基于情绪词与情感词协作学习的情感分类方法研究

代大明 李寿山 李培峰 朱巧明

(江苏省计算机信息处理技术重点实验室 苏州 215006) (苏州大学计算机科学与技术学院 苏州 215006)

摘 要 情感分类任务旨在自动识别文本所表达的情感色彩信息(例如,褒或者贬、支持或者反对)。提出一种基于情绪词与情感词协作学习的情感分类方法:在基于传统情感词资源的基础上,引入少量情绪词辅助学习,只利用大规模未标注数据实现情感分类。具体来讲,基于文档-单词二部图的标签传播算法框架,利用情绪词与情感词构建两个视图,通过协作学习的方法从大规模未标注语料中抽取高正确率的自动标注样本作为训练数据,然后训练分类器进行情感分类。实验表明,该方法在多个领域的情感分类任务中都取得了较好的分类效果。

关键词 情绪词,情感词,二部图,标签传播算法,协作学习

中图法分类号 TP391 **文献标识码** A

Research on Sentiment Classification of Collaborative Learning Based on Emotion Words and Sentiment Words

DAI Da-ming LI Shou-shan LI Pei-feng ZHU Qiao-ming

(Provincial Key Lab of Computer Information Processing Technology of Jiangsu, Suzhou 215006, China)

(School of Computer Science & Technology, Soochow University, Suzhou 215006, China)

Abstract Sentiment classification aims to distinguish the expressed sentiment categories by the text, such as positive vs. negative and agree vs. disagree. We used a opinion lexicon, together with a small scale of emotion key words to conduct sentiment classification with unlabeled data. Specifically, a document-word bipartite graph was builts, and then the opinion words and emotion words were served as labeled points while the documents were regarded as unlabeled points in the graph. Label propagation algorithm was used to propagate the label information of the words to the documents. Finally, the high confident automatically-labeled samples were used as training data for sentiment classification through collaborative learning method. Experimental results demonstrate that our approach achieves a good performance for sentiment classification across multiple domains.

Keywords Emotion words, Sentiment words, Bipartite graph, Label propagation algorithm, Collaborative learning

1 引言

随着 Web 2.0 的发展,互联网上相继出现了大量关于观点的评论文本,迫切需要计算机帮助商业公司或用户自动分析和获取这些文本的情感信息。情感分析(Sentiment Analysis)即为在该背景下出现的一个面向文本情感信息处理的新兴研究方向[1]。

情感分类(Sentiment Classification)是情感分析研究中的一个基本任务,该任务旨在将文本按照情感倾向进行褒贬分类。与传统基于主题的文本分类相比,情感分类被认为更具有挑战性[3]。到目前为止,大多数针对情感分类的研究是基于监督学习的,这种方法虽然取得了较好的分类效果,但需要大量人工标注语料,从而导致构建分类器的时间和经济代价比较大。因此,后续研究出现了一些基于少量人工标注数据的半监督学习方法,其并取得了不错的成绩[4]。

为了进一步减少标注代价,无需人工标注语料的情感分类方法一直以来都受到青睐。通常,大多数这类方法是基于情感词进行学习的,但是由于情感词歧义比较大且部分情感词是领域相关的,因此取得的效果并不理想。本文引入情绪词,情绪词一般是用于描述个人内心感受的词语,具有如下特点:(1)数量少,情感色彩强烈;(2)表达的情感极性往往是领域独立的。本文提出一种基于情绪词与情感词协作学习的情感分类方法:在基于文档-单词二部图的标签传播算法框架下,使用少量情绪词集与情感词资源构建两个视图;然后分别学习标注系统对未标注数据进行标注,从中挑选具有相同标签的样本加入训练样本集,通过协作学习的方法多次迭代获得高正确率的训练样本集;最后采用最大熵分类方法在训练样本集上进行情感分类。

