

北京理工大学

本科生毕业设计（论文）外文翻译

外文原文题目：A rule-based approach to emotion cause detection
for Chinese micro-blogs

中文翻译题目：对中文微博基于规则的情绪原因检测

毕业设计（论文）题目：基于神经网络的文本情绪原因识别研究

姓 名：胡佳星

学 院：计算机学院

班 级：07111402

指导教师：史树敏

通过使用情绪原因提取方法的基于文本的情感分类

摘要

近年来，社交网络在人们形成观点和做出决定上所起到的日益增长的巨大作用已经引起了各界的关注。微博作为时下最流行的社交网络应用之一，允许所有人对不同的话题分享自己的观点并且参与讨论，因此它也被看做获取人们的观点和情感数据的丰富来源。在这篇论文里，我们提出并且实现了一种新的用于识别微博帖子中包含的情感的方法。不同于传统的主要基于统计学的方法，我们尝试通过引入像社会学等其他领域的知识和理论，来推测和抽取产生情感的原因。根据一个理论：相应的触发事件是产生情绪的必不可少的一部分，情绪原因提取技术被看做是提高选择特征的质量的关键性一步。首先，在经过对样本数据的全面的分析之后，我们构建了一个基于规则的自动化系统用于检测和提取每一个情绪化的帖子的情绪触发事件。我们同时建立了一个通过人工标注的中文微博帖子情绪语料库。之后，我们根据已经提取出来的触发事件训练出一个分类器用于对微博帖子进行情感分类。我们所构建的系统的整体表现使我们相信其具有很好的应用价值和未来。实验结果表明我们的方法在选择有用且包含充足语义信息的特征方面非常有效。我们的系统在大多数方面都大大超越了基本要求，这也表明了它的巨大潜力。这一探索将会对情感分类的研究任务提供一种新的方法并且为未来的文本情感处理打下坚实的基础。

1. 简介

微博是博客的一种形式，它运行用户发表简短的文本贴（通常对长度做了严格限制，不得超过 200 个字符）。微博博主可以发表对很多话题的看法，这其中包括了日常生活、对电影或书的评论以及对社会事件的观点。因为微博的简单和随意的特点，微博用户的数量近年来增长的非常迅速。微博用户可以很快的更新他们的内容，这使得微博服务也扮演着实时新闻发布的中心的角色。一些机构比如公司、慈善团体和政府部门把微博作为市场营销和处理公共关系的工具。微博正在逐渐变成一个信息、想法和观点汇聚的平台。如今，许多人做决定的时候都

会受到他们在微博中关注的博主的影响。微博中的帖子被看做是挖掘情感和观点数据的丰富资源。挖掘微博社区中的用户情感在实现公共观点追踪、信息过滤和顾客关系管理等问题上引起了广泛的关注。

文本中的情感分析时现在自然语言处理领域中一个很热的研究方向。文本情感的探测或分类是许多学者和研究人员关注的任务。尽管细节部分可能存在差异,但通常而言,情感分析的目标是相同的——检测并且识别出给定文本中所传达的情感类型,比如高兴、生气和惊讶。由于基于统计学的分类方法的固有性质,通常研究人员所采用的最常用的方法都是基于统计的模型。特征选择算法例如 InfoGain 和 χ^2 测试,分类算法例如支持向量机(SVM)和 K-最近邻算法(K-NN),

是一些传统的用于文本分类任务的方法。然而,这些方法在下面两个问题上表现的并不好。第一个是对于带有否定或反问的问题的复杂语句,这些方法并不能很好的处理。第二个是对于像为什么会有这样的情绪或者这样的情绪是如何产生的这样的更深层次的信息,在这些方法中通常是被忽略的。但是,这些更深层次的信息是非常有趣的并且有时候可以更好地反映句子中的隐含的情感。如果我们想像一下一个普通人是怎样感受和理解一篇文章中所隐含的情感,我们就会注意到经常是那些在统计学上并不十分明显的因素比如事件、人物的反应,才真正决定了我们对文章所表达情感的理解。在对人们发布到网络上的帖子中所包含的情感进行分类的时候我们通常出于直觉就会把这些因素考虑进去。

在很多被他人提出并且接受的与情感如何产生问题有关的因素中,触发事件经常都被看做是最关键的因素之一。许多不同学科的学者一直在研究情感和触发事件之间的关系和相互作用。从心理学的角度来看,有理论相信触发事件本身应该是情感经历中不可缺少的一部分。在社会学领域,Kleres 提出了“叙事分析”这一方法论,它通过找到发生了什么来系统的分析情感。Lee, Chen, 和 Huang 设计了一个基于规则的系统用于检测产生情感的原因。尽管这些研究人员并没有专注于完成自动化情感分类的任务,但它们为我们的工作打下了坚实的基础。

