**DOI**:10.16452/j.cnki.sdkjzk.2020.06.015 **文章编号**:1672-3767(2020)06-0115-09
**引用格式**:赵华,邹若飞. 基于 Tree LSTM+CRF 的属性级观点挖掘[J].山东科技大学学报(自然科学版), 2020,39(6):115-122.
ZHAO Hua,ZOU Ruofei.Tree LSTM+CRF for aspect-level opinion mining [J].Journal ofShandong University of Science and Technology(Nature Science),2020,39(6):115-122.

# 基于 Tree LSTM+CRF 的属性级观点挖掘

**赵 华,邹若飞**
(山东科技大学 计算机科学与工程学院,山东 青岛 266590)

**摘 要**:评价对象与观点内容的提取是观点挖掘中非常重要的任务。本研究提出了一个树结构长短期记忆网络(Tree LSTM)结合条件随机场(CRF)的联合模型抽取评价对象和观点内容。首先对评论句进行依存句法分析,根据句子的依存分析树构建 Tree LSTM,并设计树结构下 LSTM 单元的计算方法;接着将 Tree LSTM 的输出作为 CRF 的输入进行序列标注,实现评价对象与观点内容的抽取。最后在 SemEval Challenge 2014 任务 4 的数据集上对模型性能进行了验证,评价对象和观点内容抽取结果的平均 F1 值在餐馆和笔记本电脑领域分别为 86.76%、83.22%和 79.86%、80.42%,优于现有的评价对象和观点内容抽取方法。实验结果表明,设计的 Tree LSTM 能很好地学习词语之间的层次关系,同时联合模型有效避免了传统 CRF 需要构建特征工程的弊端。

**关键词**:观点挖掘;评价对象抽取;观点内容抽取;树结构长短期记忆网络;条件随机场
**中图分类号**:TP183 **文献标志码**:A

## Tree LSTM+CRF for aspect-level opinion mining

ZHAO Hua, ZOU Ruofei
(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

**Abstract**: The extraction of aspect terms and opinion terms is a significant task in opinion mining. In this paper, a tree-structured long short-term (Tree LSTM) memory network combined with Conditional Random fields (CRF) is proposed to extract aspect terms and opinion terms. Initially, dependency parsing of commentary sentences is carried out, Tree LSTM is built according to the dependency parsing tree of sentences, and calculation method of LSTM unit under the tree structure is designed. Afterwards, the output of Tree LSTM is tagged as the input of CRF to realize the extraction of aspect terms and opinion terms. This paper validates the performance of the model on the data set of SemEval Challenge 2014 Task 4. The average F1 scores of aspect terms and opinion terms extraction results are 86.76%, 83.22% and 79.86% and 80.42% respectively in restaurants and laptops, which are superior to the existing aspect terms and opinion terms extraction methods. The experimental results show that the Tree LSTM design presented in this paper can learn the hierarchical relationship between words well, and the joint model effectively avoids the drawbacks of traditional CRF which needs to construct feature engineering.

**Key words**: opinion mining; aspect terms extraction; opinion terms extraction; tree LSTM; CRF

随着互联网的发展,互联网上的评论信息越来越多,属性级观点挖掘因能够挖掘出评论中包含的各个评

**收稿时间**:2019-11-12
**基金项目**:青岛市哲学社会科学规划研究项目(QDSKL1901124);教育部人文社会科学研究青年基金项目(16YJCZH154)
**作者简介**:赵 华(1980—),女,山东泗水人,副教授,博士,主要从事智能信息处理等方面的研究.
E-mail: huamolin@163.com
邹若飞 (1993—),男,山东泰安人,硕士研究生,主要从事人工智能等方面的研究.E-mail: crysisstuart@163.com

价对象及观点内容、提取评论句中有价值的信息、快速准确地得出用户的关注点而备受关注。观点由属性(aspect)、观点内容(opinion)、持有者(holder)及情感(sentiment)组成,评价对象(aspect terms)是出现在评论句中涉及属性的单词或词组。例如:在评论句"The service of the restaurant is good, but the food tastes general."中,service 和 food 是具体的评价对象,good 和 tastes general 是其对应的观点内容。本研究的目的是抽取评论中包含的评价对象和观点内容。