本文第 2 节介绍相关工作;第 3 节给出基于情绪词与情感词协作学习的情感分类方法;第 4 节是实验结果与分析;最后给出总结及下一步工作展望。

到稿日期:2012-02-16 返修日期:2012-06-08 本文受国家自然科学基金(60970056,61070123,61003155),高等学校博士学科点专项科研基金(20093201110006),模式识别国家重点实验室开放课题基金资助。

代大明(1985-),男,硕士生,主要研究方向为自然语言处理、情感分析,E-mail:mingroady@gmail.com;**李寿山**(1980-),男,副教授,主要研究方向为自然语言处理、情感分析;**李培峰**(1971-),副教授,硕士生导师;**朱巧明**(1963-),男,教授,博士生导师。

2 相关工作

目前，主流的情感分类研究主要集中在基于机器学习的分类方法上。该方法一般可以分为3种类型：全监督学习方法[4]、半监督学习方法[4,7]、非监督学习方法[7,10]。此外，为了解决领域间分类性能损失的问题，领域适应学习方法在情感分类方法研究中得到了充分的发展[12]。

一直以来，无需人工标注语料的情感分类方法一直受到许多研究者的青睐。传统的方法大多基于生成或已有的情感词典或者相关资源进行词汇计算(Term-Counting)来识别情感类别。Turney[10]首次提出基于种子词(excellent，poor)的非监督学习方法，其使用"excellent"和"poor"两个种子词与未知词在搜索网页中的互信息来计算未知词的情感极性，并用以计算整个文本的情感极性。朱嫣岚等人[14]将一组已知极性的词语集合作为种子，基于HowNet对未知词语与种子词进行语义计算，从而判别未知词的极性。后来又出现了基于机器学习的情感分类方法，例如Dasgupta and Ng[7]提出了使用谱聚类把文本按照情感维度聚类，其中仍需要人工判别情感维度的极性。与以上研究不同的是，本文在基于情感词资源的基础上引入少量的情绪词集，使用机器学习的方法进行情感分类。

3 基于情绪词与情感词协作学习的情感分类方法

3.1 概述

情绪一般是指人的内心反应与感受，如喜、怒、哀、乐等。情绪词是描述情绪的词语，是情绪表现最明显的特征形式。情感词是具有情感倾向的词语。情绪词与情感词间既存在差异又存在联系，例如"漂亮"，它体现出对某事物的正面的情感倾向，认为是情感词，因其不是关于人的内心活动的描述，所以不认为是情绪词；而"平静"描述了人的情绪，但没有体现明显的情感倾向，只认为是情绪词；"高兴"既表达了人的情绪，也具有明显的情感倾向，即正面，所以可认为既是情绪词，也是情感词。情绪词一般是少量的，并且大都领域独立，即其情感倾向一般不随领域的不同而改变。

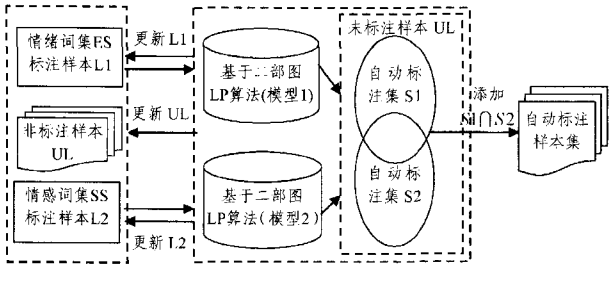

图1 基于情绪词与情感词协作学习的情感分类方法系统框架图

传统的基于资源的情感分类研究一般借助于情感知识，如情感词典，并通过词的权值计算(Term-Counting)得到文档的情感倾向。基于情感词资源，本文尝试机器学习的方法计算文档的情感倾向：采用基于文档-词的二部图的标签传播(LP)算法构建识别系统识别未标注样本，由已知词或文档的情感倾向计算其它文档的情感倾向，从而创建自动标注训练集，然后采用最大熵分类方法训练分类器进行情感分类。但是这样获取的训练样本可靠性不高，分类性能较词汇计算(Term-Counting)提升也不够理想。为此，希望基于情感词资源的同时，引入少量情绪词集，采用协作学习的方法构建高效的自动标注训练集。图1给出了本文方法的系统框架图。

