

文章编号:

基于领域情感词典特征表示的细粒度意见挖掘*

郁圣卫¹, 卢奇¹, 陈文亮¹

(1. 苏州大学 计算机科学与技术学院, 江苏省 苏州市 215006)

摘要: 细粒度意见挖掘的主要目标是从观点文本中获取情感要素并判断情感倾向。目前现有方法大多基于序列标注模型, 但很少利用情感词典资源。本文提出一种基于领域情感词典特征表示的细粒度意见挖掘方法, 使用领域情感词典在观点文本上构建特征表示并将其加入序列标注模型的输入部分。首先构建一份新的电商领域情感词典, 然后在电商评论文本真实数据上, 分别为条件随机场 (CRF) 和双向长短期记忆-条件随机场 (BiLSTM-CRF) 这两种常用序列标注模型设计基于领域情感词典的特征表示。实验结果表明, 基于电商领域情感词典的特征表示方法在两种模型上都取得了良好的效果, 并且超过其他情感词典。

关键词: 细粒度意见挖掘; 情感词典; 特征表示; 序列标注模型

中图分类号: TP391

文献标识码: A

Fine-grained Opinion Mining Based on Feature Representation of Domain Sentiment Lexicon

YU Shengwei¹, LU Qi¹, CHEN Wenliang¹

(1. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu, 215006, China)

Abstract: Fine-grained opinion mining aims at detecting sentiment units and determining sentiment polarity from opinion text. Recent previous methods are mostly based on sequence labeling models but rarely use the information of sentiment lexicon resources. This paper proposes a fine-grained opinion mining method based on feature representation of domain sentiment lexicon. This method generates feature representation by using domain sentiment lexicon, puts it into the input of sequence labeling model. Firstly, we build a new sentiment lexicon in E-commerce domain. Then we design feature representation of domain sentiment lexicon for CRF and BiLSTM-CRF. Finally, we evaluate our systems on E-commerce reviews. Experimental results show that our proposed method performs well on both models and outperforms the method based on other lexica.

Key words: Fine-grained opinion mining; Sentiment lexicon; Feature representation; Sequence labeling model

1 引言

细粒度意见挖掘的主要目标是从观点文本 (如用户评论文本等) 中获取情感要素 (如观点持有者、观点对象、观点表达等) 并根据这些信息判断观点在文本中表达的情感倾向^[1]。举个例子, 在观点文本“我觉得质量不错”中, 观点持有

者“我”针对观点对象“质量”使用观点表达“不错”来表达“正面”情感倾向。细粒度意见挖掘的结果通常可应用于观点摘要和观点检索^[2]。

细粒度意见挖掘获取的情感要素根据具体任务需求和具体领域而定。以电商评论文本为例, 其观点持有者大多为第一人称, 并且观点对象以及观点表达都是出现在文本中的连续序列片段。因此, 对于这类文本通常只获取观点对象及观点

收稿日期:

定稿日期:

基金项目: 江苏省高校自然科学研究重大项目 (16KJA520001); 国家自然科学基金面上项目 (61572338)

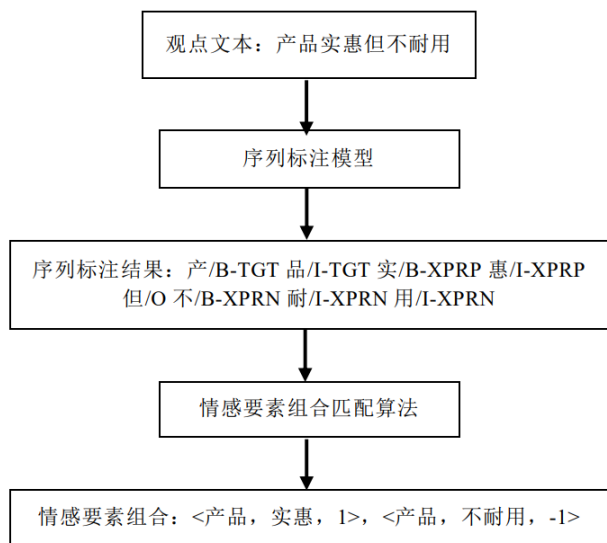


图1 本文采用的细粒度意见挖掘任务框架

表达，并将其看作序列标注问题。但是，观点表达和观点对象在同一语句中未必是一一对应的关系，同一语句中的情感倾向也未必是单一的。例如“产品实惠但不耐用”，观点对象“产品”对应多个观点表达和多种情感倾向。如果单纯使用序列标注模型，很难准确判断情感倾向的所属。

为了解决上述问题，我们采用如图1所示的任务框架。对于观点文本，我们首先使用序列标注模型识别其中的情感要素序列，然后采用匹配算法将其整合为结构化的情感要素组合。在序列标注部分，序列标签采用 IOBES 标签体系^[3]，其中 TGT 代表观点对象，XPR 代表观点表达，对 XPR 使用追加标签 P、M、N 分别表示正面、中性、负面倾向的观点表达。针对序列标注的结果，我们以三元组 <观点对象, 观点表达, 情感倾向> 来生成情感要素组合。其中，情感倾向部分使用数值 1, 0, -1 分别表示正面、中性、负面倾向。

