 该项扩展有助预测学生从一个知识到另外一个知识点的能力。

BKT+FSA: 该模型同时扩展了知识遗忘、技能发现、潜在能力。

2.4 结果分析

本节对在不同数据集下, 不同模型的 AUC 结果进行比较分析, 目的在于通过实验验证 DKTwMF 模型效果。

对于 Assistments 数据集, 各模型的 AUC 评价结果如图 4 所示。

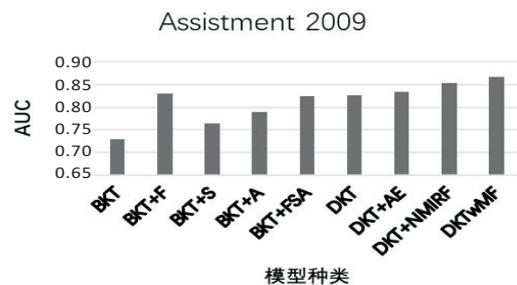


图 4 Assistments 数据集 AUC 结果

从图 4 中可以发现, 基本 BKT 的 AUC 值仅有

0.73,通过不同的扩展方式模型效果均得到了不同程度的提升;扩展知识遗忘之后的模型 BKT+F 达到了略优于基本 DKT 的效果,AUC 值达到 0.83;BKT+FSA 由于过多参数导致该模型效果略差于 BKT+F。从图 5 中可以发现,通过加入多特征(DKT+NMIRF)、对输入进行降维(DKT+AE)都能在一定程度上提升原模型 DKT 的效果;DKTwMF 在 Assistments 数据集上达到了最优的结果 0.867,达到所有模型中最优效果。

对于 KDD 数据集,各模型的 AUC 评价结果如图 5 所示。

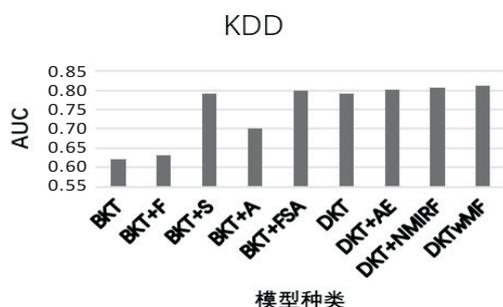


图 5 KDD 数据集 AUC 结果

从图 5 中可以发现,BKT 在该数据集上仅达到 0.62 的 AUC 值,BKT+F 相比基本 BKT 模型效果没有明显提升,扩展知识遗忘后 AUC 值仅增加 0.01;在 BKT 系列模型中,BKT+FSA 达到最好的预测效果,并且 AUC 值稍大于基本 DKT 模型;DKT 模型的 AUC 值为 0.79,在扩展多特征后 DKT+NMIRF 达到与 BKT+FSA 相当的预测效果。不同于 Assistments 数据集,在该数据集上 DKTwMF 模型相比基本 DKT 模型虽然没有明显的效果提升,但仍然达到了所有模型中的最优效果。

对于 OLI 数据集,各模型的 AUC 评价结果如图 6 所示。

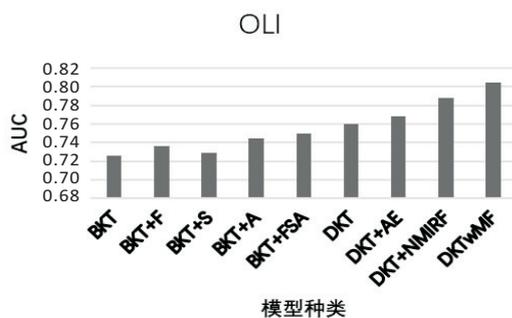


图 6 OLI 数据集 AUC 结果

从图 6 中可以发现,BKT 系列模型中 AUC 值最好的是 BKT+FSA,但是该模型仍不如未经改进的 DKT 模型。而 DKTwMF 模型的 AUC 值相比 DKT 模型又增加了 0.04,DKTwMF 模型相比 BKT 模型效果

提升了 11%。

通过上述实验可以发现:对于 BKT 系列模型,针对不同数据集不同模型的表现效果大不相同,想要得到适合该数据集的 BKT 模型需要不断的尝试验证。虽然 BKT+FSA 能够达到不错的效果,但是过多的参数可能导致效果下降(Assistments 数据集)。对于 DKT 系列模型,可以发现通过加入多特征输入、数据降维都可以在不同程度上提升模型效果。DKTwMF 模型同时具备了这两种优化方法,因此模型具有更高的 AUC 值、能够更准确地预测学生表现情况。无论是多知识点生成多条记录(Assistments)的场景,还是多知识点生成一条记录(KDD)的场景,该模型均能达到不差于其他模型的表现。并且,相比于 BKT 系列模型,该模型不需要专家知识的指导就可以完成建模,建模效果优于一般手动选择特征。综上所述,该文提出的模型能够智能、高效、准确地对学生知识水平进行建模。

