【电子与信息科学 / Electronics and Information Science】

基于注意力机制的双层 LSTM 自动作文评分系统

夏林中, 罗德安, 刘俊, 管明祥, 张振久, 龚爱平

深圳信息职业技术学院人工智能技术应用工程实验室, 广东深圳 518172

摘要: 研究一种基于新型神经网络结构的自动作文评分模型, 该模型包括双层长短时记忆 (two-layer long short-term memory, LSTM) 神经网络层和注意力机制层, 模型输入层的词向量通过 word embedding 预训练谷歌文本库生成. 相较于基于本地文本数据集预训练, 预训练谷歌文本库生成的词向量含有更丰富的上下文语义信息及依赖关系; 双层长短时记忆网络的下层抽取上下文语义信息及隐藏的上下文依赖关系, 上层捕获更深层次的上下文依赖关系; 注意力机制依据双层长短时记忆网络的输出计算注意力概率, 以突出关键信息在文本中的重要程度. 模型所使用数据集由 Hewlett 基金提供, 并以二次加权 kappa 系数作为模型的评估指标. 实验结果表明, 对比其他基准模型 (如双向 LSTM 模型和 SKIPFLOW-LSTM 模型等), 基于注意力机制的双层 LSTM 模型所获二次加权 kappa 系数平均值最好.

关键词: 人工智能; 自然语言处理; 自动作文评分; 长短时记忆网络; 注意力机制; 二次加权 kappa 系数
中图分类号: TP18 **文献标志码:** A **doi:** 10.3724/SP.J.1249.2020.06559

Attention-based two-layer long short-term memory model for automatic essay scoring

XIA Linzhong, LUO De'an, LIU Jun, GUAN Mingxiang,
ZHANG Zhenjiu, and GONG Aiping

Engineering Applications of Artificial Intelligence Technology Laboratory, Shenzhen Institute of Information Technology,
Shenzhen 518172, Guangdong Province, P. R. China

Abstract: We propose a neural network architecture-based automatic essay scoring model which contains a two-layer long short-term memory (LSTM) and an attention mechanism layer. The Google word vector dataset, which includes the richer word information and contextual information than the local-trained word vector dataset, is used to generate the embedding word vector of the input layer of the model by pre-training. The lower layer of the two-layer LSTM network captures the context semantic information and hidden context dependency, and the upper layer extracts the deeper context dependency. The attention mechanism layer focuses on the information extracted from the upper hidden layer of two-layer LSTM and calculates the attention probability to highlight the importance of key information in the text. The dataset used for automatic essay scoring task is provided by the Hewlett Foundation, and the quadratic weighted kappa coefficient is used as the evaluation index of the model. The experimental results show that the proposed method outperforms other automatic essay scoring baseline models such as bidirectional LSTM, SKIPFLOW-LSTM, and so on, in terms of the value of quadratic weighted kappa coefficient.

Received: 2019-12-26; **Accepted:** 2020-02-12

Foundation: Shenzhen Science and Technology Basic Research Foundation (JCYJ20190808093001772); Engineering Applications of Artificial Intelligence Technology Laboratory of Shenzhen Institute of Information Technology (PT201701); Research Fund of Shenzhen Institute of Information Technology (ZY201708)

Corresponding author: Associate professor XIA Linzhong. E-mail: 43966506@qq.com

Citation: XIA Linzhong, LUO De'an, LIU Jun, et al. Attention-based two-layer long short-term memory model for automatic essay scoring [J]. Journal of Shenzhen University Science and Engineering, 2020, 37(6): 559-566. (in Chinese)

Key words: artificial intelligence; natural language processing; automatic essay scoring; long short-term memory; attention mechanism; quadratic weighted kappa coefficient

一篇高质量作文不仅要关注单词、短语、句子及句法结构的使用,还要注重作文局部关键信息和上下文之间的逻辑结构与关系^[1].要写出一篇高质量作文,大量练习必不可少,由此产生了大量需要即时评价并反馈的作文练习稿,为作文评分员带来极大的工作负担^[2].

