

基于触发词语义选择的 Twitter 事件共指消解研究

魏 萍¹ 巢文涵¹ 罗准辰² 李舟军¹

(北京航空航天大学计算机学院 北京 100191)¹ (军事科学院军事科学信息研究中心 北京 100142)²

摘 要 随着社交媒体的发展与普及,如何识别短文本中事件描述的共指关系已成为一个亟待解决的问题。在传统的事件共指消解研究中,需要从 NLP 工具和知识库中获得丰富的语义特征,这种方式不仅限制了领域的扩展性,而且还导致了误差传播。为了打破上述局限,提出了一种新颖的基于事件触发词来选择性表达句子语义的方法,以判断短文本中事件的共指关系。首先,利用双向长短记忆模型(Bi-LSTM)提取短文本的句子级语义特征和事件描述级语义特征;其次,通过在句子级特征上应用一个基于事件触发词的选择门来选择性表达句子级语义,以产生潜在语义特征;然后,设计了触发词重叠词数和时间间隔两个辅助特征;最后,通过融合以上特征形成一个分类器来预测共指关系。为评估上述方法,基于 Twitter 数据标注了一个新的数据集 EventCoreOnTweets(ECT)。实验结果表明,与两个基准模型相比,提出的选择性表达模型显著提升了短文本共指消解的性能。

关键词 事件共指消解,短文本,双向长短记忆模型,神经网络

中图法分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.12.020

Selective Expression Approach Based on Event Trigger for Event Coreference Resolution on Twitter

WEI Ping¹ CHAO Wen-han¹ LUO Zhun-chen² LI Zhou-jun¹

(School of Computer Science and Engineering,Beihang University,Beijing 100191,China)¹

(Information Research Center of Military Science,PLA Academy of Military Science,Beijing 100142,China)²

Abstract With the development and popularization of social media,how to recognize the coreference relation between two event mention in short texts is an urgent issue. In traditional researches about event coreference resolution,a rich set of linguistic features derived from pre-existing NLP tools and various knowledge bases is required,which restricts domain scalability and leads to the propagation of errors. To overcome these limitations,this paper proposed a novel selective expression approach based on event trigger to explore the coreference relationship on Twitter. Firstly,a bi-direction long short term memory (Bi-LSTM) is exploited to extract the features at sentence level and at mention level. Then,the latent features are generated by applying a gate on sentence level features to make it selectively express. Next,two auxiliary features named the overlapped words of trigger and time interval are designed. Finally,all these features are concatenated and fed into a simple classifier to predict the coreference relationship. In order to evaluate this method,this paper annotated a new dataset EventCoreOnTweet (ECT). The experimental results demonstrate that the selective expression approach significantly improves the performance of coreference resolution of short texts.

Keywords Event coreference resolution,Short text,Bi-direction long short-term memory,Neural networks

1 引言

近年来,诸如 Facebook, Twitter 此类的社交网络已经成为人们发表意见、传播新闻和相互交流的平台。由于人们通常每天都会发布数百条关于重要活动、重大事件的推文,因此亟需能够识别短文本中不同事件描述之间的共指关系的系统,以便更好地整合和利用 Twitter 等社交平台的信息。目

前,大多数解决事件共指消解的方法主要集中在被正式书写的文本语料上,如 MUC、ACE2005 语料库^[1]、ECB^[1]、OntoNotes 和 IC 语料库^[2]。在诸如 Tweet 的社交文本中,通常存在表情符号、URLs、垃圾信息、广告信息、拼写错误等,为我们的任务带来了巨大的挑战。

事件共指消解任务旨在识别描述现实世界中同一事件的文本集合。事件共指消解分为文档内的事件共指消解和跨文

¹⁾ <http://projects.ldc.upenn.edu/ace/>

到稿日期:2018-01-24 返修日期:2018-04-13 本文受国家自然科学基金-青年科学基金项目(61602490),国家重点研发计划:众智化专业知识协同开发技术及应用(2017YFB1402403)资助。

魏 萍(1992—),女,硕士生,主要研究方向为自然语言处理;巢文涵(1979—),男,博士,讲师,CCF 会员,主要研究方向为自然语言处理、机器翻译,E-mail:chaowenhan@buaa.edu.cn(通信作者);罗准辰(1984—),男,助理研究员,主要研究方向为自然语言处理;李舟军(1963—),男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为数据挖掘与人工智能、网络与信息安全。

档的事件共指消解,该任务被广泛应用于许多自然语言处理任务中,包括话题检测与跟踪^[3]、信息抽取^[4]、自动问答^[5]和语义相似度计算^[6]等。

传统的事件共指消解模型通常从多种语言知识库和其他资源抽取丰富的语义特征,这些特征来自于自然语言处理工具的输出,如词性标注、句法分析、语义角色标注等,这种方式限制了领域的可扩展性,并且可能导致误差传播。因此,亟需一种不依赖传统自然语言处理工具的新方法来获取事件的语义特征。从神经网络中获取文本的潜在语义是很有吸引力的解决方案^[7-10]。

与事件共指消解相比,实体共指消解得到了更加广泛的研究。实体共指消解任务主要负责连接和聚合描述同一个真实实体的语段。已有的实体共指消解任务^[11-12]通过结合上下文中的动词和语义角色取得了不错的效果。Rao 等^[13]提出了一种流式聚类算法来识别 Twitter 上的实体共指关系,但事件比实体有着更为丰富的语义表达和更为灵活的语言结构,因此这种方法并不适用于事件共指消解。

