



可在 ScienceDirect 上获得的内容列表

信息处理与管理

期刊主页: www.elsevier.com/locate/inforoman



HEMOS:一种新颖的基于深度学习的用于社交媒体情感分析的细粒度幽默检测方法

Da Li^{a,*}, Rafal Rzepka^{b,c}, Michal Ptaszynski^d, Kenji Araki^b

北海道大学信息与科学学院 *Department of Information Science, Faculty of Science, Hokkaido University, Japan* *deRI* *Center for Information Science, Intelligent Information Systems (AIS), 4 Technology, Japan*

条信息

关键词:

情感分析幽默极性
社交媒体表情符号

深度学习

2010 硕士:

00-01

99 -

00

摘要

本文介绍了一种基于深度学习方法的中文细粒度情感分类系统 **HEMOS** (hu mor- emoji - slangbased)。我们研究了识别幽默、象形图和俚语对社交媒体情感处理任务的影响的重要性。在第一步中，我们收集了 576 个频繁出现的网络俚语表达作为俚语词典；然后，我们将 109 个微博表情符号转换为文本特征，创建中文表情符号词典。下一步，通过在标准的“积极”和“消极”情感类别中添加新的“乐观幽默类型”和“悲观幽默类型”进行两种极性标注，我们将这两种词典应用于基于注意力的双向长短期记忆循环神经网络 (AttBiLSTM)，并在规模较小的标记数据上测试其性能。实验结果表明，所提方法能够显著提升微博 (中国最大的社交网络) 情感极性预测的性能。

1.介绍

随着 web 2.0 应用的广泛流行，人们可以在 Twitter、Facebook 或微博 (2009 年推出的中国最大的社交媒体网络) 等社交网络上自由发表意见。这些新的沟通渠道每天产生大量的数据，这些数据成为各种研究目的的有洞察力的信息源，通常使用自然语言处理 (Kavanaugh et al., 2012)。其流行且必不可少的任务之一是情感分析，它主要关注从在线用户生成的内容中自动预测情感，但也吸引了心理学、认知语言学或政治学等其他领域的兴趣。情感分析可以在选举结果预测、股票预测、意见挖掘、商业分析和其他数据驱动的任务中发挥作用 (Li, Rzepka, & Araki, 2018a)。情感分析已被广泛应用于分析在线用户生成数据的现实世界应用中，如意见挖掘、产品评论分析和业务相关活动分析 (Zhao et al., 2018)。近二十年来，英语情感分析取得了值得关注的成功。另一方面，对其他语言的情感进行分类，例如中文，仍然处于早期阶段 (Wang et al., 2013)。目前的情感分析主要集中在基于文本的用户生成内容上。然而，社交网络上表情符号、图片、表情包等各种新形式的符号信息，可以让用户更有创造性地表达自己。表情包和俚语都有

* 相应的作者。

E-mail: lida@ist.hokudai.ac.jp (李丹)。1 <https://www.weibo.com>

<https://doi.org/10.1016/j.ipm.2020.102290>

2020 年 1 月 9 日收到;修订后的表格于 2020 年 4 月 28 日收到;2020 年 5 月 3 日接受, 2020 年 6 月 20 日在线可用

0306-4573/©2020 Elsevier Ltd.版权所有。

成为社交网络上在线非正式表达中引人注目的一部分。

幽默是定义我们作为人类和社会实体的个人方面之一。它是一个非常复杂,但又无处不在的概念,我们可以简单地通过有趣的效果的存在来定义它,例如大笑或幸福的感觉,这是一组在我们的生活中发挥相关作用的现象(Reyes, Rosso, & Buscaldi, 2012)。在我们有口语或书面语言之前,人类在某些情况下用笑来表达我们的享受或加入(Mathews, 2016)。在今天的社会,人们倾向于写有趣的笑话,玩文字游戏,用表情包或表情包来评论时事的“白色讽刺”,²和缓解生活压力的自我嘲讽,往往以非常有创意的方式。

表情符号是电子信息和网页中使用的表意文字和笑脸符号。这种象形文字起源于 1997 年的日本手机,在被添加到多个移动操作系统后,在 2010 年 10 月开始在全球范围内越来越流行。2015 年,牛津词典将“笑 cry”表情命名为“年度词汇”。在我们看来,在情感研究中忽略表情符号是不合理的,因为它们传达了一条重要的情感信息,并在社交媒体中发挥着表达情绪和观点的重要作用(Guibon, Ochs, & Bellot, 2016;Li, Rzepka, Ptaszynski, & Araki, 2019b;诺瓦克, Smailović, 斯鲁班, & Mozetič, 2015)。

网络俚语在互联网上同样无处不在。微博、讨论组、社区问答网站和社交网络等新的社会语境的出现,使得俚语和非标准表达在网络上大量出现。尽管如此,俚语在传统上一直被视为非标准语言的一种形式,并不是语言分析的重点。因此,它在很大程度上被忽视了(Kulkarni & Wang, 2017)。

此外,我们还注意到,传统的情感分析一般只考虑三个维度(积极、消极和中性)进行文本分类,而在情感分析中使用一些象形图很难简单地归类为积极或消极。特别是,许多表情符号和俚语似乎纯粹是用于笑、自嘲或开玩笑,这传达了一种隐含的反讽,这种反讽在社交网络上的中文对话中被大量使用(Li et al., 2019b)。因此,我们认为,幽默作为自然语言中最常见的表达方式之一,也可以被视为一个有用的维度,用于 emoji 数据集的情感分析。

在本文中,为了填补上述空白,我们重点关注微博上使用网络俚语和表情符号的幽默帖子,以建立:

- 1)俚语和表情包是否都通过识别具有挑战性的两极分化的幽默条目来提高情感分析;
- 2)添加新的“乐观幽默类型”和“悲观幽默类型”类别是否改善了情感极性预测的结果。

为了进行实验,我们收集了 576 个频繁出现的中文网络俚语作为一个俚语词典。然后,我们将 109 个微博表情符号转换为文本特征,创建了一个中文表情符号词典。

我们采用深度学习的方法,利用基于注意力的双向长短期记忆循环神经网络对社交媒体进行情感分析,并将帖子分为“积极”、“消极”、“乐观幽默型”和“悲观幽默型”四类。我们的实验结果表明,提出的方法可以显著提高微博情感极性预测的性能。