为了解决上述问题,研究发明了许多基于机器学习的自动作文评价方法^[3-10].近些年出现的基于深度神经网络(deep neural network, DNN)的自动作文评分(automatic essay scoring, AES)方法,在自动作文评分领域取得了显著成效.为了进一步提升 DNN 处理 AES 任务的性能,研究也提出多种改进的 DNN 方法^[11-19].DONG 等^[20]提出一种基于词向量的双层卷积神经网络(convolutional neural network, CNN)模型,该模型的第 1 个卷积层用来提取句子特征,第 2 个卷积层基于句子向量学习篇章级特征.TAGHIPOUR 等^[21]提出一种在本地数据集上训练词向量的基于双层长短期记忆(two-layer long short-term memory, LSTM)和隐层输出平均值(mean over time, MoT)的 LSTM-MoT 模型,本地数据集仅包含最常用的 4 000 个词汇,其他词汇均被定义为停用词,且该模型以所有 LSTM 隐藏单元输出之和的平均值作为最终作文表示.DONG 等^[22]提出一种基于注意力(attention-based)的 LSTM-CNN 模型,该模型充分利用 CNN 和循环神经网络(recurrent neural network, RNN)的优势,并利用注意力机制(attention mechanism, AM)自动学习不同词或句子在作文中所占的相对权重,结果表明,该模型优于未使用注意力机制的方法,从而证明注意力机制的有效性.TAY 等^[23]提出 SKIPFLOW-LSTM 模型,当模型读取作文的各个词向量时,SKIPFLOW 机制可以捕捉不同时间步对应的 LSTM 隐层输出间的关系,从而为 AES 提供特征依据.CHEN 等^[24]提出一种 Topic-BiLSTM-attention 模型,该模型不仅利用作文的主题相关性特征,还通过双向 LSTM 捕捉作文的上下文依赖关系,更好地为 AES 提供依据.LIANG 等^[25]提出一种连体双向长短期记忆神经网络结构(siamese bidirectional LSTM neural network architecture, SBLSTMA)模型,该模型能同时捕捉到作文的上下文语义和评分标准信息.

本研究为 AES 任务设计一种新颖的 DNN 结构,

该 DNN 结构是一种增强的 LSTM 结构,由 word embedding 层(使用谷歌文本库进行 word embedding 预训练生成词向量 GWV)、双层(two-layer) LSTM 层和 AM 层构成,称为基于注意力机制双层长短期记忆(attention-based two-layer LSTM with Google word vector as the input layer, GAT-LSTM)神经网络自动作文评分模型.使用这种结构的主要考虑如下:①由于本地数据集规模较小,进行 word embedding 预训练生成的词向量(名称为 LWV)在上下文语义信息表达能力上远弱于 GWV,因此,本研究使用 GWV 作为双层 LSTM 层的输入,以提高上下文语义信息的表达能力;②多层神经网络从数据中学习有用特征的能力更强大^[26],因此,双层 LSTM 结构的下层可以抽取上下文语义信息和隐藏的上下文依赖关系,上层可以捕获更深层次的上下文依赖关系;③利用注意力机制增强对作文中关键字词的关注,并为其分配更大的注意权重.本研究主要依据二次加权 kappa 系数分析模型性能.

1 自动作文评分模型

基于 GAT-LSTM 结构的 AES 模型如图 1.该模型由 word embedding 层、双层 LSTM 层、AM 层和 softmax 层组成.Word embedding 层将作文中的单词转换成词向量,并将其作为 LSTM 层的输入;双层 LSTM 层用来抽取上下文语义信息和隐藏的上下文依赖关系;AM 层自适应关注不同局部信息对评分的影响,并依据影响大小为不同局部信息分布相应注意权重,使作文特征在保留最有效局部信息的基础上,最大程度解决信息冗余的问题;softmax 层是分类层,此处不同类别对应相应的作文得分.图 1 中,word embedding 层的 $x_i(t = 1, 2, \dots, L)$ 代表输入的单词, $e_i(t = 1, 2, \dots, L)$ 代表对应单词的词向量;双层 LSTM 层中 h_i^1 代表双层 LSTM 的下层隐藏状态, h_i^2 代表双层 LSTM 的上层隐藏状态;AM 层中 a_i 代表每个隐藏状态的注意力权重系数, h' 代表综合计算后的注意力权重系数;softmax 层的 y_{score} 是模型最后预测的作文分数.

