Computer Science and Application 计算机科学与应用, 2021, 11(3), 770-781
Published Online March 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.113079

Hans汉斯

基于LSTM/GCN的在线学习文本特征提取方法

温创斐[1]，曾　安[1]，潘　丹[2*]

[1]广东工业大学计算机学院，广东　广州
[2]广东技术师范大学电子与信息学院，广东　广州
Email: 173537017@qq.com, zengan@gdut.edu.cn, *2656351065@qq.com

收稿日期：2021年2月28日；录用日期：2021年3月24日；发布日期：2021年3月31日

摘　要

传统方法对在线学习文本进行特征筛选往往费时费力且迁移性较差。针对这一问题，根据在线学习文本短，专业词汇多，文本间结构信息丰富等特点，提出基于LSTM/GCN对Doc2Vec所得文本向量中文本－文本关系进行强化的文本嵌入方法，以解决传统方法中文本在投影到嵌入空间后结构信息丢失的问题。并提出指标MeanRank用于量化文本向量中结构信息的留存情况。实验结果表明，方法在指标MeanRank和文本分类精度上优于传统方法。可视化结果表明，增加结构向量使得文本向量在课程内部具有一致连贯性，在课程间更有区分度。

关键词

特征提取，学习分析，图神经网络

Online Learning Text Feature Extraction Method Based on LSTM/GCN

Chuangfei Wen[1], An Zeng[1], Dan Pan[2*]

[1]School of Computers, Guangdong University of Technology, Guangzhou Guangdong
[2]School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou Guangdong
Email: 173537017@qq.com, zengan@gdut.edu.cn, *2656351065@qq.com

Received: Feb. 28th, 2021; accepted: Mar. 24th, 2021; published: Mar. 31st, 2021

*通讯作者。

文章引用: 温创斐, 曾安, 潘丹. 基于 LSTM/GCN 的在线学习文本特征提取方法[J]. 计算机科学与应用, 2021, 11(3): 770-781. DOI: 10.12677/csa.2021.113079

Abstract

Traditional methods for feature filtering of online learning text are often time-consuming and poorly migratory. To address this problem, and based on the characteristics of short texts of online learning text , many specialized vocabularies, and rich structural information between text, an end-to-end text feature extraction method is proposed. The method emphasizes the text-text relationship based on LSTM/GCN by obtaining the text vector based on Doc2Vec model to solve the phenomenon that the traditional method text loses structural information after projection to the embedding space. And the metric MeanRank is proposed to quantify the retention of structural information in the text vector. Experimental results on the Yale Open Course dataset show that the method outperforms traditional methods in terms of metrics MeanRank and text classification accuracy. Visualization of t-distributed stochastic neighbor embedding of text vectors shows that adding structural vectors makes text vectors consistently coherent within courses and more discriminative between courses.

Keywords

Feature Extraction, Learning Analytics, Graph Neural Network

1. 引言

学习分析(Learning Analytics, LA)用于描述技术增强学习(Technology Enhanced Learning, TEL)研究领域[1]，该领域的目标是开发用于检测教育系统中数据模式的方法，并使用这些方法来提高学习体验。学习分析所使用的机器学习算法大多依赖于数据良好的特征表示，而教育系统中的特征常基于专家知识进行人为设计，导致教育系统之间特征构造方法往往不同，难以复用，海量的在线教育数据无法得到充分利用。

目前，基于深度学习实现端对端的特征提取因其便捷性受到了广泛的关注。将文本转换为实数向量属于自然语言处理(Natural Language Processing, NLP)的研究领域，通常被称为文本嵌入方法。然而，课堂的简介通常篇幅较短并含有大量的专业性词汇，使用传统的文本嵌入方法来提取特征效果欠佳。

近期有研究者的工作[2]表明，使用图卷积神经网络可以高效地提取单词－单词和文本－单词特征用于文档分类，证明了结构信息对文本嵌入的增益。基于此，本文提出了基于 LSTM [3] /GCN [4]的在线学习文本特征提取方法，方法分为三个模块：数据预处理模块，语义嵌入提取模块和结构嵌入提取模块。通过利用文本－文本结构信息对文本的语义嵌入向量进行补充和增强来得到课堂的表示向量。其中结构嵌入被定义为包含当前文本与其上下文文本结构信息的嵌入向量。方法大致过程如下：利用文本嵌入提取模块将课堂描述文本转化为语义嵌入向量；再将语义向量输入到结构嵌入提取模块进行更新融合得到结构嵌入向量；最后将这两部分向量结合得到最终的课堂嵌入向量。

2. 相关研究

文本嵌入是自然语言处理中将文本映射到实值向量空间技术的统称。一种常见的做法是使用词嵌入

模型得到所有单词的嵌入，再用平均或求和操作作为后处理得到文本嵌入。Siamese CBOW [5]选择将平均操作纳入到训练过程中，通过预测当前句子邻近句子的嵌入得到了更鲁棒的方法。Arora [6]使用平滑的反频率加权机制取代平均操作，并参考 PCA/SVD 对词嵌入的平均向量进行修改，结果表明词嵌入加权平均是一条简单但难以超越的基线。随着人工智能的发展，神经网络在各基准任务的突出表现，部分研究者将 Word2vec [7]和 Doc2Vec [8]的思想与神经网络优秀的特征提取能力相结合，获得了良好的效果。Sent2Vec [9]扩展了 Word2Vec 的 PV-CBOW 结构，利用 N-Gram 进行句子表示的学习，并使用平均算子得出句子向量。Doc2VecC [10]结合了 Doc2Vec 的 PV-DM 结构和词嵌入平均，添加了一些噪声以产生更健壮的文本嵌入。Vo [11]在 Doc2Vec 的 PV-CBOW 架构中加入了句法成分，丰富了文本嵌入的情感信息。一些研究者通过使用更复杂的编码器以提取更丰富的信息。Quick-thought [12]使用了两个编码器来增强提取能力，S-BERT [13]则选择语言模型 BERT 进行微调作为编码器。

