

基于 LSTM 和多特征组合的电影评论专业程度分类

吴 璠 李寿山 周国栋

(苏州大学计算机科学与技术学院 江苏 苏州 215006)

摘 要 社交网络上的电影评论通常既有专业评论家写的专业评论,也有普通观众写的非专业评论,区分网络电影评论是否为专业评论对于电影质量评估有着重要的价值。由于电影评论属于短文本,用词不规范,特征稀疏,因此传统的文本特征选择方法以及传统的分类模型并不能完全适用于电影评论专业程度的分类。为此,文中主要研究基于神经网络模型电影评论的专业程度分类,即判断其是专业评论还是非专业评论。首先通过基于神经网络的 LSTM 模型学习不同特征的表示,包括基于词的表示、基于词性的表示,以及基于依存关系的表示,然后通过融合不同特征表示来学习和捕捉有效的文本特征,从而帮助评论专业程度分类。该方法在美国著名的影评网站烂番茄网(Rotten Tomatoes)数据集上进行实验,实验结果表明,在融合了词性和依存关系特征的模型的分类正确率达到了 88.30%,比仅使用词特征的基准模型提高了 3.66%。这说明在模型中引入词性特征、依存关系特征能够有效提升评论专业程度分类的效果。

关键词 多特征组合,评论专业程度分类,神经网络,LSTM,SVM

中图法分类号 TP391 文献标识码 A

Movie Review Professionalism Classification Using LSTM and Features Fusion

WU Fan LI Shou-shan ZHOU Guo-dong

(Institute of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract Movie Reviews on social networks usually include professional reviews written by professional critics, as well as non-professional reviews written by ordinary audience, and it is of great value to distinguish whether online film reviews are professional reviews for film quality evaluation. Due to the fact that film review is a short text book with irregular words and sparse features, the traditional text feature selection method and traditional classification model cannot fully apply to the classification of film review's professional level. Therefore, the paper mainly studied movie review professionalism classification based on neural network model, that is judging whether it is professional review or non-professional review. The representation of different features is learned through neural network-based LSTM model, including word-based representation, part-of-speech representation, and representation based on dependencies, and valid text features are learned and captured by fusing different feature representations to help review professionalism classification. The method was experimented on the Rotten Tomatoes dataset of the famous American film review website. The experimental results show that the classification accuracy rate of the model combining part-of-speech and dependency is 88.30%, which is 3.66% higher than the benchmark model only using word features. This shows that the method of introducing part-of-speech features and dependency features into the model can effectively improve the effectiveness of professional classification of reviews.

Keywords Multi-feature fusion, Review professionalism classification, Neural networks, LSTM, SVM

1 引言

近年来,随着快节奏生活压力的增大,看电影成了人们放松减压、休闲聚会的一种娱乐方式,电影娱乐行业也迅速发展起来。而且随着各大社交平台的发展,越来越多的人开始在各种社交网站上通过撰写影评来分享自己的观影感受,这些社交网站上的影评质量参差不齐,其中不仅有专业影评家写的评论,也有普通观众写的评论。大部分观众在观看一部电影之前,都会在互联网上先搜索这部电影的介绍查看其他已

看用户的评论信息,因此,用户评论数据对电影票房有着重要影响。当我们根据评论来评价一部电影时,我们需要更多地考虑专业影评家的评论,其更加客观,参考价值更高,而非专业的普通观众写的评论相对来说带有更多主观的感情色彩,会与个人喜好的影片类型等个人因素有关,但是在一定程度上也能够反映大众的口味与倾向。因此如何快速地区分专业影评家撰写的评论和非专业普通观众的评论就显得尤为重要。

目前关于评论专业程度的研究还很少,比较相关的研究

本文受国家自然科学基金(61331011,61672366)资助。

吴 璠(1994—),女,硕士,CCF 会员,主要研究方向为自然语言处理;李寿山(1980—),男,教授,主要研究方向为自然语言处理,E-mail:lishoushan@suda.edu.cn(通信作者);周国栋(1967—),男,教授,主要研究方向为自然语言处理。

方向是情感分类和评论质量检测。但是,与上述研究不同的是,在评论专业程度分析中,我们更加关注于人的因素的影响,通过分析作者的不同表达方式来更好地获取评论的专业程度。为了获取作者表达意见的不同方式,需要从评论中抽取不同的特征来体现两种评论不同的表达方式。本文基于神经网络模型主要构建 3 种特征的表达:词特征、词性特征和依存特征,从而构建作者表达意见的方式。通过一个联合学习模型学习所有特征的表达,并最终输出整个文本的表示和对应的专业程度分类标签。实验表明,本文提出的基于神经网络的 Multi-LSTM 模型能够有效地提高电影评论分类的准确度。

