

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2015.041

基于自动编码特征的汉语解释性意见句识别

贺宇 潘达 付国宏[†]

黑龙江大学计算机科学技术学院, 哈尔滨 150080; [†] 通信作者, E-mail: ghfu@hotmail.com

摘要 提出一种基于自动编码特征的汉语解释性意见句识别的分类方法。首先从汽车和手机两个领域的产品评论中构造一个解释性意见语料库。然后, 采用分类的方法进行解释性意见句识别。特别地, 采用自动编码技术表示和学习解释性意见句分类的词嵌入特征。最后, 在支持向量机框架下通过实验优选解释性词语嵌入维度, 并同一些传统特征表示方法进行比较。实验结果表明, 与传统的卡方、信息增益和 TF-IDF 及其组合方法相比, 自动编码特征的引入能有效提升汉语解释性意见句识别性能。

关键词 意见挖掘; 解释性意见句识别; 自动编码

中图分类号 TP391

Chinese Explanatory Opinionated Sentence Recognition Based on Auto-Encoding Features

HE Yu, PAN Da, FU Guohong[†]

School of Computer Science and Technology, Heilongjiang University, Harbin 150080; [†] Corresponding Author,
E-mail: ghfu@hotmail.com

Abstract An auto-encoding feature based classification method to Chinese explanatory opinionated sentence recognition was presented. An explanatory opinion corpus is built firstly from online product reviews in cellphone and car domains. Then, word embeddings are learned from product reviews using the auto-encoding technique. Finally, the learned word embeddings are used as features for explanatory opinionated sentence classification under the framework of supported vector machines. Experimental results show that word embeddings are more effective than some traditional representations of features like chi-square, TF-IDF and information gains for explanatory opinionated sentence classification.

Key words opinion mining; explanatory opinionated sentence recognition; auto-encoding

随着 Web2.0 技术的快速发展, 网络上出现大量的用户生成的意见文本。意见挖掘 (opinion mining) 也因此成为自然语言处理 (natural language processing, NLP) 相关领域的热门研究, 并在情感分类、意见抽取和意见摘要等方面取得很大进展^[1]。但是, 目前的意见挖掘研究大多集中在不同语言层面的情感倾向性识别和意见信息抽取方面, 很少有人关注这些意见产生背后的原因, 如对情感倾向的解释或说明等^[2]。比如意见句“iPhone 的屏幕很好”,

虽然给出 iPhone 属性“屏幕”的正面评价, 但没有明确说明“屏幕”好的原因是因为分辨率高、屏幕大还是材质好, 这样很容易对用户行为产生误导, 可能有人想买大屏幕手机, 而选择了 iPhone。因此, 用户还需要更多的相应产品评价, 以便了解相关情感倾向背后的原因 (即解释性意见信息), 进而作出正确的决策。解释性意见句识别正是为了适应这一需求而产生的一个新的研究问题, 其主要任务是从意见文本中识别蕴含解释性意见信息的意见句^[2]。显

国家自然科学基金(61170148, 60973081), 黑龙江省人社厅留学人员科技活动项目, 哈尔滨市科技创新人才研究专项(2009RFLXG007)资助

收稿日期: 2014-06-28; 修回日期: 2014-11-02; 网络出版时间: 2014-11-28 11:04

然,解释性意见句识别在意见摘要等领域具有十分重要的意义和应用前景^[2]。

目前,解释意见句子识别相关研究还处于起步阶段。Kim 等^[2]在 SIGIR'13 发表的论文中首次提出解释性意见句抽取问题,但他们针对的是英语产品评论文本。而有关汉语解释性意见句识别的研究还未见报道。

基于以上原因,本文在前人研究基础上,面向产品评价领域着重研究汉语解释性意见句识别问题。特别地,本文将汉语解释性意见句识别看做一个分类问题,同时采用自动编码技术表示和学习解释性意见句分类特征,并融入基于支持向量机(supported vector machines, SVMs)^[4]的分类器。与基于词袋模型(bags of words)的传统特征选择和表示方法相比,基于自动编码技术的词向量特征不仅可以发掘词语潜在的语义,可以降低参数维度,还可以避免人工选择特征中的问题。在手机和汽车两个领域上的对比实验表明:基于自动编码特征的解释性意见句分类方法明显优于卡方(chi-square, χ^2)、TF-IDF 和信息增益(information gain, IG)等基于传统特征表示的分类方法,而且特征选择过程大大简化。

1 相关工作

目前,与本研究相关的工作主要体现在 3 个方面,即意见摘要^[1-2,5-8]、解释性意见抽取^[2]和自动编码技术^[3,9-15]。

意见摘要的主要任务是从一组相关的意见文本中发掘人们关于某个评价对象的意见信息,并以某种自然的方式把这些散落在不同文档中的多样的、冗余的意见信息聚集或融合,形成一个简明的摘要。意见摘要研究目前主要涉及情感分类、意见抽取、意见聚集和意见融合等任务。根据摘要表示形式的不同,意见摘要可分为结构化的意见摘要^[5-6]和非结构化的文本意见摘要^[7-8]。虽然意见摘要,特别是基于属性的结构化意见摘要系统取得重大发展,甚至进入实用化阶段^[1,5-6],但是现有的摘要系统大多只能提供人们对评价对象属性的正面或负面情感倾向信息,无法提供对用户决策更有用的关于这些情感倾向产生原因的解释性信息。

