短文本情感分析的研究现状

——从社交媒体到资源稀缺语言

拥措^{1,2} 史晓东² 尼玛扎西¹

(西藏大学信息科学技术学院 拉萨 850000)¹ (厦门大学信息科学与技术学院 福建 厦门 361005)²

摘要 随着社交网络的逐渐成熟,各类语种的文本出现在社交网络上。而这些非规范的短文本蕴藏着人们对事物的褒贬、需求等意见,是国家政府和企业了解公众舆论的重要参考信息,具有重大的研究价值和应用价值。首先,对目前互联网短文本情感分析领域常用的神经网络、跨语言和应用语言学知识等研究方法进行归纳和总结;其次,对当前短文本情感分析研究的热点领域——社交媒体和资源稀缺语言的情感分析进行现状分析;最后,对短文本情感分析研究的趋势进行总结,分析存在的问题,并对未来进行展望。

关键词 情感分析,短文本,社交媒体,资源稀缺语言

中图分类号 TP391.1 文献标识码 A

Research Status of Sentiment Analysis for Short Text

——From Social Media to Scarce Resource Language

YONG Tso^{1,2} SHI Xiao-dong² NyimaTrashi¹

(School of Information Science and Engineering, Tibet University, Lhasa 850000, China)¹

(School of Information Science and Engineering, Xiamen University, Xiamen, Fujian 361005, China)²

Abstract With the gradual maturity of social networks, texts of various languages appear on social networks. These short texts contain praise and demand of people. They have important reference for the government and enterprises to understand the public opinion, which have significant value in research and application. First of all, the current research methods of sentiment analysis for Internet short text were summarized, including neural network, cross language and applied linguistics knowledge. Secondly, the current situation analysis was carried out in the hot spot field of sentiment analysis for short text. Finally, the research trend of sentiment analysis for short text was summarized, and the future was prospected.

Keywords Sentiment analysis, Short text, Social media, Scarce resource language

1 引言

近年来,不断发展与壮大的互联网不仅丰富和方便了人们日常生活的方方面面,也给人们的言论自由提供了平台。人们在网络上通常以文本形式表达自己的观点,进而产生了关于各种事物的海量评论文本。随着社交网络的逐渐成熟和移动终端技术的迅猛发展,微博、微信等新型社交媒体和交流平台使得人们能够更加方便地点评热点事件和表达自己的情感。这些短文本蕴藏着人们对事物的褒贬、需求等意见,是国家政府了解公众舆论的重要参考信息。互联网各类平台上网民发布的信息中包含着不同趋向的情感特征,深度挖掘这些特征对于舆情监控、市场营销都有着重要意义。为此,自动地从用户产生的海量短文本中挖掘出用户的情感信息,使政府和企业能够做出正确的决策行为,成为了一个迫切的需求。情感分析就是在这样的大背景下应运而生的。

初期的研究工作主要是在英文语境下进行;随着互联网国界的不断扩张以及各种语言信息化的迅速发展,不同国家

和地区的语言文字逐渐出现在了互联网上。规范文本的情感分析在国内外已经发展得比较成熟,但对社交媒体领域和资源稀缺语言的情感分析研究在最近几年才开始。

本文首先对短文本情感分析的研究现状进行了综述,分别对利用神经网络、跨语言方法和应用语言学知识进行情感分析的研究方法进行了归纳和总结;其次,对当前短文本情感分析研究的热点领域——社交媒体和资源稀缺语言的情感分析研究进行了现状分析;最后,对短文本情感分析的研究现状进行了总结,提出存在的问题,并对未来进行了展望。

2 情感分析

理解情感是人工智能一直以来的目标之一。情感分类作为对情感理解的一个小步骤,其目的是将情感分类为积极/消极情感或者更为细腻的情感类别。为了解情感,研究者提出了很多方法:基于情感词典分类^[1],早期的基于机器学习的方法^[2],以及最近的神经网络模型,如卷积神经网络^[3]、递归自编码网络^[4]、长短期记忆神经网络^[5]。尽管这些神经网络

本文受国家自然科学基金项目(61262086)资助。

拥措(1974—),女,硕士,副教授,主要研究方向为藏语自然语言理解, E-mail: yongtso@163.com(通信作者);史晓东(1966—),男,博士,教授,主要研究方向为自然语言处理、人工智能;尼玛扎西(1964—),男,博士,教授,主要研究方向为藏语自然语言处理。