本文第 2 节主要介绍了评论专业程度分类的相关工作;第 3 节阐述了提出的分析方法;第 4 节介绍实验设置并分析实验结果;最后,总结全文并提出研究方向。

2 相关工作

尽管评论专业程度分类的研究工作刚刚起步,但该任务与情感分类和评论专业程度分类有很多相似之处,与情感分类一样,评论专业程度分类也是一个分类问题,从情感分类的研究中可以得到启发。情感分类是指从用户主观评论文本中提取出用户对所评价对象的情感倾向性。最初由 Pang 等^[1]提出的有关情感分类的方法主要应用于在线评论,他们按照不同的方法提取特征,如 Unigram,Bigram,POS,Position 等,然后对比分析支持向量机(Support Vector Machine,SVM)、最大期望算法(expectation Maximization Algorithm,EM)、朴素贝叶斯(Naïve Bayes,NB) 3 种机器学习方法的实验结果,最终发现在 IMDB 影评数据集上 SVM 结合 Unigram 特征的提取方法效果最好,正确率达到了 82% 以上。之后,Turney^[2]提出了一种基于点互信息值的方法来分析特定短语的情感极性,进一步判断整个评论的情感倾向性,该方法首先将文本进行分词和词性标注,并提取出形容词或副词短语,然后使用“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,用以计算整个文本的情感极性。以上都是基于传统机器学习方法的研究。近年来,随着深度学习和神经网络的迅速发展,Teng 等^[3]建立了一个基于上下文情感词典的神经网络模型来进行情感分类,结果表明该模型能进一步提高分类的准确度。Qian 等^[4]提出了一种句子级别注释训练的简单模型,该模型能够捕捉情感词汇、否定词汇和强度词汇在情感表达中的语言角色,在情感分类中取得了不错的效果。最近,随着微博等新兴媒体的迅速发展,越来越多的人开始对微博评论这种短文本进行情感分析,刘志明等^[5]将 SVM 和 IG(Information Gain),以及 TF-IDF(Term Frequency-Inverse Document Frequency)方法作为特征项权重,通过三者结合来对微博进行情感分析。近年来,对情感分类的研究热度一直不减,Songpan 等^[6]提出了一种新的情感分析方法-情感短语模式匹配(Sentiment Phrase Pattern Matching,SPPM),SPPM 与其他算法相比较,准确率、召回率和 F1 值都有较大幅度的提高,该方法可以根据学生的意见来改善教学策略。Chang 等^[7]提出了一种基于扩展特征和动态合并的双通道卷积神经网络的文本情感分类算法,该算法比传统的单通道卷积神经网络算法具有更好的分类效果。

专业评论和非专业评论的评论质量显然是不同的,目前,国内外不少学者已经在研究评论质量检测。评论质量也称为评论的有用性,其主要是在统一的范围内量化评论的质量或衡量评论对用户的有用程度。评论质量的相关研究如下。文献[8-11]采用分类的方法预测评论质量,他们将评论质量检测看作一个二分类问题,即预测评论质量的高或低。之后,Lu 等^[12]提出利用作者的身份和社交网络的背景信息来提高评论的质量检测性能,并且提出了一个通用的框架,通过在基于文本的预测中添加正则化约束来融合社会背景信息。该实验证明,社会背景信息确实能提高评论质量检测的准确性。Zeng 等^[13]定义了一个三分类模型,解决了意见挖掘问题,能够从在线消费者的评论中快速找到有用的评论。Yang 等^[14]使用语义特征训练了一个模型来识别评论的质量,实验结果表明使用语义特征的模型更容易推广到不同类别的产品评论。Meng 等^[15]研究了产品评论质量与销售量之间的关系,结果表明,积极评价和消极评价的总体质量水平会与销售量产生交互影响。Zhang 等^[16]研究评论审稿人的专业知识对评论质量的影响。Siering 等^[17]研究发现评论内容相关信号(即特定评论内容和写作风格)和评论者相关信号(即评论者专业知识和非匿名性)都会影响评论的质量。Yang 等^[18]提出了一种利用产品类别信息的模型,其利用双层回归器来做评论质量检测,实验结果表明,该模型在亚马逊网站的 5 个热门产品类别中,都取得了不错的预测效果。Xiong 等^[19]为了调查同行评论的质量,发现结构特征、Unigrams 和元数据组合在模拟同行评论和产品评论的质量方面是有用的。

