基于序列标注反馈模型的方面信息提取方法

范守祥[1]，姚俊萍[1+]，李晓军[1]，马可欣[2]

（1. 火箭军工程大学 301 教研室，陕西 西安 710025；2. 61486 部队，上海 200072）

摘 要：针对已有方面信息提取方法存在信息利用效率低、易受错误传递影响的问题，提出一种基于编码器-解码器架构的序列标注反馈模型，将文本分类特征分为语义特征、词性特征、依赖特征 3 类，通过双路编解码、门控机制，将语义、词性、标签等信息多次融合，获取并提高最终特征表示能力；为降低错误传递问题对模型性能的不良影响，提出数据增强与反馈方法，将模型判断错误的样本经变换后生成新样本，反馈到训练样本集合中并融入下一轮训练流程，提高模型对各种语言现象的识别接受能力。在两个数据集上进行实验并与现有方法进行对比，实验结果表明，该方法能有效强化信息利用水平，降低错误传递问题的影响，具有更好的性能。

关键词：方面提取；深度学习；数据增强；双路编解码；信息融合

中图法分类号：TP391 **文献标识号**：A **文章编号**：1000-7024（2020）09-2643-07

doi：10.16208/j.issn1000-7024.2020.09.036

Aspect extraction method based on sequential label feedback model

FAN Shou-xiang[1], YAO Jun-ping[1+], LI Xiao-jun[1], MA Ke-xin[2]

(1. No. 301 Faculty, Rocket Force University of Engineering, Xi'an 710025, China;
2. 61486 PLA Troops, Shanghai 200072, China)

Abstract: To solve the problem of low information utilization efficiency and vulnerability caused by error transmission in the existing method of aspect extraction, a sequential label feedback model based on encoder-decoder architecture was proposed. Text classification features were divided into semantic features, part-of-speech features and dependent features, final feature representation was got and enhanced by fusing information such as semantics, part-of-speech, and labels multiple times through dual codec and gate controlling mechanism. To overcome the problem of error transmission, a sample generation mechanism was proposed, and the samples that the model judged wrongly were transformed to generate new samples, they were fed back to the training sample set and integrated into the next training process to improve the model's ability to recognize and accept various language phenomena. Results of experiments on two data sets and comparison with existing methods show that the proposed method can effectively enhance the level of information utilization, reduce the impact of the problem of error transmission, and has better performance.

Key words: aspect extraction; deep learning; data enhancement; dual codec; information fusion

0 引 言

方面信息提取的目标是从给定原始文本序列中提取表征实体、实体属性或反映实体某一侧面的信息。方面信息是情感信息的直接受体，通常为一个词语或短语。比如在笔记本电脑评论“They don't just look good; they deliver excellent performance.”中，“look”和“performance”分别表征了笔记本电脑两个不同侧面且被赋予正向的情感极性，故应作为方面信息提取出来。目前已出现许多不同类型的提取方法，如基于主题模型的方法[1-3]以及基于条件随机场的方法[4,5]等。随着深度学习在各个领域的广泛运用与发展，基于深度学习的方面信息提取方法[6-9]正受到越来越多的关注，并且取得不错的成果。

在使用深度学习方法进行方面信息提取时，常作为序

收稿日期：2020-03-23；**修订日期**：2020-05-13

作者简介：范守祥（1986－），男，辽宁丹东人，硕士研究生，研究方向为智能信息处理技术；＋通讯作者：姚俊萍（1978－），女，陕西渭南人，博士，副教授，硕士生导师，研究方向为信息系统与数据工程；李晓军（1981－），男，河北秦皇岛人，博士，副教授，硕士生导师，研究方向为数据质量技术、信息系统人因工程；马可欣（1990－），女，辽宁沈阳人，硕士，助理工程师，研究方向为大数据与信息系统。E-mail: junpingy200225@163.com

列标注任务，重点是使模型能够自动提取有用特征，比如 Yin 等[10]通过在词向量空间中构建词与词之间的句法依赖路径向量，将句法依赖信息融入到词的最终特征表示中，Ma 等[11]用序列到序列的生成模型将目标词的语义特征、上下文特征、文本整体语义特征以及前词标签特征等融入到目标词最终特征表示中。除上述特征外，目标词以及前词的词性信息对于判断目标词是否为方面信息具有重要价值。比如在上例中，如果知道“performance”为名词，“excellent”为形容词，获取修饰关系，将能够更加高效判断“performance”为方面信息。目前，方面信息提取任务中存在两个突出问题：一是如何将包括词性信息在内的各类信息更有效融入到最终特征表示中，比如当前多数研究是将词性信息嵌入到词向量中，构造包含词性特征的向量空间，但利用效果不明显。二是如何解决错误传递问题。当前大多词性或句法解析系统准确率不高，且评论语言具有随意性，存在省略表达、语法错误、拼写错误等多种问题，会进一步降低词性解析准确率，融入此类特征有可能造成特征混乱，甚至产生负面效果。

