序列特征与学习过程融合的知识追踪模型

李子杰[1]，周菊香[1,2*]，韩晓瑜[1]，甘健侯[1,2]，鹿泽光[1,3]，王俊[1,2]

（1. 云南师范大学民族教育信息化教育部重点实验室，云南 昆明 650500；
2. 云南师范大学云南省智慧教育重点实验室，云南 昆明 650500；3. 中科国鼎数据科学研究院，北京 100085）

摘　要：知识追踪是人工智能技术与教育相结合的新兴领域，旨在通过学生过去完成习题的交互序列对学生的知识状态进行评估，是实现大规模个性化学习服务的关键核心技术。随着深度学习在计算机视觉、自然语言处理、推荐系统等领域的广泛应用，知识追踪领域也出现了大量基于神经网络的方法，简称深度知识追踪（DKT）模型。针对目前已有 DKT 模型在可解释性和准确性方面的不足，提出一种序列特征与学习过程融合的知识追踪模型 SLKT，模型包括知识状态模块、序列特征模块、预测模块。知识状态模块用以模拟学生学习过程，序列特征模块捕捉学习者近期学习状况。通过序列特征和学习过程的融合，有效解决了基于知识状态建模方法无法考虑学习者近期学习状况的问题，同时提出一种带约束的动态 $\boldsymbol{Q}$ 矩阵表示练习和知识点之间的关系，从而更好地进行学习者学习过程建模，在确保较好可解释性的同时有效提升模型的准确性。在 3 个知识追踪领域公共数据集上的实验结果表明，相比 DKT、动态键值记忆网络（DKVMN）、自注意力的知识追踪（SAKT）、卷积知识追踪（CKT）等深度追踪模型，SLKT 模型在曲线下面积（AUC）、准确率指标评估中表现较优。

关键词：智慧教育；深度学习；知识追踪；学习过程建模；$\boldsymbol{Q}$ 矩阵

中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0067528

Knowledge Tracing Model Based on the Fusion of Sequence Features and Learning Processes

LI Zijie[1], ZHOU Juxiang[1,2*], HAN Xiaoyu[1], GAN Jianhou[1,2], LU Zeguang[1,3], WANG Jun[1,2]

(1. Key Laboratory of Educational Informatization for Nationalities, Ministry of Education, Yunnan Normal University, Kunming 650500, Yunnan, China;
2. Yunnan Key Laboratory of Smart Education, Yunnan Normal University, Kunming 650500, Yunnan, China;
3. Guoding Institute of Data Science, Beijing 100085, China)

【Abstract】 Knowledge tracing is a novel field that merges artificial intelligence and education with the aim of evaluating the knowledge state of students through an interactive sequence of exercises that they have completed in the past. This is a core technology for implementing large-scale personalized learning services. With the widespread application of deep learning in computer vision, Natural Language Processing (NLP), recommendation systems, and other fields, a significant number of neural network-based methods, known as Deep Knowledge Tracing(DKT) models, have emerged in the field of knowledge tracing. To address the shortcomings of existing DKT models in terms of interpretability and accuracy, this study proposes a knowledge tracing model called SLKT that integrates sequence features with learning processes and includes knowledge state, sequence feature, and prediction modules. The knowledge state module is used to simulate the learning process of students, and the sequence feature module captures their recent learning status. Through the integration of sequence features and the learning process, the SLKT model effectively solves the problems of knowledge state modeling methods that cannot consider learners′ recent learning status. A dynamic $\boldsymbol{Q}$ matrix with constraints is also proposed to represent the relationship between exercises and knowledge points to better model the learning process of the learner. This ensures better interpretability, and effectively improves model accuracy. In this study, the superior performance of the proposed model is verified using the Area Under Curve (AUC) and accuracy metrics, comparing multiple sets of experiments using depth tracing models such as DKT, Dynamic Key-Value Memory Network (DKVMN), Self-Attentive Knowledge Tracing (SAKT), and Convolutional Knowledge Tracing (CKT) on public datasets in three knowledge tracing fields.

【Key words】 smart education; deep learning; knowledge tracking; learning process modeling; $\boldsymbol{Q}$ matrix

收稿日期：2023-04-30　　**修回日期：**2023-10-07

基金项目：国家自然科学基金（62266054）；云南省教育厅科学研究基金资助（2022Y180）。

通信作者 E-mail：* zjuxiang@126.com

计算机工程 www.ecice06.com

0 引言

知识追踪是智慧教育场景下最重要的算法之一,其拥有追踪学生知识状态的能力[1],这种特性使得知识追踪被广泛应用到各种智能教学系统和 MOOC 中。凭借追踪学生的知识状态,平台与教师通过分析学生的知识水平与其变化判断学生当前的学习状态并分析存在的问题,从而评估该学生目前的学习需求并为该学生制定未来学习规划,最终实现个性化学习。知识追踪作为大规模个性化学习服务平台构建中不可或缺的一环[2],大量研究人员将目光集中在该研究领域,目前已经形成稳定的研究方向和专用的数据集。

随着深度学习的广泛应用,尤其是在自然语言处理与计算机视觉多种任务上取得的惊人突破,神经网络已经成为解决人工智能问题的主流方法。深度学习的效果主要取决于目标任务的复杂程度与训练神经网络的数据体量。因此,随着在线教育的发展和人们对教育数据的挖掘愈发重视,深度学习应用到教育领域已经成为一种风向[3],包括课堂表情识别[4]、辍学预测[5]、学习资源推荐[6]在内的多项人工智能的教育应用已经广泛采用了神经网络的方法。深度知识追踪是深度学习在知识追踪的首次尝试,其使用循环神经网络(RNN)或长短时记忆(LSTM)网络作为基础结构,通过学生交互序列学习学习者特征,进而预测结果,深度知识追踪(DKT)[7]在知识追踪领域取得了巨大的提升,在 Simulated-5、Khan Math 等数据集上相比基于贝叶斯方法的 BKT 模型在曲线下面积(AUC)指标上拥有超过 15%的提升。该值的提升使得知识追踪领域全面进入了深度学习时代,国内研究者也积极推进该领域的研究[8-10]。