为了避免上述局限,本文基于神经网络设计了一个新的面向 Twitter 数据的事件共指消解模型。首先,使用双向长短期记忆模型为 Tweet 文本中的事件描述生成句子级语义特征和事件描述级语义特征;其次,采用选择门结构,根据事件触发词从句子级特征中过滤不重要或不相关的单词的语义表达,再使用注意力机制^[14-15]来为每个词的特征表达赋予不同权重,从而产生更为准确的潜在特征;然后,根据事件触发词的重叠词数和 Tweet 的发布时间构建局部特征;最后,连接潜在特征、事件描述级特征和局部特征作为共指消解分类模型的输入来判断两个事件描述是否具有共指关系。为了测验所提方法的有效性,创建了一个新的语料库 EventCore-OnTweets(ECT),语料标注了事件描述的触发词和共指索引。实验表明了所提选择性表达模型在事件共指消解任务中的有效性。

2 相关工作

2.1 基于聚类的方法

早期的研究工作^[4,16]提出了一种在特定场景中用于事件共指消解的一般方法,如“选举”“间谍”“辞职”等场景。Chen 等^[17]在 ACE¹⁾中正式陈述事件共指消解任务,并提出了凝聚层次聚类算法(Agglomerative Clustering Algorithm)来进行事件共指消解;此外,还探讨了不同特征对事件共指消解模型的影响。Chen 等^[18]针对事件共指提出了一种谱聚类(Spectral Graph Clustering)的解决方案,并且通过真实数据评测了该方法的效果。Liu 等^[19]通过随机森林(Random Forest)训练出分类器,然后通过处理每一对事件描述的共指分数来决定最终的聚类类别。Peng 等^[20]提出了语义相关性函数来发现事件,使用左链接贪婪聚类算法(Left-linking Greedy Algorithm)来挖掘事件描述的共指关系。

2.2 基于贝叶斯的方法

Bejan 等^[1]基于先前的监督学习方法^[4,18]依靠领域内的

大量语言特性和人力来标注数据,提出了两个基于层次狄利克雷过程(Hierarchical Dirichlet Process)^[21]和无限阶乘隐马尔可夫模型(Infinite Factorial Hidden Markov Model)^[22]的生成式无参贝叶斯模型,来进行无监督的文档内和跨文档的事件共指消解。Yang 等^[23]提出了一种层次距离依赖的贝叶斯生成式模型,对可用的监督信息进行编码来指导该模型,以获得更好的效果,即结合了监督方法和无监督方法各自的优势,充分使用基于特征、可学习的距离函数作为聚类先验来扩展距离依赖的中餐馆过程(Distance-dependent Chinese Restaurant Process,DDCRP)^[24]。

2.3 基于回归的方法

已有的很多方法都只是处理事件共指消解任务,Lee 等^[25]注意到实体共指消解中的名词短语可以成为事件,而事件的参与者可以是一个实体,为了探测这种联系给这两个任务带来的影响,其沿用 OntoNotes^[26]语料的标注和标准,修改并完善了 Bejan 等^[1]创建的 EventCoreBank(ECB)语料库,使用线性回归模型训练了一个跨文档的联合处理事件共指和实体共指的消解模型。Araki 等^[27]注意到发现事件及其子事件的关系可以减少共指消解的难度,因此其他引入多类别的逻辑回归模型在解决事件共指消解的同时,还发现和改善了层次子事件的结构。

2.4 基于神经网络的方法

传统的解决方案从大量的知识库中抽取丰富的语义特征,限制了领域扩展性,使用上游 NLP 组件(如词性标注、句法分析等)还会导致传播误差。Krause 等^[10]针对此问题,提出利用卷积神经网络(Convolutional Neural Networks,CNN)来自动抽取句子的语义特征。该模型首先通过处理共指候选文本及各自的上下文来生成潜在的特征表达(Latent Feature Representations);然后将该特征与词级别(Lexical-level)的特征、两个事件描述共有的特征(Pairwise Features)连接起来,作为可训练相似度函数的输入来得到共指分数;最后使用闭包的方式为每个事件描述分配共指链。本文提出的方法使用神经网络来表示 twitter 文本中事件描述的语义特征便是受其工作的启发。

3 问题定义

面向 Twitter 的事件共指消解任务旨在确定 Tweets 中两个事件描述之间的共指关系。事件描述由事件触发词、参与者、时间和地点组成。我们将判断两个事件描述之间是否存在共指关系作为一个二分类任务,给定两个 Tweets(事件描述的上下文)T1 和 T2,以及它们各自的事件触发词作为输入,其中事件触发词可以是不连续的,并且限定在名词、名词短语、动词和动词短语之间,根据分类器的输出来判断 T1 和 T2 是否存在共指关系。如以下两条推文:

T1: @RedPillTweets Hillary’s team is paying crazy people to attack Trump supporters! <https://t.co/nsG8Ay6oHB>

T2: @MSNBC undercover journalists expose DNC paid agitators to cause violence @ trump rallies. video evidence <https://t.co/ZumXQQamMi>

¹⁾ <http://www.nist.gov/speech/tests/ace/>

可以从上述推文中发现 3 个事件描述 (事件描述的触发词由粗体标识,XXX 表示上下文文本):

- EM1:XXX **is paying** XXX **to attack** XXX
- EM2:XXX **expose** XXX
- EM3:XXX **paid** XXX **tocause**XXX

事实上,EM1 和 EM3 具有共指关系,描述的是同一个事件,EM1 和 EM2、EM2 和 EM3 都不具有共指关系,换言之,EM2 是一个单独的事件(Singleton Reference)。

4 模型

本文模型由两部分组成,一部分用于生成 Tweet 中已识别事件触发词的事件描述的语义表示 (见图 1),另一部分用于判断两个事件描述是否具有共指关系 (见图 2)。本节将对此进行详细介绍。