本文提出一种基于编码器-解码器架构的序列标注反馈模型，针对上述第一个问题，在 Ma 等[11]提出方法基础上，模拟人对事物做出判断时融入各类信息且不断择优选择的过程，将目标词标签分类特征分成语义特征、词性特征、依赖特征 3 类构成，并利用双路编解码、门控机制等策略，对各类信息进行多次融合，生成依赖词性特征以及不依赖词性特征的两类特征表示，并利用该特征分别进行目标词类别判断，最后在两类判断结果的基础上再次进行结果选择，得出最佳判断结果。针对第二个问题，提出一种数据增强与反馈方法，在训练模型时，利用类似遗传算法的思想，通过对判断错误的样本进行随机删词、随机交换词语位置以及随机交换当前错误样本与其余训练样本中词语 3 种方式，生成新训练样本，融入到下一轮训练过程中，增加拼写错误、语法错误、随机错误以及简化语言表达等语言表达形式的样本数量，弥补训练样本不平衡问题，提高模型对各类语言形式的适应能力。实验结果表明，通过上述模型设计与训练方式的改进，训练效果得到较大提升。

1 相关工作

近年来，基于神经网络的深度学习方法在自然语言处理各领域任务中得到广泛应用，如机器翻译、问答系统、情感分析、文本摘要等等。亦有许多基于深度学习的方法被提出，用以解决方面信息提取任务。在这些方法中，常使用基于 RNN（recurrent neural network）、CNN（convolutional neural network）、注意力机制（attention mechanism）、预训练词嵌入（pretrained word embedding）等模型和技术。基于深度学习方法可以自动捕获词、短语或句子的特征而不需要人为设计。比如 Poria 等[6]将 CNN 引入到方面提取任务中，Liu 等[7]提出一种基于 RNN 的判别模型，将不同语料训练出的词嵌入作为不同类型 RNN 的输入，展示出利用神经网络方法提取方面信息的有效性。

目前使用基于深度学习的方法进行方面信息提取主要有以下 3 种思路。①利用预训练词嵌入本身被编码的语义信息。词的编码表示是使用神经网络处理自然语言问题的基础，早期使用独热编码（one-hot）表示词语的方式不仅造成严重的数据稀疏问题，且无法很好表达词语的含义以及词语之间的关系。通过大量文本资源训练得出的词嵌入能较好保留词语的语义特征，且语义相近的词语在词嵌入空间中位置较为接近，能有效保留词语间的关系。比如 He 等[8]提出一种无监督模型，使用独立的方面词嵌入重构句子表示，并将其与原始句子表示之间的差异作为最小化目标，学习得到方面信息词嵌入，依据方面信息词嵌入与原始词嵌入在同一词嵌入空间中的特点，通过距离找到合理的方面信息。②利用句子中词语间的句法依赖关系。句法依赖信息是句子中不同成分之间的依赖关系，方面信息往往是特定句法依赖结构中的特定成分。比如 Yin 等[10]提出在词嵌入空间中，通过学习依赖路径向量将两个词联系起来，构建包含词嵌入、目标词的线性上下文表示、依赖路径上下文表示等特征，并作为 CRF（conditional random field）的输入提取方面信息。③利用观点信息和方面信息共现特征。方面信息是观点表达对象这一事实表明，方面信息和观点信息经常同时出现在一个完整表达中，缺少其中一项的另一项都不能称为方面信息或者观点信息，利用方面信息和观点信息的共现关系可以加强两者提取效果。比如，Wang 等[9]提出一种多层注意力网络，每一层都使用一对注意力机制分别学习方面信息和观点信息的特征，且两者间相互学习注意力权值；Li 等[12]使用两个带有记忆模块的 LSTM（long short-term memory）网络分别处理方面信息和观点信息抽取，观点和方面之间的关系则通过两个网络的记忆模块间的信息交互建立。