通常,知识追踪的效果主要取决于对交互序列的学习与模拟学生的学习行为。基于交互序列可以利用学习序列的隐藏状态以预测结果,如 DKT、基于上下文注意的知识追踪(AKT)[11]等,而动态键值记忆网络(DKVMN)[12]、DKT+Forget[13]等模型则通过模拟学生的学习行为构建学生的知识点状态从而预测结果。然而,当前方法常常基于其中一种,可能存在如下问题:1)基于学习过程的方法通过建模知识状态来预测结果,但无法通过知识状态重现交互序列,因此对序列的短期依赖并不敏感;2)基于序列的方法通常使用 RNN 等支持序列输入的模型构建,这种通用结构像是黑盒,无法可解释地掌握学生具体的知识状态,同时也导致性能不佳。

同时,本文发现知识点状态的改变取决于每次学生的做题情况,但现有模型并没有根据已掌握的知识建立,从而导致知识点间的学习迁移性不足。这类方法仅使用学习者掌握的知识点状态进行预测,无法模拟学生对题目类型的偏好、猜测等非知识能力的影响因素。为了解决上述问题,本文设计序列特征与学习过程融合的知识追踪模型 SLKT,包含知识状态模块、序列特征模块和预测模块 3 个部分。该模型能有效解决基于知识状态建模方法无法考虑学习者近期学习偏好的问题,同时提出一种带约束的动态 $\boldsymbol{Q}$ 矩阵表示练习和知识点之间的关系,从而更好地对学习者学习过程进行建模,在确保较好可解释性的同时有效提升模型的准确性。

1 相关工作

1.1 知识追踪

知识追踪是实现教育智能化的重要工具之一。目前知识追踪模型主要分为传统的知识追踪模型[14-15]和基于深度学习的知识追踪模型[16]。本文主要研究基于深度学习的知识追踪模型,对传统的知识追踪模型将不再赘述。目前,基于深度学习的知识追踪研究主要分为 2 个方向:1)将学生的做题表现作为预测学生知识点掌握状态的唯一依据,通过降低预测值和实际值之间的损失来优化模型,该类模型更加注重序列预测,即通过数据的波动状态来获取数据的特征,以达到精准预测的效果;2)通过模拟学生的学习过程来提升模型的准确率,该类模型一味地追求模型的精确率,导致模型建模的学生知识掌握情况和学生真实学习过程不符,从而使得模型难以达到更高层次的精确率。目前,基于深度学习的知识追踪模型的 2 类研究方向都不同程度地取得了新的进展[17-20]。

1.2 捕捉序列特征的知识追踪模型

捕捉序列特征的知识追踪模型主要有 DKT、深度知识追踪的改进 DKT+[21]、自注意力的知识追踪(SAKT)[22]、基于 Transformer 的知识追踪(SAINT)[23]以及对 SAINT 进行改进的 SAINT+[24]等。2016 年,研究人员提出的 DKT 模型是第 1 个将深度神经网络应用到知识追踪模型,相较于传统的知识追踪模型在预测准确率上取得了巨大的提升。DKT 模型将学生以往的信息存储在 RNN 中,通过 RNN 的记忆能力以达到预测学生知识掌握程度的效果。DKT+模型是针对 DKT 存在无法重构观测到的输入和预测状态波动 2 个问题提出的,通过加入正则化项解决以上问题。SAKT 针

计算机工程 www.ecice06.com

对以往模型难以处理稀疏数据的问题，提出使用自注意力机制计算对于当前预测真正有用的数据，以达到使用较少数据就能得到较优的准确率。基于 Transformer 的 SAINT 模型是针对以往模型无法捕捉练习和学习者作答情况之间复杂关系而提出的，将练习和学习者作答情况分别输入到 Transformer 的编码器和解码器。SAINT＋是对 SAINT 模型进行改进的，整合 2 个时间特征嵌入来增强 SAINT。该模型认为学生回答问题的时间和相邻学习活动之间的时间间隔对模型的预测也有影响。SAINT 和 SAINT＋在目前公开的最大教育基准数据集 EdNet 上验证是有效的。

1.3 模拟学生学习过程的知识追踪模型

模拟学生学习过程的知识追踪模型主要使用 DKVMN 的知识追踪、基于顺序键值记忆网络(SKVMN)的知识追踪[25]、基于 AKT 以及学习过程一致的知识追踪(LPKT)[26]等。DKVMN 模型是针对基于深度学习的知识追踪模型缺乏可解释性的问题提出的，利用 1 个静态矩阵 **key** 和 1 个动态矩阵 **value** 分别存储知识概念和更新相应概念的掌握水平，使得该模型可以直接输出每个概念的掌握程度。SKVMN 模型是在 DKVMN 的基础上进行改进的。DKVMN 虽然为基于深度学习的知识追踪模型的可解释性做出了重大突破，但是该模型是通过最近的练习获得知识增长，因此无法捕捉练习序列中长期依赖关系。SKVMN 利用 Hop-LSTM 解决这一问题，并改进了 DKVMN 中写的过程。Hop-LSTM 与注意力有异曲同工之妙，Hop-LSTM 是通过计算练习的相似性来捕获重要信息，使较远的重要信息也可以得到重视，从而解决 DKVMN 无法捕捉练习序列中长期依赖关系的问题。AKT 受认知科学关于遗忘机制的启发，研究人员提出一种新的单调注意力机制，利用指数衰减曲线来降低序列中较早时刻练习的重要性。文献[26]认为只有模型真正模拟学生的学习过程，才能在准确率上有新的突破。LPKT 由 3 个部分组成，分别为学习模块、遗忘模块和预测模块，并利用时间信息模拟学习与遗忘过程，取得较优的效果。

从上述捕捉序列特征的知识追踪模型中可以看出，该系列的模型更加注重对序列特征的捕获，无论是利用全部的序列还是利用自注意力机制捕捉相关性高的数据序列，都是利用模型捕捉学生之前的习题序列数据特征，以达到预测学生知识点掌握程度的目的。从模拟学生学习过程的知识追踪模型中可以看出，该系列的模型更加注重对学生学习过程的探索，例如学生在学习时会经历哪些过程、学生的学习会受哪些因素的影响、如何更好地模拟学生学习过程等。近年来，知识追踪在这 2 个研究倾向上都取得了一定的进展，但将这 2 种方法结合的研究相对较少。