本文受到情感分类、评论质量检测的启发,在进行电影评论专业程度分类时也尝试引入 Unigram,Bigram 和词性等特征,同时利用句法的依存关系来构建模型,以获得较好的分类结果。

3 分析方法

影评,就是电影的观看者在观看完电影以后将所感所想以文字语言、语录等形式表现出来,并在社交网站上发表的一种评论、看法、见解等。而在这些影评中,有的是由专业影评家写的影评,他们的评论往往更为客观,可信度更高;有的只是普通观众写的影评,往往带有更多的主观性,参考价值相对更小。因此,对评论进行专业程度分类任务具有一定的研究价值。对电影评论进行专业程度分类任务本质上是一个二分类问题,其目标类别只有专业和非专业两种。本文采用文本分类中性能较好的 LSTM 神经网络模型来实现多特征融合的评论专业程度分类。

我们在处理后的数据集上,随机抽取专业影评人写的电影评论和非专业普通观众写的电影评论各 10 000 条,并与词性(POS)和依存关系(Dep)等特征相结合,利用 LSTM 构建模型进行分类。具体来说,首先用单个词 Unigram 来构建基准模型,然后在单个词 Unigram 信息的模型中逐步添加词性(POS)信息和依存关系(Dep)信息来对电影评论专业程度进行分类以提高准确度。本文之所以选择上述 3 种类型的文本信息来构建模型是因为专业影评人写的电影评论和普通观众写的评论在用词的词性上有很大的不同。普通观众的评论更多地使用形容词,而专业影评人则更多地使用电影名、场景规模、演员演技等名词来进行电影之间的分析比较。同时专业

的评论和非专业的评论在句子结构上也具有明显的差异,专业评论的句法结构更加完整,所以本文在实验中引入了句子依存关系特征。

3.1 特征选择

文本特征被认为是文本表示时可处理的最小单位,特征提取的优劣间接影响分类模型的分类效果。受到传统的情感分类特征选择的启发,为了训练出一个较好的实验模型,实验选取了 3 类文本特征,包括单个词(Unigram)、词性(POS)和依存关系(Dep),通过多特征组合的实验,评估出不同特征对

评论专业程度分类的作用,并找出最优的特征组合。

3.1.1 词性特征

在电影评论中,专业影评家的评论相对客观,主要从电影的演员演技、情节设置、特效效果以及与之前类似电影进行比较来评价,评论中大多出现名词。而普通观众在写评论时则更倾向于表达主观情感,分享自己的观影感受,更多地使用带有强烈情感色彩的形容词,因此,词性在专业评论和非专业评论中具有一定的区分性。表 1 给出了专业评论和非专业评论词性标注后的具体示例。

表 1 词性标注示例

专业评论	非专业评论
superior_JJto_TOWarrior_NNPIn_IEvery_DTway_NN	Interesting_JJ,_,puzzling_JJand_CCsurprising_JJ,_,
Besides_INthe_DTacting_JJand_CCoriginal_JJscreenplay_NN,_,the_DTclassical_JJmusic_NNscore_NNis_VBZanother_DTfavorable_JJcomponent_NNof_INthe_DT crowd-pleasing_JJescapism_NNoffered_VBNby_INthis_DTfilm_NN,_,	Slow_JJbut_CCinteresting_JJ

注:JJ 表示形容词,NN 表示名词,CC 表示连词,DT 表示限定词,VB 表示动词,IN 表示介词

从表 1 词性标注的例子可以看出,在非专业的评论中形容词(JJ)的出现频率更高,更多的形容词如 Interesting, puzzling 等用来直接表达作者的感情倾向,而专业的评论则更多地出现名词(NN),用来比较分析不同电影的特效、配乐等,所以在 LSTM 分类模型中加入词性特征应该是有效的。实验使用 Stanford Parser 工具¹⁾进行词性标注。