3.2 情绪词的收集与标注

Plutchik[14]按照主要情绪(primary emotions)与复合情绪(complex emotions)的分类提供了一系列英文情绪词，而Turner[16]的情绪分类系统允许更多的主要情绪之间的复合情况。所以本文采用Turner的系统，该系统有4类主要情绪与26类复合情绪，每类情绪有一定数量的情绪词。Turner的情绪识别系统并没有包括Plutchik的一些频率较高的情绪词。因此，本文把Plutchik中的一些情绪词加到Turner的情绪分类系统中。情绪词列表按照以下方式再次修正：(1)手工添加词缀(POS)信息，例如sadden，saddens，saddened，saddening，sadness；(2)移除具有语义歧义的词语。表1显示了本文使用的情绪词的统计情况。

表1 情绪词的统计信息

	正面 (Positive)	负面 (Negative)	举例 (Examples)
原始情绪词	55	165	enjoy，happy，dissatisfy，hate，…
歧义情绪词	49	62	obliging，surprise，shy，…
消歧并POS后的情绪词	112	301	enjoy，enjoying，enjoyed，happy，happily，happiness，…

3.3 基于二部图的标签传播算法

3.3.1 基于文档-词二部图(bi-graph)模型

在许多关于情感分类的研究中，文档通常用词袋(bag-of-words)模型化并用向量形式描述[4,10]。在这些设置中，单词与文档间的关联是不清晰的。为了更好地捕捉单词和文档之间的关系，本文采用基于文档-词的二部图表述文档与单词的关系。图2描述了文档-词的二部图[17]，$n\times V$ 矩阵 X：维度为 n 的文档向量与维度为 V 的词向量。如果文档 d_i 包含词 w_k，并且权值 x_{ik} 对应于文档 d_i 中词 w_k 的权重，设置无向边 (d_i,w_k)，则文档 d_i 到单词 w_k 的转移概率为 $\frac{x_{ik}}{\sum_k x_{ik}}$；同理，单词 w_k 到文档 d_j 的转移概率为 $\frac{x_{jk}}{\sum_k x_{jk}}$。因此文档 d_i 到文档 d_j 的转移概率 $t_{ij}=\sum_k \frac{x_{ik}}{\sum_k x_{ik}}\cdot\frac{x_{jk}}{\sum_j x_{jk}}$。所有文档到其他文档的转移概率构成了整个二部图里面的文档转移概率矩阵 $M=\{t_{ij}\}$ ($n\times n$ 矩阵)。

