# 基于语料资源的情感词典扩展方法研究

谢松县，刘博，王挺

（国防科学技术大学 电子科学与工程学院，湖南 长沙 410073 ）

摘要：为了对雷达天线扫描方式进行自动识别，改进开发了天线扫描方式模拟器，并分别研究了电子扫描和机械扫描的特征提取和识别方法。基于最大主瓣脉冲序列的特征参数实现电子扫描和机械扫描的区分，基于单个天线扫描周期脉冲序列的特征参数实现8种机械扫描方式的自动识别。仿真结果表明，该方法能够区分一维电扫、二维电扫和机械扫描，并且采用支持向量机决策树对机械扫描方式的识别正确率高于决策树方法。（小五）

关键词**：**天线扫描方式；自动识别；电子扫描；机械扫描；支持向量机决策树（小五）

中图分类号**：**TN95 文献标志码**：**A　　　文章编号： （小五）

**Automatic recognition method of radar antenna scan type（三号，加粗）**

LI Cheng1, WANG Xuesong2（小四）

(1.College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China；2.College of Science，National University of Defense Technology, Changsha 410073, China) （小五）

**Abstract:** In order to recognize radar antenna scan type automatically, an improved antenna scan pattern simulator is developed in this paper, and features extraction and automatic recognition methods for electronic scan type (EST) and mechanical scan type (MST) are studied respectively. EST is firstly distinguished from MST based on the characteristic parameters extracted from the maximum main beam pulse sequence; then 8 MSTs are automatically recognized based on the parameters extracted from the pulse sequence in a scan period. The simulation results show that it is able to distinguish between one-dimensional EST, two-dimensional EST and MSTs by this method. Moreover, the correct recognition ratio of MSTs by support vector machine decision tree in this paper is higher than that by decision tree. （小五）

**Keywords:** antenna scan type; automatic recognition; electronic scan type; mechanical scan type; support vector machine decision tree（小五）

人们在信息交流中总是很自然地对周围事物表达情感，意见和喜好。当人们接触到关于某个主题或实体的信息，通常都会形成对这些话题（实体）的个人倾向和看法。这些带有主观性内容的信息，无论是书面或者口头的，理解其表达的正确的观点（情感）是极其重要的。因此，意见和情感的理解是人际交往的一个重要方面。多年以来，为了理解人类行为，个体或者群体情感一直得以研究。对情感或观点的群体或者社会化分析迎合了对于某事件或实体对一群人所造成的影响或极化度量需求。社会情绪进行了研究，政治理解和预测选举结果，以及在市场营销，来预测某一个产品的成功，并建议他人。比如在政治领域，可以研究社会化情感来预测选举的结果；在商品市场，通过研究对某些商品社会化观点的可以预测畅销与否，并且向客户推荐。

在社交媒体出现以前，收集带有观点信息的数据费力且达到的规模比较小。社交网络出现后，大规模的用户开始频繁使用社交媒体进行信息交流，收集用户的观点数据就变得简单了。从这方面来说，社交媒体开启了用户之间交互新的可能。尤其是微博平台，可以允许用户实时分享评论和观点。比如Twitter，作为极度受欢迎的微博平台，吸引了上百万用户每天分享数以百万计的帖子。这种数量巨大的丰富的用户产生数据（User Generated Content，UGC）为研究用户行为提供了无尽的机会。

能够从文本中自动识别出观点和情感等主观性信息近年来吸引了学术界和工业界的广泛关注。人工对数量庞大的文本进行分析发现主观信息是不可行的，因此出现了所谓观点挖掘（或成为情感分析，opinion mining, sentiment analysis）研究，可以从自然语言文本中自动识别出用户的观点信息。采用计算方法进行观点情感分析会在不同维度上对情感进行度量。其中使用极性（porlarity）分类方法[1-4]将情感分析问题简化为分类问题，使用有监督或无监督方法将用户的的观点进行积极（positive）、消极（negative）以及中性（neutral）分类。极性分类相关的一类问题时主客观分类问题，就是将主观文本和客观文本分开，一般使用的是有监督的方法[5]。此外，观点强度分析也是研究的方向，比如，SentiStrength[6]，能够从句子层面计算出积极和消极的观点强度值。情绪也可以通过创建情绪词典来进行计算，Plutchik情绪轮盘[7]，提出了四对对立的情绪状态：joy-trust，sadness-anger，surprise-fear和anticipation-disgust。Mohammad等[8]根据Plutchik情绪类别标注了一系列的词发展出了NRC词情绪关联词典。以上的情感分析方法都是在句法层面，还有其他的基于语义知识库在语义层面进行情感分析的方法[9-11]。