2 预备知识

2.1 问题定义

知识追踪的目标是根据学习者过去的交互序列预测学生在未来交互中的表现，其中，给定某一学习者交互序列 $s \in S$，在 t 时刻已发生的交互表示为 $s=[r_0,r_1,\cdots,r_t]$，其中每个交互记录了学习者进行答题时的练习编号及学习者是否作答正确，即 $s=[r_0,r_1,\cdots,r_t]=[(e_0,a_0),(e_1,a_1),\cdots,(e_t,a_t)]$，$e \in E$，$a \in \{0,1\}$。知识追踪的目标是通过该序列 s 预测下一时刻进行练习 e_{t+1} 时，学生的作答情况 a_{t+1}。

2.2 符号描述

本文中，使用以下符号对模型中的重要变量进行描述，如表 1 所示。

表 1 变量符号及其描述

Table 1 Variable symbols and their descriptions

符号	描述
$e_t \in E$	t 时刻学习者所进行的练习
$a_t \in \{0,1\}$	t 时刻学习者练习的作答是否正确
R_t	t 时刻学习者的交互，包括了 e_t 和 a_t
E	所有练习的集合，总个数为 $\lvert E\rvert$
C	所有概念(知识点)的集合，总个数为 $\lvert C\rvert$
$\boldsymbol{Q} \in \mathbb{R}^{\lvert E\rvert \times \lvert C\rvert}$	练习和知识点之间的关系矩阵
$\boldsymbol{q}_t \in \mathbb{R}^{1\times \lvert C\rvert}$	t 时刻学习者所进行的练习与各知识点之间的关系
$\boldsymbol{K}_t \in \mathbb{R}^{1\times \lvert C\rvert}$	t 时刻学习者的知识状态，每个知识点由 D 维向量表示
$\Delta \boldsymbol{K}_t^{e_t} \in \mathbb{R}^D$	t 时刻从 r_t 中获得的 e_t 对应概念的知识状态增量
$\Delta \boldsymbol{K}_t \in \mathbb{R}^{\lvert C\rvert \times D}$	t 时刻从 r_t 中获得所有概念的知识状态增量
D	嵌入维度

3 本文模型

本文模型的目标是使模型能够建模学生的知识状态，从而模拟学生在完成一系列练习时的学习变化情况，在此基础上，利用最近几次交互帮助模型预测最终的结果，使得模型能够掌握学生的近期情况。本文模型主要包括知识状态模块、序列特征模块、预测模块，其结构如图 1 所示。在学生每次作答后，通过知识状态模块对学生掌握的知识进行更新，同时

计算机工程 www.ecice06.com

将最近几次交互输入到 1 个由 Self-Attention 构成的序列特征模块，以得到学生的近期表现，最后利用学生掌握的知识状态和近期表现预测学生在下一时刻的作答情况。因此，该设计的优势及特点体现在以下方面：

1）具备完整的学生知识能力建模过程，通过记忆组件、遗忘组件对学习者的知识状态进行更新，该模块模拟了学习者学习的真实过程，记忆组件实现增益效果，遗忘组件模拟知识的遗忘现象，配合动态 $\boldsymbol{Q}$ 矩阵实现精准的知识点状态更新。

2）仅依靠学生知识点状态进行知识追踪存在无法掌握学生近期学习偏好的问题（如学习状态波动、完成练习时的非知识因素、其他学习活动带来的知识变化），将序列特征模块作为补充，隐式地对学生近期学习序列进行表征，解决单模块下无法应对学生状态波动的问题。

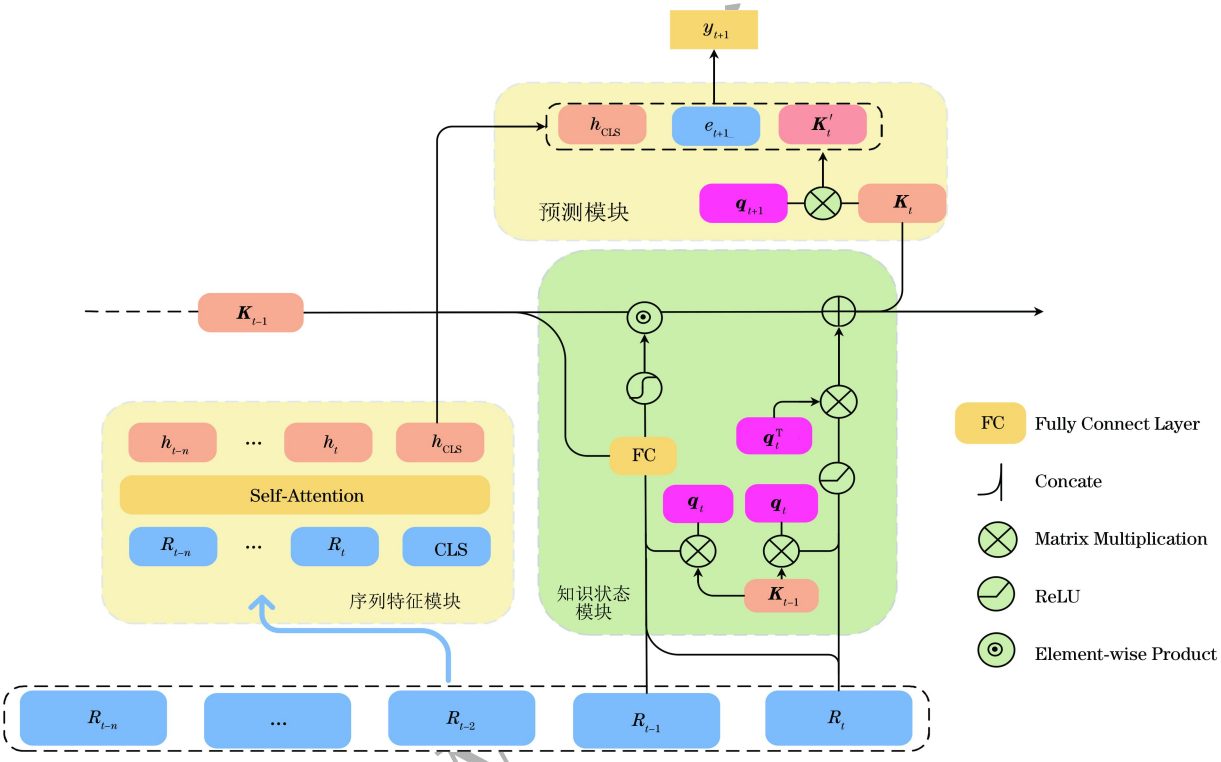


图 1 基于序列特征与学习过程相融合的知识追踪模型

Fig. 1 Knowledge tracing model based on the fusion of sequence features and learning processes