上述工作为将文本映射到实值向量空间提供了各种创新解决方案，并在基础任务的表现上有了显著的突破。然而这些模型对文本之间的关系的利用并不充分。如 Doc2Vec 使用了一个独热向量来表示模型当前关注并训练的文本向量，即在当前文本嵌入的更新阶段不涉及与其相关的文本。S-BERT 通过对正负文本样本进行采样作为训练数据进而将文本之间的关系考虑在内，并使用了微调过后的 BERT 模型作为嵌入编码器。这种方法很好地利用了 BERT 提取语义信息的能力，但是只考虑二元关系可能是不充分的。“上下文无关的方法”常使得文本之间隐藏的信息没有得到充分的利用[14]，Angelova 等通过改进松弛标记技术以充分探索数据的上下文关系；Liu [15]通过在 LSTM 使用双注意力机制利用单词上下文信息提高分类精度；曾[16]通过在注意力层后增加填充层确保每个单词都具备上下文信息；Yao [2]通过在单词－文档异构图上运用图卷积技术将文本分类任务转化为节点标注任务。以上的工作都利用了结构信息，Angelova 利用了文本－文本之间的信息，但因为传统方法特征提取能力有限，算法过程仅对文本可能的类别进行预测，未形成其它任务可复用的文本嵌入。Liu [15]、曾[16]和 Yao [2]更进一步，使用目前更流行的 BiLSTM 和 GCN 对单词向量进行处理得到了文本嵌入，但只探讨了单词－单词和单词－文本之间的关系。

3. 基于 LSTM/GCN 的文本特征提取方法

在学习网站中，学习者浏览的每一节课堂都是一门课程的一部分，课程之间存在复杂的先修关系，每门课程又可能同时属于多个教育计划。这就导致了课堂之间存在复杂的结构信息。利用这些结构信息对语义信息进行补充，可以更好地服务于学习网站的学习分析。本节描述了构建课堂嵌入的流程，图 1 展示了将原始文本数据转换为嵌入数据的流程，其中 ⊕ 符号表示将两个流程得到的向量连接在一起。

数据预处理模块将从在线学习平台获得的原始文本数据进行清洗、分词和标注等操作，数据预处理模块之后通过语义嵌入提取模块将文本数据转化为实数向量。随后是本方法的核心模块：结构嵌入提取(SEE)模块，如流程图所示，SEE 模块有两种架构，每种架构对应一种课程内课堂可能的结构，分别为时序式和图式。提取器输出的内容作为该课堂的结构嵌入，为课堂嵌入提供更丰富的特征。最后，课堂嵌入的计算公式为

$$V_{\mathrm{SE}}^{i}=\mathrm{Concatnate}\left(V_{\mathrm{DE}}^{i},V_{\mathrm{SEE}}^{i}\right) \tag{1}$$

其中，特定课堂 i 的课堂嵌入 V_{SE}^{i} 是一个实数向量，它由连接文本嵌入 V_{DE}^{i} 和结构嵌入 V_{SEE}^{i} 得到。函数 $\mathrm{Concatnate}(\cdot,\cdot)$ 的作用是将传递给它的两个向量连接在一起。V_{DE}^{i} 和 V_{SEE}^{i} 的值分别由文本嵌入模块和 SEE 模块获得。