在这篇论文中,我们采用了一种全新的方法对文本所隐含的情感进行分类。我们提出了一种利用情感原因抽取技术实现的情感分类技术,这种技术是整合了跨学科的知识和对微博数据的研究分析而得出的。我们把它看做一种新的方法是因为我们并不知道以前有任何的情感分类工作用到过这种技术。我们专注于识别从微博这一最受欢迎同时影响力最大的中文博客社区中抽取出的帖子中所隐含的情感。我们将情感触发事件作为切入点以克服传统方法中的一些缺陷。由相同的事件所触发的情感和反应被假定是相似的,因此由反问句导致的问题可以被减少。另外,我们也将深层次的语义信息考虑了进去。实验结果表明我们的系统可以以一个很好的精确度从微博帖子中提取出情感触发事件。基于有效的触发事件提取技术,我们的系统在情感分类上的效果有了巨大的提升。

本文的其余部分结构如下。第 2 节讨论了情感分析的相关工作,包括传统方法和新探索。第三部分简要介绍了中文微博平台微博,并介绍了利用情感原因提取技术提出的方法。在第 4 节中,我们的方法的两个阶段的实验结果被报告和讨论。第 5 节介绍了结论和我们今后的工作。

2. 相关工作

在这一部分，我们将展示并简要介绍情感分析任务的相关工作。

2.1 情感等级

分类主要有两种方式：二元分类（情感极性的粗粒度分类）和多级分类（细粒度多级分类）。

大多数先前的研究工作都集中在二元分类，即正面和负面。但是，能显示更详细信息的多级分类系统通常有更多实际的意义。例如，如果用户确切的情感状态是已知的，商业广告就可以实现更精确地推送并且不那么烦人。更多地了解用户的当前感受也会帮助社交网站创造一种更温暖和友好的氛围。尽管当前众多的理论中对情感多级分类问题并没有达成统一的共识，但一般有几种假定的基本的情感类型。一些主要的情绪类型，比如高兴和愤怒在类似的研究中是直观且常见的。在我们的研究中，为了实现性能与类别的丰富性之间的平衡我们采用了 Ekman 和 Friesen (1971) 的六个基本情绪类别（快乐，愤怒，厌恶，恐惧，悲伤，惊喜）。应该指出的是，没有分隔正确与错误的分界线。情感类别的设置应该根据不同的情况进行适当的设置。

2.2 文本情感分类

一般来说，对于文本中的情感处理的研究和应用仍然处于一个非常初级的阶段。自然语言固有的二义性和微妙性是许多使这项任务非常具有挑战性的因素中的一部分，尤其是在社交网络环境中，句子往往不完整、语无伦次。有很多的研究都专注于在不同类型的文本中识别和划分情绪类别，如新闻、报告、儿童童话故事、产品评论和客户反馈。通常有两种常见的方法来解决这个问题，即基于规则的和基于机器学习的方法。Chaumartin (2007) 提出并实现了一种基于规则的用于标记新闻头条中的情绪的系统。它根据语言知识和预定义规则来计算单词的情绪极性。尽管这个系统达到了高精度，但召回率很低。当涉及到基于机器学习的方法时，Tan 和 Zhang (2008) 探讨了四种特征选择方法（MI，IG，CHI 和 DF）和五种学习方法（质心分类器，K 近邻，winnow 分类器，朴素贝叶斯和支持向量机）。实验结果表明 IG 和 SVM 表现最好。他们还指出分类器严重依赖于域和主题。Tokuhisa, Inui 和 Matsumoto (2008) 采用 k-最近邻 (k-NN) 方法和两步分类模型。基于从网络提取的非常大量的数据，这个系统的表现明显要优于基准线。方法和数据集上也会存在差异。Ghazi, Inkpen 和 Szpakowicz (2010) 对分层和平坦的分类方法进行了比较。唐和陈 (2011) 从作者角度，读者视角以及使用 Plurk 数据集的组合视角对情感挖掘进行建模。Ye, Zhang 和 Law (2009) 采用了情感分类技术进行评估挖掘并在一个大型的训练数据集上达到超过 80% 的准确率。Kontopoulos, Berberidis, Dergiades 和 Bassiliades (2013) 提出了利用基于本体技术对 Twitter 帖子进行的更有效的情绪分析

我们的工作与其他人不同。我们首先引入一些来自其他领域，比如社会学的人类情感知识，并调查研究了微博上的帖子，以揭示原因事件和某些特定类型的情绪之间的一些联系。基于由 Lee 等人提出的框架（2010a），我们还开发了一个适用于对微博数据进行情感原因提取的子系统。另外，我们不只是检测情绪原因。这一能有效的挖掘深层次信息的情感原因提取子系统，将被整合为我们的情绪分类系统的一部分，这会使得我们的分类结果变得更好。我们的探索将为情绪处理问题提供一种全新的方式。

3. 利用情绪原因检测技术进行的情感分类

我们的方法背后的基本思想是寻找对情绪是“有意义”的特征，而不是简单地选择具有较高的共现度的单词。图 1 描述了我们的情感分类方法的总体框架。在以下小节中，我们将扩展系统的重要部分，解释我们如何利用情绪原因对微博上的帖子进行情绪分类。