1.1 Word embedding

本研究中 word embedding 层分别生成 LWV 和 GWV 词向量.其中, LWV 通过 word2vec 方法使用

图1 GAT-LSTM 结构图

Fig. 1 The architecture of the GAT-LSTM

skip-gram 模型训练本地数据集获得, LWV 向量维度取 100; GWV 通过 word2vec 方法使用 skip-gram 模型训练谷歌文本库获得, GWV 向量维度取 300.

1.2 长短时记忆神经网络

RNN 适于处理类似文本的序列信息, 其不仅可以对上下文关系建模, 还可处理长度可变的序列信息. 但在实际应用中, 梯度消失导致 RNN 仅有短期记忆, 无法实现信息的长期保存. LSTM 是 RNN 的变体, 通过添加内部的门控机制维持信息的长期保存, 从而解决信息的长期依赖问题.

LSTM 单元包括输入门 i_t 、遗忘门 f_t 及输出门 o_t , 如图 2. LSTM 隐藏层的输入和输出向量分别为 x_t 和 h_t , 记忆单元为 C_t . 输入门用于控制当前输入数据 x_t 流入 C_t 的数量, 即保存到 C_t 中的输入信息数量; 遗忘门控制信息的保留及遗忘, 并以特定方式避免当梯度随时间反向传播时引发的梯度消失问题, 即上一时刻信息 C_{t-1} 对当前时间步对应 C_t 的影响; 输出门可以控制记忆单元 C_t 对当前输出值 h_t 的影响, 即在时间步 t 时控制记忆单元 C_t 的输出信息. 输出 h_t 不仅受 C_t 和 x_t 的影响, 还受 h_{t-1} 的影响; 记忆单元 C_t 不仅受 h_{t-1} 和 x_t 的影响, 还受 C_{t-1} 影响. LSTM 单元的计算方法为

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t * \tanh C_t \quad (6)$$

其中, i_t 是输入信息和过去信息经过输入门的中间计算结果; f_t 是输入信息和过去信息经过遗忘门的中间计算结果; o_t 是输入信息和过去信息经过输出门的中间计算结果; $\tilde{C}_t$ 是输入信息和过去信息经过激活函数 $\tanh$ 所得的中间结果; σ 是 sigmoid 函数; $*$ 是哈达玛积; W 和 U 是神经网络的权重矩阵; b 是神经网络的偏置向量.

双层 LSTM 网络如图 3. 双层 LSTM 层的下层可以抽取上下文语义信息和隐藏的上下文依赖关系, 上层可以捕获更深层次的上下文依赖关系. 下层的隐藏状态 h_t^1 与上层的隐藏状态 h_t^2 进行全连接, 最终可得上层隐藏状态输出为

$$h_t^2 = W_m \cdot h_t^1 + b_m \quad (7)$$

图2 LSTM 单元示意图

Fig. 2 The LSTM cell

其中, W_m 和 b_m 分别是下层 LSTM 和上层 LSTM 之间的网络全连接权重矩阵和偏置向量.

图 3 双层 LSTM 模型示意图

Fig. 3 The two-layer LSTM model

1.3 注意力机制

在自然语言处理中, 基于 LSTM 隐藏状态进行作文分类的方法有两种: 一种是将 LSTM 最终隐藏状态作为作文分类器的输入, 但会丢失部分远离最终隐藏状态的信息; 另一种是将所有时间步的隐藏状态相加后取平均值作为作文分类器的输入, 但无法区分每个时间步的输入信息对作文分类影响的大小. 为此, 本研究提出使用注意力机制学习每个时间步输入信息的重要程度, 从而捕获作文中的关键词语, 并加大这些关键信息的注意力权重.

