第7章　基于三支决策的多粒度文本情感分类

Multi-Granularity Sentiment Classification Based on Three-Way Decisions

张志飞[1,2]　王睿智[1,2]　苗夺谦[1,2]

1. 同济大学计算机科学与技术系

2. 同济大学嵌入式系统与服务计算教育部重点实验室

文本情感分类主要研究如何从文本中挖掘用户关于实体或其属性的观点、态度、情绪等主观信息。因文本的语言粒度有所不同，其表达的情感粒度也存在差别。本章将文本情感分为词语级、句子级和篇章级三个不同的粒度级别，对多粒度文本情感分类从情感的不确定性角度给予全新的解释，并基于三支决策提出若干文本情感分类方法，以解决情感分类中的上下文有关、主题依赖和情绪分类等问题。

7.1　引　　言

互联网的飞速发展催生了大量的用户生成内容，能够体现用户的观点、态度、立场、情绪等。这些内容的表现形式大都是非结构化或半结构化的文本，如产品评论、股票评论、影视评论、新闻评论、微博等。文本不再局限于客观事实的描述，而是侧重于观点的表达。社交媒体的兴起使得电子形式的主观性文本易于获取，文本情感分类技术受到广泛关注。文本情感分类是指通过挖掘文本中的立场、观点、看法、情绪、好恶等主观信息，识别文本的情感色彩[1,2]。

从语言粒度看，文本情感分类主要分为词语级、句子级和篇章级[3,4]。词语情感分类是文本情感分类的基础，难点在于上下文有关情感分类问题[5,6]，即同一个情感词语在不同的上下文表达不同的极性。句子情感分类认为情感词语和主题词语两者紧密联系，情感词语能够表示主题[7]，而主题词语能够辅助情感分类[8]，因此，将主题信息转化为情感先验信息以指导主题依赖的句子情感分类。篇章情感分类相对简单，假设篇章只评论一个对象。情绪和观点两个概念接近但不等同，如含有情绪但不一定有极性，而且一个篇章往往含有多个情绪[9]，因此，将篇章情绪分类看成复杂的篇章情感分类，即多标记学习问题[10]。

将上述问题归结为情感的不确定性。例如，词语的情感不确定性来自上下文，句子的情感不确定性来自表达的主题，篇章的情感不确定性来自情绪重叠。因此，本章将文本情感分类看成情感的不确定性分析，从词语、句子和篇章三个语言粒度进行情感分类研究[11]。

粗糙集作为一种不确定性分析的工具，主要优势之一是不需要任何预备或额外的

在其他指标上性能较差。篇章情绪不等价于词语或者句子情绪的累加，因为篇章体现的是更大粒度或者总体的情绪。例如，111 个不含有情绪标记的篇章中存在具有情绪标记的句子。

7.6 本章小结

本章从情感不确定性角度对词语、句子和篇章三个粒度的文本情感分类给予全新的解释。以三支决策作为不确定性分析的工具，解决情感分类中的上下文有关、主题依赖和多标记情绪等问题。

词语情感分类方法结合上下文有关的反义词对及其所在正域提出双向规则和单向规则。8 对上下文有关的反义词对上的实验表明，基于三支决策的词语情感分类方法能够显著提高情感分类性能。但是，需要研究如何将三支决策分类方法推广到更大范围的上下文有关词语。

句子情感分类方法将主题转化的情感先验和三支决策分类器相结合，验证决策阈值和情感先验之间的关系。微博情感语料上的实验表明，理想的主题分类将显著提高情感分类的性能，同时给出使用该方法的建议性原则。因此，需要将主题分类细化，达到情感分类错误代价最小的目的。

篇章情绪分类方法利用 DW-ML-KNN 算法的多标记实值函数定义三支决策区域，提出标记共现定理和标记互斥定理。结论是 DW-ML-KNN 属于 0.5-概率粗糙集模型，而严格意义上属于(0.75,0.25)-决策粗糙集模型，性能总体较优。但是，需要结合风险代价以寻找更合适的决策阈值。

致　谢

感谢评阅专家提出的宝贵意见。本章工作获得了国家自然科学基金项目（项目编号分别为 61273304、61202170）和高等学校博士学科点专项科研基金项目（项目编号：20130072130004）的资助。

参 考 文 献

[1] Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.

[2] 赵妍妍, 秦兵, 刘挺. 文本情感分析. 软件学报, 2010, 21(8): 1834-1848.

[3] Feldman R. Techniques and applications for sentiment analysis. Communications of the ACM, 2013, 56(4): 82-89.

[4] 黄萱菁, 张奇, 吴苑斌. 文本情感倾向分析. 中文信息学报, 2011, 25(6): 118-126.