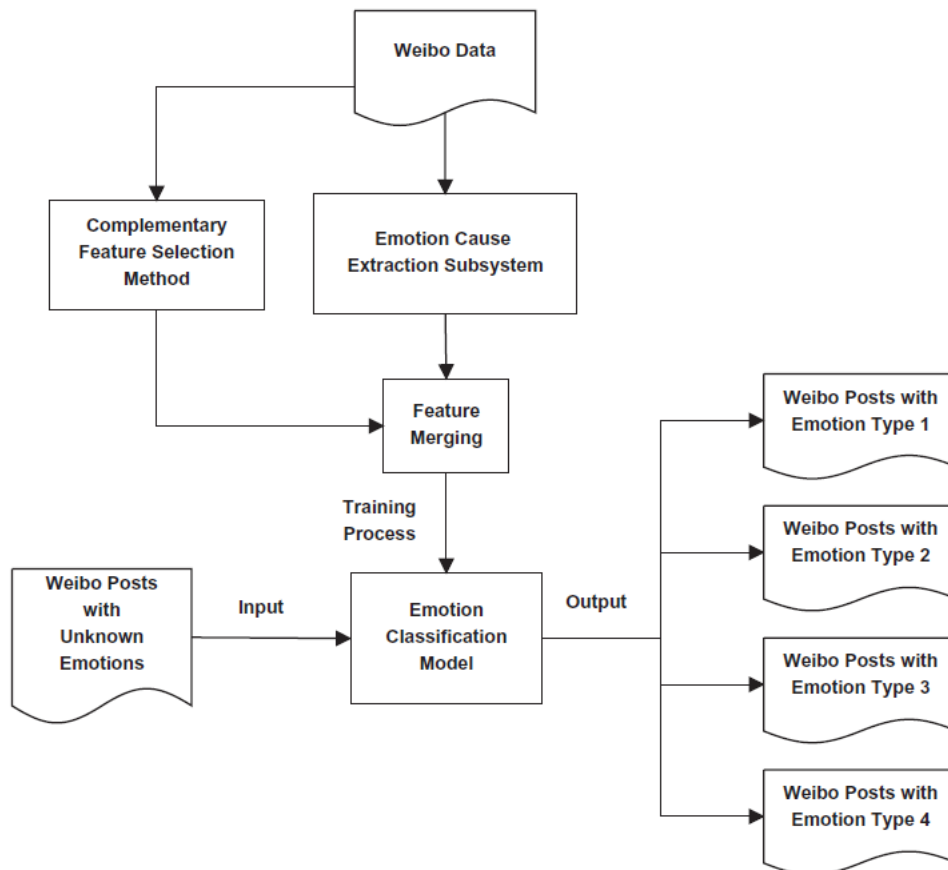


图 1. 利用情绪原因检测技术的情绪分类的一般工作流程图

3.1. 微博社交社区

微博，中文名称为“微博”，是中国的一个基于互联网的社交网络服务。微博上的内容在格式是与Twitter类似的。每个帖子的长度都限制为140个字符。用户可以使用“@ UserName”格式来提及或与他人交谈，使用“#TagName#”主题标签格式来标记主题或关键点。此外，微博还提供了一些先进的功能。用户可以在他们的帖子中插入各种图形表情和网页链接或附加视频或图片。帖子可以被评论并使用“// @ UserName”格式‘重新发布’。它也作为即时通讯服务用来在用户之间交换实时消息。在2009年推出的微博迅速吸引了巨大的用户数量，现在已经成为中国是最流行的网站之一，超过30%的互联网用户正在使用它。如图2显示了微博的搜索热度。



图 2. 微博的谷歌搜索度趋势

典型的微博帖子是由用户所提交的简短的非正式文本。微博被选为我们的研究中的测试数据的目标来源是出于以下原因。首先，它是最受欢迎的中国互联网用户在线发表意见的平台。因此假设包含情绪和情绪触发事件的帖子的数量是很巨大是没有任何问题的。其次，研究现代社交网络的环境中的情感分类有很多实际效益。第三，以前的研究很少关注来自网络的中文帖子。

在自动从微博上获得随机数目的帖子后，我们需要执行一些必要的预处理步骤。像网络链接和地理标记这样会造成干扰的内容会被删除。然后我们使用ICTCLAS工具包可以用适当的词性（POS）标签来解析，分割和标记中文帖子。

3.2.利用情绪原因实现的特征选择子系统

从社会学的角度来看，叙事是不可分割且带有自身情感的结构。叙事分析可以作为一个研究和检测情绪的方法（Kleres, 2011）。人类的情感经历具有关键性的叙述层面。这种技术背后的关键思想是如果已知“谁的行为如何，谁发生了什么”（Sarbin, 1989）那么我们就能够对情感进行推断并分析。除此之外，

外部事件通常被视为某些情绪的触发器 (Wierzbicka, 1999)。Kleres (2011) 也词汇和句法层面列举了一些语言表现形式, 这可以让我们更好的理解情绪是如何被表达和理解的。例如, 就整个句子的层面来讲, 在一个带有感情色彩的含有中立的成分和对情感和双重命题的直接引用是很常见的