假设 LSTM 的输入向量为 $x_1, x_2, x_3 \dots x_t$, 对每个向量输入对应的 LSTM 隐藏状态分别是 $h_1, h_2, h_3 \dots h_t$, 则注意力机制的核心思想就是计算某个时间步对应的隐藏状态 h_i 与最终隐藏状态 h' 的相似度 S_i .

$$S_i = \tanh(W_n \cdot h_i + b_n) \quad (8)$$

其中, W_n 和 b_n 为网络的权重矩阵和偏置向量. 每个时间步输入信息的注意力权重 a_i 为

$$a_i = \frac{\exp(S_i * h')}{\sum_{j=1}^t \exp(S_j * h')} \quad (9)$$

其中, h' 是被随机初始化的超参数, 在训练过程中通过学习获得. 加入注意力权重后的 LSTM 隐藏状态向量 v 为

$$v = \sum_{i=0}^t a_i \cdot h_i \quad (10)$$

最后, 将 v 输入分类层 (作文不同得分对应不同类别), 即使用 softmax 函数对作文成绩进行预测, 预测函数为

$$y_{\text{score}} = \text{softmax}(W_s \cdot v + b_s) \quad (11)$$

其中, W_s 和 b_s 为网络权重矩阵和偏置向量.

2 实验

2.1 实验设置

2.1.1 数据集

本研究所使用数据集由 Hewlett 基金会支持的学生作文自动评分大赛 (automated student assessment prize, ASAP) 提供, 将发布的数据集划分为 3 部分: 60% 作为训练集; 20% 作为验证集; 20% 作为测试集. 该数据集由 8 个子集组成, 每个子集具体情况如表 1.

表 1 ASAP 数据集概况

Table 1 Description of ASAP dataset

子集	作文篇数/篇	平均长度/单词数	得分范围/分
1	1 783	350	2 ~ 12
2	1 800	350	1 ~ 6
3	1 726	150	0 ~ 3
4	1 772	150	0 ~ 3
5	1 805	150	0 ~ 4
6	1 800	150	0 ~ 4
7	1 569	250	0 ~ 30
8	723	650	0 ~ 60

2.1.2 评估方式

实验结果及 ASAP 中均使用二次加权 kappa (quadratic weighted kappa, QWK) 系数 κ 进行评估. κ 定义为

$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}} \quad (12)$$

$$w_{i,j} = \frac{(i-j)^2}{(N-1)^2} \quad (13)$$

其中, i 和 j 代表不同的评分; N 为评分等级数; $O_{i,j}$ 表示被评分员 1 评为 i 分以及被评分员 2 评为 j 分的作文篇数, $\{O_{i,j}\}$ 是一个 $N \times N$ 矩阵; 假设评分之间不相关, 是随机生成的, 以相同方法可以构造出另一个 $N \times N$ 的矩阵 $\{E_{i,j}\}$. κ 是评估作文不同评分之间一致性的重要系数, 其取值范围为 $[-1, 1]$, 当 $\kappa = 0$ 时, 表示作文不同评分之间的一致性完全随机; 当 $\kappa = 1$ 时, 表示作文不同评分之间的一致性完全相同.

2.1.3 参数设置

本研究采用 Adam 随机优化器及交叉熵损失函数. 损失函数计算为

$$\text{Loss}_{\text{total}} = - \sum_i \{ y' \ln y_{\text{score}} + (1 - y') \ln (1 - y_{\text{score}}) \} \quad (14)$$

其中, y' 和 y_{score} 分别为作文的真实分数和预测分数. GAT-LSTM 模型的超参数设置见表 2.

2.2 基准模型

采用表 3 的 7 种自动作文评分模型作为基准模型, 对比研究 GAT-LSTM 模型性能. 其中, 所有基准模型均使用 ASAP 提供的数据集; HR1-HR2 代表由 2 个人工评分员进行评分; DP + SCM 模型^[1]使用统计学方法抽取篇章级特征和语义连贯特征实现自动评分; 其他模型基于深度神经网络方法进行自动评分.