2 基于编码器-解码器架构的序列标注反馈模型

本文提出一种基于编码器-解码器架构的序列标注反馈模型，由编码器、解码器以及数据增强与反馈组件三部分组成。在编码器中，利用双层双向 GRU（gated recurrent unit）网络对词向量与词性向量分别进行编码，得到融合上下文的词与词性的向量特征表示；在解码器中，构建了双路解码组件以及门控信息融合单元，分别对目标词标签进行解码，而后融合两路解码结果并再次解码，得到最终最优标签解码结果；在数据增强与反馈组件中，按照预定样本生成规则，对模型判断错误的样本进行新样本生成，并反馈到下一轮训练样本中。这一部分，将分别对模型中的编码器、解码器中的双路解码组件、双路门控信息融合单元，以及数据增强与反馈组件进行详细描述，模型整体结构如图 1 所示。

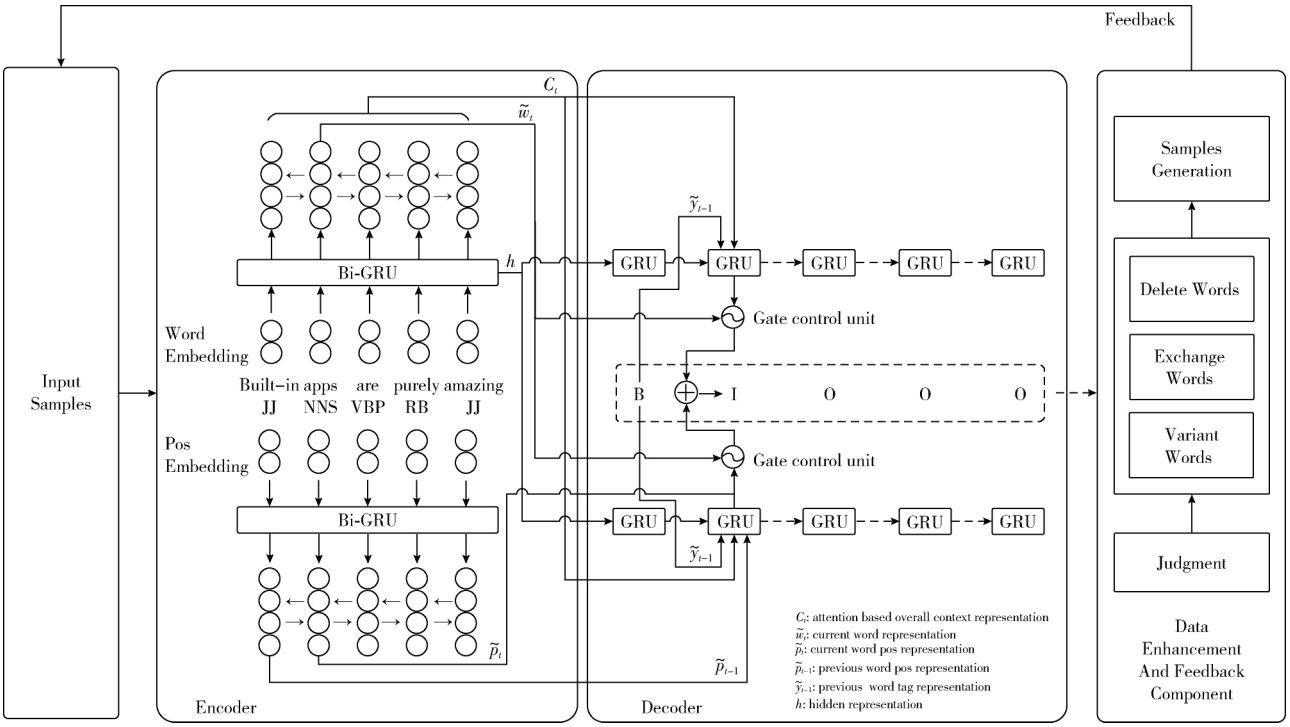

图 1　模型总体结构

2.1　问题描述与符号说明

方面信息提取是在给定含有方面信息及对应情感表达的文本序列

$$S = \{x_1, x_2, \cdots, x_n\} \tag{1}$$

条件下，提取方面信息

$$A = \{x_{n-i}, x_{n-i+1}, \cdots, x_{n-i+k}\} (k \leqslant i) \tag{2}$$

的问题。为便于模型描述，文中使用 $w = \{w_1, w_2, \cdots, w_n\}$ 表示文本序列对应词向量序列，使用 $p = \{p_1, p_2, \cdots, p_n\}$ 表示文本序列对应的词性向量序列，使用 $y = \{y_1, y_2, \cdots, y_n\}$ 表示文本序列对应的标签索引序列，采用 $\{B, I, O\}$ 三级方面信息分类方案，其它符号说明见图 1 注释。

