面向微博的中文反语识别研究 *

邓钊 [1]，贾修一 [1]，陈家骏 [2]

（1.南京理工大学 计算机科学与工程学院，江苏 南京 210094；

2. 南京大学 计算机科学与技术系，江苏 南京 210023）

摘要：反语识别已成为当前研究的热点，但当前对于中文反语识别研究报道较少。针对于此，本文主要研究面向社交网络的中文反语识别。在借鉴外文相关工作的基础上，结合中文语言和社交网络的特性，构建了六种特征，通过信息增益对比了各种特征有效性，并检测不同分类器在该特征体系中的稳定性。实验结果表明本文构建的特征在识别反语的任务中有显著的效果。

关键词：反语识别；特征构建；微博

中图分类号：TP391　　**文献标识码：**A

Research on Chinese Irony Detection in Microblog

DENG Zhao[1], JIA Xiuyi[1], CHEN Jiajun[2]

(1.School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China;

2. Department of Computer Science and Technology, Nanjing University, Nanjing, Jiangsu 210023, China)

Abstract: Irony detection has drawn much attention in recent years. However, most studies are based on foreign languages. In this paper, we focus on the Chinese irony detection in microblog. By considering the characteristics of both Chinese language and social networks, we build a set of discriminating features for Chinese irony detection. For these features, information gain is applied to compare their efficiency, and several classifiers are also applied to test their stability. Experimental results show the efficiency of our proposed features.

Key words: irony detection; feature construction; microblog

1 引言

反语通常又称为“说反话”，其字面意思和所要表达的意思相反，是一种带有强烈情感色彩的修辞手法。在社交网络里，反语已成为一种普遍的语言表达方式。在微博这类包含符号，图片和短文本等信息的分享传播平台，针对热门话题及争议话题，用户常常使用反语表达如嘲弄或讽刺等强烈情感倾向。而反语的使用增加了微博情感分析的难度，为提高微博情感分析的准确率，我们需要对反语识别进行研究。

目前反语识别的可计算化研究已引起一些学者的关注，但主要集中在英文为代表的外文短文本反语识别。据我们所知，对于中文反语研究，目前还处于起步阶段，只有 Tang 等人针对繁体字构建了一个反语语料库并分析了反语常见的句式结构 [1]。反语的识别需要正确理解该话语发生的具体语境，而当前研究很难形式化给出语境的计算表达式，特别是在缺少自然会话中的语气、身体姿势等用于视听理解的辅助手段情况下，这就给反语识别带来了极大的困难。此外，和英文反语识别相比，中文通常使用谐音词或歧义词等来表达反语情感，这也使得中文反语识别在词语层面上就比英文反语识别具有更深的难度，使得我们无法直接将针对外文反语识别的研究简单的运用到中文反语识别。

和自然会话相比，社交网络上的语言表达虽然缺少一些语气或肢体行为等辅助手段，但社交网络平台本身的一些特性也有助于反语的使用和识别，如连续标点符号和表情符号的使用等等，这在一定程度上能够帮助我们理解反语所在的语境。有鉴于此，我

* **收稿日期：**　　**定稿日期：**

基金项目：国家自然科学基金（61403200）；江苏省自然科学基金（BK20140800）

们在参考外文相关工作的基础上，考虑中文语言的特性和微博平台的特点，对识别反语的特征构建做了初步的研究。

本文主要使用基本词汇情感、标点符号、谐音词、微博长度、动词被动化和文本情感模糊度六种特征构建反语识别特征体系。在此基础上，通过信息增益方法对比了各特征对反语识别的影响程度。此外，还实验验证了在该特征体系下不同分类器的分类性能及稳定性。

2 相关工作

反语作为一种修辞现象，受到语言学家，心理学家和认知学家的广泛关注[2]。随着情感分析技术的深入研究，反语识别也得到了自然语言处理领域学者们的重视。对于反语识别的研究，我们依据研究角度不同，将相关工作分为两类。