属性级观点挖掘最早由 Hu 等[1-2]提出,而后引起了诸多研究者的关注。目前,常用的属性级观点挖掘方法可以分为无监督学习方法和有监督学习方法。无监督学习方法中,Hu 等[1]对数据词性标注后以 Apriori 算法进行关联规则挖掘找到频繁名词及名词短语,然后对错误词语进行剪枝后得到要抽取的评价对象。Popescu 等[3]在文献[1]的基础上,将 PMI(point-wise mutual information)加入剪枝策略中,计算频繁项与预定义的判别短语的 PMI 值,确定是否为要抽取的评价对象。刘鸿宇等[4]根据依存句法模板和规则抽取频繁项,通过剪枝处理得到要抽取的评价对象。江腾蛟等[5]提出了基于浅层语义与语法分析相结合的评价搭配抽取方法。廖祥文等[6]利用词对齐模型抽取候选评价对象与评价搭配组合,建立多层情感关系图,利用随机游走方法计算置信度,最后选取置信度高的候选评价对象与观点内容作为输出。这些无监督方法相对来说可操作性强,无需大量标注数据,但人工干预过多,需要提前建立模板,适用于目标领域较小的数据集。有监督学习方法中,常采用 PLSA(probabilistic latent semantic analysis)和 LDA(latent dirichlet allocation)等主题模型;而另外一些研究者将该任务看作是文本序列标注问题,采用隐马尔可夫模型(hidden Markov model, HMM)和条件随机场(conditional random fields, CRF)等方法。Wei 等[7]先建立词集对评论文本进行标注,再使用 HMM 进行训练,抽取评价对象和观点内容并判断极性。刘全超等[8]利用 CRF,选择句法特征、语法特征、语义特征及相对位置特征,抽取评价对象与观点内容的搭配。丁晟春等[9]采用 CRF 选取词、词性、情感词以及本体四个特征抽取评价对象。这类有监督的方法准确率较高,但由于需要大量人为设计特征,所以领域局限性强。

最近的研究中,研究者们开始尝试基于深度学习方法的属性级观点抽取方法。Irsoy 等[10]使用深层双向循环神经网络抽取观点内容。Liu 等[11]提出使用循环神经网络(recurrent neural network, RNN)结合词向量的方式抽取评价对象和观点内容。Yin 等[12]提出一种无监督的方法,利用循环神经网络学习融合依存路径信息的词向量,然后用词向量作为 CRF 的特征来抽取评价对象。Wang 等[13]提出基于注意力的 LSTM 模型进行属性级的情感分类。Giannkopoulos 等[14]提出 B-LSTM(bidirectional long short-term memory),结合 CRF 的分类器从有监督和无监督两类研究方向抽取评价对象信息。Wang 等[15]提出一种名为 RNN-CRF(recursive neural conditional random fields)的联合模型抽取评价对象和观点内容,首先根据句子的依存句法关系构建依存树递归神经网络,将评价对象与观点内容的信息编码到递归神经网络(recursive neural network, RNN)中学习更高级的隐层表示,然后将结果输入 CRF 进行序列标注。

本研究提出一个树结构长短期记忆网络(tree-structured long short-term memory networks, Tree LSTM),结合条件随机场的联合模型来抽取评价对象和观点内容,在很好地表征词语之间的层次关系的同时,有效避免传统 CRF 需要大量人工定义特征并且编写特征模板的弊端。

## 1 Tree LSTM+CRF

以评论"iPhone is pretty good."为例,本研究提出的联合模型如图 1 所示。模型共分为三层,底层是各词的词向量,中间层是 Tree LSTM 模块,顶层是 CRF 模块。

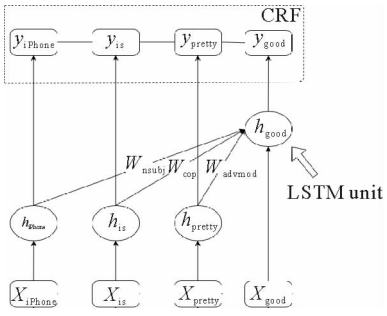



**图 1 Tree LSTM+CRF 联合模型结构**

Fig. 1 Tree LSTM+CRF joint model structure

### 1.1 Tree LSTM

为了更好地理解本文模型,首先给出 Tree LSTM 中包含的各个参数含义(表 1)。其中 $v$ 为词典大小,包含所有在评论语句中出现的词,$d$ 为词向量维度。

构建 Tree LSTM 的过程:

1) 对所有评论语句进行依存句法分析,得到每个句子的依存分析树。

2) 按依存分析树的结构,以 LSTM 单元为节点为每个句子生成 Tree LSTM 模块。

图 2 为根据依存句法分析得到的例句的依存分析树和 Tree LSTM。下面将基于依存句法关系,由叶子节点到内部节点逐个计算各节点的隐向量。以图 2(b)为例,首先计算叶子节点单词"iPhone"的隐向量值:

**表 1　Tree LSTM 各参数代表内容**

Tab. 1　Tree LSTM parameters representing content

| 参数名 | 含义 |
|---|---|
| $\boldsymbol{W}_e \in \mathbf{R}^{d\times v}$ | 词向量矩阵,每个词的 $d$ 维词向量都存储于此 |
| $\boldsymbol{W}_r \in \mathbf{R}^{d\times d}$ | 依存关系矩阵的统一表达形式,$r$ 为具体的依存关系 |
| $\boldsymbol{W}_i,\boldsymbol{U}_i \in \mathbf{R}^{d\times d},\boldsymbol{b}_i \in \mathbf{R}^{d\times 1}$ | LSTM 输入门 $i$ 的权重参数及偏置项 |
| $\boldsymbol{W}_f,\boldsymbol{U}_f \in \mathbf{R}^{d\times d},\boldsymbol{b}_f \in \mathbf{R}^{d\times 1}$ | LSTM 遗忘门 $f$ 的权重参数及偏置项 |
| $\boldsymbol{W}_u,\boldsymbol{U}_u \in \mathbf{R}^{d\times d},\boldsymbol{b}_u \in \mathbf{R}^{d\times 1}$ | LSTM 候选值 $u$ 的权重参数及偏置项 |
| $\boldsymbol{W}_o,\boldsymbol{U}_o \in \mathbf{R}^{d\times d},\boldsymbol{b}_o \in \mathbf{R}^{d\times 1}$ | LSTM 输出门 $o$ 的权重参数及偏置项 |

$$\boldsymbol{i}_{\text{iPhone}} = \text{sigmoid}(\boldsymbol{W}_i \cdot \boldsymbol{x}_{\text{iPhone}} + \boldsymbol{b}_i), \tag{1}$$

$$\boldsymbol{u}_{\text{iPhone}} = \tanh(\boldsymbol{W}_u \cdot \boldsymbol{x}_{\text{iPhone}} + \boldsymbol{b}_u), \tag{2}$$

$$\boldsymbol{c}_{\text{iPhone}} = \boldsymbol{i}_{\text{iPhone}} \odot \boldsymbol{u}_{\text{iPhone}}, \tag{3}$$

$$\boldsymbol{o}_{\text{iPhone}} = \text{sigmoid}(\boldsymbol{W}_o \cdot \boldsymbol{x}_{\text{iPhone}} + \boldsymbol{b}_o), \tag{4}$$

$$\boldsymbol{h}_{\text{iPhone}} = \boldsymbol{o}_{\text{iPhone}} \odot \tanh(\boldsymbol{c}_{\text{iPhone}}). \tag{5}$$

其中,$\odot$ 代表逐元素乘积,参数含义如表 1 所示,在经过 LSTM 多个门计算之后即可得出单词"iPhone"的隐向量值 $\boldsymbol{h}_{\text{iPhone}}$。其他叶子节点的隐向量值同样方法计算得到。例如单词"is"的隐向量值计算如下:

$$\boldsymbol{i}_{\text{is}} = \text{sigmoid}(\boldsymbol{W}_i \cdot \boldsymbol{x}_{\text{is}} + \boldsymbol{b}_i), \tag{6}$$

$$\boldsymbol{u}_{\text{is}} = \tanh(\boldsymbol{W}_u \cdot \boldsymbol{x}_{\text{is}} + \boldsymbol{b}_u), \tag{7}$$

$$\boldsymbol{c}_{\text{is}} = \boldsymbol{i}_{\text{is}} \odot \boldsymbol{u}_{\text{is}}, \tag{8}$$

$$\boldsymbol{o}_{\text{is}} = \text{sigmoid}(\boldsymbol{W}_o \cdot \boldsymbol{x}_{\text{is}} + \boldsymbol{b}_o), \tag{9}$$

$$\boldsymbol{h}_{\text{is}} = \boldsymbol{o}_{\text{is}} \odot \tanh(\boldsymbol{c}_{\text{is}}). \tag{10}$$

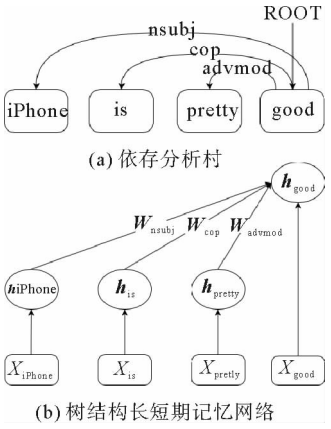



**图 2　例句的依存分析树及其生成的 Tree LSTM**

Fig. 2　Dependent analysis tree for example sentences and its generated Tree LSTM

当所有叶子节点计算完毕后根据依存关系计算内部节点的值,单词"good"的隐向量值计算如下:

$$\boldsymbol{i}_{\text{good}} = \text{sigmoid}(\boldsymbol{W}_i \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_i \cdot \boldsymbol{W}_{\text{nsubj}} \boldsymbol{h}_{\text{iPhone}} + \boldsymbol{U}_i \cdot \boldsymbol{W}_{\text{cop}} \cdot \boldsymbol{h}_{\text{is}} + \boldsymbol{U}_i \cdot \boldsymbol{W}_{\text{advmod}} \cdot \boldsymbol{h}_{\text{pretty}} + \boldsymbol{b}_i), \tag{11}$$

$$\boldsymbol{f}_{\text{good-iPhone}} = \text{sigmoid}(\boldsymbol{W}_f \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_f \cdot \boldsymbol{W}_{\text{nsubj}} \cdot \boldsymbol{h}_{\text{iPhone}} + \boldsymbol{b}_f), \tag{12}$$

