

## 主要内容

- 介绍了此前的工作主要是基于topic的文本分类，而基于sentiment的文本工作较少
- 作者解释了，使用电影评论来作为语料库的好处在于不需要手动的标注，根据评分或者星级可以直接得出三类评级(positive, neutral, negative)，但这里作者只关注了正面和负面的评价。corpus,label
- 相比于对主题的分析，人更加擅长于对文本情感进行分析，而机器做的好的方面和人是相反的
- 直观的策略是基于选定了两类特征(positive and negative)的，对于文本的词频进行统计，作者基于此做了preliminary experiment来制造了3个baseline。

在这个实验上，人类直观想出来的特征(introspection)往往不如机器基于统计再加人的提取。

这里作者只是做一个baseline所以没有特别的选择出最优的分类器。

- 疑问：人工语料分类器选择的人太少，只有两个人
- 疑问：基于主题的文本分类任务，机器能达到的准确率作者没有提及。其实是提及了的，只是在后文 6.2 Results。基于主题的多分类任务的结果(容易达到90%+)都是远超情感二分类的。
- 然后，作者表示，这篇论文主要处理的问题是，探索基于主题的文本分类方法能否直接应用于情感分类任务上。即把情感分类任务视为positive and negative两个主题。
- 作者应用了3中传统机器学习方法，朴素贝叶斯，最大熵，支持向量机。
- 疑问：没有尝试使用条件随机场进行检查(CRF)，尝试检查的模型太少。
- 简单的介绍了三种机器学习模型在本实验中的应用方式（因暂时未学，先略过了）

这篇基本很可能已经算是情感计算的开山之作了，从过去基于主题的(topic)分类，到作者他们尝试的基于情感(sentiment)的分类。当然，由于算是介绍性质的论文，作者他们应用的算法都很基本，他们这篇论文的关注点也不在于对准确率提高，而是对情感计算这个问题的初步引入，并且和过去基于主题的文本分类进行对比。

- 最终使用的是没有偏置的语料，简化实验的考虑。  
作者首先验证了情感分类和主题分类的不同，并且验证了机器学习算法应用于情感分类是可行的，得到的分类结果会好于人直觉想出的分类标准。在测试的结果中，Naive Bayes表现最差，SVM表现最好，但差别并不显著。
- 接下来作者详细的介绍了他们做的工作，从数据集划分，数据来源和处理手段介绍，一些他们处理时的非传统手段(unconventional step)，来源于过去的一项研究。  
除此以外，在实验过程中得到了两个和主题分类相比显著的不同：
  - 有一段没有太看懂，"For this study, we focused on features based on unigrams (with negation tagging) and bigrams. Because training MaxEnt is expensive in ..."
- 1. 已经验证了的相同的传统机器学习算法在topic categorization和sentiment classification上面的表现差异是明显的，情感分类的准确度均大大低于主题分类的结果
- 2. 在所有尝试的特征中，竟然只有unigram是有着最好的分类结果的特征，加入特征词汇的频率也不能改善分类结果，这和主题分类中引入特征词汇的频率得到的结果截然相反。
- 介绍了他们最初的设想的探索结果，说明了文本主题分类任务的方法，不可以直接迁移到情感分类任务上。对于机器而言，主题分类任务是相对简单的，但情感分类任务是困难的，在前文作者提出的一点解释是，情感分类更需要understanding，即使机器能比人类对于主题分类细小差别掌握的更好，并不能有助于机器理解文本。  
当然，作者通过重新查看语料库对这些问题进行了解释，并用"thwarted expectations"来描述这一现象，先抑后扬。主要体现在评论性文字中大量的某一情感方面的词汇并不一定是作者真正想表达的内容，换句话讲，评论者可能会先对电影的各方方面面进行批评，但最后的总结却明显表达对这一电影本身的喜爱之情。这一点在评论中是很常见的。

- 然后作者介绍了他们更多更深入的研究结果，作者发现，只考虑feature出现与否，而不是出现的频率能够得到更好的结果，尤其是在SVM上。作者发现，这一特点和过去曾经报道过的关于topic categorization的结论相反。原来的结果是关于naive bayes的。这样尝试的主要原因是最大熵模型的限定，具体需要在补学之后再思考。

作者最后提出一点可能的未来改进的方向，就是分辨出选用的features，是否是真的在对评价的关注对象在进行描述。

We speculate that this indicates a difference between sentiment and topic categorization — perhaps due to topic being conveyed mostly by particular content words that tend to be repeated — but this remains to be verified.

Hence, we believe that an important next step is the identification of features indicating whether sentences are on-topic (which is a kind of co-reference problem)

在这个考虑下，作者不再使用频率作为NB和SVM的特征。

当然，作者反复的在提及Turney他们的工作，等下补充阅读。

- 接下来作者使用了bigrams，尽管显然不再满足朴素贝叶斯的假设，但研究表明，即使不满足假设也不意味着朴素贝叶斯表现会更差。这一点我比较感兴趣。