2.2　编码器

在编码器中，使用两个独立的双层双向 GRU 网络分别对词性量和词性向量进行编码。GRU 网络是循环神经网络的一种，可以有效保留序列信息，并且可以解决梯度消失与爆炸问题。编码器依次读取文本序列中每一个词语对应词向量与词性向量，输出目标词的输出向量及隐状态向量

$$\widetilde{w}_t, h_t^e = \mathrm{GRU}(w_t, h_{t-1}^e) \tag{3}$$

$$\widetilde{p}_t, h_t^{e\prime} = GRU(p_t, h_{t-1}^{e\prime}) \tag{4}$$

其中，$h_t^e, h_t^{e\prime} \in R^{n^e}$ 表示目标词及词性的隐状态向量，n^e 表示编码器 GRU 网络隐状态大小，$\widetilde{w}_t, \widetilde{p}_t \in R^{2n^e}$ 表示目标词及词性的输出向量，输出向量作为目标词编码结果输入解码器，文本序列末尾词语的隐状态向量作为解码器的隐状态初始值。在编码器中对词向量以及词性向量进行编码如下考虑：一是获得融合上下文信息的目标词向量表示，使词语的向量表示对不同上下文具有一定的区分能力；二是在使用词性信息时，由于词性信息在解析获得过程中正确率不高，存在错误词性信息问题，通过重新编码融入上下文信息，可以一定程度上稀释词性信息错误，提高模型容错率与鲁棒性。

2.3　双路解码组件

在双路解码组件中，重点解决解码过程中的信息依赖问题，生成目标词的标签依赖特征。直觉上，在目标词分类特征形成过程中，目标词的语义特征、整体语义特征、前词标签特征、前词词性特征等都能够影响目标词的特征表示，在 Ma 等[11]设计的方法中使用了前 3 类特征，词性特征对于解码目标词特征表示具有重要作用，同时这种作用是动态的，在面对不同的语言表达时，词性信息所起的作用强弱不同。所以，基于对词性信息的这一认识，构建了两路解码组件，一路用以形成包含词性信息的依赖特征，另一路则相反，最终通过两路信息融合出能够体现词性信息不同作用的依赖特征表示。在使用编码器-解码器架构时，其突出优点是可以对文本进行预先编码，获得文本的整体语义信息，同时可以作为记忆模块在解码过程中随时调用记忆信息，这些信息可满足解码过程中所需各类信息。整个模块从编码器获取目标词条件下的整体语义信息、词性信息以及隐状态信息，同时输入前词标签信息，得到解码输出信息及隐状态信息

$$o_t^d, h_t^d = \mathrm{GRU}(C_t \oplus \tilde{y}_{t-1}, h_{t-1}^d) \tag{5}$$

$$o_t^{d\prime}, h_t^{d\prime} = \mathrm{GRU}(C_t \oplus \tilde{y}_{t-1} \oplus \tilde{p}_{t-1}, h_{t-1}^{d\prime}) \tag{6}$$

其中，GRU 网络为双层单向网络，$\oplus$表示向量的合并操作，C_t 表示目标词基于注意力的文本整体语义信息，具体构建方法参照文献［11］，$\tilde{y}_{t-1}$ 为前词标签的向量特征，采用随机初始化产生，在训练流程中由标准标签类别信息转换得到，在测试流程中由前词标签分类结果转化得到。o_t，$h_t \in R^{n^d}$ 分别表示解码输出向量与隐状态向量，n^d 表示解码器 GRU 网络隐状态大小。解码输出向量表征了目标词的依赖信息，并用于下一步信息融合。

2.4 双路门控信息融合单元

双路门控信息融合单元重在解决生成目标词特征表示问题，该特征表示包含目标词语义信息、词性信息、依赖信息，3 类信息通过合并门进行融合。同样生成两路分别包含与不包含词性信息的目标词表示特征。这两路表示特征分别进行目标词标签分类判断，得出依赖词性信息与不依赖词性信息的两个一级分类结果，再通过参数矩阵控制两类结果融合，最终产生目标词的二级分类结果，这样建模突出了词性特征对最终结果的影响是一种具有强弱控制的影响。计算过程如下

$$g_{t1}, g_{t2} = \mathrm{softmax}(W_1 \widetilde{w}_t, W_2 o_t^d) \tag{7}$$

$$g'_{t1}, g'_{t2}, g'_{t3} = \mathrm{softmax}(W_1 \widetilde{w}_t, W_3 o_t^{d\prime}, W_4 \widetilde{p}_t) \tag{8}$$