$$\boldsymbol{f}_{\text{good-is}} = \text{sigmoid}(\boldsymbol{W}_f \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_f \cdot \boldsymbol{W}_{\text{cop}} \cdot \boldsymbol{h}_{\text{is}} + \boldsymbol{b}_f), \tag{13}$$

$$\boldsymbol{f}_{\text{good-pretty}} = \text{sigmoid}(\boldsymbol{W}_f \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_f \cdot \boldsymbol{W}_{\text{advmod}} \cdot \boldsymbol{h}_{\text{pretty}} + \boldsymbol{b}_f), \tag{14}$$

$$\boldsymbol{u}_{\text{good}} = \tanh(\boldsymbol{W}_u \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_u \cdot \boldsymbol{W}_{\text{nsubj}} \cdot \boldsymbol{h}_{\text{iPhone}} + \boldsymbol{U}_u \cdot \boldsymbol{W}_{\text{cop}} \cdot \boldsymbol{h}_{\text{is}} + \boldsymbol{U}_u \cdot \boldsymbol{W}_{\text{advmod}} \cdot h_{\text{pretty}} + \boldsymbol{b}_u), \tag{15}$$

$$\boldsymbol{c}_{\text{good}} = \boldsymbol{i}_{\text{good}} \odot \boldsymbol{u}_{\text{good}} + \boldsymbol{f}_{\text{good-iPhone}} \odot \boldsymbol{c}_{\text{iPhone}} + \boldsymbol{f}_{\text{good-is}} \odot \boldsymbol{c}_{\text{is}} + \boldsymbol{f}_{\text{good-pretty}} \odot \boldsymbol{c}_{\text{pretty}}, \tag{16}$$

$$\boldsymbol{o}_{\text{good}} = \text{sigmoid}(\boldsymbol{W}_o \cdot \boldsymbol{x}_{\text{good}} + \boldsymbol{U}_o \cdot \boldsymbol{W}_{\text{nsubj}} \cdot \boldsymbol{h}_{\text{iPhone}} + \boldsymbol{U}_o \cdot \boldsymbol{W}_{\text{cop}} \cdot \boldsymbol{h}_{\text{is}} + \boldsymbol{U}_o \cdot \boldsymbol{W}_{\text{advmod}} \cdot \boldsymbol{h}_{\text{pretty}} + \boldsymbol{b}_o), \tag{17}$$

$$\boldsymbol{h}_{\text{good}} = \boldsymbol{o}_{\text{good}} \odot \tanh(\boldsymbol{c}_{\text{good}}). \tag{18}$$

在计算内部节点隐向量值时,输入该节点的除了该词词向量外,还有该词与其多个子节点的依存关系信息。每个子节点都会有一个遗忘门去处理该子节点传来的信息,经过 LSTM 多个门单元计算后即得出此内部节点的隐向量值,内部节点的一般计算公式总结如下:

$$\boldsymbol{i}_j=\text{sigmoid}(\boldsymbol{W}_i\cdot\boldsymbol{x}_j+\boldsymbol{U}_i\cdot\sum_{k\in C(j)}\boldsymbol{W}_r(jk)\cdot\boldsymbol{h}_k+\boldsymbol{b}_i)\,, \tag{19}$$

$$\boldsymbol{f}_{jk}=\text{sigmoid}(\boldsymbol{W}_f\cdot\boldsymbol{x}_j+\boldsymbol{U}_f\cdot\boldsymbol{W}_r(jk)\cdot\boldsymbol{h}_k+\boldsymbol{b}_f)\,, \tag{20}$$

$$\boldsymbol{u}_j=\tanh(\boldsymbol{W}_u\cdot\boldsymbol{x}_j+\boldsymbol{U}_u\cdot\sum_{k\in C(j)}\boldsymbol{W}_r(jk)\cdot\boldsymbol{h}_k+\boldsymbol{b}_u)\,, \tag{21}$$

$$\boldsymbol{c}_j=\boldsymbol{i}_j\odot\boldsymbol{u}_j+\sum_{k\in C(j)}\boldsymbol{f}_{jk}\odot\boldsymbol{c}_k\,, \tag{22}$$

$$\boldsymbol{o}_j=\text{sigmoid}(\boldsymbol{W}_o\cdot\boldsymbol{x}_j+\boldsymbol{U}_o\cdot\sum_{k\in C(j)}\boldsymbol{W}_r(jk)\cdot\boldsymbol{h}_k+\boldsymbol{b}_o)\,, \tag{23}$$

$$\boldsymbol{h}_j=\boldsymbol{o}_j\odot\tanh(\boldsymbol{c}_j)\,。 \tag{24}$$

其中，$C_{(j)}$ 代表当前节点 $j$ 的所有子节点的集合，$\boldsymbol{W}_r(jk)$ 代表单词 $j$、$k$ 之间的依存关系矩阵。当该句所有词的隐向量值计算完毕后，即将结果输入到条件随机场中进行序列标注。