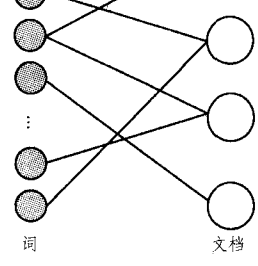

图2 文档-词的二部图

3.3.2 标签传播(LP)算法

图3描述了用于自动标注未标注样本的LP算法。在LP算法获取自动标注样本后，以这些标注样本作为训练样本集训练分类器进行情感分类。

输入：

P：n×2 标注矩阵，同时 p_{ij} 标识文档 i 属于类别 j(j=1,2)的概率

P_0^L：P_0 的前 m 行对应的 m 个标注实例

P_0^U：P_0 的后 n−m 行对应的 n−m 个未标注实例

T：T 中每一项 t_{ij} 表示从向量 i 到向量 j 的转移概率

过程：

1) 初始化：

a) 根据标注样本设定 P_0^L 的值；

b) 初始化 P_0^U，迭代次数标记 t=0；

2) 循环迭代 N 次直到收敛；

a) 传播实例的标注信息到相邻的实例依据公式 $P^{t+1}=T \cdot P^t$；

b) 还原标注实例的标注信息，即用 P_0^L 替代 P_{t+1}^L；

3) 对于每个未标注实例，根据 $\arg\max_j P_{ij}$ 得到它的正负标签。

图 3　用于标注未标注数据的 LP 算法

3.4　协作学习方法

为了得到高精度的训练样本，进一步提高系统的分类性能，本文提出了协作学习的方法。其使用情绪词与情感词分别构建两个视图(views)，以基于二部图的 LP 算法作为分类方法分别构建两个分类系统，然后对大量未标注的数据进行标注，从中选取置信度最高并具有一致标签(两个视图的分类器所给分类标签一样)的部分样本添加到自动标注样本集，从剩余样本中挑选置信度最高的样本分别添加到各视图，迭代重复这个过程直到达到某一结束条件。图 4 中给出了本文提出的协作学习方法的具体实现流程，即图 1 系统构架图的具体方法实现。

输入：

S_E：情绪词集，S_S：情感词集

L：自动标注训练数据集，U：未标注数据

过程：

1)标注样本集 L_1 为空，L_2 为空；

2)以情绪词集 S_E 与标注集 L_1 构成视图 $View_1$，以情感词集 S_S 与标注集 L_2 构成视图 $View_2$；

3)迭代直到 $L' \cap L''$为空为止：

a)在 $View_1$ 中，根据 LP 算法对未标注集 U 进行标注；从结果中选取置信度最高的数量为 p 的样本加入 L′；

b)在 $View_2$ 中，根据 LP 算法对未标注集 U 进行标注；从结果中选取置信度最高的数量为 p 的样本加入 L″；

c)添加一致标签样本 $L' \cap L''$到 L 中，即 $L=L+L' \cap L''$；

d)从 L′的剩余样本 $L'-L' \cap L''$中添加 2n 个置信度最高的正负样本到 L_1，即 $L_1=L_1+(L'-L' \cap L'')2n$；

e)从 L″的剩余样本 $L''-L' \cap L''$中添加 2n 个置信度最高的正负样本到 L_2，即 $L_2=L_2+(L''-L' \cap L'')2n$；

f)更新 U，即 $U=U-(L' \cap L'')-(L'-L' \cap L'')2n-((L''-L' \cap L'')2n)$。

图 4　使用自动标注样本与未标注数据的协同训练方法

4　实验设计与分析

4.1　实验设置

本文实验语料来源于 Mark Dredze 收集的多领域情感评论语料[1)]，分别为 Book，DVD，Electronics，Kitchen 4 个领域的产品评论。从中选取正、负评论各 800 篇作为未标注数据，各 200 篇作为测试数据。情感词来源于 MPQA[2)] 词典，该词典包括 2721 个正面倾向词和 4913 个负面倾向词。

实验中，根据所表达的情感强度不同，情感词被赋予了不同的权值：强(2，−2)、弱(1，−1)。对于情绪词，给予强的权值。设置 LP 算法中的迭代次数为达到收敛为止。在实验过程中，协作学习方法中的参数 p 分别取 150、200、250、300；p 值的设定主要鉴于：在本试验中，低于 150 或过低的数值，以及高于 300 或者过高数值，都将会导致协作学习过程迭代次数极少而快速结束，从而失去协作学习迭代过程的效果。参数 n 分别取 0、2、4、6、8、10，用以考察参数变化对分类结果的影响。n 的取值主要是鉴于：在迭代过程中，加入的样本数量尽量小，以保证加入样本后，性能变化波动不至于过大而难以观察其真实可靠性，所以在本实验当中，取到 10 为止。