我们的主要贡献如下:

- 我们收集了 576 个常用的中文网络俚语作为中文俚语词典。
- 我们将 109 个微博表情符号转换为文本特征,创建了中文表情符号词典。
- 我们根据经验确定了微博上可见的中国文化固有的幽默特征,并将微博分为四篇
- 我们将俚语词汇和表情符号词汇应用于基于注意力的双向长短期记忆循环神经类别:积极、消极、乐观幽默和悲观幽默。

network (AttBiLSTM), 并提出了 HEMOS(见第 6 节), 一种用于社交媒体情感分析的细粒度幽默检测方法。

2.相关研究

传统上,情感分析是一种二元方法,将文本情感分为积极、消极和中性。Peng, Cambria, and Hussain(2017)对中文情感分析进行了全面的文献综述。根据他们的发现,中文情感分析的方法大致可以分为有监督的机器学习方法和无监督的基于知识的方法。对于这项研究,我们主要关注有监督的机器学习方法。

基于机器学习的情感分析任务通常分为三个独立的阶段,即文本分割、特征提取和情感分类。文本分割将文本分割成有意义的 token。特征提取同时检索情感特征和原始分词特征,并将它们表示在词袋(BoW)中。最后,将数据集输入到机器学习模型中,用于将情感分数分配给给定的文本。常用的算法有 Naïve 贝叶斯、最大熵、SVM、神经网络等(Dhande & Patnaik, 2014;Khan, Baharudin, Lee, & Khan, 2010;Vinodhini & Chandrasekaran, 2012)。

此类研究的一个典型例子是 Tan 和 Zhang 的实证工作,他们对中文文档中的情感进行了分类(Tan & Zhang, 2008)。四个特征,即互信息、信息增益、卡方和文档频率,分别在五种不同的算法上进行了测试,包括质心分类器、k-近邻、Winnow 分类器、

²与“善意的谎言”类似,我们将这种特殊的善意讽刺定义为“白色讽刺”。

Naïve 贝叶斯(NB)和支持向量机(SVM)。在这些算法中,发现信息增益和 SVM 特征在主题依赖分类器下产生最佳性能。Chen 等人提出了一种新颖的情感分类方法,该方法融合了现有的中文情感词典和卷积神经网络(Chen, Xu, Gui, & Lu, 2015)。他们的结果表明,所提出的方法优于仅使用词嵌入特征的卷积神经网络(CNN)模型(Kim, 2014)。

这些中文情感分类方法虽然取得了令人满意的结果,但只是从极性基的角度考虑了一个纯文本数据集。除了经典的二标签分类问题(积极、消极或中性),Liu 和 Chen 提出了微博的多标签情感分析原型(Liu & Chen, 2015)。他们还在两个微博数据集上比较了 11 种最先进的分类方法(BR、CC、CLR、HOMER、RAkEL、ECC、MLkNN、RF-PCT、BRkNN、BRkNN-a 和 BRkNN-b)的性能。虽然 Liu 和同事考虑了多标签分类,但没有一篇论文将幽默作为情感分类的一个维度。

最近,对于英语语言的情感分析,更多的研究人员意识到了 emoji 和俚语数据集的价值。Wu 等人构建了一个用于情感分析任务的英语俚语词典(命名为 SlangSD),并证明了它的易用性(Wu, Morstater, & Liu, 2016)。在 Soliman, Elmasry, Hedar 和 Doss(2014)的研究中,作者构建了一个情感词的俚语情感词和习语词典(俚语 sentiment Words 和习语词典, SSWIL)。他们还提出了一种针对阿拉伯俚语的高斯核 SVM 分类器,用于对 Facebook³上的阿拉伯新闻评论进行分类。提出的分类器实现了 88.63%的精确率和 78%的召回率。Manuel 等人提出了一种方法,用于查找在互联网上的博客、评论和论坛文本中新发现的俚语情感词的情感得分(Manuel, Indukuri, & Krishna, 2010)。在他们的论文中,还提出了一种利用俚语词汇,借助 Delta 词频和加权逆文档频率技术来计算文档情感得分的简单机制。

Chen 等人提出了一种额外关注表情符号的 Twitter 情感分析新方案(Chen, Yuan, You, & Luo, 2018)。他们首先分别学习了正面和负面情感推文下的双极性表情嵌入,然后通过基于注意力的长短期记忆网络(LSTM)关注这些双极性表情嵌入,训练了一个情感分类器。他们的实验表明,双极性嵌入对于提取情感感知的表情符号嵌入是有效的。在中国社交媒体微博方面,Zhao 等人构建了一个名为 MoodLens 的系统(Zhao, Dong, Wu, & Xu, 2012),这是第一个用于微博中文帖子情感分析的系统。在 MoodLens 中,95 个表情符号被映射为四类情绪(愤怒、恶心、快乐和悲伤),这些情绪作为词条的类别标签。他们收集了 350 多万条带标签的帖子作为语料库,并训练了一个快速的 Naïve 贝叶斯分类器,经验精度为 64.3%。然而,他们的方法中的精度仍然比较低,很多表情包的情绪已经发生了巨大的变化。更具体的微博表情分析在第 5 节讨论。

2017 年, Felbo, Mislove, Søgaard, Rahwan 和 Lehmann(2017)在他们的 Twitter 情感分析模型 DeepMoji 中提出了一个强大的系统,利用表情符号。在这项研究中,通过双向长短期记忆(Bi-LSTM)模型训练了包含 64 个常见表情符号之一的 1.46 亿条推文,以解释在线推文内的情感。DeepMoji 也很好地完成了讽刺检测任务,验证准确率达到了 82.4%。他们的系统甚至超过了人类检测器,后者成功获得了 76.1%的准确率。虽然讽刺和反讽一般倾向于传达负面情绪,但由于这两个因素可能会逆转给定文本的整体情绪得分,我们发现,在中国社交媒体(我们的例子是微博)中,除了所代表的正面和负面情绪外,用户往往会表达一种逃脱传统两极化的隐性幽默。