### 1.2 CRF

条件随机场是序列标注任务中的主流方法之一，是一种判别式概率模型。本研究使用线性链条件随机场，输入是 Tree LSTM 各个节点求出的值，输出是标签，联合模型如图 1 所示。

对每个句子，将经过 Tree LSTM 计算并输入到 CRF 中的隐向量序列表示为 $H=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_n\}$，模型输出的标签序列表示为 $Y=\{y_1,y_2,\cdots,y_n\}$，单词的标注标签本文选用标准的 BIO 标注方式，即 $y_i\in\{\text{BA},\text{IA},\text{BO},\text{IO},\text{O}\}$，其中 BA 代表评价对象的开始部分，IA 代表评价对象的内部，BO 代表观点内容的开始部分，IO 代表观点内容的内部，O 代表其他词。例如评论句“The service of the restaurant is good, but the food tastes general.”，CRF 的输入为 $H=\{\boldsymbol{h}_{\text{The}},\boldsymbol{h}_{\text{service}},\boldsymbol{h}_{\text{of}},\boldsymbol{h}_{\text{the}},\boldsymbol{h}_{\text{restaurant}},\boldsymbol{h}_{\text{is}},\boldsymbol{h}_{\text{good}},\boldsymbol{h}_{\text{but}},\boldsymbol{h}_{\text{the}},\boldsymbol{h}_{\text{food}},\boldsymbol{h}_{\text{tastes}},\boldsymbol{h}_{\text{general}}\}$，模型的输出结果为 $\boldsymbol{Y}=\{O,BA,O,O,O,O,BO,O,O,BA,BO,IO\}$。

在给定输入 $\boldsymbol{H}$ 的条件下，$\boldsymbol{Y}$ 的条件概率分布计算：

$$P(\boldsymbol{Y}\mid\boldsymbol{H})=\frac{1}{Z(\boldsymbol{H})}\prod_c\psi_c(Y_c\mid\boldsymbol{H})\,。 \tag{25}$$

其中：$Z(H)$ 为规范化因子，用于归一化；$\psi_c(Y_c\mid H)$ 为势函数；$P(Y\mid H)$ 是所有最大团 $C$ 上势函数的乘积。此处最大团包含两类，一是 Tree LSTM 输入到 CRF 的代表状态特征的团，二是输出序列中代表转移特征的团。在计算状态特征势函数时，额外融合上下文窗口大小为 3 的信息，则词“is”处的状态特征势函数计算示例如图 3 所示：

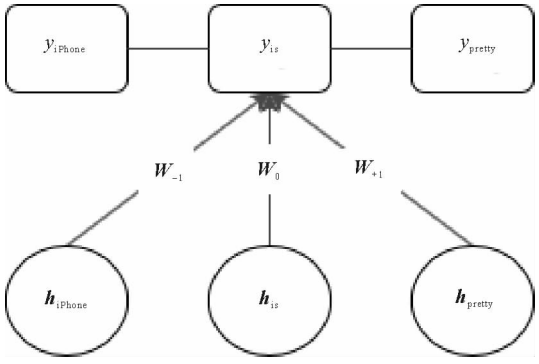



**图 3 词“is”处窗口大小为 3 的状态特征势函数计算示例**

Fig. 3 Example of state feature potential function calculation with window size 3 at word “is”

在上下文窗口为 3 时，节点 $\boldsymbol{k}$ 处的状态特征势函数计算公式：

$$\psi=\exp(\boldsymbol{W}_0^k\cdot\boldsymbol{h}_k+\boldsymbol{W}_{-1}^k\cdot\boldsymbol{h}_{k-1}+\boldsymbol{W}_{+1}^k\cdot\boldsymbol{h}_{k+1})\,。 \tag{26}$$

$\boldsymbol{W}_{-1}^k,\boldsymbol{W}_0^k,\boldsymbol{W}_{+1}^k$ 为节点 $k$ 处上下文窗口权重矩阵。以长度为 4 的例句为例，该句的条件概率计算公式：

$$\text{p}(y\mid h)=\frac{1}{Z(h)}\exp\Big(\sum_{k=1}^{4}\boldsymbol{W}_0^k\cdot\boldsymbol{h}_k+\sum_{k=2}^{4}\boldsymbol{W}_{-1}^k\cdot\boldsymbol{h}_{k-1}+\sum_{k=1}^{3}\boldsymbol{W}_{+1}^k\cdot\boldsymbol{h}_{k+1}+\sum_{k=1}^{3}\boldsymbol{V}_{k,k+1}\Big)\,。 \tag{27}$$