目前，较为常用的序列标注模型有条件随机场（CRF）^[4]和双向长短期记忆-条件随机场（BiLSTM-CRF）^[5]。其中，CRF 属于传统机器学习方法，BiLSTM-CRF 属于深度学习与传统机器学习方法的结合。Nakayama 和 Fujii^[6]基于 CRF，通过构建丰富的特征模板，在日语细粒度意见挖掘上取得了一定效果。但是，需要人工构建复杂的特征模板。Liu 等^[7]使用预训练词向量，并在循环神经网络（RNN）的隐层输出结果中加入了语言特征表示如词性和词块，在英语细粒度意见挖掘上取得了很好的效果。然而，这些特征表示仅仅用到文本中的字、词及其语言信息，但缺乏丰富的情感信息。因此如何使用外部情感词典资源来改进细粒度意见挖掘的性能可以作为值得研究的课题。

的课题。

针对上述现有方法的不足，我们在使用序列标注模型进行细粒度意见挖掘的基础上，提出一种基于领域情感词典特征表示的方法。我们首先基于标注数据构建了一份新的电商领域情感词典，然后在电商评论文本真实数据上，分别为 CRF 和 BiLSTM-CRF 这两种常用序列标注模型设计基于领域情感词典的特征表示。实验结果表明，这种基于领域情感词典特征表示的方法在两种模型上都取得了良好的效果，并且超过其他情感词典。

本文的其余部分作如下安排：第2节对相关工作进行介绍；第3节介绍情感词典资源；第4节详细介绍我们提出的方法；第5节介绍实验和结果分析；第6节是结论介绍。

2 相关工作

细粒度意见挖掘属于情感分析与意见挖掘任务。在情感分析与意见挖掘相关任务被提出的早期，有关细粒度意见挖掘的研究工作很少。起初，较为常见的相关任务是对文档级别文本进行情感倾向分类，如 Pang 等^[8]使用传统机器学习方法进行情感分类。随着相关任务的拓展研究，研究人员需要在标注粒度更为细致的情感分析语料上开展研究。Wilson 等^[9]构建了短语级别的情感分析标注语料，综合考虑若干短语反映的情感倾向并对文档实例进行情感分类。随着进一步的研究，仅进行情感倾向分类难以满足检索任务的需求，因而研究人员进行扩展，提出情感要素抽取任务。情感要素可以包括观点持有者^[10]、观点对象^[11]、观点表达^[12]、情感倾向和强度^[13]等。情感要素抽取任务由于获取的是情感文本中具有实际意义的细粒度要素，因此可看作细粒度意见挖掘。早期阶段，细粒度意见挖掘的相关工作大多采用 CRF 模型进行序列标注^[14-15]。然而，基于 CRF 模型的序列标注模型通常需要人工构建复杂的特征模板。近三年来，随着深度学习的发展，更多的研究人员采用深度学习的序列标注方法进行建模，如 Irosy 等^[16]基于 RNN 模型，Katiyar 等^[17]用到 LSTM 模型。目前，深度学习方法与传统机器学习方法的结合模型如 BiLSTM-CRF 模型、BiLSTM-CNN-CRF 模型等在序列标注任务上取得了很好的效果^[18]。然而，目前基于深度学习的细粒度意见挖掘工作鲜有利用情感词典资源。

情感词典资源的构建对情感分析与意见挖掘相关任务的研究也格外重要。Kaji 等^[19]从大规

模 HTML 文本中构建用于情感分析任务的词典资源, Jijkoun 等^[20]使用 Bootstrapping 方法构建话题特定的情感词典。目前, 常用的英文情感词典有 SentiWordNet、LIWC、ANEW 等; 常用的中文情感词典有知网的“情感分析用词语集”(下文简记为“知网情感词典”)、大连理工大学信息检索实验室的“情感词汇本体”^[21](下文简记为“大连理工情感词典”)等。这些词典对情感分析与意见挖掘任务提供了一定帮助。然而, 它们的词条较为通用, 将其应用于领域特定的观点文本中时效果不佳。

基于上述相关工作和不足, 本文在使用序列标注模型进行细粒度意见挖掘的基础上, 结合情感词典资源, 提出一种基于领域情感词典特征表示的方法。

3 情感词典资源

3.1 电商领域情感词典的构建

本文主要针对电商领域观点文本进行细粒度意见挖掘。依照本文提出的方法, 需要面向电商领域的领域情感词典。对于电商领域的观点文本, 其特殊之处主要有以下几点:

- 短文本居多, 行文或正规或随意。
- 观点持有者大多为第一人称。
- 通常针对电商产品或产品相关属性, 送货服务或售后服务等观点对象进行观点评价。
- 包含电商领域特有的观点对象词条及观点表达词条。

这些特性决定了电商领域观点文本中情感要素的构成和分布。对于电商领域观点文本, 我们主要针对观点表达和情感倾向的对应关系构建相应的情感词典。观点表达可以只反映单一的情感倾向, 例如“好吃”只反映正面倾向; 也可以反映多种情感倾向, 例如“高”在“价格高”中反映负面倾向而在“性价比高”中反映正面倾向。因此, 我们将观点表达按照反映出的情感倾向的不同定义为下列四个类别:

- 正面观点表达
- 中性观点表达
- 负面观点表达
- 多极性观点表达

其中, 前三种类别的观点表达只反映单一的情感倾向, 第四种类别反映多种情感倾向。

除上述观点表达词条以外, 我们也在构建情感词典时加入了与观点表达具有密切关联的词条:

- 否定词

否定词的功能是让观点表达的情感倾向发生反转, 其加入可使情感词典中未出现的带有否定前缀的观点表达的情感信息也能够被检索出来。否定词的追加间接地扩充了情感词典的情感信息。

电商领域情感词典的构建要点在于获取电商评论文本中观点表达和情感倾向的对应关系。由于领域资源受限, 本文直接对实验部分用到的训练数据中的情感要素组合进行抽取, 从而构建得到电商领域情感词典。首先, 我们从情感要素组合中抽取得到观点表达和情感倾向的对应关系并进行汇总, 得到候选领域情感词条。对于候选结果, 我们进行人工检验并纠正了一些错误。最终, 将校正后的观点表达词表作为电商领域情感词典。表 1 给出了电商领域情感词典中各个类别的词条示例:

表 1 电商领域情感词典词条示例

正面观点表达	...物有所值, 实惠, 舒服...
中性观点表达	...美中不足, 一般, 凑合...
负面观点表达	...伤不起, 差评, 贵...
多极性观点表达	...水水的, 大, 高...
否定词	...不, 不怎么, 没那么...

3.2 与其他情感词典的比较

我们选取了两种情感词典, 分别是知网情感词典和大连理工情感词典, 和本文构建的电商领域情感词典进行比较。

为了便于比较, 我们将两种情感词典的条目和电商领域情感词典进行了统一。对于知网情感词典, 由于其仅将观点表达分为正面倾向和负面倾向, 因此我们将其等价于正面观点表达和负面观点表达, 同时将其他三类置空。对于大连理工情感词典, 其将观点表达分为正面、中性、负面以及正负都有可能这四类, 因此我们将其对应等价到定义中的四类, 并将其否定词类别置空。

我们从两个方面对这三种情感词典进行比较, 分别基于情感词典的条目统计信息和条目中具体词条的比较。表 2 给出了三种情感词典的观点表达类别及相关统计信息:

表 2 三种情感词典的观点表达类别及相关统计信息

情感词典	领域	知网	大连理工
正面	844	4566	11229
中性	82	N/A	5375
负面	2084	4370	10784
多极性	99	N/A	78
否定词	29	N/A	N/A

条目统计信息仅仅是衡量情感词典的一个方面。表 3 给出了三种情感词典共有词条和电商领域

表3 情感词典词条异同性比较

	类别	条目数	示例
共有词条	正面	164	...名不虚传, 清脆, 气派...
	中性	9	...将就, 勉强, 普通...
	负面	110	...乱七八糟, 低劣, 累赘...
	多极性	0	N/A
	否定词	0	N/A
电商领域情感词典独有词条	正面	680	...物超所值, 杠杠滴, 五星...
	中性	73	...将就着用, 一般, 中评...
	负面	1974	...毫无信用, 杂牌货, 易碎...
	多极性	99	...水水的, 大, 高...
	否定词	29	...不, 不怎么, 没那么...

情感词典独有词条的比较结果。从表中可以看出：

- 电商领域情感词典在各个类别都拥有大量和电商领域高度相关的独有词条。
- 两种通用词典中和电商领域高度相关的词条数量较少。

针对上述三种情感词典，我们为其定义情感词典标签。表4给出了不同情感词典的情感词典标签：

表4 不同情感词典的情感词典标签

情感词典	领域	知网	大连理工
正面	DoUP	HowUP	DuUP
中性	DoUM	N/A	DuUM
负面	DoUN	HowUN	DuUN
多极性	DoP	N/A	DuP
否定词	DoN	N/A	N/A

4 基于领域情感词典特征表示的方法

4.1 基于领域情感词典的数据表示

我们提出的细粒度意见挖掘方法以电商评论文本为研究对象。针对电商评论文本，我们使用第3节构建的电商领域情感词典来生成数据表示。

对于数据表示的生成，我们采用如下方式：首先使用领域情感词典对电商评论生文本进行最大正向匹配以得到具体观点表达在生文本中的位置，然后采用IOB2标签体系生成情感词典标签。最终，将得到的情感词典标签和生文本进行结合，构成基于情感词典的数据表示。给出基于领域情感词典的数据表示实例如表5所示。在表5的例子中，观点表达“实惠”与“不耐用”都出现在领域情感词典中。其中，DoUP代表领域情感词典的正面观点表达标签，DoN代表否定词标签。