最近,在我们之前的研究(Li, Rzepka, Ptaszynski, & Araki, 2018b)中,我们分析了微博上使用的带有面部表情的表情符号的使用情况。我们将表情符号极性应用在一个长短期记忆循环神经网络(LSTM)中,并将微博帖子分为三类:积极的、消极的和幽默的。在 Li, Rzepka, Ptaszynski, and Araki (2019a)中,我们提出了一种基于注意力的 GRU 网络模型,利用表情极性来改进较小标注数据集上的情感分析。我们的实验结果表明,所提出的方法可以显著提高情感极性预测的性能。这两项研究都为深度学习模型的 softmax 输出的概率分配了一个超参数,并将 Li et al. (2018a)工作中标记的表情符号作为表情极性。然后,我们为表情极性分配一个超参数,以计算情感分类的最终概率输出。但我们发现,很多微博词条都带有自嘲等幽默的一面,也包含一些乐观或悲观的情感负荷。简单地把它们归为“幽默”一类,而不进行两极分化,很容易做出错误的预测。因此,为了解决这个问题,我们提出了一种细粒度的幽默分类方法,以改善双极性情感分析的结果。

3. 中国文化中的幽默

幽默是一种引发笑声和提供娱乐的体验倾向,它也可以被定义为一种可靠的激发兴奋的方式(Ruch, 1993)。然而,关于幽默是什么以及它的社会功能,存在着许多理论。在漫长的历史进程中,不同文化背景的地区对幽默有着不同的解读。古希腊时期,柏拉图首先在《斐利伯斯》中提出了这样一种观点:幽默的本质是弱者的无知,因而在被嘲笑时无力报复。后来,亚里士多德在其著名的《诗学》中提出,一种不令人厌恶的丑陋是幽默的根本方面(Jones, 2005)。在名为“Natya Shastra”的古梵文戏剧中,Bharata Muni 将幽默定义为九种情绪反应之一,可以通过演员表演的情绪模拟在观众中唤起(Sharma, 2011)。幽默的表达

³<https://www.facebook.com/>



图 1 所示。微博贴出网络俚语和表情包的例子。这条微博写道：“慢跑后，我饿了。有人给我发了一张烤串的照片。我太累了，不想谈恋爱。”

最早记载于公元前 2500 年中国出现第一批中国诗歌和文学书籍的时候(Yue, 2010)。在中国古代，在儒家文化中，幽默传统上被认为是颠覆性的或不得体的。它通常被认为是讽刺和讽刺。另一方面，道教非常重视幽默。道教的核心思想和表达方式都体现了幽默的智慧(Yue, 2014)。目前中国人在社交媒体上倾向于表达隐式幽默，一些帖子可以被视为悲观的讽刺，而其他条目则可以被认为是纯粹的玩笑或乐观的自嘲。对这些现象的更具体分析，将在下一段讨论。如今，幽默成为一种无处不在的人类现象，出现在所有类型的社会互动中(Vieweg, Hughes, Starbird, & Palen, 2010)。它也是一种沟通的形式(通常是创造性的)，弥合了不同语言、文化、年龄和人口统计学之间的差距(Furnham, 1984)。我们大多数人在一个典型的一天中，会因为一些有趣的事情笑很多次(Martin, 2006)。笑可以释放内啡肽，放松身体，帮助缓解压力。虽然幽默是一种游戏形式，但它具有多种认知、情感和社交功能(Martin & Ford, 2018)。然而，幽默是一种难以定义的情感。语言学家、心理学家和人类学家都认为幽默是一个包罗万象的范畴，涵盖了任何令人发笑的事件，或者被认为是令人发笑的事件(Attardo, 2010)。不同的人不一定会觉得同样的事情同样有趣。许多事情在一个群体看来是有趣的，但在另一个群体看来却可能令人生厌，而有些笑话是私人的或个人的，往往仅限于一个或很少几个人觉得好笑。正如拉斯金(2012)所说，“许多笑话或情景会让很多人(如果不是所有人)感到有趣，这一事实进一步强化了幽默的普遍性。因此，我们面对的是一种普遍的人类情感——幽默。对幽默做出反应是人类行为、能力或能力的一部分，而其他部分又构成了人类语言、道德、逻辑、信仰等重要社会心理表现。”因此，我们假设对幽默情感的识别对情感分析具有重要意义。

通过分析微博，我们发现了一个有趣的现象，那就是人们倾向于表达隐性幽默，这种幽默摆脱了传统的正负两极性。而且，表情符号似乎在增强这种效果方面起到了至关重要的作用。图 1 展示了一个例子

一个微博，包含了表情包和网络俚语。在微博的第二行，(lei jue bu ai)是(hen lei, gan jue zi ji le累觉不爱 le，意思是“太累了，不想谈恋爱 很累，感觉自己不会再爱了

这种表达形式通常被缩短为由四个汉字组成的“短语”。这些表达方式与英语中的用法类似，如“lol”(大声笑出来)、“idk”(我不知道)、“ASAP”(尽快)，但对于中文表达来说，它承载着更有创意、复杂和模糊的含义。

一般来说，这种特殊的表达形式可以从中文的“成语”中看出，这是一个四字短语，其含义是从老一辈继承下来的，包含着道德观念、智慧之珠或前人的经验。如今，“成语”在汉语对话和教育中仍然扮演着重要角色。然而，年轻一代倾向于采用这种形式，并赋予它新的含义，在数字语境中带有一种含蓄的幽默、自嘲或娱乐感。通常，网络俚语

⁴ 本文用斜体表示汉语的罗马化(拼音)。

表 1
我们的中文网络俚语词汇的例子。

| 类型 | 例子(起源) | 英语翻译 |
|-----------|--------|----------|
| 数字 | | “笑” |
| 拉丁字母 | | “该死的” |
| 缩写 | | “生活太艰难了 |
| 中国的收缩 | | 有些谎言 |
| 新词 | | 最好不要揭穿。” |
| 带有改变的或的短语 | | “失败者” |
| 延伸的含义 | | “低俗大亨” |
| 双关语和文字游戏 | | “和谐” |
| 俚语来源于 | | “兄弟” |
| 外语 | | |