模型取得了巨大的成功,但已有研究中还存在一些缺陷。1)递归自编码网络和基于句法树的长短期记忆神经网络^[6]等树结构模型依赖解析树结构和昂贵的短语级别标注,当仅用句子级标注训练时,这些模型的性能会大幅下降。2)序列模型,如 CNN 和 RNN,不容易产生竞争的结果^[6]。3)语言学方面的知识还未完全应用在神经网络里。

2.1 神经网络情感分析

近年来,有许多神经网络被用于情感分类。最前沿的模型可能是递归自编码神经网络,它从句子的子短语递归建立一个句子表示。这样的递归模型通常依赖于输入文本的树结构,为了产生竞争的结果,通常需要在每个子短语上做标注。序列模型不依赖于特殊的树结构,比如 CNN 是另一个广泛应用于情感分类的模型。由于 LSTM 模型具有较强的上下文建模能力,它也经常用于学习句子级别的表示中。

2.2 在情感分类上应用语言学知识

语言学知识和情感资源对情感分类非常有用,比如情感词典、否定词和强度词。情感词典,如 Hu and Liu 情感词典^[30]和 MPQA 情感词典,被广泛用在情感分类上。否定词在改变语篇情感中起着关键作用,一些早期的否定模型旨在反转修改后文本的情感值的符号^[7]。Zhu 等^[8]将否定词作为特征加入到神经网络中。Kiritchenko 等^[9]将否定词和其他语言学知识融入到了 SVM 分类器中。强度词能改变一个词的程度,比如情感强度。一个短语的情感强度表示相关情感的强度,这对于细粒度情感分类非常重要。Taboada 等^[10]直接改变修饰词的极性,或通过一个固定的价值改变情感强度。Wei^[11]通过使用一个线性回归模型为内容词预测情感强度值。Malandrakis 等^[12]通过引入一个核函数并结合语义信息来预测情感得分。Qian 等^[13]为了开发简单的序列模型,也试图充分利用语言资源,以利于情感分类,作者试图开发不依赖于句法分析树和短语级别标注的一个简单模型,且为了获得好的性能,简单的模型可以受益于语言资源;该作者将情感词典、否定词和表示程度的词等语言学资源以约束条件的形式融入到现有的句子级别的 LSTM 分类模型中,取得了不错的性能效果。

2.3 跨语言情感分析

目前,最先进的情感分类方法大多是基于监督学习算法,需要大量手动标记的数据。然而,在不同的语言里,标记的资源通常是不平衡的。跨语言情感分类通过将资源丰富的语言的情感资源适应到资源贫乏的语言中来处理这个问题。

利用一种语言(语言 A,例如英文或中文)中的情感资源(词典或标注语料)对其他语言(语言 B,例如藏文)中的文本进行情感分类,可被看作是一种特殊的跨领域情感分类。现有的跨语言情感分析研究的主要思想是:首先,为了克服语言障碍,建立源语言和目标语言之间的关系;然后,开发一个适当的知识迁移方法,监督和半监督学习的方法都利用源语言标注数据为目标语言训练一个情感分类模型。

这里有两种供跨语言情感分析选择的传统解决方案:1)结合多分类器的集成学习^[14],这是集成学习解决方案的先驱,通过从不同的训练数据集学习多个分类器;2)迁移学习,为从一种语言知识适应到另一种语言而发展的策略。目前,大多数研究都探索迁移学习,专注于知识的适应。比如,Wan 等^[15]应用一个监督的协同训练框架反复地适应从两种语言

学习到的知识。所有这些方法都依赖于机器翻译来建立语言连接。

同时,未标注的并行数据也拉近了两种语言之间的差距。为了用 EM 算法解决特征覆盖问题,文献^[16]利用未标注的并行数据学习未登录情感词。同样,文献^[17]为了缓解数据稀疏性问题,使用未标注的并行数据进行特征聚类。他们也利用未标注并行数据来减少由于机器翻译、知识迁移产生的噪音和不正确的情感标记带来的负面影响。然而,并行数据也是一个稀缺资源。