第一类工作主要从语言学和心理学角度出发。对于英文的反语识别，Gibbs 等人从心理学角度分析了口语中反语的形成和实用性[3]。Utsumi 从语言学角度分析了反语的本质，定义了反语的三大要素，提出了一个统一识别反语的计算模型[4]。对于中文的反语识别，刘正光通过对反语在中文对话产生过程的研究，尝试从语言学和心理学角度分析反语的本质[2]。Li 也从语言学角度分析了中英文中反语使用的差异性[5]。

第二类工作主要从反语识别的可计算化角度出发。该类工作又可细分为两种：第一种是研究反语识别的特征构建。对于英文反语识别，González-Ibáñez 等人仅通过字典中的词汇和“@<用户>”标签等简单的特征识别反语，发现仅通过一些简单的词汇特征无法准确有效的识别反语[6]。Reyes 等人从不同角度研究了电商评论和社交媒体中的反语识别工作，构建了包含 n 元文法、POS 的 n 元文法、滑稽程度、词汇褒贬程度、情感复杂度和欢乐程度等抽象复杂的特征体系[7,8,9]。Burfoot 等人针对新闻语料，在基本词袋特征基础上讨论了标题，脏话和俚语等特征[10]。对于葡萄牙文反语识别，Vanin 等人研究了固定词汇、标点号、词性序列和特殊的葡萄牙语表达方式等特征识别反语[11]。Barbieri 等人针对 Twitter 研究了意大利语的反语识别[12]。

第二种主要从分类算法的研究角度出发。González-Ibáñez 等人使用支持向量机和逻辑斯蒂回归两种经典的分类算法识别反语，发现支持向量机算法表现普遍好于逻辑斯蒂回归[6]。Reyes 等人在文献[8]中使用朴素贝叶斯和决策树两种算法识别反语，分别研究了在数据平衡和数据不平衡状态下分类器的性能。Reyes 等人在文献[9]中使用了朴素贝叶斯、支持向量机和决策树三种经典算法识别反语，研究了三种分类在不同数据集上识别反语的性能。Tsur 等人提出了一种基于模式匹配的半监督学习方法识别反语[13,14]。

反语识别的可计算化研究主要集中在英文为代表的外文语料上，而基于中文短文本的反语识别研究只有 Tang 等人针对繁体字进行了语料库构建和分析了反语的常用句式结构[1]，对于反语识别所需的特征和分类算法等则没有涉及。由于中英文语言差异性，相关外文的工作无法直接应用于本文的工作中，例如文献[3]中的“@<用户>”标签特征未出现在本文的特征体系中，因为在中文社交平台中用户之间的关系是松散的。表 1 统计了我们构建的语料库中反语集和 10000 条非反语微博中含有“@<用户>”标签的微博比例。如表 1 描述，反语集和非反语集的“@<用户>”标签比例相差微小。

表 1 “@<用户>”标签比例数

	“@<用户>”标签比例数
反语集	0.196
10000 条反语微博	0.182

3 面向微博的中文反语识别特征体系

本章针对中文反语的特点，在相关工作的基础上，考虑微博自身的特点，构建了用于微博反语识别的特征体系，主要包括基本词汇情感，中文特有的谐音词，连续的标点符号，微博的长度，动词被动化，双引号内外情感模糊度等六种特征。

a．基本词汇情感。在自然语言处理领域，通常使用 n 元文法来表示基本的词汇特

征，是指将相邻的 n 个单词作为一个特征。文献[3]研究发现在 Twitter 反语识别的任务中二元文法和三元文法不但比一元文法复杂而且实验结果比一元文法差，所以在同为短文本的中文微博的反语识别任务中，本文的基本词汇情感特征只应用一元文法。在一元文法的特征表示中，中文首先需要使用分词工具将整条微博分词，然后建立词典构建特征。在分词过程中，由于微博约束比较少，所以微博中经常出现病句、错别字以及网络用词，这些问题往往会导致分词错误。由于错误词汇出现频率不高，针对该问题，故将一些低频词汇从词典中过滤掉。此外，本文主要研究面向中文的反语特征体系，非中文词汇也不予考虑。

