**情感分类国内外现状**

情感分类将句子（或文档）的情感极性分为正或负，是情感分析领域的主要研究方向（Pang和Lee，2008; Liu，2012; Feldman，2013）。现有的情感分类方法主要由**两个主流方向所主导**。 **基于词汇的方法**(Turney, 2002; Ding et al., 2008; Taboada et al., 2011; Thelwall et al., 2012) 通常利用情感词的词典，其中每一个词都用情感极性或情感强度注释。**语言规则**，如集约化和否定，通常被纳入以聚合句子（或文档）的情感极性。基于语料库的方法将情绪分类作为文本分类任务的一种特殊情况（Pangetal.，2002）。 他们主要从句子（或文档）中用手动注释的情感极性或通过情感信号（例如表情符号）收集的远程监督语料库来构建情感分类器，(Go et al., 2009; Pak and Paroubek, 2010; Kouloumpis et al., 2011; Zhao et al., 2012).

大多数现有的方法遵循Pang et al. (2002)，并采用基于语料库的方法进行情绪分类。 Pang et al.（2002）作为文本分类问题的特殊情况来处理评论的情感分类的开创者，并且首先研究**机器学习方法**。他们使用**朴素贝叶斯，最大熵和支持向量机（SVM）**与一组不同的功能。在他们的实验中，通过**具有词袋特征的SVM实现了最佳性能**。在这个角度，许多研究集中于设计或学习有效的特征，以获得更好的分类性能。在电影或产品评论中，Wang和Manning（2012）提出了**NBSVM**，它在朴素贝叶斯和NB特征增强的SVM之间进行交易。 Kim和Zhai（2009）和Paltoglou和Thelwall（2010）通过调查信息检索的变量加权函数来学习特征权重。 Nakagawa et al。 （2010）使用依赖树，极性转移规则和条件随机场（Lafferty等人，2001）与隐藏变量来计算文档特征。在Twitter上，Mohammad et al.（2013b）在SemEval 2013中开发了一个最先进的Twitter情绪分类，使用各种情绪词典和手工制作的功能。

随着深度学习（表示学习）的复兴（Hinton和Salakhutdinov，2006; Bengio等人，2013; Jones，2014），最近的研究集中于**学习低维，密集和实值向量作为情感分类的文本特征**。 Glorotetal. 2011）调查堆栈**去噪自动编码器**来学习文档向量，用于情感分类中的域适配。 Yessenalina和Cardie（2011）将每个词表示为矩阵，并使用迭代矩阵乘法组成词。 Socher et al. 提出**递归自动编码器**（RAE）（2011），**矩阵向量递归神经网络**（MV-RNN）（2012）和**递归神经网络**（RNTN）（2013b），以学习基于其可变长度短语及其子群的组成。为了学习句子表示，Kalchbrenner et al。 （2014）利用动态卷积神经网络和Le和Mikolov（2014）调查段矢量。为了学习情感分析的单词向量，Maas et al。 （2011）提出了Blei等人的**概率文档模型**。（2003），Labutov和Lipson（2013）从现有词嵌入中重新嵌入词和Tang等（2014b）开发三个神经网络来学习从包含正/负表情符号的tweets单词向量。

**情感分类算法比较**

•DistSuper：收集了10M正负表情的微博作为训练数据，并使用LibLinear和ngram特征构建分类器（Go等人，2009; Zhao等人，2012）。

•SVM：n-gram特征和支持向量机是广泛使用的基础方法来构建情感分类器（Pang et al。，2002）。我们使用LibLinear来训练SVM分类器。

•NBSVM：NBSVM（Wang和Manning，2012）在朴素贝叶斯和增强SVM的 NB之间交易，。我们使用NBSVM-bi，因为它在评论的情感分类上表现最好。

•RAE：递归自动编码器（Socher et al，2011）已被证明通过学习句子表示对情感分类是有效的。我们使用从100M tweets中学习的预训练短语嵌入来训练RAE。

•SentiStrength：Thelwall et al（2012）建立一个基于词典的分类器，它使用语言规则来检测tweet的情感强度。

•SSWEu：Tang et al（2014b）提出从收集的10M tweets表情符号来学习情感特定词嵌入（SSWE）。他们应用SSWE作为Twitter情绪分类的特征。

•NRC：NRC在SemEval 2013 Twitter情感分类轨道中建立了最先进的系统，融合了各种情感词典和手工制作特征（Mohammadetal。，2013b）。

除了DistSuper，其他基线方法以受监督的方式进行。