表5 基于领域情感词典的数据表示举例

产	品	实	惠
O	O	B-DoUP	I-DoUP
但	不	耐	用
O	B-DoN	B-DoUP	I-DoUP

4.2 基于领域情感词典的CRF模型

CRF是一种判别式模型，在序列标注任务中通常采用线性链条件随机场（Linear-chain CRF）。根据该模型的定义，可以将序列标注任务转化为如下形式：给定输入序列形如： $\mathbf{x} = x_1, x_2, \dots, x_n$ ，任务目标是预测与该输入序列等长的标签序列： $\mathbf{y} = y_1, y_2, \dots, y_n$ ，标签序列中每个位置与输入序列相对应。然后，由公式(1)计算条件概率 $P(\mathbf{y}|\mathbf{x})$ ：

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, \mathbf{x}) + \sum_{i,k} \mu_k g_k(y_i, \mathbf{x})) \quad (1)$$

其中， $Z(\mathbf{x})$ 是归一化因子， f_k 和 g_k 分别是转移特征函数和状态特征函数，它们的输出值都为布尔值。对于 f_k ，当 y_{i-1}, y_i, \mathbf{x} 满足转移特征函数的具体数值时输出为1，否则为0； g_k 与其类似。 λ_k 和 μ_k 分别是对应特征函数的权值。在训练过程中，由输入序列和标签序列构成的每一组实例通过最大化公式(1)的对数似然概率来训练模型中的各个变量。测试时，给定测试数据中的一组输入序列实例 \mathbf{x}' ，选取满足公式(2)的输出序列 \mathbf{y}^* 作为最佳预测标签序列：

$$\mathbf{y}^* = \operatorname{argmax} P(\mathbf{y}'|\mathbf{x}') \quad (2)$$

对于CRF模型特征模板的设计，我们采用如表6所示的特征模板：

表6 CRF模型采用的特征模板

F1	char[n], n ∈ {-3, -2, -1, 0, 1, 2, 3}
F2	char[n]%char[n+1], n ∈ {-2, -1, 0, 1}
F3	char[n-1]%char[n+1], n ∈ {0}
F4	dict[n], n ∈ {-3, -2, -1, 0, 1, 2, 3}
F5	dict[n]%dict[n+1], n ∈ {-2, -1, 0, 1}
F6	dict[n-1]%dict[n+1], n ∈ {0}

表7 CRF模型生成特征举例

f1	产; 品; 实; 惠; 但; 不; 耐
f2	品%实; 实%惠; 惠%但; 但%不
f3	实%但
f4	O; O; B-DoUP; I-DoUP; O; B-DoN; B-DoUP
f5	O%B-DoUP; B-DoUP%I-DoUP; I-DoUP%O; O%B-DoN
f6	B-DoUP%O

其中，n代表与当前位置的偏移量（0表示当前位置），char代表字符，dict表示情感词典标签取值，具体取值与4.1小节保持一致。模板F4、F5、F6生成的特征即作为情感词典特征。以表5中的当前位置字符“惠”为例，其按照表6生成的特征如表7所示。

4.3 基于领域情感词典的 BiLSTM-CRF 模型

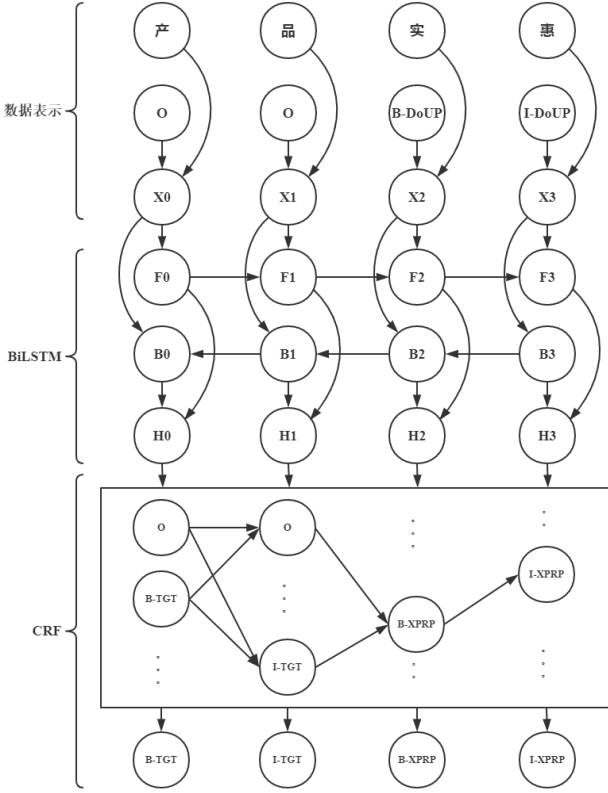


图2 BiLSTM-CRF 框架图

图2是本实验采用的 BiLSTM-CRF 框架。

第一层是数据表示层。其作用是将我们给出的字序列和情感词典标签序列的输入映射到向量级别的输入表示。本实验中，共有两种输入表示，一种是字表示，另一种是情感词典表示，分别通过事先建立的索引关系得到。然后，将字表示和情感词典表示进行拼接，得到最终的输入表示。