b．中文特有的谐音词。谐音词是中文特有的，意思是和正确词汇发音相同或者相似的词汇，例如“河蟹”是“和谐”的谐音词。微博的内容往往偏向口语化，很多用户使用谐音词代替相应词汇表达反语，讽刺等情感倾向。实际上，大部分谐音词作为单独的词汇已包含于基于一元文法的词典，但是有些特定谐音词因不是正式词汇无法被分词工具准确的识别，所以需要通过导入用户自定义常用谐音词词典使分词工具识别这些词汇。

c．连续的标点符号。Vanin 等人和 Dmitry 等人都提及连续的标点符号在识别反语任务中的重要性 [11,14]，Carvalho 等人也通过模式匹配方法统计连续标点符号在反语语料中出现次数验证了连续标点符号是识别反语的重要线索 [15]。由于微博的随意性，用户经常使用连续的标点符号表达自己的情感。Vanin 和 Dmitry 等人在反语识别任务中将连续标点符号的个数作为特征值，但是我们在分析语料时发现大多数连续的两个标点符号反映用户情感并不明显，只有 3 个及 3 个以上的标点符号同时出现时才能表达用户情感，而且用户情感并未随着标点符号个数增加而波动，所以本文只提取 3 个及以上的连续的标点符号作为特征，并且使用布尔值表示该特征。

d. 微博的长度。张林等人发现 app 短文本评论的长度会影响情感的判别 [16]，评论越长其中包含的非情感信息越多，而这些非情感信息会影响情感的判别。因此，我们认为同为短文本的微博的长度也可能会影响反语的识别。本文根据微博长度将微博分为 3 个等级，分别为：短微博，中等长度的微博和长微博。

e．动词被动化。在中文中许多动词用法很特殊，这些特殊动词被动化之后情感会发生巨大的反转。例如“就业”是个中性动词，但是如果在“就业”前加上“被”字，比如“我被就业了”，那么情感将发生极大的反转。因为动词的这种用法通常不会出现在正式文献中，所以通过统计动词和该动词被动化之后在正式文献中的频率可以自动识别这些特别的动词。

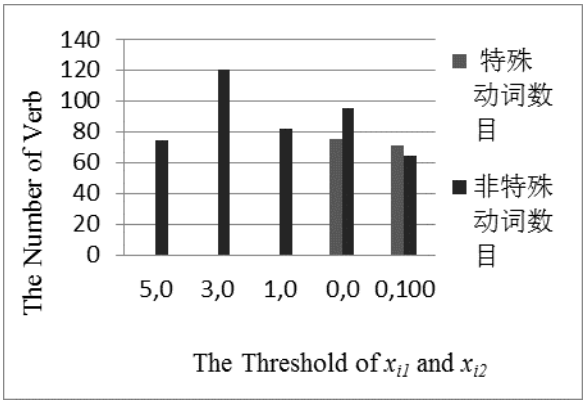

图 1. x_{i1}, x_{i2} 取不同阈值时整个数据集特殊动词的统计

识别这种特殊动词实验的正式文献语料是搜狗实验室收集的 48.2M 新闻语料，主要来自搜狐新闻网站。图 1 是 x_{i1}, x_{i2} 取不同阈值整个数据集特殊动词的统计，其中 x_{i1} 是动词被动化后在正式文献出现的次数，x_{i2} 是动词原形在正式文献出现的次数。横坐标是的 x_{i1}, x_{i2} 阈值，例如(5,0)中的 5 是 x_{i1} 的值，而 100 表示 x_{i2} 大于等于 100，纵坐标是当 x_{i1}, x_{i2} 取具体的阈值时，通过手动统计