3.1.2 依存特征

实验中选取的另一个特征为语义依存关系,即句子中词对之间的二元关系,其中一个称为核心词,另一个为依存词。依存关系反映的是核心词和依存词之间语义上的依赖关系,

更直观地说,依存关系反映的是句子中的“主谓宾、定状补”这些语法成分之间的关系。选择句子的依存关系作为特征是因为专业的评论句法结构会更加规范,依存关系会更加完整。图 1 给出了依赖解析树示例,表 2 给出了实验中专业评论和非专业评论的依存关系示例。

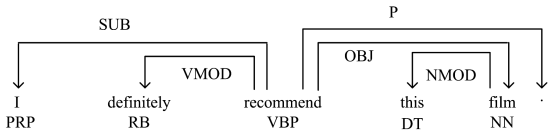


图 1 依赖解析树示例

表 2 依存关系示例

专业评论	非专业评论
Concentrates_victimson_consequences_the_consequencesdire_consequences consequences_Concentratesit_victimshas_victimsfors_victimsthe_victims victims_	Interesting_ puzzling_Interestingand_ puzzlingsurprising_Interestingsurprising_puzzling
A_sequelbloated_sequelaaction_sequelsequel_packedso_packed_packed_ with_ effectsspecial_effectseffects_packedthat_cartoonit_cartoon 's_cartoonvirtually_ cartoona_cartooncartoon_packed	Nice_ graphics_plotand_graphicsenjoyable_ graphicsenjoyable_plotplot_Nice

图 1 中,SUB 为主谓依存关系,VMOD 为动词性修饰关系,OBJ 为宾语关系,NMOD 为复合名词修饰关系。再将依赖解析树关系转化为核心词和依赖词之间的二元关系如表 2 所列,表 3 给出了不同特征示例对比。

表 3 特征示例

特征类型	示例
Unigrams	A fast and twisty thriller.
POS	A_DT fast_JJand_CCtwisty_JJthriller_NN,_,
Dependencies	A_thrillerfast_thrillerand_fasttwisty_fasttwisty_thriller_thriller_

3.2 基于 LSTM 的评论专业程度分析

为了解决 RNN 的梯度消失问题,1997 年,Hochreiter 和 Schmidhuber 等提出了 LSTM 神经单元。当今比较流行的是将传统 RNN 中的隐含层以 LSTM 单元替代,也就是 LSTM 网络。LSTM 单元利用记忆单元(Memory Cell)保留历史信息,历史信息由 3 个门控制,即输入门(Input Gate)、遗忘门(Forget Gate)和输出门(Output Gate)。

LSTM 的更新过程如下:首选决定哪些单元信息需要从单元状态中抛弃,这项任务是由遗忘门(Forget Gate)完成的。以当前层的输入 X_t 和上一层的输出 H_{t-1} 作为输入,在 $t-1$ 时刻的细胞状态输出为:

$$f_t = \sigma(W_f \cdot (H_{t-1}, X_t) + b_f)$$
 (1)

其中, i_t, f_t, o_t 和 C_t 分别表示时刻对应的 3 种门结构和细胞状态。然后需要决定在单元状态中需要存储哪些新信息,这需要输入门(Input Gate)的 sigmoid 层和 tanh 层共同完成。输入门(Input Gate)的 sigmoid 层决定哪些值需要更新,记为 i_t ,由 tanh 层新创建的向量 C_t 将其添加到细胞状态中。将旧的细胞状态乘以 f_t ,用以遗忘信息,其与新的候选信息 $i_t \cdot \tilde{C}_t$ 的和用以更新细胞状态。

$$i_t = \sigma(W_i \cdot (H_{t-1}, X_t) + b_i)$$
 (2)

$$C_t = \tanh(W_c \cdot (H_{t-1}, X_t) + b_c)$$
 (3)

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
 (4)

最后,由输出门决定输出信息,运行一个 sigmoid 层来决

¹⁾ <https://nlp.stanford.edu/software/lex-parser.shtml>

定单元状态中哪些部分需要输出。然后将单元状态输入到 tanh 函数(将值转换成 -1 到 1 之间)中,再乘以输出的 sigmoid 门限值,这样就只输出需要输出的那部分。

$$O_t = (W_o \cdot (H \cdot , X \cdot) + b_o)$$

(5)

$$h_t = O_t \cdot \tanh(C_t)$$

(6)

采用 LSTM 模型避免了传统机器学习模型中的特征工程,并且能取得与当前基于传统机器学习相当甚至更好的性能。LSTM 等深度学习模型可借助 GPU 来训练,大大缩短了训练时间,同时,LSTM 的网络模型在情感分类任务^[20-23]中已经表现了优异的效果,也更容易推广到其他文本分类的任务中去,具有一定的通用性。