除了心理学家和社会学家所阐述的的意义之外, 这个方法论方法也描述了一个分析框架, 这个框架使得我们能够以一种不同的方式来看待情绪处理问题。由于情绪和事件之间有很强的联系, 我们很自然地想到找出引发情绪的事件可能有助于挖掘情绪。所以在我们的工作中我们尝试通过找出为什么会产生情绪以及他们如何被人类读者感受到来检测情绪。首先, 我们采用情绪事件的概念。如 Talmy (2000) 所述, ‘触发事件’ 应该是情感产生的原因。应该指出原因事件并不总是我们通常理解上的实际触发事件。它也可能是导致某人产生某种情绪或者某种看法的事件。这里有些例子, 在这些例子中, 原因事件部分被着重强调了。

1. Shi3 Zai4 Tai4 Gao1 Xing4 Le, Wo3 De Fen3 Si1 Chao1 Guo4 Bai3 Wan3 Le!
(I am so thrilled that I have more than one million followers now!)
2. Zu3 Qiu3 Zhi3 Shi4 Yi1 Xiang4 You3 Xi4. Yan3 Bian4 Cheng2 Ru2 Ci3 Xue4 Xing1 De4 Si1 Sha1, Bei1 Ai1 A4!
(Soccer is just a game. How could it turn into such a violent fight. What a tragedy!)
3. Zui4 Ai4 Zhou1 Mo1 De4 Hao3 Shi2 Guang1 Le. He2 Qin1 You3 Yi4 Qi3 Chi1 Fan4, He1 Cha2. Zhen1 Shi4 Tai4 Qie4 Yi4 La.
(Love the good time of weekend most. It is such a pleasure to have dinner and drink some tea with family and friend.)

Lee 等人 (2010a) 证明了挖掘像情绪原因事件这类深度信息的可行性, 这恰好情感叙事性的社会学理论相符。他们介绍一个精心设计的多才多艺的框架用于情绪原因事件检测。该框架由以下关键部分组成:

1. 标记列表: 标记情感原因事件出现的的关键字列表。
2. 情感关键字列表: 通常用于表达情绪或感受的单词和短语的列表。
3. 语言模式集: 描述的情绪如何表达以及元素如何组织的语言模式表。

由于和在文本中检测和抽取情绪原因事件具有相似的目标, 我们决定引入基于规则的方法的一般结构。然而, Lee 等人 (2010a) 的工作是基于“中国科学院现代汉语平衡语料库”的, 简称“科学语料库” (Chen, Huang, Chang, & Hsu, 1996 年)。这个语料库收集的文本包括新闻报道、小说和诗, 这些文本通常很长且规范。而我们的工作正好相反, 我们的重点是在微博上发布的简短 (有时甚至不完整) 并且不正式的帖子。在文本类型上存在的明显的区别意味着算法的原始设置不在适用于我们的任务。

为了克服这个问题, 我们进行了彻底的调查研究以重新设计系统。首先, 我们基于微博研究了一个单独的包含超过 1000 个随机情绪实体的开发数据集。我们测试了整个语言暗示词列表 (Lee, Ying, & Huang, 2010b) 中的词汇, 以研究每个单词在识别情绪原因事件的出现上的效果。一些形式词和短语在微博中极为罕见, 一些日常的随意性表达更有可能被互联网用户使用。基于手动检查, 我们从标记列表中删除了并不经常与原因事件搭配出现的词, 并添加一些新的词汇。表 1 是适应微博帖子后的标记列表。

表 1. 语言标记词列表

Group No.	Cue Words
I	'let/make': Rang4, Ling4, Shi3, Gao3De2, Nong3De2, Shi3De2
II	'to think about': Xiang3, Xiang3Dao4, Xiang3Qi3, Yi1Xiang3 'to talk about': Shuo1Dao4, Jiang3Dao4
III	'to feel': Gan3Dao4, Jue2De2, Gan3Jue2
IV	'to see': Kan4Dao4, Kan4Jian4, Jian4Dao4 'to hear': Ting1Dao4, Ting1Shuo1 'to know': Zhi1Dao4, De2Zhi1, Fa1Xian4
V	'for': Wei4Le, Dui4Yu2
VI	'because': Yin1, Yin1Wei4, You2Yu2
VII	'is': Shi4, De4Shi4 'can': Neng2, Neng2Gou4, Ke3Yi3

出于同样的原因，传统的情感词汇列表比如由知网提供的并不适用于微博数据。相反，我们使用在经过长期工作后建立的情感词汇列表。它总共包含 1845 个通常用来表达感受，意见和情绪的单词和短语。该清单涵盖了许多中文流行的网络词汇和短语像''Niu2''（超棒）和''Gei3Li4''（真棒）。它还包括大部分的同义词和常见关键字变体。

然后，在将 Kleres（2011）描述的互联网用户典型的表达习惯和的结构模式考虑在内之后，我们独立地归纳出了一组语言模式，用于精确定位情感原因。由于长度限制，每个帖子通常都只专注于一个主题。简洁而富有表现力的特征通常暗示着它很可能成为情感原因事件如果事件是从帖子中提取的话。另外，我们删除了子句约束使模式匹配过程更加灵活。