第二层是 BiLSTM 层。其作用是将第一层得到的输入表示转化为隐层表示输出。长短期记忆网络 (LSTM) 是一种特殊的循环神经网络 (RNN) 模型，其主要结构如公式 (3) 所示：

$$\begin{aligned} I_t &= \sigma(X_t \cdot W_{xi} + H_{t-1} \cdot W_{hi} + b_i) \\ F_t &= \sigma(X_t \cdot W_{xf} + H_{t-1} \cdot W_{hf} + b_f) \\ O_t &= \sigma(X_t \cdot W_{xo} + H_{t-1} \cdot W_{ho} + b_o) \\ \tilde{C}_t &= \tanh(X_t \cdot W_{xc} + H_{t-1} \cdot W_{hc} + b_c) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\ H_t &= O_t \odot \tanh(C_t) \end{aligned} \quad (3)$$

其中， X_t 代表 t 时刻的输入表示， σ 代表 sigmoid 函数， \odot 是元素点积运算， I_t , F_t , O_t 分别是 t 时刻的输入门、遗忘门、输出门的结果， \tilde{C}_t , C_t , H_t 分别代表 t 时刻候选细胞状态、细胞状态以及隐层状态， W 和 b 分别表示对应的权重和偏差。最终，我们以隐层状态表示作为 LSTM

模型的输出结果。

LSTM 对于一段输入时间序列通常从左至右进行计算得到隐层表示，将其记为 \vec{F}_t 。为了更好地刻画上下文的关系，BiLSTM 会对输入时间序列分别计算从左至右的隐层表示 \vec{F}_t 以及从右至左的隐层表示 \overleftarrow{F}_t ，然后进行拼接，得到最终的隐层表示 $H_t = [\vec{F}_t, \overleftarrow{F}_t]$ 。

第三层是 CRF 层。假设序列标注的标签个数为 k ，对于输入序列 $x = x_1, x_2, \dots, x_n$ ，经过前三层的计算后可以得到维数为 $n \times k$ 的分值矩阵 P ，矩阵中的某一个元素 $P_{i,j}$ 代表第 i 个输入状态标注为第 j 个标签的得分。对于一组预测标签序列 $y = y_1, y_2, \dots, y_n$ ，定义它的得分形如公式 (4)：

$$Score(x, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \quad (4)$$

其中， A 是转移得分矩阵， $A_{i,j}$ 代表从标签 i 转移到标签 j 的得分。 y_0 和 y_n 分别是标签序列中的起始和结束标签，需要加入到原有标签集合中。因此 A 是 $k+2$ 阶方阵。由此，我们得到基于所有可能的标签集合 Y_x 下的条件概率 $P(y|x)$ 形如公式 (5)：

$$P(y|x) = \frac{e^{Score(x, y)}}{\sum_{\tilde{y} \in Y_x} e^{Score(x, \tilde{y})}} \quad (5)$$

在训练过程中，最大化形如公式 (5) 中正确标签序列的对数似然概率。测试时，选取满足公式 (6) 的结果 y^* 作为最佳预测标签序列：

$$y^* = \underset{\tilde{y} \in Y_x}{\operatorname{argmax}} Score(x, \tilde{y}) \quad (6)$$

4.4 情感要素组合匹配算法

使用上述序列标注模型对实验数据进行标注后，便可获取其中含有的观点对象、观点表达以及情感倾向并进行情感要素组合匹配。不同于关系抽取任务中头实体和尾实体都不为空的情形，<观点对象，观点表达，情感倾向> 中的观点对象可以为空。这样的问题给匹配带来了困难。因此，本实验采用如表 8 所示的算法进行匹配。

对于这个算法，我们给出如图 3 所示的具体实例。对于语句“很差劲，质量不好”的序列标注结果，我们首先以中间的逗号作为分隔符将其分隔为两个短句。对于短句“很差劲”的情感要素“<差劲, XPRN>”，锁定得到观点表达“差劲”，但是其相邻左侧和相邻右侧都不存在观点对象，因此返回结果“<NULL, 差劲, -1>”。而对于短

表 8 情感要素组合匹配算法

输入 1: 所有序列标签集合 $L=\{B-TGT, I-TGT, B-XPR, I-XPR, O\}$
输入 2: 序列标注结果 $A=\{<x_1, y_1>, <x_2, y_2>, \dots, <x_n, y_n>\}$, 其中 $y_i \in L, 1 \leq i \leq n$
输入 3: 短句标点符号集合 P
输入 4: 情感倾向函数 $p(x) = \begin{cases} 1, x = XPRP \\ 0, x = XPRM \\ -1, x = XPRN \end{cases}$
输出: 情感要素组合构成的集合 $T=\{<T_1, X_1, P_1>, \dots, <T_z, X_z, P_z>\}$, 其中 z 为情感要素组合总数
<pre> 1: for $<x_i, y_i>$ in A do 2: if x_i in P then $y_i \leftarrow B-PUNCT$ 3: end for 4: 对 A 的 IOB2 标签进行 chunk 操作, 得到标签块内组合结果 $B=\{<c_0, l_0>\}$, $l_i \in \{TGT, XPR \cdot, PUNCT\}, 1 \leq j \leq o$ 5: 根据 PUNCT 标签对 B 进行短句切分操作, 将每个短句的 二元组重新构成集合, 得到 $C=\{S_1, S_2, \dots, S_p\}$, p 为短句标点 个数, $S_k=\{<u_1, v_1>, <u_2, v_2>, \dots, <u_i, v_i>, <u_{s+1}, PUNCT>\}$, s 为每个短句中的情感要素总数 6: for S_k in C do 7: for $<u_m, v_m>$ in S_k do 8: if $v_m=XPR \cdot$ then 9: if $v_{m-1}=TGT$ then 将 $<u_{m-1}, u_m, p(v_m)>$ 加入 T 10: else if $v_{m+1}=TGT$ then 将 $<u_{m+1}, u_m, p(v_m)>$ 加入 T 11: else 将 $<NULL, u_m, p(v_m)>$ 加入 T 12: end if 13: end if 14: end for 15: end for </pre>