一些现有的基于跨语言情感分析方法的迁移学习试图通过检查标签一致性来解决由错误标记引起的噪声知识问题。例如,为了过滤掉中文里不自信的标签,文献^[18]提出了一种监督学习方法,通过对英文中人工标注的标记和翻译的中文中被分类器预测的标记来检查标记的一致性。文献^[19]的工作虽然考虑了两种语言之间的知识迁移,但利用的是相同的思路。现有的大多数迁移学习方法由于迁移知识的错误,其性能并没有达到一个较好的水平。相反,为了滤除由错误的翻译导致情感变化的噪声知识,文献^[20]提出了融入迁移知识的验证模型,针对知识迁移的不可靠性提出一种半监督学习模型 CredBoost,在该模型中加入了知识验证和适应性识别过程,进一步保证了迁移知识可信性和适应性。该方法在 NLPCC2013 英-中跨语言情感分类数据上获得了最高性能。

近年来,文献^[21-22]提出用表示学习方法来解决跨语言分类的问题。这些方法的目的是为不同的语言学习统一特征表示。随着深度学习的兴起,共享深层表示(或嵌入)被应用在跨语言情感分析研究中^[23]。通常,利用平行语料库中的成对句子来学习跨语言的词向量,从而消除了对机器翻译系统的依赖^[24]。双语词向量将训练数据和测试数据映射到一个统一的空间中。文献^[25]将双语词向量运用于英文-中文跨语言情感分析中。上述工作在双语词向量学习过程中都使用了句子对齐。然而在情感分类过程中,只有在源语言的表示被用于训练,而在目标语言的表示被用于测试,忽略了源语言和目标语言之间的语义信息的相互作用。但是,高质量的双语词向量依赖大型任务相关的平行语料库,这并不总是容易获得的。同时,在双语词向量学习过程中虽然获取了跨语言的语义相似性,但忽略了文本的情感信息。即,由于缺乏明确的情感信息,从未标注的平行语料库学习到的双语词向量不是足够有效的。双语词向量学习算法专注捕获跨语言语法和语义的相似性,忽略了情感信息。到目前为止,许多针对情感分类问题的词向量学习算法被提出,这些算法将情感信息融入词向量中。文献^[26]提出了一种概率模型,其结合无监督和监督的技术来学习词向量,捕捉语义信息以及情感信息。文献^[27]将情感标签引入到神经网络语言模型中,提高了词向量的情感表达能力。文献^[28]首次提出了双语情感词向量模型,其对原始训练数据和相应的翻译进行线性映射,而不是采用深度学习技术。文献^[22]重点聚焦用深度学习技术学习双语情感词向量 B-SWE。针对跨语言情感分析,文献^[29]提出了一种以降噪自动编码器为基础的方法来学习双语词向量,将文本的情感极性融入到双语词向量中。文献^[29]不同于采用平行语料来学习双语词向量,该技术仅仅使用训练数据和对应的翻译来学习双语情感词向量。更重要的是,在跨语言情感分析里,为了增强性能,将情感信息集成到了双语词

向量中。提出了一种同时为源语言和目标语言学习文本向量表示学习的方法,建设了一个直接学习文档来表示的双语文档表示学习模型 BiDRL,语义和情感的相关性被用来映射双语文本到相同的向量空间中,该方法在一个多语言多领域的亚马逊评论数据集上的跨语言情感分析实验性能良好。文献[30]还提出了一个基于注意力机制的双语表示学习模型,用于在源语言和目标语言中学习文档的分布式语义。在每种语言中,使用了长短期记忆神经网络来对文档进行建模,同时,还为双向长短期记忆神经网络提出了一种层次化注意力机制。句子级别的注意力模型学习一个文档中哪个句子在确定整体情感方面更为重要,而这个词级别的注意力模型学习在每一个句子中哪个词具有决定性的作用。这个模型在一个基准数据集上的英-中跨语言情感分析任务上取得了较好的结果。

3 社交媒体文本的情感分析

推特包含在不同话题方面的各种评论文本,对于情感分析,这些表达不同观点的文本是一个丰富的资源。由于社交媒体和在推特中使用的是非正式语言,对社交媒体和推特进行情感分析是一项具有挑战性的任务。社交媒体上的语言含有俚语,拼写不规范,且词汇随着时间进化,因此为标准语言建设的资源不能很好地应用于社交媒体文本的分析中。随着社会媒体网站的激增,挖掘社会媒体文本的需求应运而生。因此,通过从社交媒体文本中抽取情感词典为社交媒体进行情感分析的很多解决方案被提出^[31]。