$$r_t = \mathrm{softmax}(W_5(g_{t1} * \widetilde{w}_t + g_{t2} * o_t^d) + b_0) \tag{9}$$

$$r'_t = \mathrm{softmax}(W_6(g'_{t1} * \widetilde{w}_t + g'_{t2} * o_t^{d\prime} + g'_{t3} * \widetilde{p}_t) + b_1) \tag{10}$$

$$P(y_t \mid y_{1:t-1}, w, p) = \mathrm{softmax}(W_9(W_7 r_t + W_8 r'_t) + b_2) \tag{11}$$

其中，$g_t \in R^{n^d}$ 为信息合并门，由目标词语义特征、依赖特征、词性特征以及变化矩阵控制，经过门方向的归一化后得到；$r_t \in R^{n^d}$ 为一级分类结果，经由目标词各 g_t 信息控制门控制信息流入量后计算得到；$P(y_t \mid y_{1:t-1}, w, p)$ 表示目标词最终标签类别概率分布，经两路一级分类结果与相应控制矩阵进一步控制信息融合比例后得到，W_1 -W_9 以及 b_0 -b_2 均为可学习参数。

最终，通过最小化负对数似然损失对整个模型进行训练

$$L(\theta) = -\frac{1}{n}\sum_{i=1}^{n} y'_t \log(P_\theta(y_t \mid y_{1:t-1}, w, p)) \tag{12}$$

式中：y'_t 表示标准标注类别索引值，θ 表示模型中所有可学习参数。纵观整个模型，将决定一个词标签类别的信息分为语义信息、词性信息、依赖信息 3 类，在信息流动过程中，多次进行信息融合，旨在保证模型在判断目标词标签分类结果过程中，对各类信息具有控制与学习的能力。

2.5 数据增强与反馈组件

在利用词性信息进行方面信息提取时，词性解析错误将传递到最终的特征表示中，严重影响模型对词性信息的利用水平，甚至阻碍模型判断准确率的提升。为降低这种错误传递问题对模型的不良影响，提出数据增强与反馈组件。该组件工作于训练流程中，是提升模型对判断错误样本中语言模式理解认识能力、强化模型提取方面信息关键特征能力的重要组件。通过深入观察分析评论文本发现：评论文本在语言表述形式上具有极大灵活性，体现在经常出现简化表达、错误表达等现象。一方面，这些现象是造成词性解析错误的重要原因，另一方面，也应该看到，这类样本虽无法借助词性信息助力得到模型的正确判断，却能够得到人类的正确理解。比如对表 1 中原始样本经过简化、乱序、变异转换后，得到的样本出现表达不完整，存在语法错误等问题，但并不影响对其整体含义的认识以及方面信息的判断。对于评论表达的这个认识启示我们，对于模型判断错误的样本，可以通过人为构造简化、乱序、变异甚至错误的表达生成新的训练样本，增加与错误样本相关语言模式样本数量，加强模型对类似样本语言模式的认识能力，降低模型在处理此类样本时对词性信息的依赖程度。具体的，对训练集中模型判断错误的样本采用如下样本生成机制。

表 1 样本形式变换示例

样式	文本内容
原始样本	They don′t just **look** good; they deliver excellent **performance.**
简化形式	**look** good; excellent **performance.**
乱序形式	They don′t just **look** excellent; they deliver **performance** good.
变异形式	They don′t just **look** cool; they deliver great **performance.**

（1）删除词：随机删除样本中除方面信息以外的词语，通过参数 α 控制删除单词数量比例，模拟样本的简化或错误表达形式；

（2）交换词：随机交换样本中除方面信息以外的词语，通过参数 β 控制交换单词数量比例，模拟样本的乱序或错误表达形式；

（3）变异词：随机将样本中除方面信息以外词语与剩余训练样本中除方面信息以外词语进行交换，通过参数 γ 控制变异单词数量比例，模拟样本的同义或错误表达形式。

通过上述 3 种方式为每个判断错误训练样本生成 3 个新样本，反馈到下一轮训练样本中。由于采用随机样本生成方式经过变换产生的新样本很可能出现无法理解或与原始样本相矛盾的情况，为降低此类生成样本对模型的不良影响，每轮训练仅对排除新增样本的原始训练样本进行样本生成与反馈操作，使模型每轮次训练尽可能学习符合原始文本语言模式的特征，提高模型对判断错误样本相关表述的理解能力。