3.1 知识状态模块

知识状态模块的作用是随着学习者交互的进行，不断更新学习者的知识状态，该模块的输出会被输入到预测模块中参与计算最终预测结果。该模块的输入是学生最近 2 次的交互和上一时刻的知识状态 $\boldsymbol{K}_{t-1}$，输出为 t 时刻进行交互后的知识状态。同时，与 RNN、LSTM、GRU 类似，知识状态模块也遵循记忆门和遗忘门的设计方法。该模块维护了 1 个表示学生每个知识点掌握情况的知识状态 $\boldsymbol{K}_t$ 以及 1 个表示练习和知识点间关系的 $\boldsymbol{Q}$ 矩阵。

3.1.1 动态且带有约束的 $\boldsymbol{Q}$ 矩阵

$\boldsymbol{Q}$ 矩阵描述了练习和知识点之间的关系，即 $\boldsymbol{Q} \in \mathbb{R}^{|E|\times|C|}$。与现有的工作不同，本文对 $\boldsymbol{Q}$ 矩阵进行了一些改动使其更加符合认知，称为动态且受约束的 $\boldsymbol{Q}$ 矩阵。在过去的工作当中，$\boldsymbol{Q}$ 矩阵是 1 个由 0 和 1 组成的矩阵，根据数据情况不同，每个题目将被 1 个或多个知识点关联，矩阵中存在关联关系的位置则置为 1，无关联则为 0。

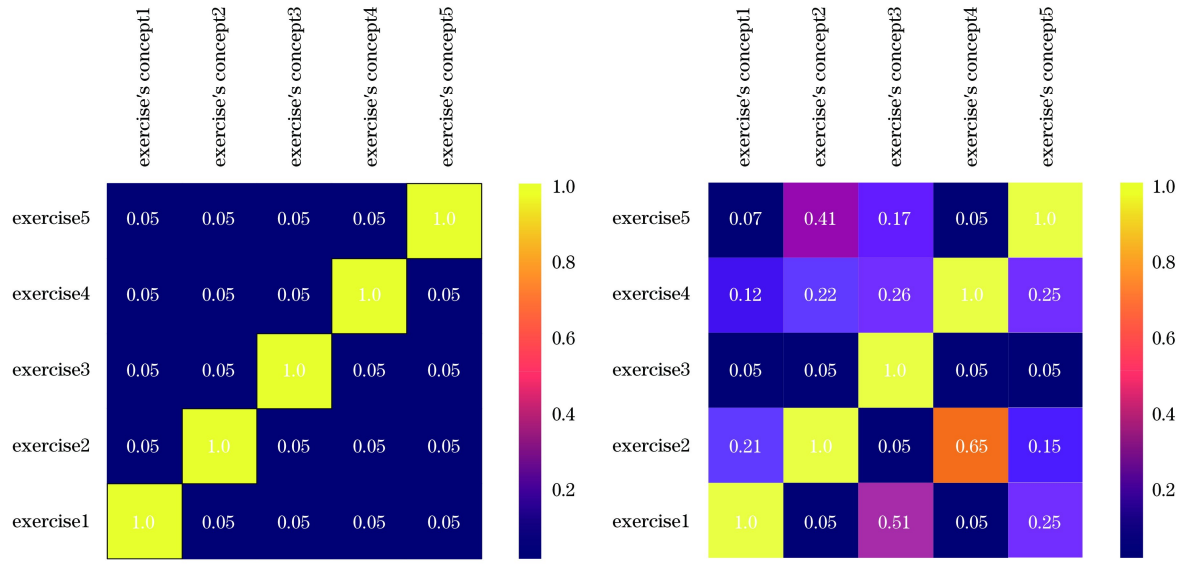
静态 $\boldsymbol{Q}$ 矩阵与动态 $\boldsymbol{Q}$ 矩阵部分习题与其知识点的关系热力图如图 2 所示（彩色效果见《计算机工程》官网 HTML 版）。目前在主流的知识追踪数据集中每个练习只关联了 1 个知识点，如 ASSISTments 系列数据集，使得 $\boldsymbol{Q}$ 矩阵通常是稀疏矩阵。这种设定存在 1 个问题：1 个习题与知识点的联系不应该是完全独立的，习题与非关联知识点间的关系不应该设定为 0，在实际使用中 $\boldsymbol{Q}$ 矩阵无关联的位置常常设置为 1 个较小值，如图 2(a)所示。但该参数也不应该被武断地设定为 1 个较小的固定值（如 0.05 和 0.1），这样虽然使得习题与每个知识点都产生了关联，但仍没有凸显出每个习题的特殊性，关联了同一知识点的不同习题在 $\boldsymbol{Q}$ 矩阵中的参数仍然一致。因此，$\boldsymbol{Q}$ 矩阵中的参数应该由算法进行动态的计算。同时，参数的取值范围应加以限制，$\boldsymbol{Q}$ 矩阵中的参数表示了习题与知识点间的关联性，因此参数应介于[0,1]之间。同时，为

www

了保证每个练习在学习时对所有知识点均产生一定影响，仍然保留较小值 γ，因此将参数范围约束在 $[\gamma,1]$。这种约束使得 $\boldsymbol{Q}$ 矩阵更符合客观规律，其训练后的 $\boldsymbol{Q}$ 矩阵如图 2(b)所示。因此，本文将 $\boldsymbol{Q}$ 矩阵设定为 1 个可训练参数的矩阵，称为 $\boldsymbol{Q}$ 矩阵的动态性。同时，在训练和推理过程中限制 $\boldsymbol{Q}$ 矩阵中的取值范围，使得 $\boldsymbol{Q}$ 矩阵约束在某一合理范围，称为 $\boldsymbol{Q}$ 矩阵的约束性，即：

$$\boldsymbol{Q}_{ij}=\begin{cases}1, \boldsymbol{Q}_{ij}\geqslant 1\\ \gamma, \boldsymbol{Q}_{ij}\leqslant \gamma\end{cases} \tag{1}$$