3.3 不同融合方法的比较

由前面的分析可知,依存关系和词性是电影评论专业程度分类任务中的有效特征,因此需要将这些特征进行融合,以更准确地进行电影评论专业分类。文中提出了两种融合方法进行比较,第一种融合方法是 LSTM-Concate-Input 融合方法,即在电影评论文本信息的基础上,将依存关系特征和词性特征连接在后面,一起作为 LSTM 模型的输入,以实现不同特征的融合。另一种融合方法是构建 Multi-LSTM 模型,为不同的特征各自构建一个 LSTM 模型,再将 3 个 LSTM 模型拼接在一起,以实现不同特征之间的融合来进行电影评论专业程度分类。实验结果表明,后者的分类效果更好。

3.4 构建 Multi-LSTM 模型

为了提高 LSTM 模型电影评论专业分类的准确度,实验中需要将 POS 词性特征和依存关系 Dep 特征融入 LSTM 模型,我们将不同种类的特征都构建为一个向量,并将每一类特征对应的向量输入 LSTM 模型,学习每一类特征的表示。词特征 $W_1 \cdots W_n$ 、词性特征 $P_1 \cdots P_n$ 和依存特征 $D_1 \cdots D_n$ 经过 LSTM 层分别输出为 $H_{w1} \cdots H_{wn}, H_{p1} \cdots H_{pn}$ 和 $H_{d1} \cdots H_{dn}$ 。在构建特征向量时,我们使用布尔值作为特征的表示。通过构建一个 Multi-LSTM 模型,最终得到了融合不同特征的向量 ∂_n ,提高了电影评论专业程度的分类准确度,Multi-LSTM 模型如图 2 所示。

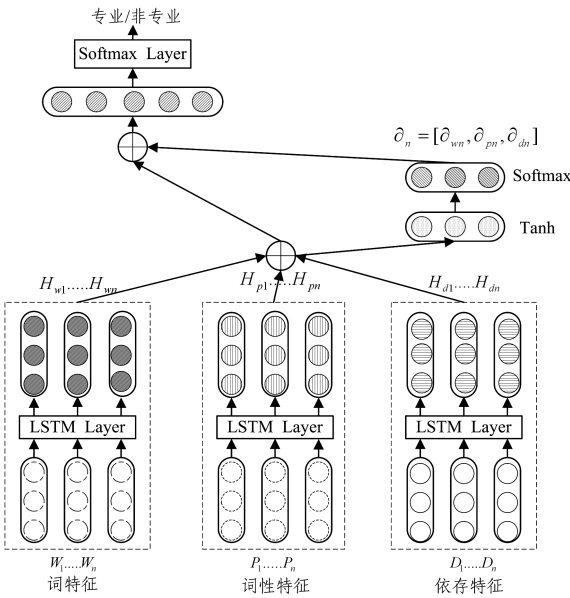


图 2 Multi-LSTM 模型

4 实验

4.1 数据集

本文使用的数据来源于美国著名的影评网站烂番茄网(Rotten Tomatoes)。该网站会跟踪一部电影的所有评论,如果正面评论超过 60% 以上,该部电影就会被标示为“新鲜”(fresh),相反的,若一部电影的正面评价低于 60%,则该作品会被标示为“腐烂”(rotten)。该网站与其他在线评论网站不同,该网站上对于一部电影的评论分为两种,一种是影评专家给出的比较专业的评论,另一种是普通观众写的评论。我们使用爬虫工具抓取其中的专业和非专业评论,经过数据预处理后,以单个词 Unigram 作为基准实验,在逐步添加词性 POS 和依存关系 Dep 特征后,通过分析对比,研究不同特征组合对实验结果的影响。

实验使用的语料包含了 58 927 条专业电影评论和 56 989 条非专业电影评论,本文随机抽取了专业和非专业评论各 10 000 条作为训练集,各 1 000 条作为测试集,表 4 统计了训练语料和测试语料中专业评论和非专业评论的分布。表 5 给出了专业评论和非专业评论的数据样例。

表 4 语料的统计结果

语料	专业语料个数	非专业语料个数	总数
训练语料	10 000	10 000	20 000
测试语料	1 000	1 000	2 000

表 5 数据样例

专业评论	非专业评论
One of the year's most stunningly feeble offerings.	Interesting, puzzling and surprising.
Undoubtedly one of the finest films of the last year.	A little scary for me
Kind of scary return to form for Carpenter.	not my type! too old!