数据集中特殊动词和非特殊动词的数目。由图 1 可知，当动词被动化后在正式文献出现的次数高于 1 时，这个动词是特殊动词的可能性几乎为 0，而 x_{i1} 为 0 时，有非常大的可能性是特殊动词。为了不丢失这种特殊动词，x_{i1}, x_{i2} 的阈值设置为(0,0)，然后手动剔除非特殊动词，最后我们将特殊动词保存在动词被动化字典中，在下文实验中我们通过布尔值方法表示该特征。

f．双引号内外情感模糊。Reyes 等人在文献[10]中强调情感模糊是反语效果的重要表现，但是他们的工作是以基于英文的 Saif 工作为基础的 [17]，而中英文语言差异大，所以无法直接借鉴他们的工作。然而在中文中用户经常将情感词放入引号中，用褒义词表达贬义或者用贬义词表达褒义，所以引号内外的情感通常不一致。本文通过设计公式 1 计算引号内外情感模糊。

$$Amb(X)=\begin{cases}1 & |X_{in}\cap P|\times|X_{out}\cap N|+|X_{in}\cap N|\times|X_{out}\cap P|>0\\ 0 & |X_{in}\cap P|\times|X_{out}\cap N|+|X_{in}\cap N|\times|X_{out}\cap P|=0\end{cases} \quad (1)$$

在公式 1 中，X_{in} 表示微博 X 引号内的词语集合，X_{out} 表示微博 X 引号外的词语集合，P 表示褒义情感词典，N 表示贬义情感词典，$|*|$ 表示集合中元素个数。例如，某微博双引号内有积极词汇而双引号外有贬义词汇或者该微博双引号内有贬义词汇而双引号外有褒义词汇，那么引号内外情感模糊 $Amb(X)$ 为 1。

4 实验

4．1 数据与实验设置

在中文微博平台的新浪微博上，用户可以发布最多 140 字的微博。一条微博除了正常的文字以外还可以包括“@<用户>”，“#主题#”，URLs 等。其中在第二章中已说明“@<用户>”在中文微博中无法作为特征识别反语。因为在新浪微博中分享功能会自动包含原网页的 URLs，所以 URLs 在本文中也作为噪声过滤掉。

和 Twitter 不同，中文微博平台的用户几乎不使用注释(#sarcasm, #sarcastic)表明该微博是反语或者其他情感分类，所以只能通过手动标注。为了检测整个特征体系的有效性，我们从新浪微博平台获取的微博中标记了 300 条反语和 28545 条非反语。

一元文法特征提取过程中的分词工具使用 java 开源分词工具 ansj[1]。微博长度特征中的短微博的长度小于 10，中等长度的微博长度介于 10 到 20，长微博的长度大于 20。我们在实验中也尝试了将微博的长度设置成其他阈值，但是阈值取 10 和 20 时实验结果最好。双引号内外情感模糊度特征提取过程中的情感词典使用台湾大学 NTUSD 实验室整理的情感词典 [2]。

4．2 各特征的信息增益

我们首先通过信息增益(IG) 对比了中文特有的谐音词，连续的标点符号，微博长度，动词被动化和双引号内外情感模糊度等五种特征对反语识别的影响程度。因为实验数据不平衡，所以实验首先从非反语集中随机抽取 300 条数据和反语集组成实验数据集，此过程重复进行 20 次，然后比较各特征对在不同数据集上的信息增益以及各特征在不同数据集上的稳定性，实验结果如图 2 所示。

图 2 中，中文特有的谐音词特征的信息增益最高，基本达到 0.05 左右，动词被动化特征的信息增益最平稳，稳定在 0.04 左右，连续的标点符号特征的信息增益基本也达到了 0.03 左右。双引号内外情感模糊度的信息增益很低，只有 0.02 左右，可能由于特征提取的方法过于简单，或者情感词典的不完整等原因导致该特征信息增益偏低。在我们人工标记反语语料时该特征是一个重要的依据，所以尽管信息增益较低，我们仍然将该特征纳入我们的识别特征体系。微博长度的信息增益很不稳定，最高可达到 0.14，最低几乎为 0。