我们的最终目标是对情感类型进行分类。情绪原因提取是为了改进传统特征选择过程而进行的尝试。很明显，如果选定的特征中有太多的噪声，分类器难以实现良好性能。所以在我们设计规则集时，精度比召回率有更高的优先级。表 2 列出了我们总结出的用于描述中文微博帖子中的元素之间的联系的语言模式，包括情绪化表达式、关键词和情绪原因。

表 2. 情绪原因检测的语言模式

No.	Linguistic pattern
1	C + I + E + K E: the nearest Noun after I C: the nearest (Noun, Verb, Noun) before I
2	E + K + II/III/IV + C E: the nearest Noun before II/III/IV C: the nearest (Noun, Verb, Noun) after II/III/IV
3	V/VI + C + E + K E: the nearest Noun before K C: the nearest (Noun, Verb, Noun) after V/VI
4	E + “yue4Cyue4K” E: the nearest Noun before “yue4Cyue4K” C: the nearest (Noun, Verb, Noun) in “yue4Cyue4K”
5	E + K + V/VI + C E: the nearest Noun before K C: the nearest (Noun, Verb, Noun) after V/VI
6	IV + C + [E] + K C: the nearest (Noun, Verb, Noun) after IV
7	E + V/VI + C + K E: the nearest Noun before V/VI C: the (Noun, Verb, Noun) between V/VI and K
8	C + E + IV + K E: the nearest Noun before IV C: the nearest (Noun, Verb, Noun) before E
9	[E + K] + VII + C + [E + K] one of the two ‘E + K’ must present C: the nearest (Noun, Verb, Noun) after VII

在表 2 中出现的缩写和符号的解释如下所示：

- C: Cause event
- I/II/III/IV/V/VI/VII: Linguistic markers in the corresponding group
- E: Experiencer
- K: Emotion Keyword
- “[]”: Optional linguistic components.

在自然语言中，动词和名词都可以用来叙述事件。Lee 等人（2010a）利用（名词，动词，名词）结构来表示事件，但是由于一些现实的困难不得不使用分句来代替。但是，因为微博很简短，有时由不完整的句子组成或者在很多情况下一个分句就可能包括了帖子中大半的内容。为了更准确地指出帖子中的情绪原因，我们设计的表达方案如下。首先，系统寻找（名词，动词，名词）的结构作为基本框架。因为情绪原因可以表示为命题或名义（Lee 等，2010a）同时由于微博具有的非正式性特点，没有必要同时用两个名词和一个动词来表示情绪原因。在实践中，如果三者中至少有一个是被认为是匹配的，那么原因（C）就会被看做是模式的一部分。在定位为这个框架之后，该框架之外的文本将被丢弃。换句话说，只有这个框架内的子串被保留并作为情感原因返回，而不是保留包含结构的整个子句。以下给出两个例子。

1. **Original Post:** "Zhong1 Guo2 Hao3 Sheng1 Yin1" Rang4 Wo3 Jing1 Tan4, Zhe4 Shou3 Ge1 Ba3 Wo3 Ye3 Chang4 De2 high Qi3 Lai Le, Shui4 Bu4 Zhao2 Jiao4.
(“The Voice” surprised me. This song really made me excited and unable to sleep.)
Linguistic Pattern: [C (Noun): (“Zhong1 Guo2 Hao3 Sheng1 Yin1”)]+[I: Rang4]+[E: Wo3] + [K: Jing1 Tan4]
([C (Noun): (“The Voice”)] [K: surprised] [E: me])
Emotion Cause Event: “Zhong1 Guo2 Hao3 Sheng1 Yin1”
 (“The Voice”)
2. **Original Post:** Zai4 Jia1 Li3 Dou1 You3 Ming2 Xian3 Zhen4 Gan3, Ke3 Shi1 Di4 Zhen4 Ju2 Que4 Mei2 Neng1 Fa1 Bu4 Yu4 Gao4, Nong4 De2 Ren2 Men2 Hen3 Shi4 Fen4 Nu4.
(We can even felt the shake in our apartments, but the Seismological Bureau failed to send any warning, which made people very angry.)
Linguistic Pattern: [C (Noun, Verb, Noun): (Di4 Zhen4 Ju2, Mei2 Neng1 Fa1 Bu4, Yu4 Gao4)] + [I: Nong4 De2] + [E: Ren Men2] + [K: Fen4 Nu4]
([C (Noun, Verb, Noun): (Seismological Bureau, failed to send, warning)] [I: made] [E: people] [K: angry])
Emotion Cause Event: Di4 Zhen4 Ju2 Que4 Mei2 Neng1 Fa1 Bu4 Yu4 Gao4
(Seismological Bureau failed to send any warning)

对于这两个示例，都会应用语言模式 1 来提取原因。在例 1 中，只有一个名词被找到，所以它被保存为情绪原因。在例 2 中，整个（名词，动词，名词），三元组被找到，所以这个结构中的子字符串被保存为情绪原因。