近年来,推特的情感极性受到了极大的关注。大部分研究工作将推特的情感分类看成是文本分类的一个特例。特别地,文献[32]在 SemEval 2013 推特情感分类任务上建立了最好的系统,使用了不同的情感词典和各种人工特征。文献[31]提出情感特定的词向量,它能提供比目前使用不同词表示的最好系统更好的性能。然而,目前针对推特情感分类的词向量模型有两个缺点:1)每个词产生一个词向量,不管这个情感词是否是多义;2)大部分方法在词向量空间里忽视了推特的主题信息,而这个信息决定了多义词在具体上下文中的特定含义。

目前,推特情感分析研究采用的方法聚焦在特征工程上。一个情感分类器的性能很大程度上依赖于特征表示的选择。为了获得更好的性能,很多不同的特征学习方法被相继提出^[33-35]。

最近的研究为情感分类探索学习低维、稠密和实数值的词向量。文献[26]提出了一个概率文档模型。文献[31]设计了3种神经网络模型,以从含有积极和消极表情符号的推特中学习词向量。这些方法与线性或神经网络分类器结合时,显示出了可喜的结果。对于推特情感分类,目前存在的大部分模型针对每个词用一个单独的分布式表示,而在情感分类问题中,因为词经常是多义,每个词在不同话题下可以含有不同的情感极性。

4 资源稀缺语言的社交媒体文本情感分析

目前,大部分情感分析研究的主要语种对象是英语,对资源稀缺语言的情感分析研究得较少,特别是在社会媒体文本上的研究更少。本文将国外阿拉伯语和国内藏语作为代表,分析了资源稀缺语言的情感分析研究现状。

4.1 阿拉伯语推特情感分析

文献[26]建设了一个阿拉伯语的社会媒体情感数据集,它包括约10000条推特数据,这些数据被分类为客观、主观积极、主观消极、主观混合;通过对数据集进行标准划分,进行了四分类的情感分析实验。文献[37]试图在阿拉伯语的推特情感分析方面弥合差距,而阿拉伯语推特缺乏非正式语言领域的情感词典。他们从大型的推特数据集中抽取产生了两个情感词典,并且仅用简单的基于词典的评价方法就在情感分类任务中取得了较好的性能。文献[38]描述了一个不依靠任何人工资源、自动学习情感词典的有效神经网络方法。该方法利用了推特中的表情符号,使用了词与表情符号之间的 PMI,定义了词的情感属性,通过使用这些来预测推特情感标签,从而学习到了更好的情感词汇。该方法在英语和阿拉伯语上取得了较好的性能。

4.2 藏文微博文本的情感分析

文献[39]对藏文文本进行情感分类,他们从藏文网站、博客、论坛中挑选感情色彩明显的文章,通过藏语三级切分体系对藏文文本进行分词和词性标注,并借助手工建立的藏文情感分析用词表,与已有的特征选择方法相结合,提取了情感特征,用相似度分类算法进行藏文文本的情感分类。文献[40]通过借鉴中文微博情感分析中较为常用的基于统计的方法和基于词典的方法来对藏文微博进行情感分析,并通过实验的方式发现基于藏文词典的藏文微博情感分析的准确率明显高于基于 TF-IDF 的藏文微博情感分析的准确率。文献[41]结合藏文句法结构和语义特征向量构建语义特征空间,提出了一种基于语义空间的藏文微博情感分类方法。该方法使用句法树构造情感语义空间,并在多维特征空间中进行聚类,从而得到语义簇,并通过计算语义簇的 TF-IDF 值实现藏文微博的情感分类。

结束语 本文对情感分析的几种研究方法和目前的研究热点领域进行了研究现状的综述。综合以上分析可知,目前大部分的研究方法已从原来的传统机器学习的方法转移到利用神经网络的方法,也开始重视结合语言学的知识,同时,在构建社交媒体文本情感词典上,目前已从人工构建方式转移到自动构建。然而,目前对社交媒体上资源稀缺语言的情感分析工作才刚起步。