4．3 微博长度对识别准确率的影响

由于微博长度的信息增益不稳定而微博长

[1]https://github.com/NLPchina/ansj_seg
[2] http://ccf.datatang.com

度确实会影响反语的识别，所以本文通过区分反语和不同长度的非反语微博验证微博长度对反语识别的影响。本文从非反语中随机抽取300条特定长度的微博和反语集组成数据集，然后使用决策树分类器和5倍交叉验证测试数据集，实验结果如图3所示。

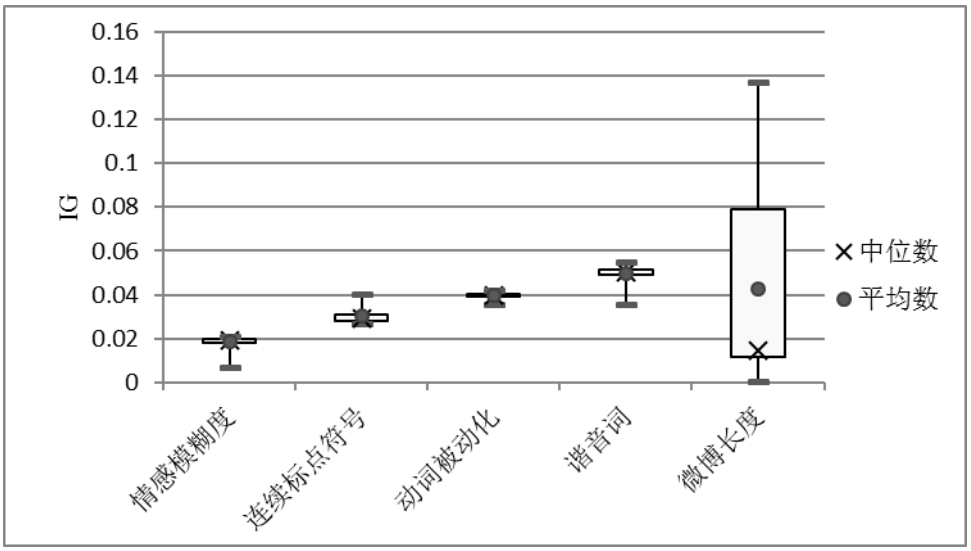

图 2.　20 组数据各特征的信息增益的箱线图

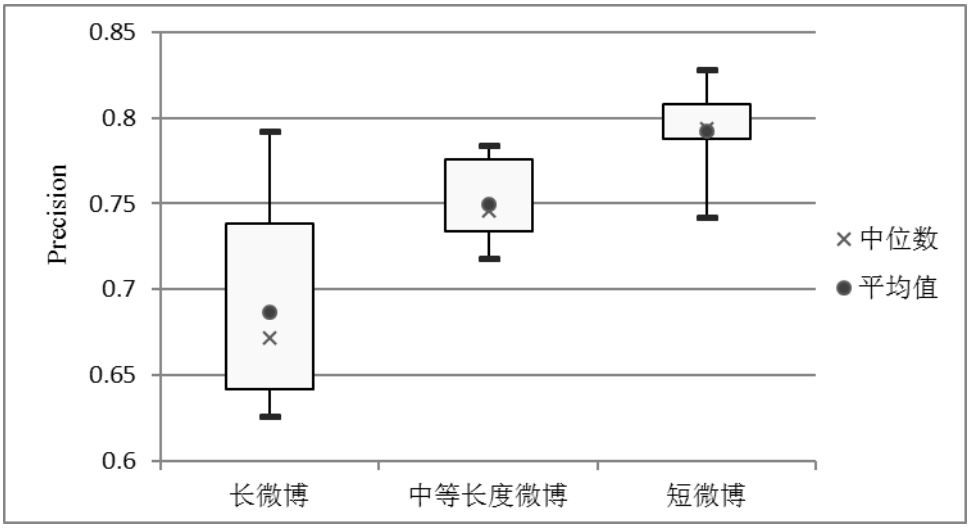

图 3.　区分反语和不同长度非反语的准确率的箱线图

图3中，区分反语和短微博的非反语集任务的准确率基本稳定于0.8左右，而区分反语和中等长度的非反语集任务的准确率徘徊于0.75左右，最后区分反语和长微博的非反语集任务的准确率却大都低于 0.7。由此可见，识别反语的难度确实和微博的长度有关联，实验结果和张林等人的结论基本一致。

4．4 不同分类器在特征集合上的有效性

最后本节将通过反语识别任务检测整个特征体系和仅有一元文法特征的有效性，实验使用五种经典的分类器：支持向量机(SVM)，决策树(C4.5)，朴素贝叶斯(NB)，逻辑斯蒂回归(LR)和随机森林(RF)。该实验数据集包括300条反语集和从非反语集中任意抽取300条数据。分类器使用5倍交叉验证进行测试。实验结果如表2和表3所示。