在已经抽取出了用于训练的帖子中的情绪原因之后，所有表示原因事件的字符串被保留。然后我们从情绪原因的集合中删除所有的停止词。我们之所以将“去除停用词”从预处理步骤移到到这里是因为我们的情绪原因抽取子系统部分依赖于停用词（语言标记）。然后我们组合所有剩余的单词和情感关键字（K），这些情绪关键字被用于在抽取阶段组成情绪原因事件集合。情绪原因集合是（C）s 和（K）s 的聚集。

对于帖子中是否包含明确的情绪和相应的原因，有三种情况：

1. 中性帖子（包括情绪类型过于模糊以至于无法判断情绪类型的帖子）；
2. 带有情绪和情绪的帖子。
3. 有情感，但没有明确表达情感的帖子。

显然，如果我们只使用情绪原因提取子系统的输出作为选定的特征，情况 1 和 3 将是难以区分，因为它们都不含可提取的情绪原因。为了将那些包含情绪却没有明确的情绪原因的句子和中性句区分开来，我们必须找出一种特征选择方法可以应用于情况 3。我们选择 χ^2 （卡方）测试。 χ^2 测试是依赖性的经典测试方法。它将观察得到的数据与根据具体假设预期得到的数据进行对比。卡方是观测

值与预期值平方差的总和与预期值的做除法得到的值。 χ^2 测试的定义式为：

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

在表达式中， χ^2 是卡方，O 是在每个类别上的观测频率，E 是期望的频率。

以 χ^2 测试为补充，我们的特征选择策略可以直接生成。由情绪原因提取和 χ^2 测试产生的两组特征被混合以最终的特征集。算法 1 描述了合并过程，其中的 `emocause_set` 是情绪原因的集合，`Chi_list` 是按 χ^2 测试分数降序排列的单词列表，`final_set` 为最终的特征集。

Algorithm 1. Feature set merging

```

1:  $d \leftarrow$  predefined dimension of final feature set
2:  $d_{\text{cause}} \leftarrow \text{sizeof}(\text{emocause\_set})$ 
3:  $r \leftarrow d - d_{\text{cause}}$ 
4:  $\text{final\_set} \leftarrow \text{emocause\_set}$ 
5: for each  $w$  in  $\text{Chi\_list}$  do
6:   if  $r == 0$  then
7:     break
8:   else
9:     if  $w$  is not in  $\text{final\_set}$  then
10:      add  $w$  to  $\text{final\_set}$ 
11:       $r \leftarrow r - 1$ 
12:   else
13:     continue
14:   end if
15: end if
16: end for

```

3.3 用于情绪分类的 SVM 和 SVR

支持向量机（SVM）已经被证明是可以被用于文本驱动的情绪分类的有效模型。基于风险最小化原则，SVM 根据从训练数据中选择的元素做出决定。它寻求一个超平面来将训练数据分为正面和负面两个类别。图 3 描绘 SVM 的基本机制。

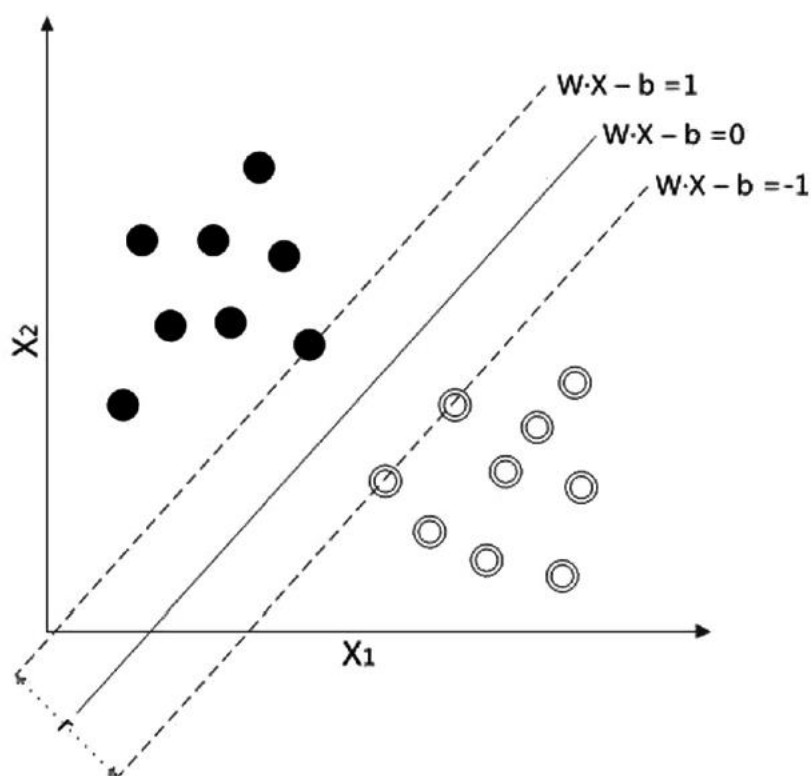


图 3. 具有余量的 SVM 最大超平面

就选定的核心函数 $K(x, y)$ 而言, SVM 有几种不同的变体。在我们的工作中, 我们将讨论限定在线性内核函数上, 因为其具有相对较高的性能。此外, 因为我们的工作集中在从微博网站随机抓取的简短和不规范的文本上, 因此我们支持向量回归 (SVR) 用于分类任务。SVR 是由 Drucker, Burges, Kaufman, Smola 和 Vapnik (1997) 提出的 SVM 的一种版本。它使用与 SVM 相同的原理, 但 SVR 的输出是一个实数。对于处理一个不平衡的数据集而言, 这是一个更好的选择。