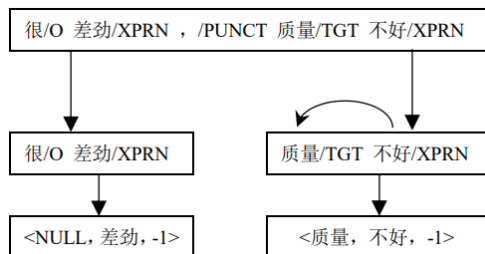


图 3 情感要素组合匹配示例

句“质量不好”的情感要素“<质量, TGT>, <不好, XPRN>”，锁定得到观点表达“不好”，其向左最近邻匹配得到观点对象“质量”，因此返回情感要素组合结果“<质量, 不好, -1>”。

5 实验

本节先介绍实验数据，其次介绍情感词典数据，然后介绍实验设置，再介绍实验的评价方法，最后是实验结果、方法对比与实例分析。

5.1 实验数据

本次实验的数据来源于 BDCI2017-TASK12 赛事¹复赛阶段的语料，所属类型为电商评论文本，共计 20000 条电商评论文本的标注实例。原始语料的标注方式如表 9 所示：

表 9 原始语料标注方式示例

原文本	情感要素组合
接触不好, 效果一般	<接触;不好;-1>,<效果;一般;0>

我们对原始数据进行初步预处理。首先，我们打乱原有语料顺序，并以 8:1:1 的比例将语料初步切分为训练集、验证集、测试集。然后，我们把初步预处理结果转化为序列标注数据。转化举例如图 4 所示。

首先，我们对每条实例中的原文本以短句标点符号为界进行切分得到若干短句，并将每一个情感要素组合和这些短句进行匹配。如果存在匹配结果，则按照情感要素组合中每个元素的位置标注于该短句上。最终得到符合序列标注规则的实验数据。

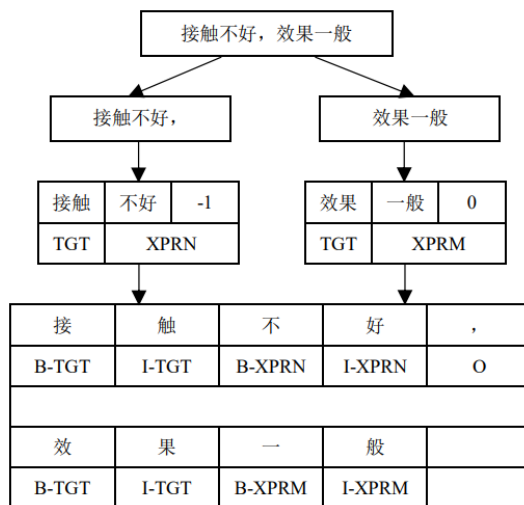


图 4 实验数据转化示例

考虑到 20000 条标注实例的句长参差不齐，在实际训练过程中，我们将训练集、验证集、测试集中的原有实例切分为短句。表 10 给出了它们的相关统计信息：

表 10 实验数据相关统计信息

	训练集	验证集	测试集
原实例数	16000	2000	2000
短句实例数	67155	8591	8562
平均短句长度	8.14	8.12	8.24
情感要素组合个数	33215	4084	4236

5.2 情感词典数据

为了进行实验比较，我们又用到知网情感词典和大连理工情感词典。由于第 4 节提出的方法可以拓展到任意情感词典上，因而在具体实验中我们使用三种情感词典对 5.1 小节得到的实验数据追加情感词典标签。不同情感词典会生成不同情感词典标签。表 11 给出了不同情感词典对同一实例的情感词典标签结果：

¹ <http://www.datafountain.cn/#/competitions/268/intro>

表 11 不同情感词典生成标签示例

	性	价	比	不	高
领域情感词典	O	O	O	B-DoN	B-DoP
知网情感词典	O	O	O	O	B-HowUP
大连理工情感词典	O	O	O	O	O

5.3 实验设置

本次实验,我们使用两种序列标注模型:CRF和BiLSTM-CRF。对于CRF,我们使用4.2小节设计的特征模板,迭代次数设置为满足收敛条件为止。对于BiLSTM-CRF,我们按照如表12所示的参数进行实验。

对于BiLSTM-CRF实验结果的选取,我们在每次epoch训练结束后测得验证集的F1值,每当得到的F1值到达极大值时存储模型。全部训练结束后,我们得到在验证集上F1值最优的模型。我们使用这个模型对测试集进行测试,得到最终的实验结果。