表 2.　在整个特征体系下五种分类器实验结果对比

	Precision	Recall	F-measure
SVM	0.710	0.610	0.656
NB	0.7013	0.4395	0.5404
C4.5	0. 7484	**0.7186**	0.7296

LR	**0.7831**	0.6915	**0.7345**
RF	0.7014	0.4867	0.5747

表 3. 只在一元文法特征下五种分类器的实验结果对比

	Precision	Recall	F-measure
SVM	0.6512	0.4886	0.5583
NB	0.6948	0.4268	0.5288
C4.5	0.7186	**0.6957**	0.7071
LR	**0.7674**	0.6629	**0.7113**
RF	0.673	0.4323	0.5264

由表 2 所知，决策树分类器在准确率、召回率和 F 值都要高于支持向量机，朴素贝叶斯以及随机森林，而逻辑斯蒂回归分类器在准确率和 F 值都比决策树分类器高。对比表 2 和表 3 可知，对特征维数不敏感的 SVM 在添加少数新特征的情况下，分类器的性能提高最多，而朴素贝叶斯的性能提高最低。

4．5 不同分类器在特征集合上的稳定性

由于反语集小而非反语集比较大，所以本文从非反语集中随机抽取 300 条数据和反语集组成实验数据集，此过程重复进行 20 次得到 20 组实验数据集，测试五种分类器在该任务中的稳定性，图 4、图 5 和图 6 统计了 20 组实验五种分类器准确率、召回率和 F 值的四分位数。