4.2 基准系统

实验过程中,为了分析不同特征值对评论专业程度分类的影响,我们分别构建了 Model_Unigram, Model_Uni_Dep, Model_Uni_Pos 和 Model_Uni_Pos_Dep 4 个模型用于比较不同特征信息对实验结果的影响。其中,Model_Unigram 代表单个词信息构建的模型,Model_Uni_Dep 代表增加了依存关系 Dep 的模型,Model_Uni_Pos 代表了增加了词性 POS 的模型,Model_Uni_Pos_Dep 代表了增加词性 POS 和依存关系 Dep 的模型。此外,在对模型输入特征时,由于每一类的特征都很多,为了防止引入噪声,对实验结果造成影响,我们均选取了出现频率最高的前 10 000 个特征,即选取 Unigram、词性、依存关系的高频特征,对特征值进行了简单的筛选处理。

4.3 实验结果与分析

本文主要进行了 3 组实验来进行评论的专业程度分类。实验 1 主要是为了与 baseline 进行比较,并且比较了不同的特征组合方法的实验结果。实验 2 主要是分析不同特征对实验结果的影响。实验 3 测试了不同数量训练样本下的实验结果。实验 1 选择利用不同特征融合方法构建 LSTM 基准模型,并与 SVM 模型和 CNN 模型进行对比,实验结果如表 6 所列。

表 6 实验结果对比

模型名称	准确率/%
SVM_Unigram	83.05
CNN_Unigram	83.69
LSTM_Unigram	84.64
LSTM_Concate_Uni_Pos_Dep	85.09
LSTM_Multi_Uni_Pos_Dep	88.30

我们发现,两种特征融合模型在加入词性特征 POS 和依存关系特征 Dep 后,实验结果与 baseline 相比有了较大的提高。这也证明在评论专业程度分类中加入依存关系和词性是有效的。在实验中,LSTM 和 CNN 单个词模型的准确度比传统的 SVM 方法更高,这在一定程度上也体现了深度学习的发展潜力。

而我们之所以选择利用评论的单个词信息、依存关系和词性的组合构建模型,而不是利用评论的 Bigram 等特征,是因为我们在各特征模型的对比实验中发现,选取单个词 Unigram、依存关系 Dep 和词性 POS 组合时模型具有更好的性能,实验结果如表 7 所列。

表 7 LSTM 中不同特征组合的实验结果

模型名称	准确率/%
LSTM_Unigram	84.65
LSTM_Multi_Uni_Dep	85.25
LSTM_Multi_Uni_Pos	87.65
LSTM_Multi_Uni_Pos_Dep	88.30

通过表 7 的结果可以看出,在 LSTM 模型中,使用单个词、词性、依存关系特征组合时实验效果达到最好,正确率为 88.30%,这说明加入词性和依存关系对模型准确度的提高是有效的。

在 LSTM 实验中通过对词信息构建的基准模型增加特征值,我们发现在引入依存关系和词性信息之后,实验结果有了显著的提升。这表明了在电影评论的分类中,专业的评论和非专业的评论在用词的词性上具有显著的差别,非专业的评论会更多地使用带有强烈主观色彩的情感词和语气词,而这些情感词大多为形容词,因此,在加入词性特征后会使实验准确度有显著提高。并且,我们也发现添加句子的依存关系对预测模型的提升也很明显,这说明了专业的影评人写的电影评论在句子的结构和用词上更加规范化,因此加入句子中词的依存关系在一定程度上也能提高模型的准确率。图 3 显示了不同数量训练样本下的实验结果,横轴为训练样本的数量,纵轴为实验准确率。

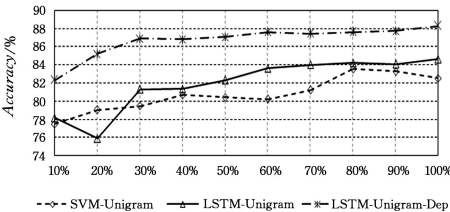


图 3 不同数量训练样本实验

结束语 本文首次利用神经网络模型 LSTM 完成评论专业程度分类的任务,使用了传统的机器学习方法 SVM 和神经网络模型 CNN 作为对比实验,其中 LSTM 方法的准确率要优于 SVM 方法和 CNN 模型。以 LSTM_Unigram 模型为依据,通过实验对比了各个特征值对预测结果的影响,从中我们发现加入评论的词性和依存关系特征确实能