3 实验部分

3.1 实验数据与设置

本实验沿用目前在方面信息提取任务领域广泛使用的两个数据集，分别是 SemEval-2014 任务 4[13] 中的 Laptop 评论数据集和 SemEval-2016 任务 5[14] 中的 Restaurant 评论数据集，详情参见表 2。所有评论数据均标记有根据字母位置索引定位的方面信息，使用 NLTK（natural language toolkit）将文本分割为单词序列，使用斯坦福大学的自然语言处理工具包 Stanford-CoreNLP 对评论文本进行词性解析，获取单词词性信息，并从训练数据集中随机抽取 150 个训练样本作为验证数据集。

表 2 样本数据信息统计

数据集	训练集		测试集	
	评论数量	方面数量	评论数量	方面数量
Laptop 评论数据集	3045	2358	800	654
Restaurant 评论数据集	2000	1743	676	622

为提高模型通用性与实验可对比性，词向量采用 Glove. 840B. 300 d 预训练词向量初始化，词性向量以及标签类别向量均采用随机初始化，向量维度设定为 300。编码器 GRU 网络隐层大小设定为 300，解码器 GRU 网络隐层大小设定为 600。使用 Adam[15] 优化算法更新模型参数，学习率设定为 0.001。采用 dropout 策略防止参数过拟合，dropout 率设定为 0.5，batch 大小设定为 24。

3.2 对比方法

为充分评估模型的有效性，将与下列方法进行对比：

（1）RNCRF[16]：基于句法依赖树构建递归神经网络，学习词语表示特征，并作为条件随机场的输入特征，判断词语标签类别；

（2）MIN[12]：基于 LSTM 的多任务学习模型，设计有记忆模块，并通过记忆模块间的信息交互同时提取方面和情感信息；

（3）CMLA[9]：耦合多层注意力模型，通过多层注意力机制分别提取方面与情感信息特征，并通过交互学习提升彼此特征表示能力；

（4）Glove-CNN[17]：利用词向量本身蕴含的信息以及多层 CNN 构建简单模型，提取方面信息特征；

（5）HAST[18]：从 LSTM 构建的记忆中提取方面检测历史以及观点摘要信息，并以此强化当前方面信息检测能力；

（6）seq2seq4ATE[11]：基于序列到序列的生成模型，借助文本整体语义特征以及前词标签依赖特征等信息解码目标词标签特征。

3.3 实验结果与分析

本文提出方法与对比方法性能情况见表 3。从表中可以看出，所有方法都存在 Laptop 评论数据测试结果好于 Restaurant 评论数据测试结果情况，这可能与多方面因素有关，比如 Laptop 评论数据中存在大量通用表述的方面信息，比如“Operating System”“Screen”等，这些方面信息较容易被模型习得，而 Restaurant 评论数据中除了一部分通用表述方面信息以外还存在大量低频方面信息，比如具体餐品名称“Lemon Chicken”、“Honey Walnut Prawns”，这些餐品名称在形式上以不规则短语为主，规律性较难被模型掌握；另外样本数据的不平衡性也可能是导致这种现象的原因之一。

表 3 方法性能对比统计（F1/%）

方法	Laptop 评论数据集	Restaurant 评论数据集
RNCRF	78.09	69.72*
MIN	77.58	73.44
CMLA	77.80	72.77*
Glove-CNN	77.67	72.08
HAST	79.52	73.61
seq2seq4ATE	80.31	75.14
OURS	**81.91**	**76.06**

注：①方法性能采用 F1 值进行比对；②带“*”数据取自文献[18]，其余数据取自各自论文中实验结果数据；③本文提出模型的 Laptop 评论数据实验结果产生于样本生成参数 $\alpha=\beta=\gamma=0.3$，Restaurant 评论数据实验结果产生于样本生成参数 $\alpha=0.5$，$\beta=0.4$，$\gamma=0.3$。

在 RNCRF 方法中使用了句法解析以及词性信息，使得 Laptop 评论数据测试结果较好，但是在 Restaurant 评论数据表现出所列方法中最低性能水平，体现出模型对 Restaurant 评论数据识别能力较差，与本文提出方法相比，在两个数据集上分别存在 3.82%和 6.34%的差距，说明模型利用句法及词性信息不够充分，应对句法以及词性解析错误的能力较弱。