4.1 句子级语义特征和事件描述级语义特征

由于推文通常会出现一些拼写错误、语法错误、缩写、URLs、Hashtags、@someone 等,因此在提取事件描述的语义特征时,首先需要对推文文本进行预处理,包括分词、过滤 URLs、过滤表情符号等。使用 word2vec^[28] 预训练词嵌入,并使用“<unknown>”来标识错误拼写的词和未登录词(Out of Vocabulary,OOV)。通常,与事件触发器距离更近的词具有更相关的语义,因此用距离嵌入来表示词与事件触发器之间的相对距离。综上所述,本文使用推文文本信息、事件描述的触发词及距离信息作为模型输入 (见图 1 中的 Tweet 和触发词层)。

使用一种特殊类型的 RNN,即长短期记忆网络 (Long Short Term Memory Networks,LSTM)^[29],来对每个单词在上下文中的语义表示进行编码。LSTM 的使用在多种任务中都取得了不错的成果^[30-31]。一个 LSTM 结构有 3 个门,分别是遗忘门 (Forget Gate)、输入门 (Input Gate)和输出门 (Output Gate)。遗忘门决定需要被丢弃的信息,输入门决定应该被存储的信息,而输出门决定输出的内容,LSTM 详细的定义

如式(1)一式(6)所示:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

(1)

(2)

(3)

(4)

(5)

(6)

其中, w_f, w_i, w_c 和 w_o 是权重矩阵。

使用单层双向 LSTM(bi-directional LSTM)来学习每个词在上下文中的语义特征,如式(7)一式(9)所示:

$$h_{f,i} = LSTM_f[x_i, h_{f,i-1}]$$
$$h_{b,i} = LSTM_b[x_i, h_{b,i+1}]$$
$$h_i = [h_{f,i}, h_{b,i}]$$

(7)

(8)

(9)

其中, $LSTM_f$ 表示前向 LSTM(Forward LSTM),从左至右读入输入数据。当 x_i 作为输入时, $h_{f,i}$ 是 $LSTM_f$ 的隐层输出,即时刻 i 的输出。如上文所述, x_i 是第 i 个词的词嵌入 (Word Embedding) 和距离嵌入 (Distance Embedding) 连接 (Concatenate)起来的向量。同理, $LSTM_b$ 是反向 LSTM,与前向 LSTM 不同,其从右至左读入输入数据, h_i 是第 i 个词在上下文中的语义特征表示。

句子级的语义特征是指当事件描述所在的 Tweet 文本和距离信息作为输入时,每个词在上下文中语义特征表示的序列 (如图 1 中左边虚线矩形框所示):

$$Sent_{level} = (h_1, h_2, h_3, \dots, h_n)$$

(10)

其中, n 是该 Tweet 句子的长度。

事件描述级的语义特征是指当事件触发词作为输入时,最后一个前向 LSTM 和最后一个反向 LSTM 隐层输出的连接 (如图 1 中右边虚线矩形框所示):

$$Ment_{level} = [h_{f,m}, h_{b,1}]$$

(11)

其中, m 是触发词的长度。

4.2 选择性表达

生成单个事件描述的语义表示如图 1 所示。

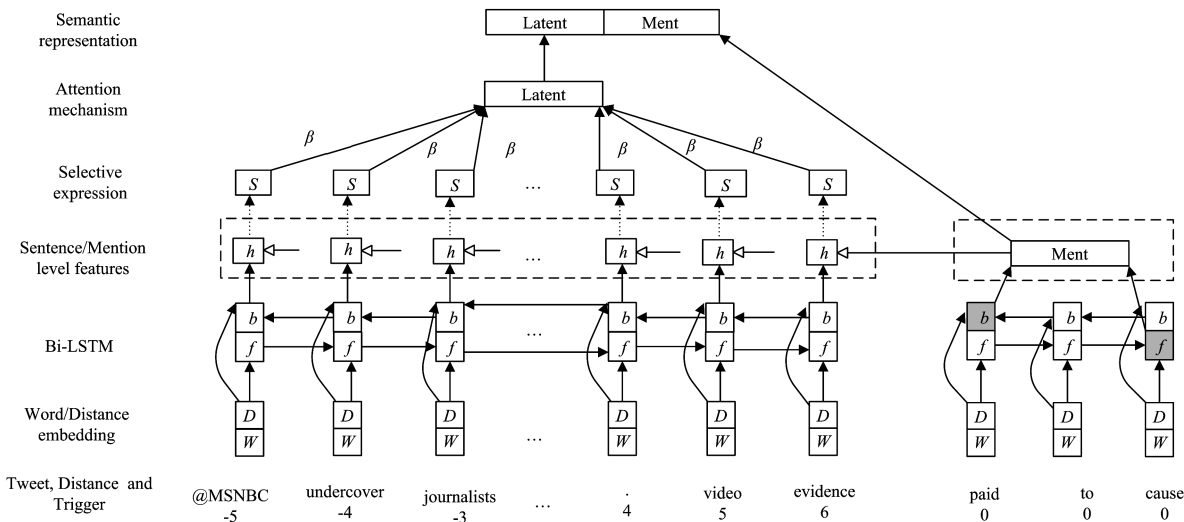


图 1 生成单个事件描述的语义表示

Fig. 1 Semantic representation for a single event mention on Twitter

我们注意到,每个单词在同一个句子中对不同的事件触发词扮演着不同的角色,如核心词、混淆词和无关词等。第 3

节的 T2 包含两个事件描述,分别是 EM2 和 EM3。在 EM2 中,可以忽视掉“@MSNBC”和“video evidence”等词,即删掉

这几个词并不会影响本文语义,同样地,对于 EM3,“@MSN-BC undercover journalists expose”都可以被忽视掉。但是在 EM2 中比较重要的词,如“undercover journalists expose”,可能会给 EM3 带来语义混淆,严重时可能会导致模型判断出 EM2 和 EM3 具有共指关系,这显然是错误的。因此,我们提出选择门结构,针对不同的事件触发词,选择与事件触发词更相关的词使得其语义得到充分表达;限制或过滤掉不相关词、不重要词、混淆词的语义表达,以得到更准确的潜在语义特征(如图 1 中的句子级语义选择性表达层):