。

4. 实验和讨论

我们的实验分两步完成:

1. 抽取情绪原因事件;
2. 训练并且测试分类器

在这一节, 首先介绍我们构建的基于微博的数据集。之后我们将说明实验的两个步骤以及分别用到的对应的评估方法。

4.1 微博情绪数据集

首先，我们从微博上随机爬取下 20000 多个帖子。删除重复的帖子之后，16485 个帖子被留在了我们的数据库中。两个人工标记员对语料库中的所有语料进行人工标注，标注出它们的情绪类型并且如果情绪原因存在也需要标注出情绪原因。为了避免歧义并简化问题，只有明确表达某种类型的情绪的语料会被标记为“具有情绪”，并且被标记上相应的情绪标签。当多个情绪存在于同一个词条中，只有主导词被选中并贴上标签。如果情绪类型不清楚或难以确定，这些条目将被视为中立。帖子注释不一致被转发给第三个注释者作最后决定。此外，所有帖子在注释后均由作者确认。

情绪原因事件由目标文件的一个子串表示。例如，在句子“Zhen1 Gao Xing4, Jin1 Tian1 Dao4 Gong1 Si1 Hou4 Ling3 Dao3 Gei3 Wo3 MenMei3 Ge4 Ren2 Dou1 Fa1 Le4 Hong2 Bao1。”（“今天真高兴在我们到公司后，我们的老板给了每个人一些奖金”），字符串“Ling3 Dao3 Gei3 Wo3 MenMei3 Ge4 Ren2 Dou1 Fa1 Le4 Hong2 Bao1。”（“老板给大家一些奖励”）被复制并标记为情绪原因事件。标注者被指示遵循最小长度的规则，这意味着只有标记足够覆盖原因的最短长度。我们假设在微博上没有超过 140 个字符的帖子，这意味具有多个情绪原因事件的情形是非常罕见的，可以被忽略，因此这种假设是合理的。。表 3 列出了在我们的语料库中每种情绪分类对应的实例的实际数量。我们保持数据集处于不平衡状态，因为我们认为这反映了实际情况。一个重要的原因在于很大比例的帖子是没有评论的重新发布或者是由第三方应用程序自动生成。此外，通常不包含情绪的新闻标题占了中性帖子的很大一部分。

表 3. 语料库摘要

Emotion	Number of posts	Number of posts with causes
Happiness	513	354 (69.0%)
Anger	478	452 (94.6%)
Disgust	153	137 (89.5%)
Fear	125	61 (48.8%)
Sadness	313	255 (81.5%)
Surprise	102	46 (45.1%)
Neutral	14801	N/A
Total	16485	1305

4.2 情绪原因抽取

4.2.1 评估标准

与其他的自然语言处理任务类似，我们用三个关键变量来评估我们的实验结果：精确度、召回率和 f-分数。在我们对情绪原因抽取子系统的评估中，精确度和召回率的定义如下所示：

$$Precision(G, S) = \frac{\sum_{P_i \in G} \sum_{k_j \in P_i} ListP(SList_j, GList_j)}{\sum_{P_i \in S} \sum_{k_j \in P_i}} \quad (1)$$

$$Recall(G, S) = \frac{\sum_{P_i \in G} \sum_{k_j \in P_i} ListR(SList_j, GList_j)}{\sum_{P_i \in G} \sum_{k_j \in P_i}} \quad (2)$$

$$ListP(SList, GList) = \frac{\sum_{SStr_j \in SList} \max_{GStr_i \in GList} StrP(GStr_i, SStr_j)}{|SList|} \quad (3)$$

$$ListR(SList, GList) = \frac{\sum_{GStr_j \in GList} \max_{SStr_i \in SList} StrR(GStr_i, SStr_j)}{|GList|} \quad (4)$$

其中 G 是人类注释的黄金标准原因集合, S 是我们的系统产生的原因集合, P_i 是在数据集中的一个微博帖子, k_j 是这篇文章中的情绪关键词, $SList_j$ 和 $GList_j$ 是对应在 P_i 中的关键字 k_j 的系统输出和黄金标准的情感原因列表, $SStr_i$ 和 $GStr_i$ 是表示情绪原因的字符串。