表 3 基于 ASAP 数据集的不同模型的实验结果¹⁾

Table 3 Experiment results of all compared models on the ASAP dataset

模 型	子集的二次加权 kappa 系数值								平均值
	1	2	3	4	5	6	7	8	
HR1-HR2	0.721	0.812	0.769	0.851	0.753	0.776	0.720	0.627	0.754
DP + SCM	0.830	0.675	0.680	0.770	0.810	0.800	0.800	0.740	0.760
BiLSTM	0.794	0.625	0.665	0.674	0.776	0.613	0.679	0.506	0.667
LSTM-CNN-Att	0.822	0.682	0.672	0.814	0.803	0.811	0.801	0.705	0.764
SKIPFLOW-LSTM	0.832	0.684	0.695	0.788	0.815	0.810	0.800	0.697	0.764
BiLSTM-Att	0.822	0.681	0.689	0.820	0.819	0.816	0.802	0.678	0.766
Topic-BiLSTM-Att	0.827	0.696	0.691	0.816	0.811	0.823	0.809	0.707	0.773
GAT-LSTM	0.846	0.704	0.720	0.826	0.823	0.833	0.802	0.703	0.782

¹⁾ 加粗数字代表最优实验结果

2.3 实验结果

2.3.1 结果对比

表 3 的所有实验结果均使用二次加权 kappa 系数进行评估, 表中加粗字体代表最好的实验结果. 对比 HR1-HR2 和 DP + SCM, 基于深度神经网络模型的最大优势是无需繁琐的特征工程, 就可以自动学习到相应的作文自动评分特征.

对比 6 种基于深度神经网络的模型可见, GAT-LSTM 模型的 κ 平均值最高; 对于子集 1~6, GAT-LSTM 模型取得最好 κ 值; 相对于 Topic-BiLSTM-Att 模型, GAT-LSTM 模型在子集 1~6 上的 κ 值分别得到 1.9%、0.8%、2.9%、1.0%、1.2% 及 1.0% 的提升. 但在子集 7 和 8 上, Topic-BiLSTM-Att 和 DP + SCM 模型的 κ 值略优于 GAT-LSTM 模型, 主要原因是子集 7 和 8 的作文分值范围太大 (表 1). 可见, GAT-LSTM 模型的 κ 平均值优于其他基准模型, GAT-LSTM 模型自动作文评分的整体性能优于表 3

表 2 超参数设置

Table 2 The hyper-parameter settings

层	参数名	参数值
word embedding	LWV 向量维度	100
	GWV 向量维度	300
LSTM 上层	隐层单元数	256
LSTM 下层	隐层单元数	256
dropout	dropout rate	0.75
其 他	epoches	50 ~ 200
	batch size	64
	学习率	0.000 009

中的基准模型.

2.3.2 GAT-LSTM 模型各部分作用

为分析 GAT-LSTM 模型每个组成部分对模型性能提升的贡献, 分别进行以下对比研究: 以 LWV 为输入的 LSTM (W-LSTM) 模型; 以 LWV 为输入的双层 LSTM (WT-LSTM) 模型; 以 LWV 为输入的基于注意力机制的双层 LSTM (WAT-LSTM) 模型; 以 GWV 为输入的 LSTM (G-LSTM) 模型; 以 GWV 为输入的双层 LSTM (GT-LSTM) 模型. 针对数据集的 8 个子集分别绘制了 G-LSTM、GT-LSTM 和 GAT-LSTM 模型的二次加权 kappa 系数值的对比图, 如图 4. 其中, 子集 1~2 比较的是前 75 个 epoches 的迭代结果; 子集 3~7 比较的是前 50 个 epoches 的迭代结果; 子集 8 比较的是前 100 个 epoches 的迭代结果. 基于上述不同模型所获取的 κ 值如表 4. 可见, 以 GWV 为 LSTM 层输入模型的 κ 值得到提升; 基于注意力机制的双层 LSTM 模型结构明显优