式中：γ 为习题与知识点的最小相关度，可设为 0.05、0.06。



(a)静态$\boldsymbol{Q}$矩阵　　(b)动态$\boldsymbol{Q}$矩阵

图 2　静态 Q 矩阵与动态 Q 矩阵部分习题与其知识点的关系热力图

Fig. 2　Thermodynamic diagram of the relationship between static Q matrix and dynamic Q matrix exercises and their concept

3.1.2　记忆组件

记忆组件的基本假设是学生的学习收益与其作答情况相关，每次交互的发生(无论学生的表现如何)都会为学习者带来收益(对应 LSTM 中的记忆门)，即学习者知识的增量与 r_t 相关。同时，该收益和学生现有的知识状态相关，即学习者知识的增量与 $\boldsymbol{K}_{t-1}$ 相关。当学生已经掌握某一知识点时，学生再做该知识点相关的题目得到的收益则会减少。同时，知识点间存在迁移的情况，当学习者在其他知识点方面获得增长时，某一知识点的情况也会随之增长。因此，当 $\boldsymbol{K}_{t-1}$ 作为上一时间步的结果输入时，首先计算学生关于 e_t 的知识点状态 $\boldsymbol{K}_{t-1}^{e_t}\in\mathbb{R}^D$：

$$\boldsymbol{K}_{t-1}^{e_t}=\boldsymbol{q}_t\times\boldsymbol{K}_{t-1} \tag{2}$$

式中：$\boldsymbol{q}_t\in\mathbb{R}^{1\times|C|}$ 描述了 t 时刻练习与所有知识点之间的关系。计算从 r_t 中习得的知识 $\Delta\boldsymbol{K}_t^{e_t}\in\mathbb{R}^D$：

$$\Delta\boldsymbol{K}_t^{e_t}=\mathrm{ReLU}(\boldsymbol{w}_1^{\mathrm{T}}[\boldsymbol{K}_t^{e_t}\oplus r_t]+b_1) \tag{3}$$

式中：$\boldsymbol{w}_1^{\mathrm{T}}$ 和 b_1 分别是神经网络层的权重和偏置，并且采用 ReLU 将学习的收益限制为 $\Delta\boldsymbol{K}_t^{e_t}\geqslant 0$；$\oplus$ 表示 Concate 操作。最后利用 $\Delta\boldsymbol{K}_t^{e_t}$ 计算各知识点的收益 $\Delta\boldsymbol{K}_t$：

$$\Delta\boldsymbol{K}_t=\boldsymbol{q}_t^{\mathrm{T}}\times\Delta\boldsymbol{K}_t^{e_t} \tag{4}$$

式中：$\Delta\boldsymbol{K}_t\in\mathbb{R}^{|C|\times D}$ 描述了在学习 r_t 后，知识状态的增加量。

3.1.3　遗忘组件

遗忘组件模拟了学生在学习过程中的遗忘现象，主要功能是将上一时间步的知识状态进行一定比率的遗忘。该组件以多个和遗忘过程相关的变量作为神经网络层的输入，再使用 Sigmoid 控制结果位于 $[0,1]$ 区间中，达到动态计算遗忘比率的效果，该比率最终与上一时间步的知识状态共同计算出长期记忆传递到下一时间步时的衰减量，这与 LSTM 中的遗忘门设计相似。本文认为遗忘的过程和相邻 2 个交互有关系，如果学生在交互过程中递进地完成了相似的题目，那么知识状态会比较稳定，不断进行练习巩固了知识点，遗忘现象不显著；相反，如果学生在进行交互时习题的差异较大，可以认为学生在进一步学习新的内容，此时遗忘现象会更加明显。具体来说，在模型中使用 r_t、r_{t-1}、$\boldsymbol{K}_{t-1}$ 计算知识状态 $\boldsymbol{K}_{t-1}$ 的衰减量，首先计算学生关于 e_t 的知识点状态 $\boldsymbol{K}_{t-1}^{e_t}\in\mathbb{R}^D$：

$$\boldsymbol{K}_{t-1}^{e_t}=\boldsymbol{q}_{t-1}\times\boldsymbol{K}_{t-1} \tag{5}$$

然后计算衰减比率 μ：

$$\mu=\sigma(\boldsymbol{w}_3^{\mathrm{T}}[\boldsymbol{K}_{t-1}\oplus(\boldsymbol{w}_2^{\mathrm{T}}[\boldsymbol{K}_{t-1}^{e_t}\oplus r_t\oplus r_{t-1}]+b_2)]+b_3) \tag{6}$$

式中：$\mu\in\mathbb{R}^{|C|\times D}$；$\boldsymbol{w}_2^{\mathrm{T}}$ 和 b_2 分别为神经网络层的权重和偏置；σ 表示 Sigmoid 函数。该式通过 Sigmoid 将 μ 限制在 $[0,1]$ 中，控制知识的遗忘量。为了保持维度一致，$\boldsymbol{w}_2$ 层神经网络输出的结果通过 Repeat 操作后可与 $\boldsymbol{K}_{t-1}$ 进行 Concate 操作。最终，通过遗忘的比率 μ 和知识状态的增加量 $\Delta\boldsymbol{K}_t$，对 $\boldsymbol{K}_{t-1}$ 进行更新：

$$\boldsymbol{K}_t=\mu\odot\boldsymbol{K}_{t-1}+\Delta\boldsymbol{K}_t \tag{7}$$

3.2 序列特征模块

序列特征模块通过输入最近的 k 个交互，实现对学生近期整体交互情况的建模。首先，在交互序列 $[r_{t-k+1},\cdots,r_{t-1},r_t]$ 的末尾添加 $[\mathrm{Token}_{\mathrm{CLS}}]$，并使用自注意力机制层学习依赖关系，即 $[h_{t-k+1},\cdots,h_{t-1},h_t,h_{\mathrm{CLS}}]=S_{\text{Self-Attention}}([r_{t-k+1},\cdots,r_{t-1},r_t,\mathrm{Token}_{\mathrm{CLS}}])$。取 $h_{\mathrm{CLS}}\in\mathbb{R}^D$ 作为近期交互状态输入至预测层。