表达方式是幽默的、讽刺的或讽刺的，这也是使它们具有吸引力和被广泛接受的关键方面。这种非正式的表达方式在中国的社交媒体平台上很流行，被广泛使用，可以找到更多的例子:*lei jue bu ai, ren jian bu chai*(生活如此艰难，有些谎言最好不要被揭穿), *xi da pu ben*(如此令人振奋的消息，每个人都在庆祝并在全世界传播)。

表情符号通常用于增强或强调特定内容的情感。然而，我们观察到的另一个有趣的现象是，某些表情符号的含义可能会随着时间的推移而发生变化。例如，最初是为了表达“再见”的手势而设计的。然而，似乎越来越多的人开始使用这个表情符号，以一种有趣的方式来表达拒绝或自嘲的假笑，这种用法目前正在中国社交媒体上流行起来。在 [Li et al. \(2018b\)](#)的研究中，作者证实了这种效果，强调这种表情往往带有更多的幽默而不是消极性。例如，在下面的帖子中：“慢跑后，我饿了。有人给我发了一张烤串的照片。我太累了，不想谈恋爱”

”。在这种背景下，《微博》试图表达一种悲观态度的幽默微妙之处。在这种情况下，表情符号和俚语似乎在表示幽默方面发挥了实用主义的作用，而不仅仅是表现积极或消极的情绪。

4.中国网络俚语词典

俚语作为一种流行的非正式语言表达形式，在社交媒体上被广泛用于发布或评论([Jones & Schieffelin, 2009](#))。它出现在人们的日常生活中，在很大程度上改变了人们相互交流的方式。然而，由于它的非结构化性质和在不同语境下含义的棘手性，给机器直接提取情感极性带来了很大的困难。

由于网络俚语不容易自动处理，在极性识别任务中会造成很大的困难。为了提高中文社交媒体情感分析的性能，我们创建了一个中文网络俚语词典(示例如[表 1](#)所示)。

我们识别了 576 个在互联网上经常使用的中文俚语短语，并存储在我们的中文网络俚语词典中。它们的来源多种多样，包括互联网新词排行榜、百度百科、⁵ 维基百科⁶ 以及百度贴吧⁷、微博⁸等社交媒体系统(处理后的数据时间跨度在 2010 年至 2019 年之间)。

经过分析，我们将词条分为以下 7 类:

- 数字:如 233(“笑/ol” :中国人用 233 来表达“笑得停不下来”，因为 233 是中文 BBS 站点中的一个情感标志⁹，这个标志是所有表情包列表中的字符数字 233);213(“一个很笨的人”);520/
- 521(我爱你)。拉丁字母缩写:中国用户通常使用带有拼音功能的 QWERTY 键盘。大写字母打字很快，不会转换成表意文字。(小写字母会自动转换成汉字)。拉丁字母缩写(而不是汉字)有时也被用来逃避审查。如 SB(“哑巴
- 女人”);中文 YY 缩略语:(“幻想/性如**左**之思”);**人**蕉(“生活 TT(“套套”))。太难了，有些谎言最好不要被揭穿”:这句话出自台湾歌手林宥嘉一首名为《谎言》的歌曲的歌词。这句俚语反映了一些人，尤其是年轻人

⁵<https://baike.baidu.com>

⁶<https://en.wikipedia.org>

⁷<https://tieba.baidu.com>

有道文档翻译
pdf.youdao.com



图 2 所示。109 个可以转化为中文字符标签的微博表情包。

因此，作为一个工作假设，我们假设所有带有表情符号的微博都传达了非中性的情绪。首先，我们将中文俚语词典和中文表情符号词典加入到一个用于匹配新词和象形文字的分词工具中。然后我们使用更新后的工具对一个大数据集的句子进行切分。其次，我们将分词输出应用到词嵌入工具中，以获取词向量。接下来，我们应用考虑网络俚语和表情符号的词嵌入模型，用训练数据训练一个基于注意力的双向长短期记忆循环神经网络模型(AttBiLSTM)来学习一个输出表示。最后一步，我们将测试数据输入到 AttBiLSTM 模型中，并使用 softmax 分类器获得预测结果并输出它们的概率。

为了解决之前研究中遇到的幽默检测问题(Li et al., 2018b)，我们应用 AttBiLSTM 模型实现了一种细粒度分类方法，将微博分类为四类:积极、消极、乐观幽默和悲观幽默。“乐观幽默”类别包括段子、自嘲和调侃等，而“悲观幽默”类别包含讽刺和反讽。我们认为，提出的细粒度分类方法可以更有效地检测幽默表情，提升“正/负”双极性情感分类的结果。在我们的实验中，我们重点研究了微博上使用网络俚语和表情包的幽默帖子，以验证：

- 1)是否俚语和表情包词汇都通过识别难以两极化的幽默词条来提高情感分析结果;
- 2)添加新的“乐观幽默类型”和“悲观幽默类型”类别是否改善了双极性情感预测的结果。

6.1.基于注意力机制的双向长短期记忆循环神经网络

基于注意力的双向长短期记忆循环神经网络模型(AttBiLSTM)的架构如图 3 所示;我们提出的模型主要由 word 编码器、attention 层和 softmax 层组成。细节如下所示。

6.1.1.词编码器

考虑到微博的条目是少于 140 个单词的句子，与 Yang et al.(2016)的相关工作相比，在我们的研究中，我们专注于句子级别的社交媒体情感分类。假设一个句子包含 n 个单词(w_1, w_2, \dots, w_n)， k w 表示句子中的第 k 个单词， n 表示句子的长度，每个单词都嵌入到一个 d 维向量中，称为单词嵌入(Bengio, Duchame, Vincent, & Jauvin, 2003)。然后，一个由 $n \times d$ 词嵌入层生成的嵌入矩阵 Mis ，其中 n 是句子的长度， d 是嵌入大小。最后，将该矩阵作为双向 LSTM 网络的输入。

双向 LSTM 网络非常适合根据时间序列数据进行分类、处理和做出预测，因为时间序列中重要事件之间可能存在未知持续时间的滞后。双向网络的性能优于单向网络，长短期记忆网络(LSTM)的速度要快得多，也比标准的循环神经网络(rnn)或时间窗多层感知器(MLPs)更准确(Graves & Schmidhuber, 2005)。