式中计算势函数时前三项代表计算窗口为 3 的状态特征势函数，第四项代表计算转移特征势函数。

### 1.3 模型训练

在对整个模型训练时，应用链式法则利用反向传播的方法学习各个参数。误差首先从条件随机场开始反向传播，沿模型结构传到 Tree LSTM 中，ROOT 指向的节点只接收到从 CRF 传来的误差，其他节点将接收到来自 CRF 的误差和来自依存关系父节点传来的误差，LSTM 单元中各门的参数也将根据链式法则学习更新。

## 2　实验及分析

### 2.1　数据集及实验设置

本节使用 SemEval Challenge 2014 任务 4 的数据集对模型进行训练与测试,该数据集包含笔记本和餐馆两个领域的评论数据,详细信息如表 2 所示,该数据仅仅包含对评价对象的标注,实验使用了 Wang 等[15]对观点内容手工标注的数据。

**表 2　SemEval Challenge 2014 任务 4 数据集**

Tab. 2　Dataset of SemEval Challenge 2014 Task 4

| 领域 | 训练句子数量 | 测试句子数量 | 总计 |
|---|---|---|---|
| 餐馆 | 3 041 | 800 | 3 841 |
| 笔记本 | 3 045 | 800 | 3 845 |
| 总计 | 6 086 | 1 600 | 7 686 |

在训练词向量时,使用 gensim word2vec 方法进行训练,餐馆领域的训练语料数据选择 Yelp challenge dataset 的评论数据,笔记本领域训练语料数据选择 Amazon 的电子产品评论数据,词向量的维数在比对后选择 350 维,对比实验在 2.2 节中叙述。评论语句的依存句法分析树使用 Stanford Parser 生成,模型中的线性链 CRF 使用 CRFSuite 实现。

词向量由 word2vec 训练得到,依存关系矩阵 $\boldsymbol{W}_r$ 和 LSTM 中权重参数($\boldsymbol{W}_i$、$\boldsymbol{U}_i$、$\boldsymbol{W}_f$、$\boldsymbol{U}_f$、$\boldsymbol{W}_u$、$\boldsymbol{U}_u$、$\boldsymbol{W}_o$、$\boldsymbol{U}_o$)的初始化由区间为 $[-\frac{\sqrt{6}}{\sqrt{2d+1}},\frac{\sqrt{6}}{\sqrt{2d+1}}]$ 的均匀分布随机生成,其他的参数均初始化为 0。在训练 Tree LSTM+CRF 联合模型前先对 Tree LSTM 模块进行预训练,在预训练时使用小批量随机梯度下降法并使用 AdaGrad 优化算法学习参数,学习率初始化为 0.02,在考虑训练句子总数和训练速度后,batch 数选择为 25,在笔记本和餐馆领域的训练数据都运行 5 代梯度下降。在训练整体 Tree LSTM+CRF 联合模型时,参数从 Tree LSTM 模块预训练的结果继承而来。本模型各个参数的初始化选择借鉴文献[15]。

由于 SemEval Challenge 2014 任务 4 的评测模型仅对评价对象进行了抽取,为了方便比较,将本模型去除观点内容标签后重新训练学习,得到只抽取评价对象的联合模型并命名为 Tree LSTM+CRF-O。

### 2.2　实验与结果分析

为了验证提出模型的有效性,本研究还实现了以下几个模型:

1) SemEval-1,SemEval-2:SemEval Challenge 2014 任务 4 评测时性能最好的两个模型。

2) CRF-1:包含基础语言特征(词、文体、词性、上下文、上下文词性)的 CRF 模型。

3) CRF-2:包含上述基础语言特征和依存关系特征(中心词和词之间的依存关系)的 CRF 模型。

4) W+L+D+B:Yin 等[12]提出的 CRF 模型包含无监督学习得到的词嵌入特征、依存关系特征,线性上下文嵌入特征以及基础特征模板。

5) LSTM,LSTM-CRF,Bi-LSTM-CRF:分别指 LSTM 为基础的长短期记忆网络,LSTM-CRF 为长短期记忆网络结合 CRF 的模型,Bi-LSTM-CRF 为双向长短期记忆网络结合 CRF 的模型。LSTM 网络中的权重通过区间[-0.2,0.2]的随机均匀分布初始化,隐层的大小设置为 50,学习率设置为 0.01。

6) RNCRF,RNCRF-O:Wang 等[15]提出的递归神经网络和 CRF 的联合模型,RNCRF-O 是为方便比较而忽略掉观点内容标注的模型。

选用 F1 值作为模型性能的评价指标,计算方法如公式(28)~(30)所示。其中,TP 是模型正确标注的数量,TP+FP 是模型标注的总数,TP+FN 是测试集中存在的标注总数。

$$Precision=\frac{TP}{TP+FP}, \tag{28}$$

$$Recall=\frac{TP}{TP+FN}, \tag{29}$$

$$F1\text{-}Score=\frac{2\times Precision\times Recall}{Precision+Recall}。 \tag{30}$$