3.3 预测模块

预测模块实现对最终结果的预测，使用经过知识状态模块更新后的知识状态 $\boldsymbol{K}_t$ 与近期交互状态 h_{CLS} 对 e_{t+1} 的表现情况进行预测。首先，计算学生关于 e_{t+1} 的知识点状态 $\boldsymbol{K}_{t+1}^{e_{t+1}}\in\mathbb{R}^D$：

$$\boldsymbol{K}_{t+1}^{e_{t+1}}=\boldsymbol{q}_{t+1}\times\boldsymbol{K}_t \tag{8}$$

然后，计算预测结果：

$$\hat{y}=\sigma(\boldsymbol{w}_4^{\mathrm{T}}[\boldsymbol{K}_{t+1}^{e_{t+1}}\oplus h_{\mathrm{CLS}}\oplus e_{t+1}]+b_4) \tag{9}$$

式中：$\boldsymbol{w}_4^{\mathrm{T}}$ 和 b_4 分别为神经网络层的权重和偏置。

最终，该模型使用二分类交叉熵损失作为损失函数，即：

$$L=-\sum_{t=1}^{T}(y_t\log_a\hat{y}_t+(1-y_t)\log_a(1-\hat{y}_t)) \tag{10}$$

式中：y_t 为学生进行交互时的真实标签，使用 Adam 优化器进行优化，训练目标为最小化 Loss。

4 实验

为了评估模型的性能，本文在 3 个主流的知识追踪数据集上进行实验，与几种主流方法进行性能比较，并且在不同模块组合和超参数设置下进行消融实验。

4.1 实验数据

本文使用 3 个数据集对模型进行评估，分别是 ASSIST2012、ASSIST2017[27] 和 EdNet[28]。数据集的统计信息如表 2 所示。本文对原始数据进行清洗，参考其他工作的做法，去除一些无效的数据，比如错误率过高（超过 90%）、序列过短（小于 4 个）等，同时对超过 100 个交互的数据进行裁剪，对于长度不齐的批次，填充特殊 Token 并输入模型中。

表 2 数据集的统计信息

Table 2 Statistical information of datasets

数据集	学生数/个	知识点/个	练习/个	交互数/条
ASSIST2012	29 018	265	53 091	2 506 769
ASSIST2017	1 709	102	3 162	942 816
EdNet	784 309	141	12 372	72 907 005

ASSIST2012 和 ASSIST2017 数据集是从 1 个在线辅助系统 ASSITments 上收集的，不同的年份形成不同的数据集。ASSIST2012 是 2012—2013 年收集的学生在 ASSITments 系统上的在线做题数据，该数据集中包含 29 018 个学生，265 个知识点，53 091 个习题，共 2 506 769 条数据。ASSIST2017 是 2017—2018 年收集的学生在 ASSITments 系统上的在线做题数据，该数据集中包含 1 709 个学生，102 个知识点，3 162 个习题，包含 942 816 条数据。ASSITments 系列数据集是知识追踪领域使用最广泛、最具权威性的数据集。

EdNet 是目前公开的最大教育基准数据集，包含 784 309 个学生，141 个知识点，12 372 个习题，共 72 907 005 条数据。该数据集收集了大型多层次学生活动的数据，不仅包含学生的做题数据，这使得 EdNet 可以很容易地扩展到不同的领域，例如辍学预测、知识追踪、学习材料推荐等，为不同领域的任务和模型进行融合提供数据支持。知识追踪可以与辍学预测结合，帮助后进生高效学习，避免辍学，也可以为学习资料推荐提供推荐依据。

4.2 训练细节

模型基于 PyTorch 进行构建，并在 1 个独立的 NVIDIA RTX3090 上进行训练。将数据集按 8∶1∶1 随机分为训练集、验证集、测试集，选择验证集中表现最优的 epoch 在测试集上进行测试。使用网格搜索法在以下的参数取值内确定模型的最优参数，使用早停策略进行训练，训练轮数 epoch∈{100,150}，batch_size∈{64,128,256,512}，learn_rate∈{0.001,0.005,0.01}，练习的嵌入维度 $D\in\{64,128,256,512\}$，前馈神经网络隐藏层维度{64,128,256,512}，为了防止过拟合设置 Dropout 为 0.2。

4.3 实验结果

4.3.1 性能对比

本文通过 AUC、准确率（ACC）指标与相关工作部分介绍的几种主流方法进行了性能比较，并且在

计算机工程 www.ecice06.com

不同模块组合和超参数设置下进行消融实验。AUC 为受试者操作特征曲线下面积，它代表了预测结果将正例排在负例前的概率，该指标在样本不均衡的情况下依然能够公正地对模型性能做出评价，AUC 越高代表模型分类效果越好。实验结果如表 3 所示。

表 3 不同方法在 3 个数据集上的性能比较

Table 3 Comparison of performance of different methods in three datasets

模型	ASSIST2012			ASSIST2017			EdNet-KT1		
	RMSE	AUC	ACC	RMSE	AUC	ACC	RMSE	AUC	ACC
DKT	0.428 3	0.720 8	0.730 1	0.450 1	0.711 3	0.685 0	0.454 8	0.673 9	0.683 9
DKVMN	0.425 5	0.723 6	0.733 7	0.449 3	0.711 9	0.685 7	0.452 7	0.679 2	0.687 3
SAKT	0.426 1	0.722 9	0.733 3	0.463 2	0.659 8	0.668 3	0.453 4	0.678 1	0.685 2
CKT	0.421 4	0.745 1	0.736 5	0.443 3	0.732 3	0.699 4	0.450 7	0.684 1	0.689 1
本文模型	0.418 1	0.769 0	0.750 1	0.423 7	0.781 8	0.725 2	0.426 7	0.769 9	0.725 9

结果表明，相比基于序列的方法或是基于模拟学生过程的方法，本文模型在 AUC、ACC 指标上均取得了最优的结果。在 ASSIST2017 的数据集中，在 AUC 指标上本文模型相比 CKT 模型提升 0.049 5，同时该模型在其他数据集中也取得比较好的效果，相比 DKVMN、SAKT 等模型，本文模型在 AUC 指标上高出 0.045 4，ACC 指标高出 0.016 4，均优于基线模型，证明了同时将序列特征和学习过程建模的方法是行之有效并且表现良好的。此外，与经典方法进行对比，DKT、DKVNM 等模型均学习其序列特征，再通过特征进行预测，由于采用了双模块组合的方式，能够较好地捕捉交互序列数据的潜在结构和特征，从而产生更准确、多样性和高质量的学习者与练习表征结果，因此具备更好的性能。同时，由于采用知识点建模的方式，该模型具备 DKT 等模型所不具备的可解释性，支持在使用过程中对学生知识点状态的可视化。此外，采用了动态 $\boldsymbol{Q}$ 矩阵后，与之前的工作相比，动态 $\boldsymbol{Q}$ 矩阵能够表述习题和所有知识点之间的关联关系，从而解决静态 $\boldsymbol{Q}$ 矩阵只能表示习题与单一知识点之间关系的缺陷，这也是性能得到提升的原因之一。