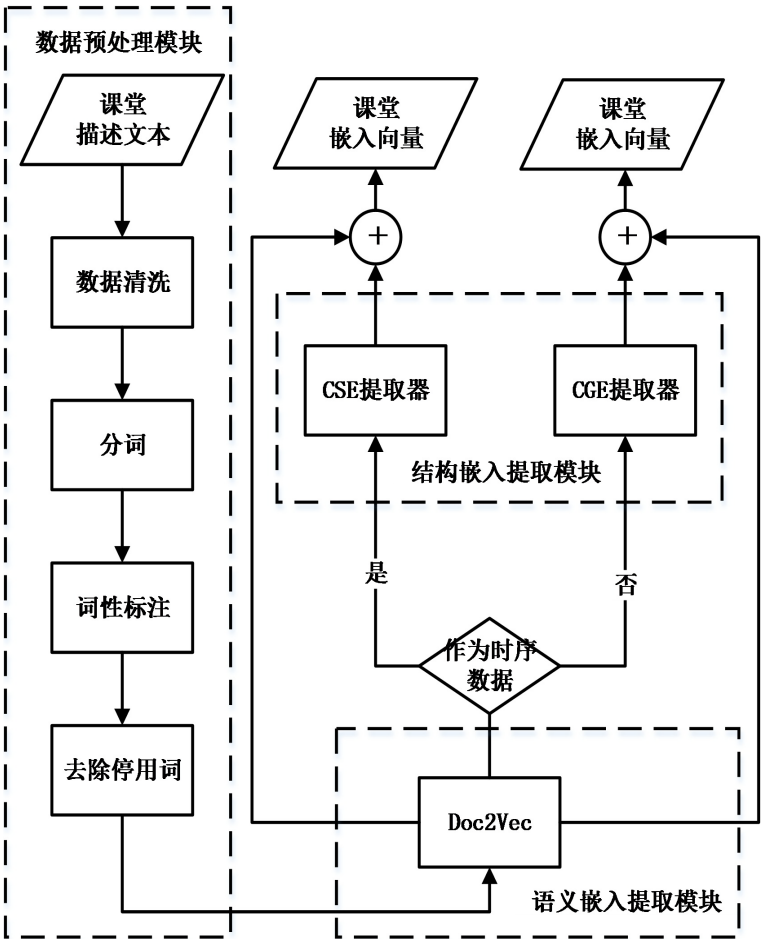


Figure 1. LSTM/GCN-based text feature extraction process for online learning
图 1. 基于 LSTM/GCN 的在线学习文本特征提取流程

3.1. 数据预处理模块

数据预处理模块主要分为以下步骤：在英文书写的过程中，为了强调或风格区分，文本中同一个单词因为大小写可能有多种形式，但往往它们表示的是同一个意思。为了消除冗余，将文本归一化至全小写；数据清洗阶段将噪音数据进行消除，例如本文实验中，消除内容为“exam”的课堂，因为它不存在有效的文本信息；分词阶段将文本分成无法再分拆的有意义的符号，对于英文文本进行分词往往指的是将每个句子分割成一系列的单词的过程；词性标注阶段给单词标上词类标签，比如名词、动词、形容词等，因为了解单词在句子中的用途有助于模型更好理解句子内容；去除停用词阶段从文本内删去“the”、“is”和“at”等对文本特征没有贡献作用的字词。

3.2. 语义嵌入提取模块

语义嵌入提取模块采用的是 Doc2Vec 的 PV-DBOW 架构。Doc2vec 受 Word2vec 的启发，目标是将可变长度的文本映射为固定长度的向量表示形式，是一种无监督方法，PV-DBOW 与 Word2vec 的 Skip-Gram 架构相似，它以段落 ID 和段落内句子的抽样作为输入，通过预测采样句子中的单词进行训练。训练结束后使用段落矩阵作为查找表来获取段落表示，本文提出的方法采用 python 的自然语言处理库 gensim 中对 Doc2Vec 的实现。

3.3. 结构嵌入提取模块

LSTM 常用于时序数据的特征提取，而图神经网络(Graph Neural Networks, GNNs)常用于基于拓扑信息的特征提取。基于此，结构嵌入提取模块将两者作为基础组件，针对在线学习数据可能具备的结构特征设计了两种结构嵌入提取器，分别是上下文序列提取器和上下文图提取器。

3.3.1. 上下文序列提取器

学习者对一门课程的学习记录通常可以看成一段关于课堂的时序数据。基于此，上下文序列提取器(Context Sequence Extractor, CSE)将 Doc2Vec 得到的课堂的文本嵌入作为输入用于提供课堂之间的差异，并使用堆叠的 Bi-LSTM 层来捕捉课堂间的顺序信息。CSE 的训练目标是将一个课堂序列准确地分类到其对应的课程中，此时中间层的输出融合了语义和结构信息，将其作为对应课堂的结构嵌入。

LSTM 单元中根据上一个隐藏状态 h_{t-1} 计算当前隐藏状态 h_t 的步骤可以描述为以下公式：

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{3}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

$$o_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] + b_o\right) \tag{6}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{7}$$

其中 W_f、W_i、W_o 和 W_c 分别表示遗忘门、输入门、输出门和 LSTM 单元的参数矩阵。C_t 是时间 t 和 $t-1$ 单元状态的加权和，sigmoid 激活函数 $\sigma(\cdot)$ 将两者权重映射到 0 和 1 之间。再经由激活函数 $\tanh(\cdot)$ 得到加权系数 o_t，最终得到隐藏状态 h_t。CSE 架构基于 Bi-LSTM 单元，如图 2 所示。