LSTM网络通过使用以下方程从 $t = 1$ 到 t 迭代计算网络单元激活来计算从输入序列 $x_T = (x_1, \dots, x_t)$ 到输出序列 $y_T = (y_1, \dots, y_t)$ 的映射。LSTM 单元的方程如下(Hochreiter & Schmidhuber, 1997):

表 2
中文表情符号词典的例子。

| Emoji | 文本特征 | 情感/含义 |
|-------|------|---------------------|
| | | “微笑” |
| | | “可爱的” |
| | | “太高兴” |
| | | “掌声” |
| | | “嘻嘻” |
| | | “watermelon-eating” |
| | | “眨眼” |
| | | “贪婪” |
| | | “无语/尴尬” |
| | | “汗” |
| | | “擦鼻涕” |
| | | “放风” |
| | | “愤怒” |
| | | “难过/下降委屈” |
| | | “可怜的” |
| | | “失望” |
| | | “悲伤” |
| | | “哭泣” |
| | | “害羞” |
| | | “肮脏的” |
| | | “爱的脸” |
| | | “想吻你的脸” |
| | | “抛媚眼” |
| | | “舔屏幕” |
| | | “渴望” |
| | | “牵狗” |
| | | “smugshug” |

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$

$$f_t = \sigma(W_f \cdot X + b_f)$$

$$i_t = \sigma(W_i \cdot X + b_i)$$

$$o_t = \sigma(W_o \cdot X + b_o)$$

$$c_t = f_t \odot c_{t-1} + i_t \tanh(W_c \cdot X + b_c)$$

$$h_t = o_t \odot \tanh(c_t)$$

(1)
(2)
(3)
(4)
(5)
(6)

其中 W 项表示权重矩阵, W_i, W_f, W_o 为对角权重矩阵用于偷看孔连接, b 项表示偏置向量(b_i 是输入门偏置向量), σ 是 logistic *sigmoid* 函数, i, f, o 和 c 分别是输入门, 忘记门, 输出门和细胞激活向量, \odot 是向量的元素乘积, 细胞输入和细胞输出激活函数, 通常(以及在我们的研究中) \tanh 。 x_t 表示 *LSTM* 单元输入的词嵌入, h_t 是隐藏状态的向量。具体示意图如图 4 所示。

双向 *LSTM* 网络包含两个独立的 *LSTM*, 它们通过合并句子的两个方向的信息来获取单词的注释(Graves & Schmidhuber, 2005)。具体来说, 在时间步骤 t , 前向 *LSTM* 根据之前的隐状态 h_{state} 和 $t-1$ 输入向量 x_t 计算隐状态 f_{ht} , 而后向 *LSTM* 计算隐状态 b_{ht}

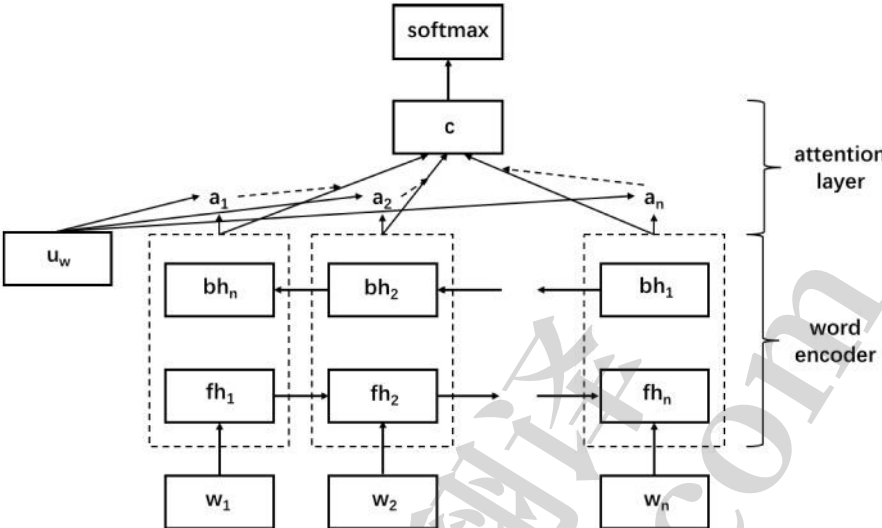


图 3 所示。AttBiLSTM 模型的架构。

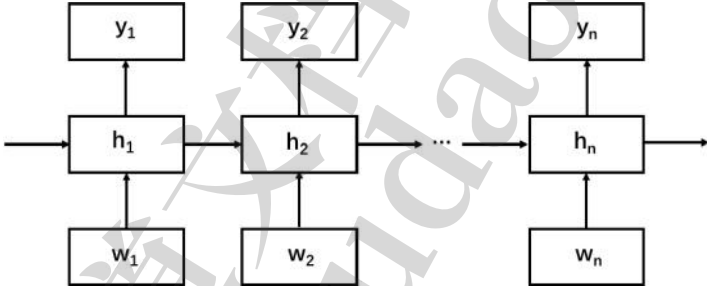


图 4 所示。LSTM 模型的架构。

基于相反 t 的隐藏状态 $bhand_{t,1}$ 的输入向量 x ，最后将两个方向的向量连接起来作为最终的隐藏状态。双向 LSTM 网络中的两个 LSTM 神经网络参数相互独立，并且它们共享句子的相同词嵌入。双向 LSTM 模型在时间步长 t 时的最终输出 h_t 定义如下：

$$h_t = [fh_t, bh_t]$$

(7)

6.1.2. 注意层

我们提出的模型将原始的微博帖子投影到向量表示中，在此基础上我们构建了一个分类器来进行情感分类。在这一小节中，我们将介绍如何通过使用注意力结构从词向量逐步构建句子级别的向量。并不是所有的单词都对微博词条含义的表示做出了同等的贡献。因此，我们引入注意力机制来提取对帖子意义重要的单词，并展示我们如何计算这些有信息量的单词的重新表示的总和，以形成句子向量。