由于社交媒体短文本与普通文本有很多不同,如未登录词多、文法不规范、数据稀疏和多语言文本混合等问题,目前很多研究中未提及对社交媒体短文本进行规范化等预处理,而短文本情感分析之前的预处理方法对情感分析有着很大的影响,例如,对社交媒体文本的新词识别。对数据集预处理方法的不同,会导致各类研究的性能结果无法直接比较。

虽然利用神经网络进行情感分析有着较明显的性能优势,但对于资源稀缺的语言,利用深度学习进行情感分析,需要大量数据和可利用的资源,然而目前资源稀缺语言的网络资源十分有限,特别是有标注的资源。为此,利用跨语言的方法对资源稀缺语言进行情感迁移学习已成为一种可行的解决方法。同时,若能进一步结合资源稀缺语种的语言学知识,如情感词典、否定词、词性等,将能提升情感分析的性能,更有利于细粒度的情感分析研究。对于资源稀缺语言而言,目前开展研究的最大障碍是缺乏权威的数据集和情感资源。为此,为资源稀缺语言构建一个权威的数据集和情感资源,将能大力推动情感分析的研究。

参 考 文 献

- [1] TURNEY P D. Thumps Up or Thumps Down? Semantic Orientation Applied to Unsupervised Classification of Reviews[C]// Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, 2002: 417-424.
- [2] PANG B, LEE L, VAITHYANATHAN S. Thumps Up? Sentiment Classification using Machine Learning Techniques[C]// Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing. Philadelphia, 2002: 79-86.
- [3] KIM Y. Convolutional Neural Networks For Sentence Classification[C]// 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). 2014: 1746-1751.
- [4] SOCHER R, PENNINGTON J, HUANG E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions[C]// 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, UK, 2011: 151-161.
- [5] MIKOLOV T, ZWEIG G. Context dependent recurrent neural network language model[C]// SLT Workshop. 2012.
- [6] TAI K, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks[J]. Computer Science, 2015, 5(1): 36.
- [7] POLANYI L, ZAENEN A. Contextual Valence Shifters[J]. Information Retrieval, 2004, 20: 1-10.
- [8] ZHU X, GUO H, MOHAMMAD S, et al. An Empirical Study on the Effect of Negation Words on Sentiment[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA, 2014: 304-313.
- [9] KIRITCHENKO, MOHAMMAD. Sentiment Composition of Words with Opposing Polarities[C]// Proceedings of NAACL-HLT 2016. San Diego, California, 2016: 1102-1108.
- [10] TABOADA M, BROOKE, TOFILOSKI, et al. Lexicon-Based Methods for Sentiment Analysis[J]. Computational Linguistics, 2011, 37(2): 267-307.
- [11] WEI W, WU C H, LIN J C. A regression approach to affective rating of chinese words from anew[M]// Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011: 121-131.
- [12] MALANDRAKIS N, POTAMIANOS A, LOSIF E, et al. Distributional semantic models for affective text analysis[J]. IEEE transactions on audio, speech, and language processing, 2013, 21(11): 2379-2392.
- [13] QIAN Q, HUANG M L, ZHU X Y. Linguistically Regularized LSTMs for Sentiment Classification[J]. arXiv: 1611. 03949v1
- [14] WAN X J. Using Bilingual Knowledge and Ensemble Technics for Unsupervised Chinese Sentiment Analysis[C]// 2008 Conference on Empirical Methods in Natual Language Processing. Honolulu, China, 2008: 553-561.
- [15] WAN X J. Co-Training for Cross-Lingual Sentiment Classification[C]// Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP. Suntec, Singapore, 2009: 235-243.
- [16] MENG X F, WEI F R, LIU X H, et al. Cross-Lingual Mixture Model for Sentiment Classification[C]// Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju, Republic of Korea, 2012: 572-581.
- [17] POPAT K, BALAMURALI A R, BHATTACHARYYA P, et al. Thehaves and the have-nots: Leveraging unlabelled corpora for sentiment analysis[C]// Proceedings of the Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria, 2013: 412-422.
- [18] XU R F, XU J, WANG X L. Instance Level Transfer Learning for Cross Lingual Opinion Analysis[C]// Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011). Portland, Oregon, USA, 2011: 182-188.
- [19] GUI L, XU R F, LU Q, et al. Cross-lingual Opinion Analysis via Negative Transfer Detection[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). 2014: 860-865.
- [20] CHEN Q, LI W, LEI Y, et al. Learning to Adapt Credible Knowledge in Cross-lingual Sentiment Analysis[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China, 2015: 419-429.
- [21] XIAO M, GUO Y H. Semi-supervised representation learning for cross-lingual text classification[C]// Proceedings of EMNLP 2013. 2013: 1465-1475.
- [22] ZHOU H, CHEN L, SHI F, et al. Learning bilingual sentiment word embeddings for cross-language sentiment classification[C]// Proceedings of 52rd Annual Meeting of the Association for Computational Linguistic. 2015: 430-440.
- [23] CHANDAR A P S, KHAPRA M M, RAVINDRAN B, et al. Multilingual deep learning[C]// Deep Learning Workshop at NIPS 2013. 2013.
- [24] SARATH CHANDAR A P, LAULY S, LAROCHELLE H, et al. An autoencoder approach to learning bilingual word representations[C]// Advances in Neural Information Processing Systems. 2014: 1853-1861.
- [25] ZHOU G Y, HE T T, ZHAO J. Bridging the Language Gap: Learning Distributed Semantics for Cross-Lingual Sentiment Classification[C]// Proceedings of Natural Language Processing and Chinese Computing. Springer Verlag, 2014: 138-149.
- [26] MAAS A L, DALY R E, PHAM P T, et al. Learning Word Vectors for Sentiment Analysis[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. 2011: 142-150.
- [27] WANG Y, LI ZH, LIU J, et al. Word Vector Modeling for Sentiment Analysis of Product Reviews[C]// Proceedings of Natural Language Processing and Chinese Computing. Springer Verlag, 2014: 168-180.
- [28] TANG X W, WAN X J. Learning Bilingual Embedding Model for Cross-language Sentiment Classification[C]// Proceedings of 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). IEEE, 2014: 134-141.
- [29] ZHOU X J, WAN X J, XIAO J G. Cross-Lingual Sentiment Classification with Bilingual Document Representation Learning[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 1403-1412.