实验结果如表 3 所示。

**表 3 各模型实验结果的 F1-Score 值**

Tab. 3 F1-Score of experimental results of each model

| 模型 | 餐馆领域 | | 笔记本领域 | |
|---|---|---|---|---|
| | 评价对象 | 观点内容 | 评价对象 | 观点内容 |
| SemEval-1 | 84.01 | — | 74.55 | — |
| SemEval-2 | 83.98 | — | 73.78 | — |
| CRF-1 | 77.00 | 78.95 | 66.21 | 71.78 |
| CRF-2 | 78.37 | 78.65 | 68.35 | 70.05 |
| W+L+D+B | 84.97 | — | 75.16 | — |
| LSTM | 81.15 | 80.22 | 72.73 | 74.98 |
| LSTM-CRF | 83.13 | 80.78 | 76.09 | 75.47 |
| Bi-LSTM-CRF | 84.12 | 82.32 | 77.96 | 76.39 |
| RNCRF-O | 82.73 | — | 74.52 | — |
| RNCRF | 84.05 | 80.93 | 76.83 | 76.76 |
| Tree LSTM+CRF−O | 86.32 | — | 78.07 | — |
| Tree LSTM+CRF | **86.76** | **83.22** | **79.86** | **80.42** |

由表 3 可知,本模型比 SemEval Challenge 的最优模型 SemEval-1 与 SemEval-2 效果好,在餐馆与笔记本领域分别比 SemEval-1 高出 2.75%和 5.31%。在与普通条件随机场的对比中,在加入依存关系特征后,CRF-2 在评价对象抽取上比 CRF-1 在餐馆和笔记本领域分别高出 1.37%和 2.14%,说明依存关系特征确实有助于评价对象的抽取,同时 CRF 模型的结果低于其他模型,说明深度学习方法比条件随机场学习信息更加有效。LSTM 模型的结果要差于 LSTM-CRF 模型,这是由于条件随机场能够纠正类似 IA、BA 这样的标注顺序错误,所以大部分模型都会在神经网络之后连接条件随机场进行标注。双向 LSTM 因为能够捕获 2 个方向的信息而比普通 LSTM 模型效果要好。本模型比 RNCRF 模型效果略好,说明将普通递归神经网络单元替换为 LSTM 单元和设计树结构下 LSTM 单元门结构的计算方法是有效的;比时间序列的 Bi-LSTM-CRF 模型的结果好,说明树结构在依存关系信息的处理中优于时间序列结构。由于依存关系是由子单词指向父单词,类似于树结构,所以在依存关系特征上树结构效果更好。Tree LSTM+CRF-O 是去除观点内容标注的模型,从实验结果可以看出,该模型性能损失不大,表明本模型鲁棒性好。

图 4 给出了 RNCRF 和 Tree LSTM+CRF 两个模型实际标注的结果示例。从图 4 中的例子中可看出,本模型比普通递归神经网络能更好地处理依存关系特征和词本身特征,对低频出现的评价对象和依存关系相对复杂的句子标注结果更好。

进一步还对不同词向量维度对模型的影响进行了实验。实验选取维度范围为 50～500 维,维度间隔为 50,实验结果如图 5 所示。

| | Definitely | try | the | taglierini | with | truffles | - | it | was | incredible | . |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RNCRF: | O | O | O | O | O | O | O | O | O | BO | O |
| Tree LSTM-CRF: | O | O | O | BA | IA | IA | O | O | O | BO | O |

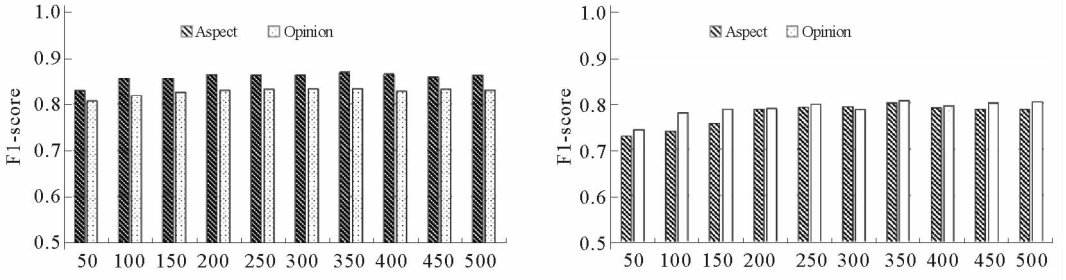
(a) 例1

| | Try | the | Peanut | Butter | Sorbet | and | the | pizza | with | soy | cheese | ! |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RNCRF: | O | O | BA | IA | IA | O | O | BA | O | BA | IA | O |
| Tree LSTM-CRF: | O | O | BA | IA | IA | O | O | BA | IA | IA | IA | O |