$$R_i = h_i * Ment_{level} \quad (12)$$

$$\alpha_i = \tanh(\omega_s \cdot R_i + b_s) \quad (13)$$

其中, ω_s 是选择门的权重矩阵; α_i 是选择门的权重参数,用于指示第 i 个词在上下文中的语义特征能够被表达的数量。对于一个具有 n 个词的推文,权重参数的表达为 $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n)$,而选择性表达是指 α 与 $Sent_{level}$ 的乘积,如式(14)所示:

$$Select = \alpha * Sent_{level} \quad (14)$$

或者:

$$Select_i = \alpha_i * h_i \quad (15)$$

$$Select = (Select_0, Select_1, \dots, Select_n) \quad (16)$$

其中, n 是该推文句子的长度。

4.3 注意力机制

因为上下文中的每个词对事件描述的语义表示都有着不同的贡献,所以我们使用注意力机制来计算每个词选择性表达的重要性分数。将重要性分数标准化后通过加权的方式产生事件描述的潜在特征:

$$u_i = V_a^T \tanh(\omega_a \cdot Select_i + b_a) \quad (17)$$

$$\beta_i = \frac{e^{u_i}}{\sum_{i=1}^n e^{u_i}} \quad (18)$$

$$latent = \sum_{i=1}^n \beta_i Select_i \quad (19)$$

其中, ω_a 和 b_a 是注意力机制的权重矩阵和偏置。将先前的

描述级特征 (Mention Level Feature) 和使用注意力机制获得的潜在特征 (Latent Feature) 连接起来,构成事件描述的语义表示:

$$V_{em} = (latent, Ment_{level}) \quad (20)$$

4.4 共指消解

事件共指消解的主要任务是判断两个事件描述是否具有共指关系,本文模型的第 2 个部分(见图 2)将用来决定两个事件描述的语义表示 V_{em}^1, V_{em}^2 属于二分类中的何种类别。

我们注意到,一个明确的事件发生在某一固定时间,对于重要程度不同的事件,人们连续讨论该事件的时间也不尽相同。对于大多数事件,两个事件描述间隔的天数越长,共指的概率就越低,因此加入了时间间隔这一特征;另外,事件触发词通常属于一类或几类事件类型,将事件描述中包含相同词的个数也作为一个特征。因此,我们增加以下两个特征作为局部特征 (Local Features) 来增强事件描述的语义表达: 1) V_w , 指事件触发词包含相同单词的数量; 2) V_D , 指两个事件描述间隔的天数(事件描述的时间等于用户发布包含这个事件描述的推文的时间)。

定义 V_{em}^1 和 V_{em}^2 成对特征 (Pairwise Features):

$$V_{pair} = (V_{em}^1, V_{em}^2, V_{local}^{1,2}) \quad (21)$$

其中, $V_{local}^{1,2}$ 为 V_{em}^1 和 V_{em}^2 的局部特征,具体计算方式为:

$$V_{local}^{1,2} = (V_w, V_D) \quad (22)$$

然后通过简单的神经网络层对这个特征进行处理,计算相似度 (Distributed Similarity):

$$V_{ds} = \text{relu}(\omega_{ds} \cdot V_{pair} + b_{ds}) \quad (23)$$

将判断两个事件描述是否具有共指关系设计为一个二分类模型;最后,应用一个 *softmax* 层来计算属于这两个类别的概率:

$$Score = \text{softmax}(\omega_{pro} \cdot V_{ds} + b_{pro}) \quad (24)$$

通过使用 Adam^[32] 来最小化交叉熵,以训练本文的模型参数。

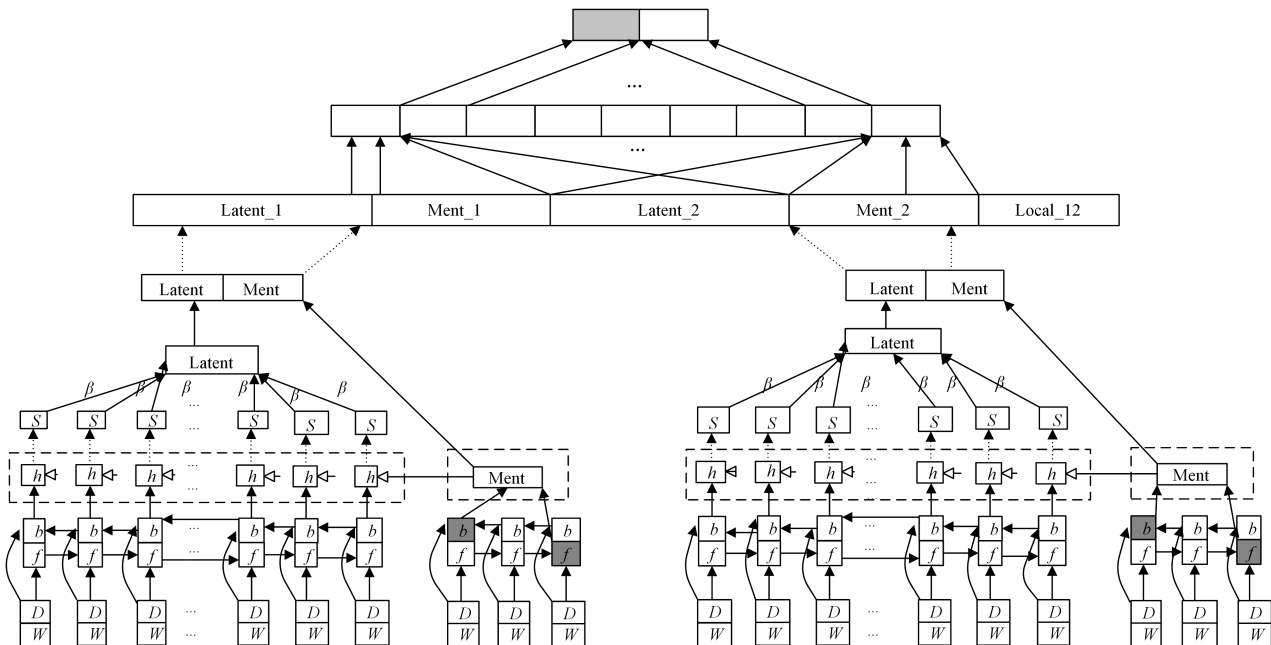


图 2 对于两个事件描述是否存在共指关系的判断

Fig. 2 Judgement of the coreferential decision of two event mentions

5 实验

5.1 数据集

Twitter 流 API 提供所有推文的 1% 的采样。我们从 Twitter 流 API 收集了连续 31 天的推文,过滤掉转发推文和词数少于 10 的推文后共获得了 1000 万条英文推文。美国大选相关话题在 Twitter 平台上得到了用户的高度关注,为了发现这段时间内的事件共指关系,使用关键字 (“trump”“hillary”“presidential election”)来过滤推文。由于标注数据集中的所有事件描述是不现实的,因此我们使用 twitter_nlp¹⁾来对每条推文进行事件抽取,对抽取出的事件触发词进行词性还原后,按频率进行排序。我们从频词中随机选择 20 个词作为我们的语料种子词,然后通过这 20 个词对我们的语料进行二次过滤,共获得 14330 条推文。之后,对过滤后的数据集进行标注,我们要求标注者判断推文是否提到某一具体事件;然后,对于提及事件的推文,即包含事件描述的推文,标注事件描述的事件触发词和共指索引。使用 Cohen’s Kappa 系数^[33]来衡量标注者之间的一致性。针对评判是否描述某一具体事件,两位标注者达到了 0.78 的 Cohen’s Kappa 系数值;针对标注触发词和共指索引,标注者达到了 0.84 的 Cohen’s Kappa 系数值。我们选择标注完全一致的 2994 条推文作为黄金标准语料库 (Corpus of Event Coreference on Twitter, ECT)。ECT 语料集的统计信息如表 1 所列。

表 1 ECT 数据集的统计结果

Table 1 Statistical results of ECT dataset

Data	Total
Tweet	2994
Event	1112
Event mention	3879

5.2 样本生成与聚类

我们参考从文档中基于已识别的事件描述生成样本的方法,来生成基于 Twitter 文本的事件共指消解任务的样本集(见图 3)。按照时间顺序迭代数据集中的事件描述,将每个事件描述与出现在其之前固定滑动窗口的推文组成一对(pair)。例如,若滑动窗口设置为 7 天,当前的事件描述发生在周五,则与其共指的候选事件描述必须发生在上周五以后。成对样本的产生的代码如图 3 所示。

```
1.  $M_t = (m_1, m_2, m_3, \dots, m_{|M_t|})$ , sorted by time
2.  $P_m \leftarrow \emptyset$ 
3. for  $i = 1, \dots, |M_t|$  do
4.   for  $j = i, \dots, |M_t|$  do
5.     if  $\text{Days}(m_i, m_j) \leq \text{Windows}$  then
6.        $P_m \leftarrow (m_i, m_j)$ 
7. return  $P_m$ 
```

图 3 成对样本的产生

Fig. 3 Generation of pair examples

采用类似于 Krause 等^[10]的传递闭包的方式,根据本文

模型判断的共指关系来组成所有的事件类簇。

5.3 评估方法

我们使用准确率(Precision, P)、召回率(Recall, R)和 F₁ 值来评测算法的性能。

$$P = \frac{\text{正确消解对象的数目}}{\text{实际消解对象的数目}}$$

$$R = \frac{\text{正确消解对象的数目}}{\text{应消解对象的数目}}$$

$$F_1 = \frac{2 * P * R}{P + R}$$

CoNLL²⁾给出了用于评测共指算法的计分器(scorer)。该计分器会计算 MUC^[34], B³^[35], BLANC^[36], CEAF_e^[37]和 CoNLL^[38]值,其中 CoNLL 值是 MUC F₁, B³ F₁ 和 CEAF_e F₁ 的平均值。

5.4 实验设置

使用 TensorFlow³⁾框架 v1.2 来实现所提模型,将数据集 ECT 按 9:1 的比例分为开发集(Dev)和测试集(Test),然后进一步将开发集按 9:1 的比例分为训练集(Train)和验证集(Valid)。表 2 列出了分割数据集的统计信息。

表 2 分割 ECT 数据集的统计信息

Table 2 Statistical results of splitting ECT dataset

Data	Train	Vali	Test	Total
Tweet	2425	269	300	2994
Event	—	—	—	1112
Event mention	3157	346	376	3879

采用两个基准模型来检验本文所提方法,一个是所提模型,但不使用选择门、注意机制和局部特征;另一个引用 Krause 的系统,但对于该系统,我们需要做一些必要的修改以适应 Twitter 文本。从论文提出的 pairwise 特征中移除了以下 3 个特征:1)事件类型是否一致;2)事件 modality 是否一致;3)先行语(Antecedent)是否是文档的第一个句子,因为这些特征不适用于我们的数据集。此外,用 Tweet 发布的时间间隔 V_d 替换两个句子之间的距离,将事件参数包含相同词的个数替换为在事件触发词中相同词的个数。其他特征的使用和模型参数与原论文相同。