表 12 BiLSTM-CRF 参数设置

参数	值
字表示维数	64
情感词典表示维数	32
隐层维数	100
优化器	RMSprop
学习率	0.001
batch size	64
epoch	100
dropout	0.5

5.4 评价方法

本次实验,我们分别对序列标注结果和情感要素组合匹配结果进行评价。对于这两个阶段的结果,我们都采用准确率(Precision)、召回率

(Recall)以及F1值进行评价。它们在公式结构上相同,但在具体含义上有所不同。给出 P (准确率)、 R (召回率)、 F (F1值)的计算方式如公式(7)、(8)、(9)所示:

$$P = \frac{|A \cap G|}{|A|} \quad (7)$$

$$R = \frac{|A \cap G|}{|G|} \quad (8)$$

$$F = \frac{2PR}{P+R} \quad (9)$$

对于序列标注结果评价, $|A|$ 代表预测集的情感要素序列(观点对象和观点表达)总数, $|G|$ 代表标准集的情感要素序列总数, $|A \cap G|$ 代表预测集与标准集完全匹配的情感要素序列总数。而对于情感要素组合匹配结果评价, $|A|$ 代表预测集的情感要素组合结果总数, $|G|$ 代表标准集的情感要素组合结果总数, $|A \cap G|$ 代表预测集与标准集完全匹配的情感要素组合结果总数。

5.5 实验结果

本次实验,我们共进行2组实验,每组实验各有4套不同的细粒度意见挖掘模型。2组实验分别基于CRF和BiLSTM-CRF,4套模型分别为BASELINE、+HOWNET、+DUTIR、+DOMAIN。其中,BASELINE仅使用字特征表示训练得到、+HOWNET在使用字特征表示的基础上加入知网情感词典特征表示、+DUTIR加入大连理工情感词典特征表示、+DOMAIN加入电商领域情感词典特征表示。表13给出了所有结果。从实验结果,我们可以看出:

表 13 实验结果

实验组	模型	序列标注结果评价			情感要素组合匹配结果评价		
		P	R	F1	P	R	F1
CRF	BASELINE	82.33	75.15	78.58	72.63	64.84	68.51
	+HOWNET	82.72	75.80	79.11(+0.53)	72.92	65.60	69.06(+0.55)
	+DUTIR	82.46	75.52	78.84(+0.26)	72.62	65.33	68.78(+0.27)
	+DOMAIN	82.37	77.10	79.65 (+1.07)	72.46	66.87	69.55 (+1.04)
BiLSTM-CRF	BASELINE	79.16	79.00	79.08	70.37	67.58	68.95
	+HOWNET	80.03	79.00	79.51(+0.43)	70.86	67.41	69.09(+0.14)
	+DUTIR	79.39	79.36	79.39(+0.31)	70.52	67.70	69.08(+0.13)
	+DOMAIN	81.01	79.90	80.45 (+1.37)	73.14	69.27	71.15 (+2.20)

表 14 与其他方法的对比

模型	序列标注结果评价			情感要素组合匹配结果评价		
	P	R	F1	P	R	F1
Yang et al.	82.97	74.11	78.29	73.00	63.83	68.11
Liu et al.	74.23	76.54	75.37	69.08	63.44	65.69
Ours	81.01	79.90	80.45	73.14	69.27	71.15

- 加入情感词典特征表示后的各模型相较于同一实验组的BASELINE, F1值都得到了提升, 表明情感词典的有效性。
- 无论是在实验组 CRF 还是在实验组 BiLSTM-CRF 上, 基于领域情感词典特征表示的模型+DOMAIN 在各项性能上均优于另外两种加入其他情感词典特征表示的模型, 表明领域情感词典可以更好地提高领域内细粒度意见挖掘任务的性能。
- BiLSTM-CRF 的整体性能优于 CRF, 表明深度学习模型更适合本任务。

5.6 与其他工作的对比

我们将本文方法得到的实验结果与其他方法进行了对比, 对比结果如表 14 所示。其中, Yang et al. 基于文献^[14], 采用 Semi-Markov CRF 作为序列标注模型; Liu et al. 基于文献^[7], 采用 RNN 作为序列标注模型。由于本文用到的实验数据缺乏预训练分布式表示以及语言特征, 实际测得性能要稍低于原作者提出的方法。Our 对应表 13 中性能最佳的模型 BiLSTM-CRF+DOMAIN。通过对比可以看出, 我们的模型取得了最好结果。

5.7 实例分析

表 15 实例分析

实例	胜在质量可靠	
标准答案	胜在[质量/TGT][可靠/XPRP]	
模型	预测结果	是否出现在情感词典中
BASELINE	胜在质量可靠	N/A
+HOWNET	胜在[质量/TGT][可靠/XPRP]	是
+DUTIR	胜在[质量/TGT][可靠/XPRP]	是
+DOMAIN	胜在[质量/TGT][可靠/XPRP]	是
实例	物流给 5 星	
标准答案	[物流/TGT]给[5 星/XPRP]	
模型	预测结果	是否出现在情感词典中
BASELINE	物流给 5 星	N/A
+HOWNET	物流给 5 星	否
+DUTIR	物流给 5 星	否
+DOMAIN	[物流/TGT]给[5 星/XPRP]	是