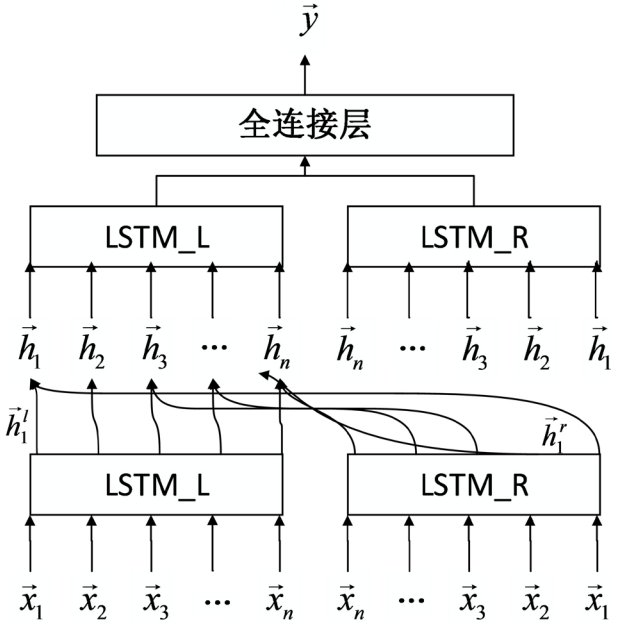


Figure 2. Contextual sequence extractor CSE Architecture
图 2. 上下文序列提取器 CSE 架构

其中 $\vec{x}_i$ 表示从 Doc2Vec 模型得到的第 i 个文本嵌入。在 Bi-LSTM 单元中，将课堂的文本嵌入序列从左到右输入 LSTM_L 得到 $\vec{h}_t^l$，使用相反的顺序输入 LSTM_R 得到 $\vec{h}_t^l$，通过将两者连接得到 $\vec{h}_t$。CSE 以固定窗口大小的文本嵌入序列作为输入，通过两层 Bi-LSTM 及全连接层后将其分类为不同的课程。在分类目标的引导下，CSE 的输出逐渐向类别的独热向量靠拢，而隐藏层则保留了课程之间的序列结构信息。因而将 CSE 的第一层输出的隐藏状态序列作为结构嵌入。由于滑动窗口之间有重叠，大部分课堂根据其在滑动窗口中的不同位置有多个隐藏状态，这里采用平均运算作为后处理来得到课堂 i 的唯一嵌入。如公式(8)中所述。

$$V_{\mathrm{SE}}^i = \mathrm{Concatnate}\left(V_{\mathrm{Doc2Vec}}^i, \mathrm{Mean}\left(\sum_{j=0}^{n} h_{\mathrm{LSTM}}^{ij}\right)\right) \tag{8}$$

其中 j 表示课堂 i 在滑动窗口中的位置，n 表示滑动窗口的大小。将课堂 i 状态向量的平均与其文本嵌入相连接，得到课堂 i 的嵌入向量。

受到 Word2Vec [7]的启发，其中 CBOW 架构将锚词 i 周围的词作为输入，并将 i 作为预测目标。基于此本文提出了上下文序列回归提取器(Context Sequence Regression Extractor, CSRE)。CSRE 将粒度从单词扩展到文本，并将预测目标从中心改为右侧。

网络整体结构与图 2 相同，但用回归层代替了分类层。其目标是根据前几节课堂的文本嵌入来拟合下一个课堂的文本嵌入。公式(9)显示了它的计算方式。

$$V_{\mathrm{SE}}^i = \begin{cases} \mathrm{Concatenate}\left(V_{\mathrm{Doc2Vec}}^i, \mathrm{Mean}\left(\sum_{j=0}^{n} h_{\mathrm{CSRE}}^{ij}\right)\right), & \text{当} i \text{不是课程的最后一节课堂} \\ \mathrm{Concatenate}\left(V_{\mathrm{Doc2Vec}}^i, \mathrm{Mean}\left(\sum_{j=0}^{n-1} V_{\mathrm{SE}}^i\right)\right), & \text{当} i \text{是课程的最后一节课堂} \end{cases} \tag{9}$$

其符号约定与 CSE 的公式一致。略有不同的是，每门课程最后一节课堂的结构嵌入无法通过模型的隐藏层获得，本文使用同一门课程中其他课堂的结构嵌入的平均值作为最后一节课堂的结构嵌入。

3.3.2. 上下文图提取器

当课堂 i-1 和 i-2 是课堂 i 的预备知识，但预备知识之间没有关联时，形成课堂 i-1 和 i-2 有一条指向 i 的弧的“V”型结构。为了利用这类隐藏的结构信息对课堂嵌入进行补充，本文引入了可以在非欧几里得空间进行卷积的 GCN [4]，并基于此提出上下文图提取器(Context Graph-like Extractor, CGE)来获得课堂的结构嵌入。GCN 层间的更新公式[4]如下所示。

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{10}$$

其中 $\tilde{A}$ 表示输入图的重归一化邻接矩阵，重归一化指的是图中的每个节点加上了一条指向自己的弧。$\tilde{D}$ 表示 $\tilde{A}$ 的度矩阵。也可以将 GCN 的更新公式(10)改写成公式(11)，其中 x_i 是输入图第 i 个节点的表示向量，$\mathcal{N}(x_i)$ 是节点 i 邻居表示向量的集合。

$$h_{x_i} = f\left(\frac{1}{\left|\mathcal{N}(x_i)\right|}\sum_{j\in\mathcal{N}(x_i)} Wx_j + b\right), \quad \forall x_i \in \mathcal{V} \tag{11}$$

公式(11)形式化地描述了更新的两个步骤：首先对每个节点 i 使用平均算子聚合其邻居节点线性变化后的特征，然后使用非线性激活函数 $f(\cdot)$ 得到下一层该节点的输入 h_i。基于 GCN 层，CGE_N 架构如图 3 所示。