- Morgan Kauffman Publishers, 1993: 416-423.
- [5] SRINIVAS N, DEB K. Multi-objective optimization using non-dominated sorting in genetic algorithms[J]. *Evolutionary Computation*, 1994, 2(3): 221-248.
- [6] HORN J, NAFPLIOTIS N, GOLDBERG D E. A niched Pareto genetic algorithm for multi-objective optimization [C] // 1st IEEE Congress on Evolutionary Computation. Piscataway: IEEE, 1994: 82-87.
- [7] ZITZLER E, LAUMANN S, THIELE L. SPEA2: Improving the strength Pareto evolutionary algorithm[C] // *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*. Berlin: Springer-Verlag, 2002: 95-100.
- [8] KNOWLES J D, CORNE D W. Approximating the non-dominated front using the Pareto archived evolution strategy[J]. *Evolutionary Computation*, 2000, 8(2): 149-172.
- [9] CORNE D W, JERAM N R, KNOWLES J D, et al. PESA-II: Region-Based selection in evolutionary multi-objective optimization[C] // *Genetic and Evolutionary Computation Conf. (GECCO 2001)*. San Francisco: Morgan Kaufmann Publishers, 2001: 283-290.
- [10] ERICKSON M, MAYER A, HORN J. The niched Pareto genetic algorithm 2 applied to the design of groundwater remediation system[C] // 1st Int'l Conf. on Evolutionary Multi-Criterion Optimization (EMO 2001). Berlin: Springer-Verlag, 2001: 681-695.
- [11] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2): 182-197.
- [12] MARTINEZ S Z, COELLO C A C. A multi-objective evolutionary algorithm based on decomposition for constrained multi-objective optimization [C] // 2014 IEEE Congress on Evolutionary Computation. Beijing, China: IEEE, 2014: 429-436.
- [13] ZHU Q, LIN Q, DU Z, et al. A novel adaptive hybrid crossover operator for multiobjective evolutionary algorithm[J]. *Information Sciences*, 2016, 345(C): 177-198.
- [14] QI Y, HOU Z, YIN M, et al. An immune multi-objective optimization algorithm with differential evolution inspired recombination [J]. *Applied Soft Computing*, 2015, 29(C): 395-410.
- [15] FRESCHI F, REPETTO M. VIS: An artificial immune network for multi-objective optimization[J]. *Engineering Optimization*, 2006, 38(8): 975-996.
- [16] JIAO L C, GONG M G, SHANG R H, et al. Clonal selection with Immune dominance and energy based multi-objective optimization[C] // 3rd Int'l Conf. on Evolutionary Multi-Criterion Optimization (EMO 2005). Berlin: Springer-Verlag, 2005: 474-489.
- [17] GONG M G, JIAO L C, DU H F, et al. Multi-objective immune algorithm with non-dominated neighbor-based selection[J]. *Evolutionary Computation*, 2008, 16(2): 225-255.
- [18] 舒万能, 丁立新, 汪慎文. 基于反馈机制的克隆反馈优化算法的稳定性研究[J]. *计算机科学*, 2012, 39(10): 187-189.
- [19] 戚玉涛, 刘芳, 常伟远, 等. 求解多目标问题的 Memetic 免疫优化算法[J]. *软件学报*, 2013, 19(7): 1529-1544.
- [20] 公茂果, 焦李成, 杨咚咚, 等. 进化多目标优化算法研究[J]. *软件学报*, 2009, 20(2): 271-289.
- [21] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2): 182-197.
- [22] WANG H, WANG W, SUN H, et al. A new cuckoo search algorithm with hybrid strategies for flow shop scheduling problems [J]. *Soft Computing*, 2017, 21(15): 4297-4307.
- [23] CASEIRO R, HENRIQUES J F, MARTINS P, et al. A non-parametric Riemannian framework on tensor field with application to foreground segmentation[C] // *IEEE International Conference on Computer Vision*. IEEE, 2011: 1-8.
- [24] ZHANG H, ZHOU A, SONG S, et al. A Self-Organizing Multi-objective Evolutionary Algorithm [J]. *IEEE Transactions on Evolutionary Computation*, 2016, 20(5): 792-896.
- [25] GIAGKIOZIS I, PURSHOUSE R C, FLEMING P J. An overview of population-based algorithms for multi-objective optimization [J]. *International Journal of Systems Science*, 2015, 46(9): 1572-1599.