4.3.2 消融实验

本文在是否采用 $\boldsymbol{Q}$ 矩阵、序列特征模块、知识状态模块的组合情况下在 ASSIST2017 数据集上进行消融实验。在 ASSIST2017 数据集上各模块消融实验结果分别如表 4 所示。从表 4 可以看出：

1)动态 $\boldsymbol{Q}$ 矩阵的影响，观察实验 4 和实验 5，引入动态 $\boldsymbol{Q}$ 矩阵特征模块后，AUC 从 77.53%提高到了 78.18%，ACC 从 71.97%提高到了 72.52%，表明动态 $\boldsymbol{Q}$ 矩阵特征模块对模型性能的提升有一定的贡献，实验 1 和实验 3 的比较也验证了其有效性。

2)序列特征模块的影响，通过实验 1 和实验 4 的对比，在知识状态模块的基础上引入序列特征模块，AUC 和 ACC 分别提升了 1.42、0.88 个百分点，这显示了序列特征模块对模型性能有所改进。序列特征模块可能通过提取序列中的关键信息，提供更全面和准确的特征，从而增强模型的预测能力。

3)知识状态模块的影响，通过实验 2 和实验 4 的对比，在使用序列特征模块的基础上使用知识状态模块时，AUC、ACC 均有较大提升，这显示了知识状态模块对模型性能有显著改进。从实验 1 来看，仅使用知识状态模块效果已经达到了 1 个较优秀的水平，但与实验 4 对比表明，知识状态模块和序列特征模块的结合使得模型达到了更高的水平。

4)综合影响。从实验 5 的结果来看，当动态 $\boldsymbol{Q}$ 矩阵特征模块、序列特征模块和知识状态模块同时使用时，AUC 和 ACC 都有进一步提升，分别达到了 78.18%和 72.52%，这表明不同模块的组合可以产生协同效应，提高模型性能。

表 4 在 ASSIST2017 数据集上各模块消融实验结果

Table 4 Results of ablation experiments for each module in the ASSIST2017 dataset %

实验	动态 $\boldsymbol{Q}$ 矩阵	序列特征模块	知识状态模块	AUC	ACC
实验 1			√	76.11	71.09
实验 2		√		62.66	63.66
实验 3	√		√	76.97	71.46
实验 4		√	√	77.53	71.97
实验 5	√	√	√	78.18	72.52

为验证动态且带有约束的 $\boldsymbol{Q}$ 矩阵的有效性，本文在是否采用动态 $\boldsymbol{Q}$ 矩阵和是否带有约束 2 种组合情况下也进行实验，结果如表 5 所示。从表 5 可以看出：

1)使用静态 $\boldsymbol{Q}$ 矩阵能够在一定程度上提高模型的性能，但引入动态 $\boldsymbol{Q}$ 矩阵后能够进一步提升模型的预测能力。

计算机工程 www.ccjournal.com

2)引入约束条件后,模型的性能得到了改善,表明约束条件对动态 Q 矩阵的应用对模型的有效性有积极的影响。

3)同时使用动态 Q 矩阵和约束条件取得了最好的结果,AUC 为 78.18%,ACC 为 72.52%,这表明动态且带有约束的 Q 矩阵的组合对于模型性能的提升效果最佳。

表 5 在 ASSIST2017 数据集上动态 Q 矩阵消融实验结果

Table 5 Results of dynamic Q matrix ablation in the ASSIST2017 dataset %

静态 Q 矩阵	动态 Q 矩阵	约束条件	AUC	ACC
√			77.53	71.97
	√		77.93	72.12
	√	√	78.18	72.52

5 结束语

本文针对目前已有的深度知识追踪模型在可解释性和准确性方面的不足,提出一种序列特征与学习过程融合的知识追踪模型 SLKT。SLKT 模型通过建模学习者学习过程并融合学习者近期交互序列的特征,有效提升了模型的准确性,同时提出带约束的动态 Q 矩阵用以表示练习和知识点之间的关系,确保了较好的可解释性。在 3 个公开数据集上的实验结果表明,SLKT 模型能够有效建模学习者学习过程,在预测学习者下一个交互表现时优于基线模型。后续将探索练习与练习间关系的建模方法,进一步挖掘“用户-练习”二部图关系,通过图神经网络较优的表征能力[29],提高练习的表征效果;同时,进一步研究序列特征与学习过程的融合方式,特别是短期序列特征与长时序特征的组合方法。

参考文献

[1] CORBETT A T, ANDERSON J R. Knowledge tracing: modeling the acquisition of procedural knowledge[J]. User Modeling and User-Adapted Interaction, 1994, 4(4): 253-278.

[2] 陈恩红,刘淇, 王士进, 等. 面向智能教育的自适应学习关键技术与应用[J]. 智能系统学报, 2021, 16(5): 886-898.
CHEN E H, LIU Q, WANG S J, et al. Key techniques and application of intelligent education oriented adaptive learning [J]. CAAI Transactions on Intelligent Systems, 2021, 16(5): 886-898. (in Chinese)

[3] 陈德鑫,占袁圆, 杨兵. 深度学习技术在教育大数据挖掘领域的应用分析[J]. 电化教育研究, 2019, 40(2): 68-76.
CHEN D X, ZHAN Y Y, YANG B. Analysis of applications of deep learning in educational big data mining [J]. e-Education Research, 2019, 40(2): 68-76. (in Chinese)

[4] 于婉莹,梁美玉, 王笑笑, 等. 基于深度注意力网络的课堂教学视频中学生表情识别与智能教学评估[J]. 计算机应用, 2022, 42(3): 743-749.
YU W Y, LIANG M Y, WANG X X, et al. Student expression recognition and intelligent teaching evaluation in classroom teaching videos based on deep attention network [J]. Journal of Computer Applications, 2022, 42(3): 743-749. (in Chinese)