情感分析通常会使用一些标注了积极和消极分数的词表的词典资源。近年来情感词典资源的构建方法研究已经得到越来越多的计算语言学和自然语言处理研究人员注意。Wilson等[12, 13]对一些英文单词进行了积极和消极类别的标注形成了OpinionFinder词典；Bradley等[1]标注了并发布了情感规范（affective norms）的英文单词词典ANEW，并且Nielsen等[3]在Twitter上应用并扩展了ANEW，形成AFINN词典。Esuli和Sebastiani[14]以及后来Baccianella[2]等在著名的英文词典Wordnet基础上采用自动计算的方式开发出了情感词典SentiWordnet。Thelwall等[6]设计实现了能对词语的情感强度进行估计的词典。Mohammad和Turney[8]根据Plutchik情绪轮分类方法使用情绪分值标注了一些词语形成NRC情绪词典。在2013和2014年举办的SemEval（Semantic Evaluation）评测中，NRC-Canada队利用NRC词典并扩展出两种新的词典，取得了最好成绩[15, 16]。为了克服以上语法层面建立的词典的上下文语境以及领域适应性问题，一些学者提出了基于概念（concept-based）构建情感词典[17]，其中SenticNet是使用常识知识库建立的公开可用的基于概念的情感词典[9]。中文情感分析研究起步较晚，缺乏普遍认可的可靠的中文情感词典。目前研究使用主要有HowNet情感词典[18]，NTUSD情感词典[19]以及大连理工大学的情感词汇本体词库[20]。这些词典主要是以手工或半自动方式编辑而成。我们前期的工作提出了根据HowNet语义关系将英文词典转换为中文情感词典的方法，并构建了比较全面的情感词典[21]。

基于语义词典的情感词典构建方法是一种通用的情感词典构建方法。采用这种方法的优势在于可以获取大量的情感词语，基于词语的语义关系也易于进行情感倾向计算。但是，基于语义词典的情感词典构建方法受限于语义词典的规模和语义关系的定义，而且对于专业领域中不断涌现的新词语，采用这种方法难以对情感词典进行扩展。随着互联网应用的不断发展，网络上的信息资源不断涌现，研究如何利用语料等信息资源对情感词典进行自动扩展具有十分重要的意义。

本文提出的基于语料资源的无监督的情感词典扩展方法是基于中科院谭松波博士提供的语料资源设计和实现的。本章将详细讨论基于语料资源的情感词典扩展的问题描述、方法设计、关键技术、具体实现和结果分析。

1 问题描述

**参考文献(References)**

[1] Bradley M M, Lang P J. Affective norms for English words (ANEW): Instruction manual and affective ratings[R]. Citeseer, 1999.

[2] Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.[C] 2010: 2200-2204.

[3] Nielsen F A R. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs[J]. arXiv preprint arXiv:1103.2903. 2011

[4] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision[J]. CS224N Project Report, Stanford. 2009: 1-12.

[5] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C] Stroudsburg, PA, USA: Association for Computational Linguistics, 2005: 347-354.

[6] Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web[J]. Journal of the American Society for Information Science and Technology. 2012, 63(1): 163-173.

[7] Plutchik R. The nature of emotions[J]. American Scientist. 2001, 89(4): 344-350.

[8] Mohammad S M, Turney P D. Crowdsourcing a word--emotion association lexicon[J]. Computational Intelligence. 2013, 29(3): 436-465.

[9] Cambria E, Olsher D, Rajagopal D. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis[C] //Twenty-Eighth AAAI Conference on Artificial Intelligence

. 2014.

[10] Grassi M, Cambria E, Hussain A, et al. Sentic web: A new paradigm for managing social media affective information[J]. Cognitive Computation. 2011, 3(3): 480-489.

[11] Olsher D J. Full spectrum opinion mining: Integrating domain, syntactic and lexical knowledge[C] 2012: 693-700.

[12] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C] 2005: 347-354.

[13] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis[J]. Computational linguistics. 2009, 35(3): 399-433.

[14] Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining[C] 2006: 417-422.

[15] Mohammad S M, Kiritchenko S, Zhu X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets[C] Atlanta, Georgia, USA: 2013.

[16] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews[C] 2014.

[17] Tsai A C, Wu C, Tsai R T, et al. Building a concept-level sentiment dictionary based on commonsense knowledge[J]. IEEE Intelligent Systems. 2013, 28(2): 22-30.

[18] 知网HowNet评价词词典[EB/OL]. 2013/2013-07-25[2013-08-15]. http://www.keenage.com/html/c\_index.html.

[19] Ku L W, Chen H H. Mining opinions from the Web: Beyond relevance retrieval[J]. Journal of the American Society for Information Science and Technology. 2007, 58(12): 1838-1850. Wiley Online Library.

[20] 情感词汇本体库[EB/OL]. 2013/2013-07-30[2013-08-15]. http://ir.dlut.edu.cn/EmotionOntologyDownload.aspx.

[21] 谢松县，刘博，王挺. 应用语义关系自动构建情感词典[J]. 国防科技大学学报. 2014, 36(3): 111-115.