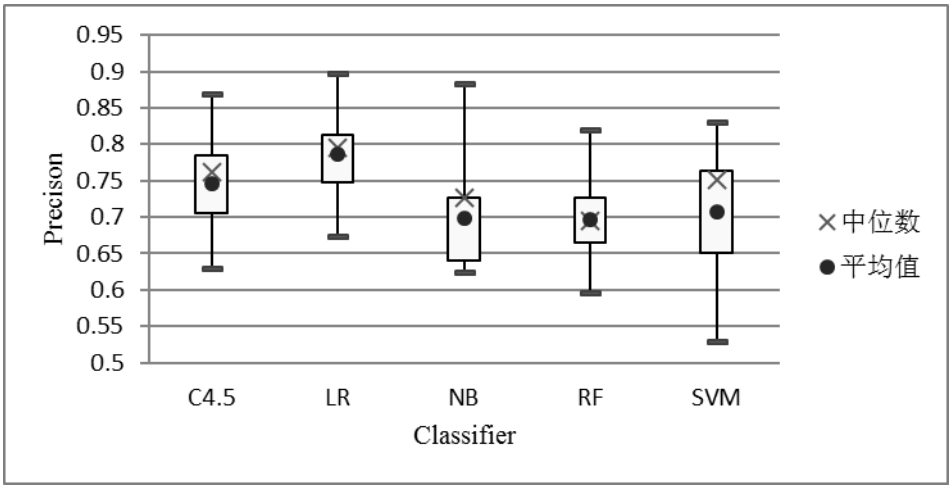

图 4. 20 组实验各分类器准确率的箱形图

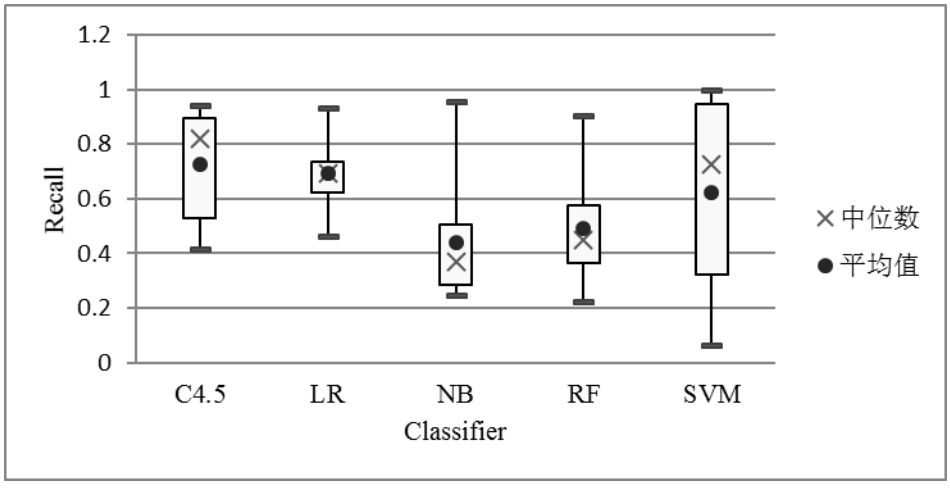

图 5. 20 组实验各分类器召回率的箱线图

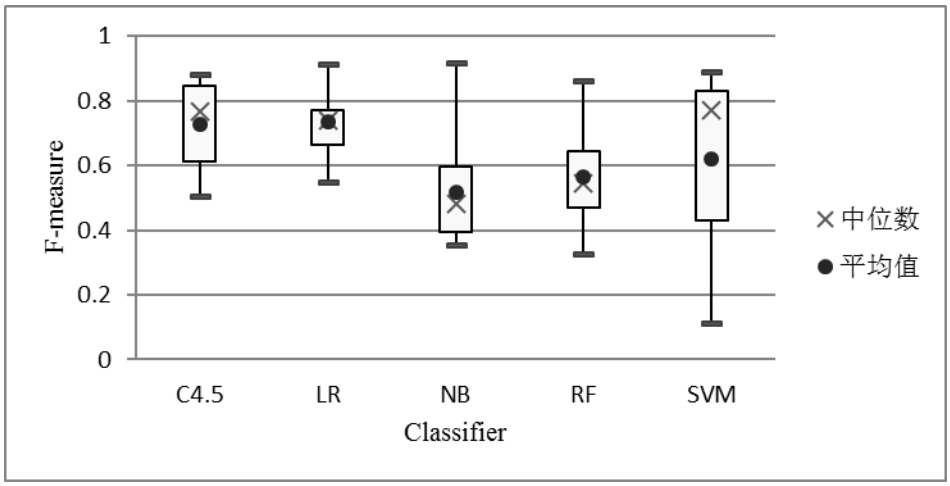

图 6. 20 组实验各分类器 F 值的箱线图

由图 4、图 5 和图 6 所示，支持向量机分类器虽有很好的准确率，但是召回率和 F 值极不稳定。决策树和逻辑斯蒂回归分类器的准确率、召回率和 F 值都要比朴素贝叶斯和随机森林分类器高。决策树分类器的召回率和 F 值比逻辑斯蒂回归高，但是决策树的准确率不及逻辑斯蒂回归分类器。