分类方法采用基于 Mallet4[3)] 工具包的最大熵分类算法。分类所使用的特征为词的 Unigram。分类的性能用正确率衡量。

4.2　实验结果与分析

4.2.1　基于二部图的 LP 算法的情感分类

本实验中分别基于情绪词(E_LP)、情感词(S_LP)、情绪词＋情感词(SE_LP)，采用基于二部图的 LP 算法获取自动标注样本集进行情感分类。图 5 显示了这些方法在 4 个领域里所获得的分类效果。从图中可以看出，基于情绪词的二部图 LP 算法的分类效果表现最差，这主要是由于情绪词的规模太小，分类的覆盖范围很小。同样，基于情感词及情绪词＋情感词的二部图 LP 算法的分类性能也较差，在 4 个领域的分类性能也并不高，接近 70%左右。导致分类性能不高的主要原因是由于这些方法获取的自动标注样本正确率低，从而影响到基于这些标注样本所训练的分类器的性能。

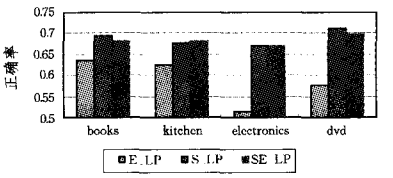

图 5　基于二部图的 LP 算法的情感分类

4.2.2　基于协作学习的情感分类

图 6 显示了本文提出的协作方法在 4 个领域训练的分类器性能，该结果显示了随着双视图添加不同数量的样本(n=0，2，4，6，8，10；p=200)，该性能的变化情况。从图中可以看出，在 n 取不同参数时，该方法都能取得较单视图更好的分类性能，说明多视图协作学习能够明显提高自动标注样本的可靠性。另外，当添加样本数过少(n=2，p=200)时，由于双视图对未标注样本的区分度降低，以致标签样本过少，获得的训练样本数量可能会过少，而且由于多次迭代造成的主动标注样本置信度低，分类效果可能会偏低；而当添加样本数过多时，会因为引入过多的置信度样本，导致分类效果可能下降，同时，不同领域因添加样本数量不同呈现不同的变化趋势，说明不同领域对参数设置是敏感而且不统一的。尽管这样，还是可以看到在局部区域内，其仍然取得了较好的效果。

1)　http://www.seas.upenn.edu/~mdredze/datasets/sentiment/

2)　http://www.cs.pitt.edu/mpqa/subj_lexicon.html

3)　http://mallet.cs.umass.edu/

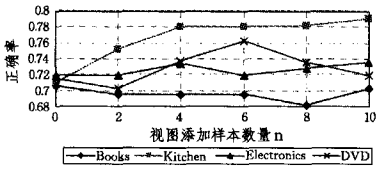

图 6 基于协作学习的情感分类性能随双视图添加样本的数量 n 的变化

图 7 显示了本文提出的协作方法在 4 个领域训练的分类器性能，该结果显示了该性能随着取样数量（$p=150,200,250,300$；$n=10$）的变化情况。从图中可以看到，在参数 p 取不同数值时，除 Book 领域外，其他 3 个领域的分类结果都超过了 70%，远远好于单独使用二部图分类方法（如图 5 所示）。在 $p=200$ 时，Book、Electronic、DVD 领域都达到了 73%以上，并且 Kitchen 领域的分类结果甚至接近 80%。同时可以发现，当取样数量过少或过多时，分类性能比最佳效果会有所下降。这是由于取样数量过少或过多可能会造成自动标注样本置信度的降低，同时过少还可能会导致获取的训练样本过小，这些都明显影响到最终的分类性能。

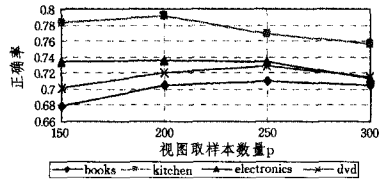

图 7 基于协作学习的情感分类性能随着视图取样本数量 p 的变化

从上述结果看出，基于协作学习的情感分类方法整体上都取得了较好的分类性能。

本文还将我们的方法同基于词计算（Term-Counting）的情感分类方法进行了比较。表 2 给出了本文所提出的协作分类与其它方法在相同资源情况下的情感分类结果。在我们的方法中，两个参数 n 及 p 分别取值 10 和 200。从表中结果可以看出，SE_LP 在各个领域的分类结果都明显好于 TC，每个领域的提升幅度在 3～5%之间。这一结果表明，使用非标注样本能够提高分类的性能。本文提出的基于协作学习的方法在 4 个领域分别取得了最佳的分类效果，在 Book、DVD 领域较 TC 提高 7～8%，较 SE_LP 提高 2～4%；在 Kitchen 领域较 TC 提高约 16%，较 SE_LP 提高 11%；在 Electronic 领域较 TC 提高约 11%，较 SE_LP 提高 6.6%。该结果充分体现了本文方法的有效性。