正如我们上面提到的，双向 LSTM 网络在每个时间步产生一个隐藏的 ht 状态。我们首先通过 t 一层 MLP 来输入单词注释 h_t 来学习一个隐藏表示 u_t ，然后将单词的重要性作为 u_t 与单词级上下文向量 u_w 的相似度来衡量，并通过 softmax 函数得到一个归一化的重要性权重 α_t 。其次，我们根据权重计算句子向量 c 作为单词注释的加权和。上下文向量 u_w 可以被理解为对记忆网络中使用的单词进行固定查询的“信息性单词”的高级表示(Sukhbaatar, Weston, Fergus 等人，2015)。方程式描述如下：

$$u_t = \tanh(W_u h_t + b_u)$$

(8)

$$\alpha_t = \frac{\exp(u_t^T u_w)}{\sum_i \exp(u_i^T u_w)}$$

(9)

$$c = \sum_t \alpha_t h_t$$

(10)

有道文档翻译
pdf.youdao.com

表 3

仅用 AttBiLSTM 方法进行两类情感分类的结果。

| 类别 | 评价 | 结果 |
|-----|-----------------|--------|
| 积极的 | 精度 | 69.07% |
| | 回忆 | 82.76% |
| | <i>F1-score</i> | 74.71% |
| 负 | 精度 | 78.05% |
| | 回忆 | 50.00% |
| | <i>F1-score</i> | 60.95% |

表 4

考虑网络俚语和表情包词典的 AttBiLSTM 两类情感分类结果。

| 类别 | 评价 | 结果 |
|-----|-----------------|--------|
| 积极的 | 精度 | 82.35% |
| | 回忆 | 84.48% |
| | <i>F1-score</i> | 83.40% |
| 负 | 精度 | 73.77% |
| | 回忆 | 70.31% |
| | <i>F1-score</i> | 71.99% |

表 5

仅用 AttBiLSTM 方法进行四类情感分类的结果。

| 类别 | 评价 | 结果 |
|-------|-----------------|--------|
| 积极的 | 精度 | 71.15% |
| | 回忆 | 70.00% |
| | <i>F1-score</i> | 70.57% |
| 负 | 精度 | 68.89% |
| | 回忆 | 67.39% |
| | <i>F1-score</i> | 68.13% |
| 乐观幽默的 | 精度 | 72.72% |
| | 回忆 | 60.61% |
| | <i>F1-score</i> | 66.18% |
| 悲观的幽默 | 精度 | 39.29% |
| | 回忆 | 61.11% |
| | <i>F1-score</i> | 47.83% |

表 6

我们提出的方法的结果。

| 类别 | 评价 | 结果 |
|----|----|----|
|----|----|----|

| | | |
|-------|-----------------|--------|
| 积极的 | 精度 | 89.79% |
| | 回忆 | 88.00% |
| | <i>F1-score</i> | * |
| 负 | 精度 | 88.89% |
| | 回忆 | 78.57% |
| | <i>F1-score</i> | 71.74% |
| 乐观幽默的 | 精度 | * |
| | 回忆 | 74.99% |
| | <i>F1-score</i> | 79.71% |
| 悲观的幽默 | 精度 | 83.33% |
| | 回忆 | * |
| | <i>F1-score</i> | 81.48% |
| | 精度 | 65.00% |
| | 回忆 | 72.22% |
| | <i>F1-score</i> | * |
| | 精度 | 68.42% |
| | 回忆 | |
| | <i>F1-score</i> | |

* $p < 0.05$ 。

这篇文章和类似的条目通常以评论的形式发布，GIF 或视频展示了一位裁判通过扣篮来展示她或他的篮球技能。这篇文章似乎表达了一种隐含的幽默细微差别，当海报看到裁判有多棒时，是一种夸张的惊讶。因为这个表情配了 emoji，提高了分类的性能，预测了隐含的幽默含义。

作为之前研究中遇到的幽默检测问题的解决方案(Li et al., 2018b)，我们提出的细粒度情感分类方法可以更清晰地检测微博帖子的情感。在图 6 中，我们展示了一个微博的例子

Post: 裁判：我当初就是因为没有对手，才选择做裁判的 🤔

Pinyin: Cai pan: Wo dang chu jiu shi yin wei mei you dui shou, cai xuan ze zuo cai pan de 🤔

Segmentation: 裁判/：/我/当初/就是/因为/没有/对手/，/才/选择/做/裁判/的/[摊手]

Translation: Referee: Because there was no opponent who could beat me, I chose to become a referee 🤔

图 5 所示。幽默帖子正确分类示例。

Post: 吃饱了就有力气减肥了 😏😏

Pinyin: Chi bao le jiu you li qi jian fei le 😏😏

Segmentation: 吃饱了/就/有/力气/减肥/了/[阴险]/[阴险]

Translation: When your stomach is full, you get the strength to reduce weight 😏😏

图 6 所示。又一个幽默帖子正确分类的例子。

它被我们提出的方法正确分类为“乐观幽默”，而基线则认为它是积极的。正如评估者所同意的那样，看来这个用户写了一个笑话只是为了好玩，而我们提出的方法正确识别了这种情绪。

误差分析显示，我们提出的方法检测负面情绪和“悲观幽默”情绪的结果还是比较低的，这与识别讽刺和反讽的难度密切相关。我们计划在未来训练更多的深度学习模型，并增加数据量，以提高悲观幽默检测的结果。

此外，由于解析器词典和我们的俚语词典中都缺少新的俚语，一些帖子被错误预测，这显然给结果带来了负面影响。在 Ptaszynski 等人(2016)的研究中，作者指出，处理网络语言的系统性能逐渐下降的一个典型原因是，网络俚语一直在变化。这一点也体现在我们的研究中;在图 7 中，我们展示了一个被错误分类为“悲观”的帖子的例子

Post: 名人英文金句翻译与当下流行语神对应, 励志心灵鸡汤秒变毒鸡汤 🐶

Pinyin: Ming ren ying wen jin ju fan yi yu dang xia liu xing yu shen dui ying, li zhi xin ling ji tang miao bian du ji tang 🐶

Segmentation: 名人/英文/金句/翻译/与/当下/流行语/神/对应/, /励志/心灵鸡汤/秒/变毒/鸡汤/[doge]