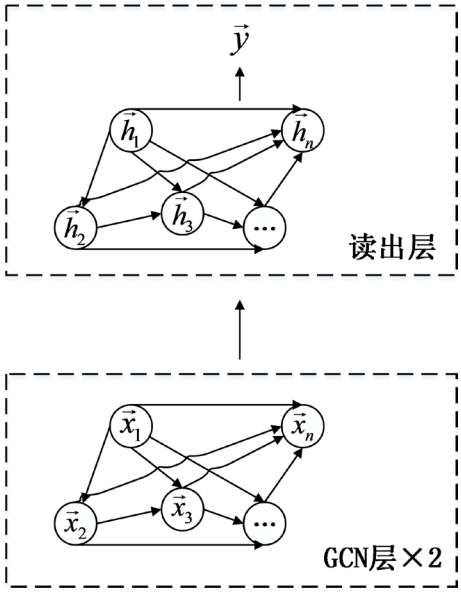


Figure 3. Context Graph Extractor CGE_N Architecture
图 3. 上下文图提取器 CGE_N 架构

CGE_N 架构中图的构造基于当前课堂只和滑动窗口内先修于它的课堂相关，在图 3 中表示为滑动窗口内当前课堂之前的所有课堂都有一条指向当前课堂的弧线。CGE_N 架构使用课堂的文本嵌入作为节点特征，在图中用 $\vec{x}_i$ 表示。CGE_N 将上述图和节点特征作为输入，在训练过程中，图中的节点通过弧线进行通信，使节点的文本嵌入进行融合。读出层合并了所有节点的隐藏表示，以获得图级表示进行分类任务。CGE_N 架构使用第二层 GCN 层的输出 h_{GCN} 作为结构嵌入，并使用平均运算对因为滑动窗口重叠而带来的多份结构嵌入进行后处理，如式(12)所示。

$$V_{\mathrm{SE}}^{i}=\mathrm{Concatnate}\left(V_{\mathrm{Doc2Vec}}^{i},\mathrm{Mean}\left(\sum_{j=0}^{n}h_{\mathrm{GCN}}^{ij}\right)\right) \tag{12}$$

在 CGE_N 架构中，权重在同一 GCN 层之间是共享的，即课堂图中节点对待每个邻居的方式是一样的。考虑滑动窗口中的最后一个课堂节点，所有节点都有一条指向它的弧线，使用相同的权重将距离不同的节点特征进行聚合是不自然的。Veličković [17]提出的 GAT 层通过将注意力机制引入图卷积层来获取一定的灵活性。

通过将 GCN 网络层替换为 GAT 网络层，提出了基于 GAT 的 CGE_T 架构，在 GAT 层中课堂节点通过注意力机制获得对各个邻居权重值用于更新自身的隐藏表示。公式(13)用于构建课堂嵌入，其中 h_{GAT} 代表第二个 GAT 层的输出。

$$V_{\mathrm{SE}}^{i}=\mathrm{Concatnate}\left(V_{\mathrm{Doc2Vec}}^{i},\mathrm{Mean}\left(\sum_{j=0}^{n}h_{\mathrm{GAT}}^{ij}\right)\right) \tag{13}$$

4. 实验与分析

4.1. 数据集

实验所用数据收集于 Open Yale Courses (https://oyc.yale.edu/courses)平台，它提供了由耶鲁大学杰出教师和学者教授的入门课程。Open Yale Course 目前包含了来自 22 个领域的 40 门课程，每门课程都有一套学习

材料，如视频片段、课程描述和音频等。数据集 Yale Open Course Ware Corpus 由南特大学 DUKe 实验室提供，其收集了 Open Yale Courses 网站的课程信息，并以 CSV 格式保存数据。数据集中有 22 个院系，每个院系的课程数量从 1 门到 7 门不等。40 门课程总共由 1058 节课堂组成，每门课程的课堂节数从 16 节到 41 节不等。

4.2. 评价指标

因课堂与课程之间的映射关系是已知信息而且对学习分析任务的帮助不大，使用常规的文本分类作为评价指标不够充分。基于特征丰富的嵌入应能保留同一课程内课堂的一致连贯性，相邻的课堂应在嵌入空间有着较小的距离，本文提出投影－检索评估算法计算 MeanRank 以评估各方法所得课堂嵌入的质量。

算法 1 投影-检索评估算法：
输入：课堂嵌入 SEs；正负采样数 sn
输出：平均检索序号 MeanRank

1: # Get the projection model
2: **for** course in SEs:
3: **for** V_{SE}^{i} in course:
4: $\mathbb{X}=\varnothing$， $\mathbb{Y}=\varnothing$
5: $\mathbb{X}_i \leftarrow \mathbb{X}_i \cup \left\{V_{\mathrm{SE}}^{i+1} \cdots V_{\mathrm{SE}}^{i+\mathrm{sn}}\right\}$
6: $\mathbb{N}_i \leftarrow \mathrm{NegtiveSampling}\left(V_{\mathrm{SE}}^{i}, \mathrm{sn}\right)$
7: $\mathbb{X}_i \leftarrow \mathbb{X}_i \cup \mathbb{N}_i$
8: **for** V_{SE}^{j} in $\mathbb{X}_i$:
9: $\mathbb{X} \leftarrow \mathbb{X} \cup \left\{\left(V_{\mathrm{SE}}^{i}, V_{\mathrm{SE}}^{j}, \left|V_{\mathrm{SE}}^{i}-V_{\mathrm{SE}}^{j}\right|\right)\right\}$
10: $\mathbb{Y} \leftarrow \mathbb{Y} \cup \{0\}$
11: **for**i in range(sn-1):
12: $\mathbb{Y} \leftarrow \mathbb{Y} \cup \{1\}$
13: **for**i in range(sn):
14: $\mathbb{Y} \leftarrow \mathbb{Y} \cup \{2\}$
15: Getmodelusingdataset ($\mathbb{X}$, $\mathbb{Y}$)
16: # Retrieval operations with model
17: **for** V_{SE}^{i} in SEs:
18:**for** V_{SE}^{j} in SEs:
19: $ipt=\left(V_{\mathrm{SE}}^{i}, V_{\mathrm{SE}}^{j}, \left|V_{\mathrm{SE}}^{i}-V_{\mathrm{SE}}^{j}\right|\right)$
20: $\mathrm{distance}_{i,j}=\dfrac{1}{\mathrm{model}(ipt)[0]}$
21: Sessions aresorted in ascending order according to distance, the ordinal number of the category '0' vector for V_{SE}^{i} is used as the rank_i
22: $\mathrm{MeanRank}=\dfrac{1}{|\mathrm{SEs}|}\sum_{i=0}^{|\mathrm{SEs}|}\mathrm{rank}_i$