图 4 模型 G-LSTM、GT-LSTM 和 GAT-LSTM 分别在数据集不同子集时二次加权 kappa 系数的异同比较

Fig. 4 The QWK value of each prompt comparison under G-LSTM, GT-LSTM and GAT-LSTM

表4 不同模型的二次加权 kappa 系数值
Table 4 The QWK value under different models

模 型	κ							
	子集 1	子集 2	子集 3	子集 4	子集 5	子集 6	子集 7	子集 8
W-LSTM	0.783	0.665	0.674	0.779	0.755	0.785	0.743	0.594
WT-LSTM	0.823	0.654	0.691	0.784	0.791	0.798	0.767	0.622
WAT-LSTM	0.833	0.667	0.697	0.815	0.799	0.799	0.787	0.681
G-LSTM	0.811	0.692	0.695	0.814	0.777	0.803	0.768	0.632
GT-LSTM	0.836	0.698	0.710	0.815	0.820	0.818	0.797	0.676
GAT-LSTM	0.846	0.704	0.720	0.826	0.823	0.833	0.802	0.703

于单纯双层 LSTM 模型结构。

由图4可见,与G-LSTM和GT-LSTM相比,GAT-LSTM的 κ 值上升速度最快,并能快速达到稳定状态。子集1与2的作文平均字数为350,其对应的迭代次数最大值是75;子集3至7的作文平均字数为150或250,其对应迭代次数最大值是50;子集8的作文平均字数为650,其对应的迭代次数最大值是100。由此推知,当 κ 值上升到稳定状态时,迭代次数与作文字数成正比。同样,子集8的 κ 值上升速度和稳定性比其他子集差,主要是因为子集8所含作文数量最少、作文平均长度最长和评分范围最广(表1)导致。

结 语

本研究的AES模型通过使用GWV、双层LSTM和注意力机制可以很好提升AES性能。与LWV相比,GWV包含更丰富的词汇语义信息和上下文信息,实验结果表明,GWV对模型性能的提升具有重要作用。双层LSTM结构可实现在下层抽取上下文语义信息和隐藏的上下文依赖关系,在上层捕获更深层次的上下文依赖关系。注意力机制可以识别作文中的关键信息,并为这些关键信息分配更多的权重。与基准模型对比,GAT-LSTM模型不仅在大多数数据集的子集上取得最好评分效果,而且其整体评分效果优于基准模型。接下来的研究将着重在如何更有效地进行词向量生成,并设计出更高效、简洁的神经网络结构。

基金项目:深圳市科技计划基础研究资助项目(JCYJ20190808093001772);深圳信息职业技术学院人工智能技术应用工程实验室基金资助项目(PT201701);深圳信息职业技术学院校级科研培育资助项目(ZY201708)

作者简介:夏林中(1980—),深圳信息职业技术学院副教授、博士。研究方向:自然语言处理。E-mail: 43966506@qq.com

引文:夏林中,罗德安,刘俊,等.基于注意力机制的双层LSTM自动作文评分系统[J].深圳大学学报理工版,2020,37(6):559-566.

参考文献 / References:

- [1] DIEGO P, ATKINSON J. Coherence-based automatic essay assessment [J]. IEEE Intelligent Systems, 2018, 33(5): 26-36.
- [2] DIKLI S. An overview of automated scoring of essays [J]. The Journal of Technology, Learning, and Assessment, 2006, 5(1): 1-35.
- [3] ATTALI Y, BURSTEIN J. Automated essay scoring with e-rater@ v. 2.0 [J]. ETS Research Report, 2005, 2004(2): 1-21.
- [4] LARKEY L S. Automatic essay grading using text categorization techniques [C]// Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Melbourne, Australia: ACM, 1998: 90-95.
- [5] WANG Huilan, ZHANG Keliang, LI Yali, et al. A new linguistic feature for automated essay scoring [C]// Proceedings of the 4th International Universal Communication Symposium. Beijing: IEEE, 2010: 340-344.
- [6] CHEN Hongbo, HE Ben, LUO Tiejian, et al. A ranked-based learning approach to automated essay scoring [C]// Proceedings of the Second International Conference on Cloud and Green Computing. Xiangtan, China: IEEE, 2012: 448-455.
- [7] YANNAKOUDAKIS H, BRISCOE T, MEDLOCK B. A new dataset and method for automatically grading ESOL texts [C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, USA: ACL, 2011: 180-189.