Translation: Translation of English familiar quotations corresponds to the current Internet slang, chicken soup becomes poisonous chicken soup in seconds 🐶

图 7 所示。一个“乐观幽默”帖子被错误分类为“悲观幽默”的例子。

幽默”的范畴,却被标注者标注为“乐观幽默”。“*苗边*”(“秒变”)、“*毒鸡汤*”(“毒鸡汤”)等俚语被解析错误,一个移位的字导致分词工具误识别。“*杜记堂*”是由“*记堂*”转化而来的俚语,意为“反励志名言”(例如:“有些人天生伟大,有些人成就伟大,有些人最后像你一样”或“我不懒,我只是有很高的动力去无所事事”)。作为“心灵鸡汤”系列丛书的简称,“*记堂*”是近年来用来表达“励志名言”的意思。社交媒体上每年都涌现出大量类似“*杜济堂*”的新词。在俚语词典中添加新词成本高昂,还不足以跟上网络俚语演变的速度。为了应对这一现象,可以在下一阶段的研究中考虑采用字符级别的上下文文化词嵌入方法(例如 BERT 预训练的中文词嵌入模型)。

9.结论及未来工作

该文提出了一种用于细粒度情感分类的 HEMOS(幽默-表情符号-俚语)系统。我们收集了 576 个常用的中文网络俚语表达,并创建了一个俚语词典;然后,我们将这 109 个微博表情符号转换为文本特征,创建了中文表情符号词典。此外,增加了新的“乐观幽默类型”和“悲观幽默类型”,我们创建了一个新的、四级的微博情感分类的基础。我们将这两个词典应用到我们新颖的深度学习方法中,即基于注意力的双向长短期记忆循环神经网络(AttBiLSTM),以对中国社交媒体进行更细粒度的情感分析。我们的实验结果表明,HEMOS 系统可以显著提高微博情感极性预测的性能。

为了实现更有效的中文情感分析方法,我们将增加标注数据量,以解决所提出的细粒度分类中“悲观幽默”和“负面”类别结果明显较低的问题。

在这项研究中,我们排除了图像和视频。然而,在进一步的研究中,将图像添加到数据源中会很有趣。我们假设这可能会增强文本情感分析,因为贴纸和表情包也携带情绪。为了利用这些额外的信息,必须在预处理阶段添加一个图像处理阶段。

此外,在数据标注阶段,我们发现,与普通用户相比,垃圾用户(传播恶意链接或商业内容的用户)使用特定表情符号的帖子的高发生率很高。处理这个问题可能是一个有趣的研究课题,我们的方法对于区分普通用户和垃圾用户可能很有用。

我们的最终目标是通过将新引入的情感相关特征输入深度学习模型,来研究这些特征对情感分析有多大的益处,这应该可以让我们构建一个高质量的情感识别器,用于更广泛的中文情感。

信用作者贡献声明

李大:概念化、形式分析、写作——原稿。Rafal Rzepka:监督、写作-审查、编辑。Michal Ptaszynski:写作-评论&编辑。荒木健二:写作-评论&编辑。

鸣谢

本文是发表在 Li, Rzepka, Ptaszynski, Araki, 2019a 上的会议论文的修订版。作者感谢匿名审稿人提出的有益的意见和建议。这项工作得到了 jsp Kakenhi Grant 编号 17K00295 的支持。

参考文献

Aldunate, N., & González-Ibáñez, R.(2017). 计算机媒介传播中的表情符号综合综述。《心理学前沿》, 7,2061。Attardo, S.(2010)。幽默的语言学理论。I.沃尔特·德·格吕特。

Batty, M., & Taylor, M. J.(2003)。六种基本面部情绪表情的早期加工。《认知脑研究》, 17(3), 613-620。

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C.(2003)。神经概率语言模型。《机器学习研究杂志》, 3(Feb), 1137-1155。

Bridle, J. S.(1990)。前馈分类网络输出的概率解释,与统计模式识别有关系。《Neuro computing》, 3, 227 - 236。

陈颖, 袁杰, 尤强, 罗杰(2018)。基于 bi-sense 表情嵌入和基于注意力的 LSTM 的推特情感分析。2018 年 ACM 多媒体会议论文集《多媒体会议》。acm117 - 125。

陈中, 徐荣, 桂霖, 陆强(2015)。结合卷积神经网络和单词情感序列特征进行中文文本情感分析。《中文信息处理学报》, 29(6), 172-178。

Dhande, L. L., & Patnaik, G. K.(2014)。使用朴素贝叶斯神经分类器分析电影评论数据的情感。《国际计算机科学新兴趋势与技术杂志(IJETTCS)》, 3(4), 313-320。

Ekman, P. & Friesen, W. V.(1971)。在面部和情感上的跨文化常量。《人格与社会心理学杂志》, 17(2), 124。

Felbo, B., Mislove, A., Sogaard, A., Rahwan, I., & Lehman, S.(2017)。使用数以百万计的表情符号出现来学习任意域的表达,以检测情绪、情感和讽刺。arXiv 预印本 arXiv:1708.00524。

弗恩汉(1984)。旅游与文化冲击。《旅游研究志》, 11(1), 41-57。

Graves, A., & Schmidhuber, J.(2005)。基于双向 LSTM 和其他神经网络架构的框架音素分类。《神经网络》, 18(5-6), 602-610。

Guibon, G., Ochs, M., & Bellot, P.(2016)。从表情包到情感分析。2016 WACAI 论文集。

Hochreiter, S., & Schmidhuber, J.(1997)。长短期记忆。《神经计算》, 9(8), 1735-1780。

Jones, g.m., & Schieffelin, b.b.(2009)。聊天短信和回话:“我的闺蜜吉尔”,从电视到 YouTube。《计算机媒介传播杂志》, 14(4), 1050-1079。

Jones, J. D.(2005)。柏拉图的幽默哲学的完整分析。 <http://www.jonathonjones.com/papers/plato.pdf>。

卡瓦诺、A·L、福克斯、e·A、Sheetz、S·D、Yang、S、Li、lt、Shoemaker、D·J、……谢,林(2012)。政府对社交媒体的使用:从常规到关键。《政府信息季刊》29(4), 480-491。