我们设置不同的下采样率和上采样率进行多次实验,并展示所有实验结果的平均值。表 3 列出了本文进行的所有实验的参数设置。

表 3 参数设置

Table 3 Setting of parameters

参数	值
rnn size	128
batch size	128
decay rate	0.9
learning rate	0.01
attention size	128
word embedding size	100
distance embedding size	14

5.5 实验结果

表 4 中的第 1 行和第 2 行显示了本文基准模型的实验结

1) <https://dev.twitter.com/docs/streaming-api>
2) <https://github.com/conll/reference-coreference-scorers>
3) <https://github.com/tensorflow/tensorflow>

果。其中,“CNN”是 Krause 等提出的模型,如此命名的原因是他们使用的是卷积神经网络(Convolutional Neural Networks,CNN)来抽取事件描述的语义特征;同理,用“RNN”来命名模型,因为我们是从小 LSTMs 中获取语义特征;“RNN”表

示不使用选择门、注意力机制和局部特征,但会使用事件描述级特征的模型;“G”是“Selective Gate”的缩写,指使用选择性门结构;“A”是“Attention Mechanism”的缩写,指使用了注意力机制;“L”是“Local Features”的缩写,指使用了局部特征。

表 4 模型与基准实验在 ECT 上的实验结果
Table 4 Experimental results of model and datum experiment on ECT

(单位:%)

Approach	BLANC			BCUB			CEAF _e			MUC			CoNLL
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
CNN	54.4	65.3	52.8	33.3	77.7	45.5	65.9	26.2	33.2	53.4	72.7	61.1	46.6
RNN	59.2	72.7	62.4	44.8	81.1	57.7	68.1	33.0	44.5	57.7	82.0	67.7	56.6
RNN+G	72.6	79.0	75.3	62.2	76.1	68.4	64.2	50.3	56.4	60.4	70.7	65.2	63.3
RNN+A	64.3	70.1	66.6	55.7	75.6	64.1	71.6	51.1	59.6	61.1	75.1	67.4	63.7
RNN+L	66.2	71.2	68.2	55.4	75.5	63.9	74.4	51.4	60.8	60.5	75.6	67.2	64.0
RNN+G+A	65.9	87.7	1.70	58.9	85.5	69.1	73.7	51.7	60.8	66.1	82.0	73.2	67.7
RNN+G+L	69.7	73.3	71.3	59.4	75.8	66.6	68.9	51.2	58.8	61.5	74.1	67.3	64.2
RNN+A+L	72.8	73.2	73.0	64.0	77.7	70.2	75.8	56.3	64.6	64.8	78.0	70.8	68.5
RNN+G+A+L	68.4	85.1	73.7	60.7	85.4	71.0	78.2	55.4	64.8	68.0	83.9	75.1	70.3

本文 RNN 模型比 CNN 模型在 BLANC F1 上提高了 9.6%,BCUB F1 提高了 12.2%,CEAF_e F1 提高了 11.3%,MUC F1 提高了 6.6%,CoNLL 提高了 10.0%。从这两组实验结果来看,RNN 可以更准确地实现事件触发词与其上下文之间的语义表达。表 4 中的最后一行对应于第 4 节中我们描述的完整模型,从表中可以看出,我们最终的模型分别比两个基准模型在 CoNLL 上提高了 23.7%(CNN)和 13.7%(RNN)。接下来,我们将详细分析选择性门、注意机制和局部特征对模型性能的影响。

表 4 中的第 3—5 行是在 RNN 的基础上只使用我们提出的 3 个特征之一的实验结果,第 6—9 行报告了使用 3 个特征的不同组合的实验结果。对比第 2 行和第 3 行可以看到,选择门的使用比 RNN 基准实验在 CoNLL 上提高了 6.7%,其作用也可以通过比较实验的第 8 行和第 9 行得到验证,这些实验结果说明由选择门提取的潜在特征可以改进本文模型,以获得更好的性能表现。注意力机制的使用在 RNN 基准实验的基础上实现了在 CoNLL 上 7.1%的提高。第 7 行和第 9 行的对比实验表明,为每个组件赋予不同的权重可以更好地表示事件描述的潜在语义。此外,局部特征的使用也在 RNN 的基础上使 CoNLL 提高了 7.4%,注意力机制和局部特征的组合使用也带来了 CoNLL 11.9%的性能提高。从这些结果中可以看到,局部特征扮演着重要的角色,这在很大程度上是因为人们在看到具体事件的报道时更倾向于使用相同的词语来复述事件或表达意见。而且,在一定程度上,我们通过关键词来构建我们的语料库,限制了触发词种类。然而,结合局部特征和选择门(第 7 行)在实验结果上并没有比仅使用局部特征(第 5 行)有较大的改善,这从侧面证明了由选择门过滤的潜在特征需要使用注意力机制来重新组合以实现更好的语义表达。完整的模型获得了 70.3% CoNLL(表 4 中的第 9 行),表明本文方法在特征提取和特征表达方面实现了显著的性能。

结束语 针对短文本事件共指消解任务,本文提出了一种新颖的基于事件触发词的选择性表达方法,其选择与事件

触发词更相关的词使得其在上下文中的语义得到充分表达;限制和过滤掉不相关词、不重要词和混淆词的语义表达,然后使用注意力机制将选择后的语义进行融合。为验证本文方法,基于 Twitter 创建了一个 Twitter 事件共指数据集,数据集标注了每个事件描述的事件触发词和共指索引。与两个基准模型相比,本文提出的选择性表达模型在特征抽取和语义表达上实现了显著的性能提升。未来可以从以下几个方面进行进一步的研究:1)可以将事件描述识别与事件共指消解任务联合起来,完成一个端到端(end-to-end)的事件共指消解系统;2)在大数据背景下,快速地从大量社交网络文本中发掘事件信息是非常有意义且极具挑战的,希望可以将本文模型扩展到流数据处理框架;3)可以提取更细粒度的特征信息,如实体共指消解、URL 指示的文本是否具有相似性来探索本文模型是否能够得到改进。

参 考 文 献

[1] BEJAN C A,HARABAGIU S. Unsupervised event coreference resolution with rich linguistic features[C]//Meeting of the Association for Computational Linguistics. Association for Computational Linguistics,2010:1412-1422.