我们对各个模型在测试集上得到的序列标注预测结果进行分析后发现:

- 在序列标注模型中加入情感词典特征表示的最大作用是改善原有基准系统 (BASELINE) 对于低频观点表达的预测效果。
- 领域情感词典相较其他情感词典具有更高的覆盖率。

以表 15 给出的两个实例对其进行说明。由于两组实验取得了一致的预测结果, 因此对其进行了合并。

在第一个实例中, 观点表达“可靠”在训练数据中的出现频次仅为 6 次, 属于低频观点表达, 因而两组实验组的 BASELINE 模型都难以对其进行识别。而“可靠”在知网情感词典、大连理工情感词典、电商领域情感词典中都有出现, 因此加入情感词典特征表示的各个模型都将其识别了出来。

在第二个实例中, 观点表达“5 星”是电商领域中产品的评分倾向, 具有一定的领域特性。由于该词条仅在电商领域情感词典中出现, 因而只有+DOMAIN 模型将其识别了出来。

6 结论

本文提出了一种基于领域情感词典特征表示的细粒度意见挖掘方法。我们首先构建一个新的电商领域情感词典, 然后基于该词典在电商评论文本上构建特征表示并将其加入序列标注模型的输入部分。该方法在不进行复杂的特征设计, 不进行过于耗时的预处理操作的前提下, 充分利用情感词典资源, 丰富了序列标注模型输入部分的特征表示。实验结果表明, 基于电商领域情感词典的方法在 CRF 和 BiLSTM-CRF 这两种序列标注模型上都取得了良好的效果, 并且超过其他情感词典。本文构建的领域情感词典已经公布在 Github² 上, 在业界共享。

参考文献

- [1] Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language[J]. Language resources and evaluation, 2005, 39(2-3): 165-210.
- [2] Liu B. Sentiment analysis and opinion mining[J]. Synthesis lectures on human language technologies, 2012, 5(1): 1-167.
- [3] Sang E F, Veenstra J. Representing text chunks[C]//Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 1999: 173-179.
- [4] Lafferty J, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[J]. 2001.
- [5] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[J]. arXiv preprint arXiv:1603.01360, 2016.
- [6] Nakayama Y, Fujii A. Extracting condition-opinion relations toward fine-grained opinion mining[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 622-631.
- [7] Liu P, Joty S, Meng H. Fine-grained opinion mining with

² <https://github.com/zeitiempo/ECSD>

- recurrent neural networks and word embeddings[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 1433-1443.
- [8] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
- [9] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C]//Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005: 347-354.
- [10] Choi Y, Cardie C, Riloff E, et al. Identifying sources of opinions with conditional random fields and extraction patterns[C]//Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005: 355-362.
- [11] Shariaty S, Moghaddam S. Fine-grained opinion mining using conditional random fields[C]//Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. IEEE, 2011: 109-114.
- [12] Breck E, Choi Y, Cardie C. Identifying Expressions of Opinion in Context[C]//IJCAI. 2007, 7: 2683-2688.
- [13] Choi Y, Cardie C. Hierarchical sequential learning for extracting opinions and their attributes[C]//Proceedings of the ACL 2010 conference short papers. Association for Computational Linguistics, 2010: 269-274.
- [14] Yang B, Cardie C. Extracting opinion expressions with semi-markov conditional random fields[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational linguistics, 2012: 1335-1345.
- [15] Yang B, Cardie C. Joint inference for fine-grained opinion extraction[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2013, 1: 1640-1649.
- [16] Irsay O, Cardie C. Opinion mining with deep recurrent neural networks[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 720-728.
- [17] Katiyar A, Cardie C. Investigating lstms for joint extraction of opinion entities and relations[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016, 1: 919-929.
- [18] Ma X, Hovy E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[J]. arXiv preprint arXiv: 1603.01354, 2016.
- [19] Kaji N, Kitsuregawa M. Building lexicon for sentiment analysis from massive collection of HTML documents[C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007.
- [20] Jijkoun V, de Rijke M, Weerkamp W. Generating focused topic-specific sentiment lexicons[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 585-594.
- [21] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造[J]. 情报学报, 2008, 27(2): 180-185.

作者简介:



郁圣卫 (1994—), 男, 硕士研究生, 主要研究领域: 自然语言处理。
地址: 江苏省苏州市 十梓街 1 号 苏州大学 计算机科学与技术学院 理工楼 217 室 邮编: 215006, 电话: 15601863738
E-mail: zeitiempo@163.com



卢奇 (1992—), 男, 硕士研究生, 主要研究领域: 自然语言处理。
地址: 江苏省苏州市 十梓街 1 号 苏州大学 计算机科学与技术学院 理工楼 217 室 邮编: 215006, 电话: 13024526987
E-mail: luqibhf@qq.com



陈文亮 (1977—), 男, 博士, 教授, 主要研究领域: 自然语言处理。
地址: 江苏省苏州市 十梓街 1 号 苏州大学 计算机科学与技术学院 理工楼 417 室 邮编: 215006, 电话: 15618055040
E-mail: wlchen@suda.edu.cn