Khan, A., Baharudin, B., Lee, L.H., & Khan, K.(2010)。文本-文档分类的机器学习算法综述。《信息技术进展》, 1(1), 4-20。

Kim, Y.(2014)。用于句子分类的卷积神经网络。arXiv 预印本 arXiv:1408.5882。

Kulkarni, V., & Wang, W.y(2017)。《TFW,DannGina,Juvie,hotsie-tosie:论网络俚语的语言和社会层面》。arXiv 预印本 arXiv:1712.08291。

Li, D., Rzepka, R., & Araki, K.(Rzepka, Araki, 2018a)。面向中国社交媒体情感分析的微博表情包初步分析, 论文集。第 32 届日本人工智能学会年会论文集 1j3- 04。

Li, D., Rzepka, R., Ptaszynski, M., & Araki, K.(Rzepka, Ptaszynski, Araki, 2018b)。面向中文情感分析的表情符号感知循环神经网络模型。第九届 IEEE 感知科学与技术国际会议论文集(ICAST 2018)161-166。

Li, D., Rzepka, R., Ptaszynski, M., & Araki, K.(Rzepka, Ptaszynski, Araki, 2019a)。基于表情感知注意力机制的双向 GRU 网络中文情感分析模型。对话代理的语言和认知方法研讨会联合论文集(LA CATO DA 2019)和关于弥合人类和自动推理之间的差距(BTG 2019)与第 28 届人工智能国际联合会议(UCAI 2019)CEUR WS Vol-2452, 论文 2

Li, D., Rzepka, R., Ptaszynski, M., & Araki, K.(Rzepka, Ptaszynski, Araki, 2019b)。一种考虑中文俚语词汇和表情符号的基于机器学习的中文社交媒体情感分析新方法。aai-19 情感内容分析研讨会论文集, AffCon 2019 CEUR WS Vol-2328, paper 10 Liu S. M., & Chen J.-H. (2015)。基于多标签分类的情感分类方法。专家系统及其应用, 42(3), 1083-1093。

Manuel K., Indukuri K. V., & Krishna, P.R.(2010)。分析网络俚语进行情感挖掘。第二届 Vaagdevi 国际现实世界问题信息技术会议论文集。IEEE 9-11。

马丁, R. A.(2006)。《幽默心理学》。爱思唯尔。

马丁, r.a., & 福特, T.(2018)。幽默心理学:一种综合方法。学术出版社。

马修斯, L.(2016)。幽默在情绪调节中的作用:适应性和非适应性幽默形式的差异效应。CUNY 学术著作。

Merity, S., Xiong, C., Bradbury, J., & Socher, R.(2016)。指针哨兵混合模型。arXiv 预印本 arXiv:1609.07843。

Mikolov, T., Chen K., Corrado, G., & Dean J.(2013)。向量空间中单词表示的有效估计。arXiv 预印本 arXiv:1301.3781。诺瓦克, P. K., Smailović, J., Sluban, B., & Mozetič, I.(2015)。表情包的情感。PLOS One, 10(12), e0144296。

彭, H., Cambria, E., & Hussain, A.(2017)。中文情感分析研究综述。认知计算, 9(4), 423-435。

Ptaszynski, M., Masui, F., Nitta, T., Hatakeyama, S., Kimura, Y., Rzepka, R., & Araki, K.(2016)。有害短语的类别最大化相关性和双过滤自动优化的可持续网络欺凌检测。国际儿童-计算机交互杂志, 8,15-30。

拉斯金, V.(2012)。幽默的语义机制。24.施普林格科技商业传媒。

Reyes, A., Rosso, P., & Buscaldi, D.(2012)。《从幽默识别到反讽检测:社交媒体的具象语言》。数据与知识工程, 74,1- 12。

鲁奇, W.(1993)。愉悦与幽默。I.吉尔福德出版社。

Rzepka, R., Okumura, N., & Ptaszynski, M.(2017)。连接面孔的世界——当代象形文字的意义和可能性(日文)。日本人工智能学会学报, 32(3), 350-355。

Shama, R.(2011)。新光文学研究中的喜剧。露露。com。

Soliman, T. H., Elmasry, M., Hedar, A., & Doss, M.(2014)。Facebook 上阿拉伯俚语评论的情感分析。国际计算机与技术杂志, 12(5), 3470-3478。

Sukhbaatar, S., Weston, J., Fergus, R.等(2015)。端到端记忆网络。2015 年神经网络信息处理系统进展论文集 2440- 2448。Tan S., & Zhang J.(2008)。中文文档情感分析的实证研究。专家系统及其应用, 34(4), 2622-2629。

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L.(2010)。两次自然灾害事件期间的微博:twitter 可能有助于情景感知。SIGCHI 计算机系统的人因会议论文集 s1079-1088。

Vinodhini, G. & Chandrasekaran, R.(2012)。情感分析与意见挖掘:研究综述。国际期刊, 2(6), 282-292。

王,X。 ,C。 ,Y。 ,太阳,L,吴,L, & 保 z(2013)。一种基于情感分析的微博社交网络抑郁检测模型。亚太知识发现与数据挖掘会议论文集。springer201 - 213。

吴, L., Monstatter, F., & 刘辉(2016)。SlangSD:构建并使用俚语情感词典进行短文本情感分类。arXiv 预印本 arXiv:1608.05129。

杨,杨,Z D,戴尔,C。 ,他,X。 ,Smola A。 ,& Hovy 大肠(2016)。用于文档分类的分层注意力网络。2016 年计算语言学协会北美分会会议论文集:人类语言技术 1480- 1489。

岳欣。(2014)。儒、佛、道文化对幽默的态度。《心理探索》, 1(5), 1-5。

岳晓东(2010)。《中国幽默的探索:历史回顾、实证发现和批判性反思》。

赵军、董林、吴军、徐坤(2012)。Moodlens:基于表情符号的中文推文情感分析系统。第十八届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集 1528- 1531。

贾,赵,P J。 ,Y。 ,梁,J。 ,谢,L。 & 罗,J。(2018)。社交媒体中表情符号使用情况分析与预测。2018 年网络大会论文集。国际万维网会议指导委员会 327 - 334。

有道文档翻译
pdf.youdao.com