- [8] ZUPANC K, BOSNIĆ Z. Automated essay evaluation augmented with semantic coherence measures [C] // Proceedings of the International Conference on Data Mining. Shenzhen, China: IEEE, 2014: 1133-1138.
- [9] CLAUDIA L, MARTIN C. C-rater: automated scoring of short-answer questions [J]. Computers and the Humanities, 2003, 37(4): 389-405.
- [10] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(4/5): 993-1022.
- [11] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. The Journal of Machine and Learning Research, 2011, 12: 2493-2537.
- [12] ALI M A, TAN G Z, HUSSAIN A. Bidirectional recurrent neural network approach for Arabic named entity recognition [J]. Future Internet, 2018, 10(12): 123.
- [13] LIAO Shiyang, WANG Junbo, YU Ruiyun, et al. CNN for situations understanding based on sentiment analysis of twitter data [J]. Procedia Computer Science, 2017, 111: 376-381.
- [14] RAN Xiangdong, SHAN Zhiguang, FANG Yufei, et al. An LSTM-based method with attention mechanism for travel time prediction [J]. Sensors, 2019, 19(4): 861.
- [15] NOWAK J, TASPINAR A, SCHERER R. LSTM recurrent neural networks for short text and sentiment classification [C] // Proceedings of the 16th International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland: Springer, 2017: 553-562.
- [16] DUAN Yanjie, LV Yisheng, WANG Feiyue. Travel time prediction with LSTM neural network [C] // Proceedings of the 19th International Conference on Intelligent Transportation Systems. Rio de Janeiro, Brazil: IEEE, 2016: 1053-1058.
- [17] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks, 2005, 18(5/6): 602-610.
- [18] BIN Yi, YANG Yang, SHEN Fumin, et al. Describing video with attention-based bidirectional LSTM [J]. IEEE Transactions on Cybernetics, 2019, 49(7): 2631-2641.
- [19] YIN Wenpeng, EBERT S, SCHÜTZE H. Attention-based convolutional neural network for machine comprehension [C] // Proceedings of the Workshop on Human-Computer Question Answering. San Diego, USA, 2016: 15-21.
- [20] DONG Fei, ZHANG Yue. Automatic features for essay scoring - an empirical study [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin, USA: ACL, 2016: 1072-1077.
- [21] TAGHIPOUR K, NG H T. A neural approach to automated essay scoring [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin, USA: ACL, 2016: 1882-1891.
- [22] DONG Fei, ZHANG Yue, YANG Jie. Attention-based recurrent convolutional neural network for automatic essay scoring [C] // Proceedings of the 21st Conference on Computational Natural Language Learning. Vancouver, Canada: ACL, 2017: 153-162.
- [23] TAY Y, PHAN M C, TUAN L A, et al. SKIPFLOW: incorporating neural coherence features for end-to-end automatic text scoring [EB/OL]. arXiv, 2017 [2019-12-01]. <https://arxiv.org/abs/1711.04981>.
- [24] CHEN Minping, LI Xia. Relevance-based automated essay scoring via hierarchical recurrent model [C] // Proceedings of the International Conference on Asian Language Processing. Bandung, Indonesia: IEEE, 2018: 378-383.
- [25] LIANG Guoxi, ON B, JEONG D, et al. Automated essay scoring: a Siamese bidirectional LSTM neural network architecture [J]. Symmetry, 2018, 10(12): 682.
- [26] LEE H, GROSSE R, RANGANATH R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C] // Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada: ACM, 2009: 609-616.

【中文责编：方 圆；英文责编：木 柯】