[2] HOVY E,MITAMURA T,VERDEJO F,et al. Events are not simple:Identity, non-identity, and quasi-identity[OL]. <http://aclweb.org/anthology/w13-1203>.

[3] ALLAN J. Topic Detection and Tracking Pilot Study :Final Report[C]// Proceedings of DARPA Broadcast News Transcription and Understanding Workshop. 1998:194-218.

[4] HUMPHREYS K,GAIZAUSKAS R,AZZAM S. Event coreference for information extraction[C]// A Workshop on Operational Factors in Practical,Robust Anaphora Resolution for Unrestricted Texts. Association for Computational Linguistics,1997: 75-81.

[5] TELLEX S,KATZ B,LIN J,et al. Quantitative evaluation of passage retrieval algorithms for question answering[C]// Inter-

- national ACM SIGIR Conference on Research and Development in Informaion Retrieval. ACM,2003;41-47.
- [6] MCCARTHY D,CARROLL J. Disambiguating Nouns, Verbs, and Adjectives Using Automatically Acquired Selectional Preferences[J]. Computational Linguistics,2003,29(4):639-654.
 - [7] ZENG D,LIU K,LAIS,et al. Relation classification via convolutional deep neural network[C] // Proceedings of COLING 2014,the 25th International Conference on Computational Linguistics; Technical Papers. 2014;2335-2344.
 - [8] NGUYEN T H,GRISHMAN R. Event Detection and Domain Adaptation with Convolutional Neural Networks[C] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2; Short Papers). 2015;365-371.
 - [9] CHEN Y,XU L,LIU K,et al. Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks[C] // The Meeting of the Association for Computational Linguistics. 2015.
 - [10] KRAUSE S,XU F,USZKOREIT H,et al. Event Linking with Sentential Features from Convolutional Neural Networks[C] // Signll Conference on Computational Natural Language Learning. 2016;239-249.
 - [11] HAGHIGHI A,DAN K. Coreference resolution in a modular, entity-centered model[C] // Human Language Technologies; the 2010 Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics,2010;385-393.
 - [12] RAHMAN A,NG V. Coreference Resolution with World Knowledge[C] // The Meeting of the Association for Computational Linguistics; Human Language Technologies. 2011; 814-824.
 - [13] RAO D,MCNAMEE P,DREDZE M. Streaming Cross Document Entity Coreference Resolution[C] // International Conference on Coling 2010. 2010;1050-1058.
 - [14] MNIH V,HEESS N,Graves A. Recurrent models of visual attention[C] // Advances in neural information processing systems. 2014;2204-2212.
 - [15] BAHDANAU D,CHO K,BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473,2014.
 - [16] BAGGA A,BALDWIN B. Cross-document coreference: Annotations,Experiments,and Observations[C] // Proceedings of ACL-99 Workshop on Coreference and Its Applications. 1999;1-8.
 - [17] CHEN Z,JI H,HARALICK R. A pairwise event coreference model,feature impact and evaluation for event coreference resolution[C] // The Workshop on Events in Emerging Text Types. Association for Computational Linguistics,2009;17-22.
 - [18] CHEN Z,JI H. Graph-based event coreference resolution[C] // The Workshop on Graph-Based Methods for Natural Language Processing. Association for Computational Linguistics,2009;54-57.
 - [19] LIU Z,ARAKI J,HOVY E H,et al. Supervised Within-Document Event Coreference using Information Propagation[OL]. http://www.lrec-conf.org/proceedings/lrec-2014/pdf/646_paper.pdf.
 - [20] PENG H,SONG Y,DAN R. Event Detection and Co-reference with Minimal Supervision[C] // Conference on Empirical Methods in Natural Language Processing. 2016;392-402.
 - [21] TEH Y W,JORDAN M I,BEAL M J,et al. Hierarchical Dirichlet Processes[J]. Publications of the American Statistical Association,2006,101(476):1566-1581.
 - [22] GAEL J V,TEH Y W,GHAHRAMANI Z. The infinite factorial hidden Markov model[C] // International Conference on Neural Information Processing Systems. Curran Associates Inc. , 2008;1697-1704.
 - [23] YANG B,CARDIE C,FRAZIER P. A Hierarchical Distance-dependent Bayesian Model for Event Coreference Resolution[J]. arXiv:1504.05929,2015.
 - [24] BLEI D M,FRAZIER P I. Distance Dependent Chinese Restaurant Processes[J]. Journal of Machine Learning Research,2011, 12(1):2461-2488.
 - [25] LEE H,RECASENS M,CHANG A,et al. Joint entity and event coreference resolution across documents[C] // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics,2012;489-500.
 - [26] PRADHAN S S,RAMSHAW L,WEISCHEDEL R,et al. Unrestricted coreference:Identifying entities and events in OntoNotes [C] // International Conference on Semantic Computing. IEEE Computer Society,2007;446-453.
 - [27] ARAKI J,LIU Z,HOVY E H,et al. Detecting Subevent Structure for Event Coreference Resolution[C] // International Conference on Language Resource and Evaluation. 2014;4553-4558.
 - [28] MIKOLOV T,SUTSKEVER I,CHEN K,et al. Distributed representations of words and phrases and their compositionality[C] // International Conference on Neural Information Processing Systems. Curran Associates Inc. ,2013;3111-3119.
 - [29] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.
 - [30] SUTSKEVER I,VINYALS O,LE Q V. Sequence to sequence learning with neural networks[C] // Advances in neural information processing systems. 2014;3104-3112.
 - [31] WU Y,SCHUSTER M,CHEN Z,et al. Google's neural machine translation system:Bridging the gap between human and machine translation [J]. arXiv preprint arXiv:1609.08144, 2016.
 - [32] KINGMA D P,BA J. A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980,2014.
 - [33] COHEN J. A coefficient of agreement for nominal scales[J]. Educational & Psychological Measurement,2016,20(1):37-46.
 - [34] VILAIN M,BURGER J,ABERDEEN J,et al. A Model-Theoretic Coreference Scoring Scheme[C] // Conference on Message Understanding, Muc 1995, Columbia, Maryland, Usa, November. DBLP,1995;45-52.