3 结束语

相比其他知识追踪方法,提出的模型主要有以下优点:(1)智能性。通过 NMIRF 自动对特征进行选择,不需要专家的领域知识的指导。模型不需要知识之间的关系图,模型中的神经网络可以通过数据自动学习各知识之间联系;(2)精准性。通过上述实验证明,DKTwMF 模型相比其他模型能够更准确地对学生的知识掌握状态进行建模;(3)高效性。DKTwMF 模型采用 NMIRF 算法对特征进行自动选择,大大节省了手动特征选择然后验证的时间。同时,使用 AE 降低数据维度也能降低模型的训练时间。但目前,DKTwMF 模型也有以下不足:(1)由于模型采用深度学习思想,需要大量数据对模型进行训练,因此该模型不适用于少量数据的情况;(2)可解释性不足。BKT 模型采用概率的方式表示学生的知识掌握状态,每个知识点都对应明确的函数表示,而神经网络会自动学习神经元之间的相互作用,没有明确的函数表明学生对每个知识的掌握程度。因此,针对这些不足,未来还有待进一步完善。

参考文献:

- [1] CZERNIEWICZ L, DEACON A, GLOVER M, et al. MO-OC—making and open educational practices[J]. Journal of Computing in Higher Education, 2016, 29(1): 81–97.
- [2] 沈宏兴,郝大魁,江婧婧.“停课不停学”时期在线教学实践与疫后在线教学改革的思考——以上海交通大学为例[J]. 现代教育技术, 2020, 30(5): 11–18.
- [3] KHAJAH M, LINDSEY R V, MOZER M C. How deep is knowledge tracing? [C]//9th international conference on

- education data mining. [s. l.] : [s. n.], 2016.
- [4] 李 婷,傅钢善. 国内外教育数据挖掘研究现状及趋势分析[J]. 现代教育技术,2010,20(10):21-25.
- [5] 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.
- [6] CORBETT A T. Cognitive computer tutors: solving the two-sigma problem[C]//User modeling 2001, international conference. Sonthofen, Germany: [s. n.], 2001:137-147.
- [7] CORBETT A T, ANDERSON J R. Knowledge tracing: modeling the acquisition of procedural knowledge[J]. User Modeling and User-Adapted Interaction, 1995, 4:253-278.
- [8] 刘恒宇,张天成,武培文,等. 知识追踪综述[J]. 华东师范大学学报:自然科学版,2019(5):1-15.
- [9] BAKER R S J D, CORBETT A T, ALEVEN V. More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing[C]//Intelligent tutoring systems. Montreal, Canada: Springer, 2008:406-415.
- [10] PARDOS Z A, HEFFERNAN N T. Modeling individualization in a Bayesian networks implementation of knowledge tracing[C]//User modeling, adaptation, and personalization. Big Island, HI, USA: Springer, 2010:255-266.
- [11] PIECH C, SPENCER J, HUANG J, et al. Deep knowledge tracing[C]//Advances in neural information processing systems. Montreal: [s. n.], 2015:505-513.
- [12] ZHANG L, XIONG X, ZHAO S, et al. Incorporating rich features into deep knowledge tracing[C]//ACM conference on learning. Cambridge, Massachusetts, USA: ACM, 2017:169-172.
- [13] MINN S, YI Y, DESMARAIS M C, et al. Deep knowledge tracing and dynamic student classification for knowledge tracing[C]//2018 IEEE international conference on data mining (ICDM). Singapore: IEEE, 2018.
- [14] SU Y, LIU Q, HUANG Z, et al. Exercise-enhanced sequential modeling for student performance prediction[C]//Thirty-second AAAI conference on artificial intelligence. New Orleans: AAAI, 2018:2435-2443.
- [15] 李 振,周东岱,王 勇. “人工智能+”视域下的教育知识图谱:内涵、技术框架与应用研究[J]. 远程教育杂志, 2019, 37(4):42-53.
- [16] GOLDSTEIN B A, POLLEY E C, BRIGGS B. Random forests for genetic association studies[J]. Stat Appl Genet Mol Biol, 2011, 10(1):1-34.
- [17] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2000, 29(5):1189-1232.
- [18] YU H, LO H, HSIEH H. Feature engineering and classifier ensemble for KDD cup 2010[C]//JMLR: workshop & conference. Hyderabad: JMLR, 2010:1-12.
- +++++
- (上接第34页)
- tion with deep recurrent neural networks[C]//2013 IEEE international conference on acoustics, speech and signal processing. Vancouver, BC, Canada: IEEE, 2013:6645-6649.
- [16] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder - decoder approaches [C]//Proceedings of SSST - 8, eighth workshop on syntax, semantics and structure in statistical translation. Doha, Qatar: ACL, 2014:103-111.
- [17] DEY R, SALEMT F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]//2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). Boston, MA, USA: IEEE, 2017:1597-1600.
- [18] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [19] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C]//Proceedings of the eighteenth international conference on machine learning. San Francisco, CA, USA: Morgan Kaufmann, 2001:282-289.
- [20] WARD R, WU X, BOTTOU L. AdaGrad stepsizes: sharp convergence over nonconvex landscapes [C]//International conference on machine learning. Long Beach, California, USA: PMLR, 2019:6677-6686.
- [21] WEBBER J. A programmatic introduction to neo4j [C]//Proceedings of the 3rd annual conference on systems, programming, and applications: software for humanity. Tucson, Arizona, USA: ACM, 2012:217-218.
- [22] 王勇超,罗胜文,杨英宝,等. 知识图谱可视化综述[J]. 计算机辅助设计与图形学学报, 2019, 31(10):1666-1676.