(上接第 49 页)

- [30] ZHOU X J, WAN X J, XIAO J G. Attention-based LSTM Network for Cross-Lingual Sentiment Classification[C] // *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas, 2016: 247-256.
- [31] TANG D Y, WEI F R, YANG N, et al. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification [C] // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 2014: 1555-1565.
- [32] MOHAMMAD S M, KIRITCHENKO S, ZHU X D. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets [C] // *Proceedings of SemEval-2013*. 2013.
- [33] RETRIEVAL F. Opinion mining and sentiment analysis[J]. *Foundations and Trends in Information Retrieval*, 2008, 2(1/2): 1-135.
- [34] OWOPUTI O, DYER C, GIMPEL K, et al. Part-of-speech tagging for twitter: Word clusters and other advances[R]. CMU, 2012.
- [35] MOHAMMAD S, KIRITCHENKO S, ZHU X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets [C] // *Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013)*. Atlanta, Georgia, USA, 2013.
- [36] NABIL M, ALY M, ATIYA A F. Astd: Arabic sentiment tweets dataset[C] // *Proceedings of EMNLP*. 2015: 2515-2519.
- [37] AL-TWAIRESH N, AL-KHALIFA H, AL-SALMAN A. AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons [C] // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. 2016: 697-705.
- [38] VO D T, ZHANG Y. Don't Count, Predict! An Automatic Approach to Learning Sentiment Lexicons for Short Text [C] // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. 2016: 219-224.
- [39] 李海刚, 于洪志. 藏文文本情感分类系统设计[J]. *甘肃科技纵横*, 2011, 40(1): 106-107.
- [40] 张俊, 李应兴. 基于情感词典的藏文微博情感分析研究[J]. *硅谷*, 2014(20): 220-222.
- [41] 袁斌, 江涛, 于洪志. 基于语义空间的藏文微博情感分析方法[J]. *计算机应用研究*, 2016, 33(3): 682-685.