5 结论及展望

本文主要研究中文微博中反语识别的可计算化问题。在考虑中文语言特性和微博语言表达特性的基础上，构建了基于一元文法的词汇特征、中文特有的谐音词、连续标点符号、微博长度、动词被动化和双引号内外情感模糊等六种特征，并实验验证了该特征体系在识别反语中的有效性和稳定性。

在未来的工作里，基于上述实验中表现出的不足，我们将改进部分特征的提取方法和条件，我们还需从更深层次挖掘识别反语的特征。研究针对不同特征空间表示的分类算法和构建更丰富的反语语料库也是我们下一步重点研究的工作。

参考文献

[1] Tang YJ, Chen H. Chinese irony corpus construction and ironic structure analysis[C]// The 25th International Conference on Computational Linguistics: Technical Papers, 2014:1269-1278.

[2] 刘正光. 反语理论综述[J].解放军外国语学院学报, 2002,22(4):16-18.

[3] Gibbs R W, Colston H L. Irony in language and thought: a cognitive science reader[M]. New York：Lawrence Erlbaum Associates, 2007.

[4] Utsumi A. A unified theory of irony and its computational formalization[C]// International Conference on Computational Linguistics, 1996:962-967.

[5] Xiang Li. Irony Illustrated: A Cross-Cultural Exploration of Situational Irony in China and the United States[M].USA:Sino-Platonic Papers,2008.

[6] González-Ibáñez R, Muresan S, Wacholder N. Identifying sarcasm in Twitter: A closer look[C]// In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Short Papers, 2011：581-586.

[7] Reyes A, Rosso P, Buscaldi D. From humor recognition to irony detection: The figurative language of social media[J]. Data & Knowledge Engineering, 2012, 74(3):1–12.

[8] Reyes A, Rosso P, Veale T. A multidimensional approach for detecting irony in Twitter[J]. Language Resources & Evaluation, 2013, 47(1):239-268.

[9] Reyes A, Rosso P. Making objective decisions from subjective data: Detecting irony in customer reviews[J]. Decision Support Systems, 2012, 53(4):754–760.

[10] Burfoot C, Baldwin T, Burfoot C. Automatic satire detection: Are you having a laugh?[C]// In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL: Short Papers, 2009:161-164.

[11] Vanin AA，Freitas LA，Vieira R, et al. Some clues on irony detection in tweets[C]// WWW 2013 Companion. ACM 978-1-4503-2038-2, 2013:635-636.

[12] Francesco B, Francesco R, Horacio S. Italian irony detection in Twitter: a first approach[C]// In Proceedings of the First Conference on Computational Linguistics, 2014: 28-32.
[13] Tsur O, Davidov D. Icwsm – a great catchy name: Semi-supervised recognition of sarcastic sentences in product reviews[C]// In Proceedings of the International AAAI Conference on Weblogs & Social. 2010:162-169.
[14] Davidov D, Tsur O. Semi-supervised recognition of sarcastic sentences in twitter and amazon[C]// In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, 2010:107-116.
[15] Carvalho P，Sarmento L. Clues for detecting irony in user-generated contents: oh...!! it's "so easy" ;-)[C]// In TSA'09 - 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement, Hong Kong, 2009:53-56.
[16] 张林, 钱冠群, 樊卫国等. 轻型评论的情感分析研究[J]. 软件学报, 2014, (12):2790-2807.
[17] Saif M, Dunne C., Bonnie D. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus [C]// In Proceedings of the 2009 Conference on EMNLP, 2009:599-608.

邓钊（1990-），男，硕士研究生，主要研究领域为自然语言处理，数据挖掘。
Email: dengzhao1hao@foxmail.com

通讯作者：贾修一（1983-），男，博士，副教授，硕士生导师，主要研究领域为机器学习、自然语言处理、数据挖掘。
Email: jiaxy@njust.edu.cn

陈家骏（1963-），男，博士，教授，博士生导师，主要研究领域为自然语言处理，软件工程。
Email: chenjj@nju.edu.cn