MIN、CMLA、HAST 这 3 种方法均利用了方面信息与观点信息之间的依赖关系提升方面信息的提取效果，其中 MIN、CMLA 明确用到显式观点标记信息，HAST 使用了隐含观点信息，3 种方法均表现出不错的效果，但是某些情况下，观点信息可能会导致方面信息的误判，比如在“Most of them were either too big, too noisy and too slow after 2 years”中，评论者并未明确提出方面信息，只是说明了一种普遍的现象，但是其中出现了很多具有强烈负面观点的信息，使模型在利用这些信息时易产生错误判断，此时需要通过文本整体语义做出更准确判断。本文虽未借

鉴上述方法利用观点信息，但是在构造依赖信息时融入了文本整体语义信息，同时利用了目标词语义信息、词性信息等，使整体性能比上述方法有较大提升。

Glove-CNN 利用通用词向量本身的蕴含的语义信息以及 CNN 的特征抽取能力提取方面信息，优势在于模型简单计算高效；seq2seq4ATE 与本文提出模型最为接近，均采用了编码器-解码器的模型架构，不同之处在于本文将词性信息多次融入到最终的特征表示中，且模拟人类进行文本语义理解的方式，融合各方信息综合判断目标词的标签类别，在两个数据集上分别比 seq2seq4ATE 提升 1.6%和 0.92%，取得所有对比方法中最佳性能。

3.4 消融实验

为进一步验证本文提出方法有效性，进行了相关消融实验，主要包括排除样本生成实验（OURS-w/o-SG）、单路解码实验（OURS-w/o-DD）、排除门控机制实验（OURS-w/o-WOG），具体实验结果见表 4。

表 4 消融实验结果

方法	Laptop 评论数据集	Restaurant 评论数据集	样本生成参数
OURS-w/o-SG	78.81	72.72	—
OURS-w/o-DD	81.16	75.74	Laptop：$\alpha=0.5$，$\beta=0.4$，$\gamma=0.3$ Restaurant：$\alpha=0.3$，$\beta=0.4$，$\gamma=0.5$
OURS-w/o-WOG	81.25	74.07	Laptop：$\alpha=0.5$，$\beta=0.4$，$\gamma=0.3$ Restaurant：$\alpha=0.5$，$\beta=0.4$，$\gamma=0.3$
OURS	**81.91**	**76.06**	Laptop：$\alpha=\beta=\gamma=0.3$ Restaurant：$\alpha=0.3$，$\beta=0.4$，$\gamma=0.5$

从表 4 中，可以明显看出，将本文所提出方法简化后，所得到的实验结果在两个实验数据集上均有所下降，表明本文提出方法的各模块均有效提高方面信息提取整体性能，具有重要作用；其次，可以看到，在排除样本生成以后，模型在两个数据集上的实验结果出现较大下降，分别达到了 3.1%、3.34%，可认为将词性信息加入模型后，由于词性解析系统本身正确率不高以及评论语言表达随意性等问题造成的词性解析结果错误，使模型的性能受到很大影响，说明样本生成对模型整体表现具有重要作用；最后，在排除门控机制实验中，Restaurant 评论数据测试结果下降 1.99%，性能降低较为明显，而在相应的 Laptop 评论数据测试中仅下降 0.66%，说明门控机制可以较为有效的应对复杂方面信息识别。

此外，对于本文提出的数据增强与反馈机制，进行了两组实验，分别为 Glove-CNN、seq2seq4ATE 两种方法添加了数据增强与反馈机制，结果显示数据增强与反馈机制在方面信息提取任务中表现出普遍有效性的倾向。具体实验结果见表 5。

表 5 样本生成对比实验结果

方法	Laptop 评论数据集	Restaurant 评论数据集	样本生成参数
Glove-CNN	77.67	72.08	—
Glove-CNN+SG	**80.98**	**74.44**	Laptop：$\alpha=\beta=\gamma=0.3$ Restaurant：$\alpha=0.3$，$\beta=0.4$，$\gamma=0.5$
seq2seq4ATE	80.31	75.14	—
seq2seq4ATE+SG	**80.58**	**75.15**	Laptop：$\alpha=\beta=\gamma=0.3$ Restaurant：$\alpha=0.3$，$\beta=0.4$，$\gamma=0.5$

3.5 数据增强参数分析

通过上述各实验数据可以看出：①3 种样本生成控制参数普遍控制在 0.3-0.5 之间时产生当前条件下最好实验结果，在随机生成条件下，样本生成参数过大会使生成样本严重偏离可能的真实语言现象，而过小会降低生成样本的演化水平；②不同方法、不同数据集下生成最好实验结果的样本生成参数略有不同，可能是由于不同模型对不同数据集样本提取的特征有所不同，且产生的错误样本不同造成的。