算法 1 步骤 9 中以 $\left(V_{\mathrm{SE}}^{i}, V_{\mathrm{SE}}^{j}, \left|V_{\mathrm{SE}}^{i}-V_{\mathrm{SE}}^{j}\right|\right)$ 对投影模型的输入进行构造，其中 V_{SE}^{i} 为当前处理的课堂嵌入，V_{SE}^{j} 为基于当前课堂采样得到的课堂嵌入，每一对 V_{SE}^{i} 和 V_{SE}^{j} 的类别由课程中两者的相对位置得到。“0”代表课堂 j 是课堂 i 的下一节课；“1”代表 j 课堂与 i 课堂的距离在 sn 以内；“2”代表 i 和 j 两节课堂距离较远，负样本由步骤 6 中函数 $\mathrm{NegtiveSampling}\left(V_{\mathrm{SE}}^{i}, \mathrm{sn}\right)$ 采样得到。在对输入进行线性变换后，采用三路 softmax 对两者的相对关系进行分类。投影模型可形式化表述为式(14)。

$$\mathrm{softmax}\left(W\left(V_{\mathrm{SE}}^{i}, V_{\mathrm{SE}}^{j}, \left|V_{\mathrm{SE}}^{i}-V_{\mathrm{SE}}^{j}\right|\right)\right) \tag{14}$$

其中 W 为投影模型参数，在步骤 15 处进行更新。Reimers [13]的实验表明这种输入的构造方式可以更好

地捕捉到两个向量之间的差异。在检索阶段，V_{SE}^{i} 课堂嵌入和所有课堂嵌入的组合作为投影模型的输入。MeanRank 指标着重评估嵌入方法是否捕捉到相邻课堂之间的关系。步骤 20 使用投影模型 model 将两个课堂嵌入向量关系分类为“0”概率的倒数作为两个嵌入之间的距离。如式(15)所示。

$$\text{distance}_{i,j}=\frac{1}{\text{model}\left(\left(V_{SE}^{i},V_{SE}^{j},\left|V_{SE}^{i}-V_{SE}^{j}\right|\right)\right)[0]} \tag{15}$$

其中 V_{SE}^{i} 和 V_{SE}^{j} 可以是方法所得 SEs 中课堂嵌入的任意组合，式中分母部分衡量了课程 j 紧随课程 i 之后的概率大小。检索结果根据距离公式所得进行升序排序，其中类别为“0”的课堂嵌入的序号 rank_i 作为这次检索的表现。将所有课堂作为 i 课堂进行检索，步骤 22 取所有检索表现的平均值 MeanRank 作为当前嵌入模型的表现。

4.3. 实验

为了方便对加入 SEE 模块前后的表现进行比较，实验中对 Doc2Vec 得到的嵌入进行了评估。在实验设置中，文本嵌入和结构嵌入的维度分别为 78 和 40。即增加 SEE 模块后，课堂嵌入的维度会发生改变，因此增加了维度为 118 的 Doc2Vec 的试验。图 4 展示了各嵌入方法的投影－检索评估结果。

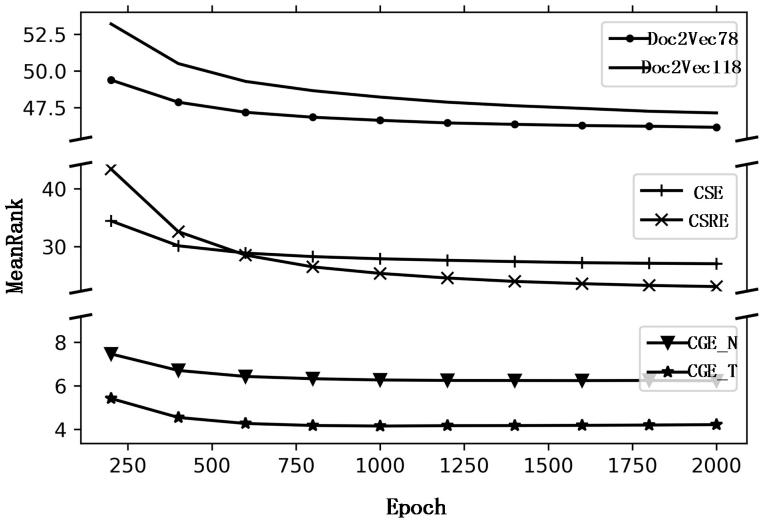


Figure 4. MeanRank of each method in the 10-fold cross-validation training in projection model
图 4. 十折交叉验证训练投影模型各方法所得 MeanRank