有两种方法可以计算 $StrP$ 和 $StrR$: 松弛匹配 1 和松弛匹配 2 (Lee 等人, 2010a)。松弛匹配 1 只考虑系统输出字符串和我们的黄金标准字符串之间是否有任何重叠。松弛匹配 2 需要考虑重叠长度。因为我们打算对原因提取系统的改进进行简要评估, 因此我们简化了这部分并采用了松弛匹配 1 的方案。

$$StrP(GStr, SStr) = \begin{cases} 1 & \text{if } GStr \text{ and } SStr \text{ overlaps} \\ 0 & \text{else} \end{cases} \quad (5)$$

$$StrR(GStr, SStr) = \begin{cases} 1 & \text{if } GStr \text{ and } SStr \text{ overlaps} \\ 0 & \text{else} \end{cases} \quad (6)$$

计算出来的 F-分数通常作为精确度和召回率的谐波平均值。

4.2.2 原因抽取的结果

为了评估我们针对微博环境重新设计的基于规则的系统的, 我们根据李等人 (2010a) 提出的相同的规则和语言线索在相同的微博数据集上进行原因抽取实验。

因为文本类型之间的差异, 我们决定采用相同的基线以实现更好的比较。这个简单的基准线可以用一条规则来描述: $C(V) + K$, 表示在情绪关键字左边包含第一个动词的子句就是情绪原因子句。

表 4 显示了基线和初始规则集的性能。最初的规则集在精确度方面明显优于基准线。但是, 其召回率远低于基线。造成这个结果的一个主要原因是初始的规则集不是基于微博帖子设计的, 这也是证明了我们调整工作的必要性。

表 4. 原始原因抽取规则集的性能

	Precision	Recall	F-score
Baseline	45.06	51.24	47.95
The original rule-set	71.63	41.51	52.56

表 5 显示了我们系统的整体性能。因为语言标记列表和专为微博帖子设计的规则集的存在，系统的精确度和召回率都大大提高。

表 5. 重新设计算法的性能

	Precision	Recall	F-score
Redesigned algorithm	75.70	51.59	61.30

表 6 显示了基准线和我们的整合了情绪原因的系统在情绪分类上的性能。符号“with EC”意味着情绪分类的结果是利用情绪原因提取技术得到的。对于大多数情绪类别（高兴，愤怒，恐惧和惊奇），我们的系统具有更高的精度同时保持同一级别的召回。对于情绪类别讨厌，虽然我们的系统精度较低，但召回率更高。就可以同时评价评估精确度和召回率的 F 评分而言，我们的系统在大多数情况下明显优于基线。具体而言，高兴的 f 分数提高了 1.6% 愤怒 2.0%，厌恶提高了 9.2%，惊奇提高了 0.8%。这表明情绪原因提取技术在选择更有效，含有更丰富信息的特征方面非常有效。对于悲伤这种情绪分类，我们的系统无法超越基准。我们的假设是不开心的时候人们不愿意陈述原因。人们在像高兴或惊奇这样的积极或强烈的情绪之下更热衷于描述当下的情况。所以从表达悲伤情绪的帖子中抽取情绪原因的效果并没有从表达其他感情的帖子中抽取情绪原因的效果好。

尽管通常精度和召回率还不是很高，改善的结果确实证明了我们提出的方法的巨大潜力。此外，随着训练数据集的规模变得越来越大，像 χ^2 测试这样的经典特征选择方法计算复杂度呈指数增长，而基于规则的系统只有线性增长趋势。

5. 结论和未来工作

情绪分类具有广泛的应用空间。一个准确和高效的分类系统是非常有趣的。在研究中，我们引入了关于从社会学中引入的关于情绪和叙事的关系，同时利用情绪原因提取技术研究了文本情绪分类的任务。我们对一个中文微博平台进行了实验。首先，自动收集了 16485 可能存在情绪的帖子。之后我们的工作分两子步骤完成：利用情绪原因提取技术的特征选择和利用 SVR 技术的情绪分类。我们基于通用的社交网络特征和其他经过详细总结的语言模式设计了一个基于规则的

子系统用来从原始文本中检测和提取原因事件。这个子系统的输出用于生成用于训练分类器的特征集。我们也说明了基准线的情况。实验结果表明，利用情绪原因提取技术，我们的系统大多数情况下明显跑赢了基准系统。具有前景的结果表明利用来自其他学科的知识和方法的模型改善传统情绪分类模型的潜力。

尽管利用情绪原因提取技术，我们成功的获得了非常好的结果，我们工作还不够全面的。仍然有很多等待着我们去研究的对人类情绪会产生巨大影响的因素存在。处理具有复杂的语言模式的文本仍然具有很大的挑战。除此之外，总结和 design 语言模式的工作仍然经过良好训练的人类专家耗费大量的人力和时间。

在未来的工作上，我们首先需要测试其他的元素以发现它们在我们的任务中的优缺点。我们也计划设计一个更复杂的原因提取系统，以更好的解决具有挑战性的情况同时减少错误。更多的语言模式也需要继续探索。最后，我们需要找到一种方法可以自动产生和修改模式集合，这样的话人类需要付出的劳动就可以最小化。