4 结束语

本文针对方面信息提取任务，提出一种基于编码器-解码器架构的序列标注反馈模型，并将文本分类特征划分为语义特征、词性特征、依赖特征 3 类特征组成。实验结果表明，本文采用的双路编解码、门控机制等策略对语义、词性、标签等信息的融合起到至关重要的作用，有效提高模型对各类信息的利用水平；所采用的数据增强与反馈机制，强化了模型对错误分类样本的认识能力，同时对模型

提取信息特征起到有益促进作用。在未来工作中，我们将继续对特征的融合表示方式进行研究，并进一步对样本生成机制进行研究，转变随机生成为有限控制生成，探索样本随机生成下的规律性。

参考文献：

[1] PENG Yun，WAN Changxuan，JIANG Tengjiao，et al. Extracting product aspects and user opinions based on semantic constrained LDA model [J]. Journal of Software，2017，28 (3)：676-693 (in Chinese). [**彭云，万常选，江腾蛟，等. 基于语义约束 LDA 的商品特征和情感词提取 [J]. 软件学报，2017，28 (3)：676-693.**]

[2] Asnani K，Pawar JD. Improving coherence of topic based aspect clusters using domain knowledge [J]. Computacion y Sistemas，2018，22 (4)：1403-1414.

[3] Shams M，Baraani-Dastjerdi A. Enriched LDA (ELDA)：Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction [J]. Expert Systems with Applications，2017，80 (SEP)：136-146.

[4] Chernyshevich M. IHS RandD Belarus：Cross-domain extraction of product features using CRF [C] //Proceedings of the 8th International Workshop on Semantic Evaluation，2015：309-313.

[5] Shu L，Xu H，Liu B. Lifelong learning CRF for supervised aspect extraction [C] //ACL 55th Annual Meeting of the Association for Computational Linguistics，Proceedings of the Conference (Long Papers)，2017：148-154.

[6] Poria S，Cambria E，Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network [J]. Knowledge-Based Systems，2016，108 (SEP)：42-49.

[7] Liu P，Joty S，Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings [C] //Conference on Empirical Methods in Natural Language Processing，2015：1433-1443.

[8] He R，Lee WS，Ng HT，et al. An unsupervised neural attention model for aspect extraction [C] //ACL 55th Annual Meeting of the Association for Computational Linguistics，Proceedings of the Conference (Long Papers)，2017：388-397.

[9] Wang W，Pan SJ，Dahlmeier D，et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms [C] //31st AAAI Conference on Artificial Intelligence，2017：3316-3322.

[10] Yin Y，Wei F，Dong L，et al. Unsupervised word and dependency path embeddings for aspect term extraction [C] //IJCAI International Joint Conference on Artificial Intelligence，2016：2979-2985.

[11] Ma D，Li S，Wu F，et al. Exploring sequence-to-sequence learning in aspect term extraction [C] //Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics，2019：3538-3547.

[12] Li X，Lam W. Deep multi-task learning for aspect term extraction with memory interaction [C] //EMNLP Conference on Empirical Methods in Natural Language Processing，2017：2886-2892.

[13] Pontiki M，Galanis D，Pavlopoulos J，et al. SemEval-2014 Task 4：Aspect based sentiment analysis [C] //Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)，2014：27-35.

[14] Pontiki M，Galanis D，Papageorgiou H，et al. SemEval-2016 Task 5：Aspect based sentiment analysis [C] //Proceedings of the 10th International Workshop on Semantic Evaluation，2016：19-30.

[15] Kingma DP，Ba JL. Adam：A method for stochastic optimization [C] //3rd International Conference on Learning Representations，ICLR 2015-Conference Track Proceedings，2015：arXiv：1412. 6980.

[16] Wang W，Pan SJ，Dahlmeier D，et al. Recursive neural conditional random fields for aspect-based sentiment analysis [C] //Conference on Empirical Methods in Natural Language Processing，2016：616-626.

[17] Xu H，Liu B，Shu L，et al. Double embeddings and cnn-based sequence labeling for aspect extraction [C] //ACL 56th Annual Meeting of the Association for Computational Linguistics，Proceedings of the Conference (Long Papers)，2018：592-598.

[18] Li X，Bing L，Li P，et al. Aspect term extraction with history attention and selective transformation [C] //IJCAI International Joint Conference on Artificial Intelligence，2018：4194-4200.