判别能力具有显著效果。

结束语 本文提出一种将主题模型与递归自编码模型相融合的情感分类模型,该模型将文本的主题信息融入到改进后的递归自编码模型中,在不同维度上对文本信息进行更深层次的分析 and 理解,显著提高了模型在互联网应用环境中对复杂文本的情感分类精度。实验结果表明了该模型的有效性。该模型下一步的发展方向应为进一步细化主题划分,使其对应到具有实际意义的不同领域,以获得更好的解释性。此外,如何提高模型的训练速度和预测精度是今后需要研究的另一个内容。

参 考 文 献

[1] ZHANG Z Q, YE Q, LI Y J, et al. Literature review on sentiment analysis of online product reviews [J]. Journal of Management Sciences in China, 2010, 13(6): 84-96. (in Chinese)
张紫琼,叶强,李一军,等. 互联网商品评论情感分析研究综述[J]. 管理科学学报, 2010, 13(6): 84-96.

[2] 赵军, 许洪波, 黄萱菁. 中文倾向性分析评测技术报告[EB/OL]. <http://www.doc88.com/p-179806395884.html>.

[3] BO P, LEE L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales[C]// Meeting on Association for Computational Linguistics. 2005: 115-124.

[4] ZHOU S C, QU W T, SHI Y Z, et al. Overview on sentiment analysis of Chinese microblogging[J]. Computer Applications and Software, 2013, 30(3): 161-164. (in Chinese)
周胜臣, 瞿文婷, 石英子, 等. 中文微博情感分析研究综述[J]. 计算机应用与软件, 2013, 30(3): 161-164.

[5] TURNEY P D. Thumbs up or thumbs down: semantic orientation applied to unsupervised classification of reviews[C]// Meeting on Association for Computational Linguistics. 2002: 417-424.

[6] LUO Y, LI L, TAN S B, et al. Sentiment analysis on Chinese Micro-blog corpus[J]. Journal of Shandong University (Natural Science), 2014, 49(11): 1-7. (in Chinese)
罗毅, 李利, 谭松波, 等. 基于中文微博语料的情感倾向性分析[J]. 山东大学学报(理学版), 2014, 49(11): 1-7.

[7] WANG S, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification[C]// Meeting of the Association for Computational Linguistics: Short Papers. Association for Computational Linguistics, 2012: 90-94.

[8] SOCHER R, PENNINGTON J, HUANG E H, et al. Semi-supervised recursive auto-encoders for predicting sentiment distributions[C]// Conference on Empirical Methods in Natural Language Processing. DBLP, 2011: 151-161.

[9] TANG D, QIN B, LIU T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2015: 1422-1432.

[10] ZHANG L, CHEN C. Sentiment Classification with Convolutional Neural Networks: An Experimental Study on a Large-Scale Chinese Conversation Corpus[C]// International Conference on Computational Intelligence and Security. IEEE, 2017: 165-169.

[11] LIANG J, CHAI Y M, YUAN H B, et al. Deep Learning for Chinese Micro-blog Sentiment Analysis[J]. Journal of Chinese Information Processing, 2014, 28(5): 155-161. (in Chinese)
梁军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析[J]. 中文信息学报, 2014, 28(5): 155-161.

[12] TANG D, WEI F, YANG N, et al. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification[C]// Meeting of the Association for Computational Linguistics. 2014: 1555-1565.

[13] LEVY O, GOLDBERG Y. Neural word embedding as implicit matrix factorization[J]. Advances in Neural Information Processing Systems, 2014, 3(4): 2177-2185.

[14] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(6): 993-1022.

[15] GRIFFITHS T L, STEYVERS M. Finding scientific topics[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(1): 5228.

[16] ZHANG S, ZHANG C, YOU Z, et al. Asynchronous stochastic gradient descent for DNN training[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 6660-6663.

[17] XIAO H, MAO M Y. NLPir Chinese character segmentation system based Chinese character segmentation tool: CN 106354714 A[P]. 2017. (in Chinese)
肖红, 毛明扬. 一种基于 nlpir 中文分词系统的中文分词工具: CN 106354714 A[P]. 2017.

[18] ORDENTLICH E, YANG L, FENG A, et al. Network-Efficient Distributed Word2vec Training System for Large Vocabularies [J/OL]. <https://arxiv.org/abs/606.08495>.

(上接第 136 页)

[35] BAGGA A, BALDWIN B. Algorithms for scoring coreference chains[C]// The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference. 1998: 563-566.

[36] RECASENS M, HOVY E, BLANC. Implementing the Rand index for coreference evaluation[J]. Natural Language Engineering, 2011, 17(4): 485-510.

[37] LUO X. On coreference resolution performance metrics [C]//

HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, Vancouver, British Columbia, Canada. DBLP, 2005: 25-32.

[38] PRADHAN S, LUO X, RECASENS M, et al. Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation[C]// Meeting of the Association for Computational Linguistics. 2014: 30.