由于维度为 78 的 Doc2Vec 嵌入包含了大部分信息，Doc2Vec78 和 Doc2Vec118 的表现几乎趋于同一点。根据 MeanRank 的计算方式，它代表了课堂嵌入检索结果中类别“0”样本的平均序号。其值越小，说明该嵌入模型得到的课堂嵌入在投影－检索评估中表现越好。图 4 显示，表现最好的是 CSRE，最差的是 Doc2Vec118。表 1 显示了各嵌入方法 MeanRank 收敛到的确切值。

本实验统一使用 Doc2Vec 作为文档嵌入模块。表 1 的前两行显示的是没有 SEE 模块的结果，其后是选择 SEE 模块的不同架构得到的结果。CSE 和 CSRE 得到了最好的结果，与 Doc2Vec118 相比，CSRE 的表现提高了十倍左右。CSE 和 CSRE 在获取结构嵌入时都考虑了课堂的局部顺序，在获取结构嵌入的过程中把课程中的课堂当作序列来处理。CSRE 的效果要好于 CSE。可能是因为 CSE 在训练过程通过文本嵌入序列预测课程类别，关注的是序列与整体之间的关系；而 CSRE 通过前面的课堂嵌入对下一节课堂的嵌入进行拟合，关注的是一节课堂与其之前的课堂之间的关系。此外，当只考虑到文本嵌入而不是

与课程信息混合时，结构性信息可能会占据更大的比例。

Table 1. MeanRank value of each embedding method
表 1. 各嵌入方法 MeanRank 值

嵌入方法	维度	MeanRank
Doc2Vec78	(78, N)	46.16
Doc2Vec118	(118, N)	47.14
Doc2Vec + CSE	(118, N)	6.23
Doc2Vec + CSRE	**(118, N)**	**4.15**
Doc2Vec + CGE_N	(118, N)	27.06
Doc2Vec + CGE_T	(118, N)	23.12
Doc2Vec78	(78, N)	46.16

CGE 试图挖掘课堂之间更复杂的关系，通过构造课堂图以及图中课堂节点之间特征的交流和融合，来使得节点获得特征更丰富的嵌入。然而，定义一个固定的图结构对局部课堂之间的关系进行建模似乎过于严格。其在投影－检索评估中的表现比没有 SEE 模块的 Doc2Vec118 提升了一倍，但比 CSE 的表现差。通过将 GCN 层改为一定程度上可以调节图结构的 GAT 层使得表现有所提升，也证实了 CGE_N 中人工定义的结构过于严格。可以预见，CGE_T 架构在更大、更复杂的结构化数据集上可以有更好的表现。

为了验证 SEE 模块对分类精度的影响，补充了在 Open Yale Course 数据集上进行文本分类任务的实验，结果如表 2 所示。

Table 2. Text classification accuracy of each embedding method
表 2. 各嵌入方法文本分类精度

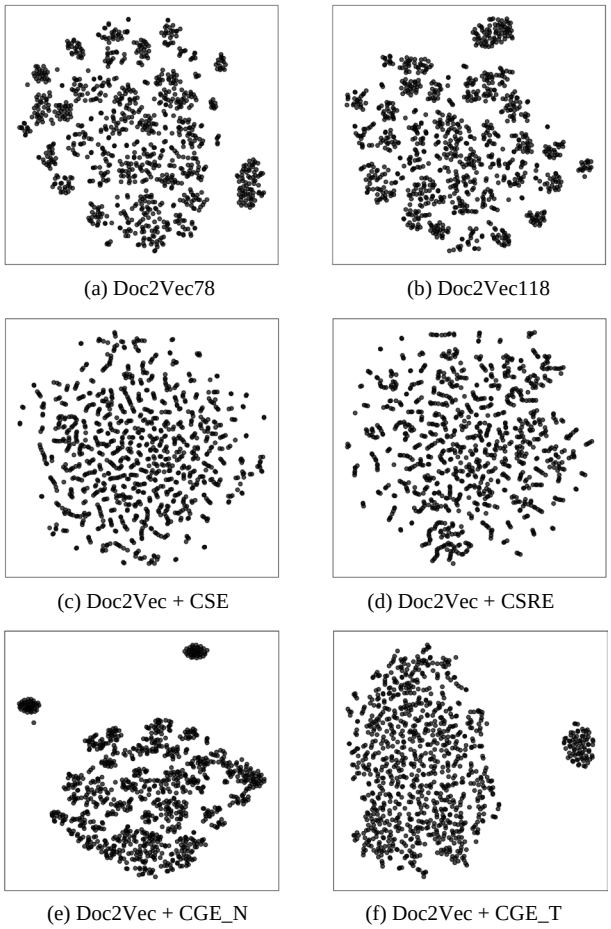
嵌入方法	维度	分类准确率
Doc2Vec78	(78, N)	90.68%
Doc2Vec118	(118, N)	94.12%
Doc2Vec + CSRE	**(118, N)**	**96.07%**

实验选取了 SEE 模块中没有用到课程类别信息的 CSRE 架构进行验证，实验结果表明加入 CSRE 提供的结构信息后可以在传统的文本分类任务中有良好的表现。

4.4. 可视化

本文使用 t-SNE (t-distributed stochastic neighbor embedding)降维可视化来探索结构嵌入对课堂嵌入的影响。根据下图 5 可以得到一些有趣的结论。