(b) 例2

**图 4　RNCRF 与 Tree LSTM+CRF 标注结果对比**

Fig.4　Comparison of RNCRF and Tree LSTM+CRF labeling results




(a) 餐馆领域词向量维度实验结果　　(b) 笔记本领域词向量维度实验结果

**图 5　词向量维度对比实验结果**

Fig.5　Word vector dimension comparison experiment results

由图 5 可以看出,在餐馆领域,评价对象的抽取结果普遍优于观点内容的抽取结果;而在笔记本领域,观点内容的抽取结果普遍优于对评价对象的抽取结果。而且模型在两个领域下都在 350 维处效果最好,模型性能随维度变化有波动但波动幅度不超过 3%,模型性能相对稳定。

## 3　总结与展望

评价对象和观点内容的抽取是观点挖掘中的重要研究内容。本研究提出一个基于 Tree LSTM 结合 CRF 的联合模型来抽取评论语句中的显式评价对象与观点内容。该模型包含两部分,第一部分是根据评论语句的依存结构树构建的 Tree LSTM,用于融合词向量和依存句法关系向量从而学习每个词的高层特征;第二部分是条件随机场,将从 Tree LSTM 得到的每个词的隐向量输入其中进行序列标注工作,将隐向量映射到代表评价对象、观点内容和其他词的标签上,实现了评价对象与观点内容的抽取。在 SemEval Challenge 2014 任务 4 的数据集上的实验结果表明本 Tree LSTM 能很好地表征词语之间的层次关系,同时联合模型有效避免了传统 CRF 需要构建特征工程的弊端。

目前本模型只是实现了简单的抽取工作,下一步将对评论句进行情感分析,深入分析用户所表达的观点;并尝试对评论中的隐式评价对象进行抽取,以全面分析用户的观点。

**参考文献:**

[1]HU M,LIU B.Mining and summarizing customer reviews[C]//10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,Seattle,Washington,USA,August.DBLP,2004:168-177.

[2]HU M,LIU B.Mining opinion features in customer reviews[C]//National Conference on Artificial Intelligence.2004:755-760.

[3]POPESCU A M, ETZIONI O. Extracting product features and opinions from reviews[C]// Hlt/emnlp on Interactive Demonstrations. Association for Computational Linguistics, 2005: 32-33.

[4]刘鸿宇,赵妍妍,秦兵,等.评价对象抽取及其倾向性分析[J].中文信息学报,2010,24(1):84-88.
LIU Hongyu, ZHAO Yanyan, QIN Bing, et al. Comment target extraction and sentiment classification[J]. Journal of Chinese Information Processing, 2010, 24(1): 84-88.

[5]江腾蛟,万常选,刘德喜,等.基于语义分析的评价对象-情感词对抽取[J].计算机学报,2017,40(3):617-633.
JIANG Tengjiao, WAN Changxuan, LIU Dexi, et al. Extracting target-opinion pairs based on semantic analysis [J]. Chinese Journal of Computers, 2017, 40(3): 617-633.

[6]廖祥文,陈兴俊,魏晶晶,等.基于多层关系图模型的中文评价对象与评价词抽取方法[J].自动化学报,2017,43(3):462-471.
LIAO Xiangwen, CHEN Xingjun, WEI Jingjing, et al. A multi-layer relation graph model for extracting opinion targets and opinion words[J]. Acta Automatica Sinica, 2017, 43(3): 462-471.

[7]WEI J, HO H H. A novel lexicalized HMM-based learning framework for web opinion mining NOTE FROM ACM[C]// International Conference on Machine Learning. ACM, 2009: 465-472.

[8]刘全超,黄河燕,冯冲.面向中文微博的评价对象与评价词语联合抽取[J].电子学报,2016,44(7):1662-1670.
LIU Quanchao, HUANG Heyan, FENG Chong. Co-extracting opinion targets and opinion-bearing words in chinese wicroplog texts[J]. Acta Electronica Sinica, 2016, 44(7): 1662-1670.

[9]丁晟春,吴婧婵媛,李霄.基于 CRFs 和领域本体的中文微博评价对象抽取研究[J].中文信息学报,2016,30(4):159-166.
DING Shengchun, WU Jingchanyuan, LI Xiao. Opinion targets extraction from Chinese microblogs based on conditional random fields and domain ontology[J]. Journal of Chinese Information Processing, 2016, 30(4): 159-166.

[10]IRSOY O, CARDIE C. Opinion mining with deep recurrent neural networks[C]// Conference on Empirical Methods in Natural Language Processing. 2014: 720-728.

[11]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.

[12]YIN Y, WEI F, DONG L, et al. Unsupervised word and dependency path embeddings for aspect term extraction[C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2016: 2979-2985.

[13]WANG Y, HUANG M, ZHU X, et al. Attention-based LSTM for aspect-level sentiment classification[C]// Conference on Empirical Methods in Natural Language Processing. 2016: 606-615.

[14]GIANNKOPOULOS A, MUSAT C, HOSSMANN A, et al. Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets[C]// Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Copenhagen: WASSA, 2017: 180-188.

[15]WANG W, PAN S J, DAHLMEIER D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: EMNLP, 2016: 616-626.

(责任编辑:傅 游)