获得较高的分类准确率。随着电影娱乐行业的快速发展以及各大社交平台的流行,如何能快速地在成千上万条评论中将专业评论和非专业评论区分出来显得尤为重要。当然,此类任务也可推广至不同领域的评论分类,例如书评、剧评等,这也是未来的研究方向之一。

同时我们在实验中也发现,专业的电影评论和非专业的电影评论在评论的长度以及使用的句式、标点符号方面也存在明显的差别。专业的评论更倾向于使用陈述句、使用逗号和句号,而非专业的评论在句式上更多使用感叹句,并且感叹号等带有情绪色彩的符号也出现得更多。本文对电影评论的研究还很有限,在特征的选择上也需要进一步完善,以上所述都有待通过未来详尽研究工作来完成。本文使用 LSTM 模型对电影评论进行分类只是一个尝试,期望对相关的需求和应用有所帮助。

参 考 文 献

[1] PANG T B,PANG B,LEE L. Thumbs up? Sentiment Classification using Machine Learning[C]// Empirical Methods in Natural Language Processing. 2002;79-86.

[2] TURNEY P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews[C]// Proceedings of ACL-02. 2002;79-86.

[3] TENG Z,VO D T,ZHANG Y. Context-Sensitive Lexicon Features for Neural Sentiment Analysis[C]// Conference on Empirical Methods in Natural Language Processing. 2016; 1629-1638.

[4] QIAN Q,HUANG M,LEI J,et al. Linguistically Regularized LSTMs for Sentiment Classification [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017;1679-1689.

[5] 刘志明,刘鲁. 基于机器学习的中文微博情感分类实证研究[J]. 计算机工程与应用,2012,48(1):1-4.

[6] PONG-INWONG C,SONGPAN W. Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SP-PM) based on association mining[J]. International Journal of Machine Learning & Cybernetics,2018,6;1-10.

[7] CHANG S,JUNZHONG J I. Text Sentiment Classification Algorithm Based on Double Channel Convolutional Neural Network [J]. Pattern Recognition & Artificial Intelligence, 2018, 31(2):158-166.

[8] LIU J. Low-quality product review detection in opinion summarization[C]// Proc. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2007;334-342.

[9] O'MAHONY M P,SMYTH B. Learning to recommend helpful hotel reviews[C]// ACM Conference on Recommender Systems. ACM,2009;305-308.

[10] WEIMER M,GUREVYCH I. Automatically assessing the post quality in online discussions on software[C]// Meeting of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics. 2007;125-128.

[11] SIERSDORFER S,CHELARU S,NEJD L W,et al. How useful are your comments?: analyzing and predicting you tube comments and comment ratings[C]// International Conference on

World Wide Web. ACM,2010:891-900.

[12] LU Y,TSAPARAS P,NTOULAS A,et al. Exploiting social context for review quality prediction[C]// International Conference on World Wide Web. ACM,2010:691-700.

[13] ZENG Y C,KU T,CHEN L P,et al. Modeling the Helpful Opinion Mining of Online Consumer Reviews as a Classification Problem[J]. 中文计算语言学期刊,2014,19(2):17-31.

[14] YANG Y,QIU M,YAN Y,et al. Semantic analysis and helpfulness prediction of text for online product reviews[C]// IEEE International Conference on Parallel and Distributed Systems. IEEE,2015:552-559.

[15] MENG Y,WANG H,ZHENG L. Impact of online word-of-mouth on sales; the moderating role of product review quality [J]. New Review in Hypermedia & Multimedia,2018,11:1-27.

[16] ZHANG J,LIN Y,HUANG T,et al. Evaluating Review’s Quality Based on Review Content and Reviewer’s Expertise[C]// International Conference on Database Systems for Advanced Applications. Springer,Cham,2018:36-47.

[17] SIERING M,MUNTERMANN J,RAJAGOPALAN B. Explaining and predicting online review helpfulness; The role of content and reviewer-related signals [J]. Decision Support Systems,2018,108:1-12.

[18] YANG Y,CHEN C,BAO F S. Aspect-Based Helpfulness Prediction for Online Product Reviews[C]// International Conference on TOOLS with Artificial Intelligence. IEEE,2017:836-843.

[19] XIONG W,LITMAN D. Automatically predicting peer-review helpfulness[C]// Meeting of the Association for Computational Linguistics; Human Language Technologies; Short Papers. Association for Computational Linguistics,2011:502-507.