结果可按使用传统方法、SEE 模块 CSE 架构和 SEE 模块 CGE 架构分为(a)(b)、(c)(d)和(e)(f)三组。子图(a)和(b)仅使用 Doc2Vec 模型，所得嵌入在降维后，课堂根据所属课程的不同形成了明显的聚簇，但课程内部课堂的分布稀疏散乱；子图(c)和(d)除了分离每门课程外，多处课堂之间连成了一条序列，即通过将文本嵌入和 SEE 模块 CSE 架构所得结构嵌入相连接，能为课堂嵌入提供序列结构信息。子图(e)和(f)中也形成了聚簇，但很难说它们遵循固定的结构。这可能是硬编码图结构和课堂间自然结构相冲突的结果。



(a) Doc2Vec78 (b) Doc2Vec118

(c) Doc2Vec + CSE (d) Doc2Vec + CSRE

(e) Doc2Vec + CGE_N (f) Doc2Vec + CGE_T

Figure 5. t-SNE dimensionality reduction visualization of embedding vectors of each method
图 5. 各方法所得嵌入向量的 t-SNE 降维可视化

5. 实验与分析

本文的目的是通过考虑上下文结构信息以获得更好的在线学习文本的嵌入。本文基于 LSTM/GCN 提出结构嵌入提取(SEE)模块，通过将结构嵌入与传统方法获得的文本嵌入进行结合对文本特征进行强化。在获得结构嵌入向量后：可视化结果表明，在线学习文本嵌入在降维空间中形成的聚簇包含更有意义的结构信息；MeanRank 指标显示在线学习文本之间相邻的关系得到更好地留存；在传统的文本分类任务中，分类精度有所提升。综上，本文提出的基于 LSTM/GCN 的文本特征提取方法在在线学习文本数据上表现优于传统方法，可以作为在线教育学习分析可信的特征来源。在后续研究中，将结合课堂多个方面的信息，如分别从视频、简介文本和音频进行特征提取并集成，从多维特征信息对课堂的特征进行补充和增强。

致　谢

在本次论文的撰写中，我得到了曾安教授和潘丹高工的精心指导，并得到了国家自然科学基金项目和广州市科技计划项目的大力支持，在此表示衷心的感谢。

基金项目

国家自然科学基金项目(61772143, 61300107)，广州市科技计划项目(201804010278)。

参考文献

[1] Chatti, M.A., Dyckhoff, A.L., Schroeder, U., *et al*. (2012) A Reference Model for Learning Analytics. *International Journal of Technology Enhanced Learning*, **4**, 318-331. https://doi.org/10.1504/IJTEL.2012.051815

[2] Yao, L., Mao, C. and Luo, Y. (2019) Graph Convolutional Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 7370-7377. https://doi.org/10.1609/aaai.v33i01.33017370

[3] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[4] Bruna, J., Zaremba, W., Szlam, A., *et al*. (2013) Spectral Networks and Locally Connected Networks on Graphs.

[5] Kenter, T., Borisov, A. and De Rijke, M. (2016) Siamese CBOW: Optimizing Word Embeddings for Sentence Representations. *Proceedings of the* 54*th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 941-951. https://doi.org/10.18653/v1/P16-1089

[6] Arora, S., Liang, Y. and Ma, T. (2016) A Simple but Tough-to-Beat Baseline for Sentence Embeddings. *ICLR* 2017, Toulon, 24-26 April 2017, 1-16.

[7] Mikolov, T., Chen, K., Corrado, G., *et al*. (2013) Efficient Estimation of Word Representations in Vector Space.

[8] Le, Q. and Mikolov, T. (2014) Distributed Representations of Sentences and Documents. *International Conference on Machine Learning*, Vol. 32, 1188-1196.

[9] Pagliardini, M., Gupta, P. and Jaggi, M. (2017) Unsupervised Learning of Sentence Embeddings Using Compositional n-Gram Features. *Proceedings of the* 2018 *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, 528-540. https://doi.org/10.18653/v1/N18-1049

[10] Chen, M. (2017) Efficient Vector Representation for Documents through Corruption.

[11] Vo, A.-D., Nguyen, Q.-P. and Ock, C.-Y. (2020) Semantic and Syntactic Analysis in Learning Representation Based on a Sentiment Analysis Model. *Applied Intelligence*, **50**, 663-680. https://doi.org/10.1007/s10489-019-01540-2

[12] Logeswaran, L. and Lee, H. (2018) An Efficient Framework for Learning Sentence Representations.

[13] Reimers, N. and Gurevych, I. (2019) Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*), Hong Kong, November 2019, 3982-3992. https://doi.org/10.18653/v1/D19-1410

[14] Angelova, R. and Weikum, G. (2006) Graph-Based Text Classification: Learn from Your Neighbors. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, 6-11 August 2006, 485-492. https://doi.org/10.1145/1148170.1148254

[15] Liu, T., Yu, S., Xu, B., *et al*. (2018) Recurrent Networks with Attention and Convolutional Networks for Sentence Representation and Classification. *Applied Intelligence*, **48**, 3797-3806. https://doi.org/10.1007/s10489-018-1176-4

[16] 曾碧卿, 韩旭丽, 王盛玉, 等. 基于双注意力卷积神经网络模型的情感分析研究[J]. 广东工业大学学报, 2019, 36(4): 10-17.

[17] Velickovic, P., Cucurull, G., Casanova, A., *et al*. (2017) Graph Attention Networks.