[5] 杨坤融,熊余, 张健, 等. 面向长短期混合数据的 MOOC 辍学预测策略研究[J]. 计算机工程与应用, 2023, 59(4): 130-138.
YANG K R, XIONG Y, ZHANG J, et al. Research on MOOC dropout prediction strategy for long-and short-term mixed data [J]. Computer Engineering and Applications, 2023, 59(4): 130-138. (in Chinese)

[6] 董永峰,王雅琮, 董瑶, 等. 在线学习资源推荐综述[J]. 计算机应用, 2023, 43(6): 1655-1663.
DONG Y F, WANG Y C, DONG Y, et al. Survey of online learning resource recommendation[J]. Journal of Computer Applications, 2023, 43(6): 1655-1663. (in Chinese)

[7] PIECH C, BASSEN J, HUANG J, et al. Deep knowledge tracing[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2015: 505-513.

[8] 郭艺,何廷年, 李爱斌, 等. 融合 GA-CART 和 Deep-IRT 的知识追踪模型[J]. 计算机工程与科学, 2023, 45(9): 1691-1700.
GUO Y, HE T N, LI A B, et al. A knowledge tracing model fusing GA-CART and Deep-IRT[J]. Computer Engineering & Science, 2023, 45(9): 1691-1700. (in Chinese)

[9] 李晓光,魏思齐, 张昕, 等. LFKT: 学习与遗忘融合的深度知识追踪模型[J]. 软件学报, 2021, 32(3): 818-830.
LI X G, WEI S Q, ZHANG X, et al. LFKT: deep knowledge tracing model with learning and forgetting behavior merging[J]. Journal of Software, 2021, 32(3): 818-830. (in Chinese)

[10] 张凯,刘月, 覃正楚, 等. 概念表示增强的知识追踪模型[J]. 计算机应用研究, 2022, 39(11): 3309-3314.
ZHANG K, LIU Y, QIN Z C, et al. Concept representation enhanced knowledge tracing [J]. Application Research of Computers, 2022, 39(11): 3309-3314. (in Chinese)

[11] GHOSH A, HEFFERNAN N, LAN A S. Context-aware attentive knowledge tracing[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM Press, 2020: 2330-2339.

[12] ZHANG J N, SHI X J, KING I, et al. Dynamic key-value memory networks for knowledge tracing[C]//Proceedings of the 26th International Conference on World Wide Web. New York, USA: ACM Press, 2017: 765-774.

[13] NAGATANI K, ZHANG Q, SATO M, et al. Augmenting knowledge tracing by considering forgetting behavior[C]// Proceedings of the World Wide Web Conference. New York, USA: ACM Press, 2019: 3101-3107.

[14] YUDELSON M V, KOEDINGER K R, GORDON G J. Individualized Bayesian knowledge tracing models [C]// Proceedings of International Conference on Artificial Intelligence in Education. Berlin, Germany: Springer, 2013: 171-180.

[15] PAVLIK P I, CEN H, KOEDINGER K R. Performance factors analysis —a new alternative to knowledge tracing [C]// Proceedings of the Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling. New York, USA: ACM Press, 2009: 531-538.

[16] SONG X Y, LI J X, CAI T T, et al. A survey on deep learning based knowledge tracing [J]. Knowledge-Based Systems, 2022, 258: 110036.

[17] LONG T, LIU Y F, SHEN J, et al. Tracing knowledge state with individual cognition and acquisition estimation [C]// Proceedings of the 44th International ACM SIGIR Conference

计算机工程 www.ecice.com

on Research and Development in Information Retrieval. New York, USA: ACM Press, 2021: 173-182.
[18] GUO X P, HUANG Z J, GAO J, et al. Enhancing knowledge tracing via adversarial training[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM Press, 2021: 367-375.
[19] WANG C Y, MA W Z, ZHANG M, et al. Temporal cross-effects in knowledge tracing[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York, USA: ACM Press, 2021: 517-525.
[20] WU J Z, HUANG Z Y, LIU Q, et al. Federated deep knowledge tracing [C]// Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York, USA: ACM Press, 2021: 662-670.
[21] YEUNG C K, YEUNG D Y. Addressing two problems in deep knowledge tracing via prediction-consistent regularization[C]// Proceedings of the 15th Annual ACM Conference on Learning at Scale. New York, USA: ACM Press, 2018: 1-10.
[22] PANDEY S, KARYPIS G. A self-attentive model for knowledge tracing[EB/OL]. [2023-03-25]. http://arxiv.org/abs/1907.06837.
[23] CHOI Y, LEE Y, CHO J, et al. Towards an appropriate query, key, and value computation for knowledge tracing [C]// Proceedings of the 17th ACM Conference on Learning @ Scale. New York, USA: ACM Press, 2020: 341-344.
[24] SHIN D, SHIM Y, YU H, et al. SAINT+: integrating temporal features for EdNet correctness prediction [C]// Proceedings of the 11th International Learning Analytics and Knowledge Conference. New York, USA: ACM Press, 2021: 490-496.
[25] ABDELRAHMAN G, WANG Q. Knowledge tracing with sequential key-value memory networks[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press,2019: 1-10.
[26] SHEN S H, LIU Q, CHEN E H, et al. Learning process-consistent knowledge tracing[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York, USA: ACM Press, 2021: 1452-1460.
[27] FENG M Y, HEFFERNAN N, KOEDINGER K. Addressing the assessment challenge with an online system that tutors as it assesses[J]. User Modeling and User-Adapted Interaction, 2009, 19(3): 243-266.
[28] CHOI Y, LEE Y, SHIN D, et al. EdNet: a large-scale hierarchical dataset in education [C]// Proceedings of International Conference on Artificial Intelligence in Education. Berlin, Germany: Springer, 2020: 69-73.
[29] 王健宗,孔令炜, 黄章成, 等. 图神经网络综述[J]. 计算机工程, 2021, 47(4): 1-12.
WANG J Z, KONG L W, HUANG Z C, et al. Survey of graph neural network[J]. Computer Engineering, 2021, 47(4): 1-12. (in Chinese)

编辑 薛晋栋

计算机工程 www.ecice06.com