[20] HUANG Q,CHEN R,ZHENG X,et al. Deep Sentiment Representation Based on CNN and LSTM[C]// International Conference on Green Informatics. IEEE,2017:30-33.

[21] LU C,HUANG H,JIAN P,et al. A P-LSTM Neural Network for Sentiment Classification [C]// In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2017:524-533.

[22] WANG Y,HUANG M,ZHU X,et al. Attention-based LSTM for Aspect-level Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2017:606-615.

[23] SHUANG K,REN X,CHEN J,et al. Combining Word Order and CNN-LSTM for Sentence Sentiment Classification[C]// International Conference. 2017:17-21.

(上接第 73 页)

定义大多建立在什么样的物理、天文条件可满足液态水存在这一条件,因此,并非所有潜在宜居的行星都可真实宜居。综上,建立在宽松宜居定义下的本研究的探索成果,以及已有相关研究成果,均不宜被过度解读。

总之,目前符合严格宜居定义的数据极少,这需要更好的机器学习算法出现,以实现更小数据集的精确近似学习;由于宇宙天体的固有自然性质,宜居行星数量远小于非宜居行星数量,这需要更优的机器学习算法,以应对数据不平衡问题。事实上,行星的外部特征包括轨道参数、质量、半径、密度等,可通过系外行星的发现及后续观测获得^[4](视向速度法、掩星法、微引力透镜法等方法则可用于发现系外行星),因此,人类掌握的严格潜在宜居行星数量(包括其各项物理、天文数据)近年来一直在增长,一旦不久的将来这个数量超过 200~300 个,机器学习的分类能力与预测能力将在这一天在文学应用领域发挥更大甚至有可能是关键性的作用。届时,机器学习可望与基于天文观测的方式有效形成互补,共同为探索宇宙、寻找外星生命做出贡献。

参 考 文 献

[1] BORA K,SAHA S,AGRAWAL S,et al. CD-HPF;New habitability score via data analytic modeling[J]. Astronomy & Computing,2016,17:129-143.

[2] SAHA S,BASAK S,SAFONOVA M,et al. Theoretical validation of potential habitability via analytical and boosted tree methods;An optimistic study on recently discovered exoplanets [J]. Astronomy & Computing,2018,23:141-150.

[3] HORA K. Classifying Exoplanets as Potentially Habitable Using Machine Learning[M]// ICT Based Innovations,2018:203-212.

[4] 田丰,胡雄,吴季. 系外行星大气与宜居系外行星研究进展及发展趋势[J]. 空间科学学报,2016,36(6):815-827.

[5] FRIEDMAN J H. Greedy function approximation;a gradient-boosting machine [J]. Annal of Statistics,2001(29):1189-1232.

[6] PENG X,SETLUR S,GOVINDARAJU V,et al. Using a boosted tree classifier for text segmentation in hand-annotated documents[J]. Pattern Recognition Letters,2008,29(1):943-950.

[7] DEMIRKIR,C,SANKUR B. Face detection using boosted tree classifier stages[C]// Signal Processing and Communications Applications Conference. IEEE,2004:575-578.

[8] PARAG T,ELGAMMAL A M. Unsupervised Learning of Boosted Tree Classifier Using Graph Cuts for Hand Pose Recognition[C]// British Machine Vision Conference 2006. Edinburgh,UK,DBLP,2013:1259-1268.

[9] WU B,NEVATIA R. Cluster Boosted Tree Classifier for Multi-View,Multi-Pose Object Detection[C]// IEEE 11th International Conference on Computer Vision. IEEE,2007:1-8.

[10] DAY M. Emotion recognition with boosted tree classifiers[C]// ACM on International Conference on Multimodal Interaction. ACM,2013:531-534.

[11] 1. 11. Ensemble methods —scikit-learn 0. 19. 1 documentation [EB/OL]. <http://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting>.

[12] Graphlabcreate™. Fast,Scalable Machine Learning Modeling in Python [EB/OL]. <https://turi.com/>.

[13] The Habitable Exoplanets Catalog-Planetary Habitability Laboratory @ UPR Arecibo [EB/OL]. <http://phl.upr.edu/projects/habitable-exoplanets-catalog>.

[14] Extrasolar Planet’s Catalogue produced by Kyoto University [EB/OL]. www.exoplanetkyoto.org.