表 2 本文方法与其他方法在 4 个领域里的分类性能对比

正确率(%)	Books	Kitchen	Electronics	Dvd
TC	62.67%	63.48%	62.44%	66.29%
SE_LP	67.89%	68.06%	67.00%	69.70%
协作学习(p=200,n=10)	70.24%	79.06%	73.60%	73.74%

结束语 本文提出了一种基于情绪词与情感词协作学习的情感分类方法。该方法不依赖于任何人工标注的语料数据，而是使用情绪词与情感词资源自动获取高精度的自动标注样本数据用以构建分类器进行情感分类。实验结果表明，本文方法在 4 个领域都获得了较好的分类性能，在 Book、DVD、Electronic 和 Kitchen 4 个领域分别取得了 70.24%、79.06%、73.60%和 73.74%的分类效果，其明显优于基于词的单视图方法和词计算（Term-Counting）方法。

由于在获取自动标注训练样本的过程中，我们的方法丢弃了许多未标注样本，因此最后获取的训练样本集的数量规模较少。在下一步工作中，我们将考虑更有效的方法以更有效地使用更多的未标注样本。此外，将尝试进一步改进协作方法，使得获得的自动标注样本具有更高的分类精度，从而进一步提高整体的分类性能。

参 考 文 献

[1] 姚天昉，程希文，徐飞玉，等. 文本意见挖掘综述[J]. 中文信息学报，2008，22(3)：71-80

[2] Pang B, Lee L. Opinion Mining and Sentiment Analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135

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

[4] Li S, Huang C, Zhou G, et al. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification[C]// Proceedings of ACL. 2010: 414-423

[5] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning techniques[C]// Proceedings of EMNLP. 2002

[6] Cui H, Mittal V, Datar M. Comparative Experiments on Sentiment Classification for Online Product Reviews[C] // Proceedings of the AI. Menlo Park: AAAI Press, 2006: 1265-1270

[7] 闻彬，何婷婷，罗乐. 基于语义理解的文本情感分类方法研究[J]. 计算机科学，2010，37(6)：261-264

[8] Dasgupta S, Ng V. Mine the Easy and Classify the Hard: Experiments with Automatic Sentiment Classification [C] // Proceedings of ACL-IJCNLP. 2009

[9] Wan X. Co-Training for Cross-Lingual Sentiment Classification [C]// Proceedings of ACL-IJCNLP. 2009

[10] Turney P. Thumbs up or thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews[C] // Proceedings of ACL. 2002

[11] 朱嫣岚，闵锦，周雅倩. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报，2006，20(1)

[12] Lin C, He Y, Everson R. A Comparative Study of Bayesian Models for Unsupervised Sentiment Detection [C]// Proceeding of ACL. 2010: 144-152

[13] Bollegala D, Weir D, Carroll J. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification[C]// Proceedings of ACL. 2011: 132-141

[14] 吴琼，谭松波，张刚. 跨领域倾向性分析相关技术研究[J]. 中文信息学报，2010，24(1)

[15] Plutchik R. Emotions: A Psychoevolutionary Synthesis [M]. New York: Harper & Row, 1980

[16] Turner J H. On the Origins of Human Emotions: A Sociological Inquiry into the Evolution of Human Affect[M]. CA: Stanford University Press, 2000

[17] Sindhwani V, Melville P. Document-Word Co-Regularization for Semi-supervised Sentiment Analysis[C] // Proceedings of ICDM. 2008