解释性意见抽取是与本文研究最直接相关的工作之一。为了弥补目前意见摘要系统无法处理解释性意见信息这一缺陷, Kim 等^[2]首次提出采用无监

督抽取方法研究解释性意见句子抽取问题。他采用改进的 BM25 模型和基于概率的模型,对英文意见文本中的解释性意见句进行打分,并按分值大小排序。实验结果表明他们的方法优于经典的 LexRank 方法。但解释性意见句子识别相关研究还处于起步阶段,迄今尚未见汉语相关研究的报道。这也是本文研究的主要目的之一。与 Kim 等^[2]方法不同,本文将汉语解释性意见句识别看做一个二元分类问题,并试图把自动编码特征融入 SVMs 来解决这一问题。

与本文研究另一个相关的工作是自动编码技术(auto-encoding)^[9]。自动编码是深度学习中广泛采用的一种无监督特征学习技术,其基本思想是一种将高维的输入信息转换为一种低维表示形式的神经网络模型。自动编码技术已经在图像分类、语音识别等领域得到广泛应用,并取得巨大成功^[10-11]。近年来,自动编码技术开始引入 NLP 领域^[12-15]。在 NLP 研究中,自动编码特征通常通过神经网络语言模型(neural networks language model, NNLM)^[12-14],将词表示为一个连续的 n 维向量的形式,即词向量(distributed word representations 或者 word embeddings)。文献[15]的实验结果表明,采用词向量特征在词性标注(part-of-speech tagging, POS)、语块分析(chunking)和命名实体识别(named entity recognition, NER)等多个 NLP 任务中均可取得目前该任务的最好性能。最近, Mikolov 等^[3]通过不同的词表示形式和不同的训练模型的对比实验证明,采用循环神经网络语言模型(Recurrent Neural Network Language Model, RNNLM) 训练的词向量表示可以更好地获取潜在的语法和语义信息。

本研究采用自动编码技术,特别是 RNNLM 来学习意见句中词语的词向量表示形式,并同 SVMs 融合进行解释性意见句分类。我们认为,自动编码特征可以发掘解释性意见句分类潜在的有用的语义信息,从而有利于分类性能的提高。因此,本文研究的另一个目的是通过与传统的分类特征选择和表示方法的实验对比来验证自动编码特征的有效性。

2 数据准备

2.1 解释性意见句与非解释性意见句

定义 1 解释性意见句: 如果一个意见句子不仅给出意见持有者对评价对象或属性的正面或负面的情感倾向性,还同时对意见持有者产生这种情

感倾向的原因进行了解释或说明,那么该意见句子称为解释性意见句;相反,如果一个意见句只给出意见持有者关于评价对象或属性的情感倾向性,而没有解释该情感倾向背后的原因,则该意见句为非解释性意见句。

从定义 1 可以看出,虽然解释性意见句识别与传统的意见句识别类似,都是识别文本中的表示主观性的带有情感倾向描述的句子;但不同的是解释性意见句识别还需要进一步判断该句是否对所表示的情感倾向进行了合理的解释。

例 1 (a) 这部手机太棒了!我最喜欢它的屏幕,赞一个!

(b) 这个手机的屏幕分辨率很高,看不到任何像素点,我非常喜欢!

例 1 中两个句子均是对某个品牌手机的屏幕的正面评价。但是句子(a)是一个非解释性意见句,它并没有明确说明意见发出者喜欢“屏幕”的原因,可能是“屏幕大”,或者是“屏幕材质好”等。相反,句子(b)不仅说明意见发出者喜欢屏幕的原因(“分辨率高”),而且进一步解释了“分辨率高”到了什么样的程度(“看不到任何像素点”)。因而,句子(b)属于比较强的解释性意见句。显然,相对于解释性意见句,非解释性意见句往往不能给网络其他用户的购买决策带来建设性的帮助。

定义 2 强解释性意见句:解释性意见句可以进一步分为强解释性意见句和弱解释性意见句。给定一个解释性意见句,如果它不仅有所表示的情感倾向的解释性内容,而且在情感倾向性程度上有进一步的说明,则该意见句是强解释性的。否则,该意见句为弱解释性意见句。

例 2 (a) 坐在这车的后排,感觉很宽敞。

(b) 这车的后排空间很宽敞,我 180 cm 的个子一点都不拥挤。

如例 2 所示,这两个来自汽车评论的意见句虽然均表示汽车的后排空间比较宽敞,但“宽敞”是一种主观感受,具体的感觉因人而异。句子(a)只是笼统地说明汽车后排感觉宽敞,但没有具体说明宽敞程度,因而解释性信息较弱。而句子(b)则在明确后排空间很宽敞的基础上,还给出宽敞的程度。因此,句子(b)蕴含的解释性信息较强,对其他用户决策更具有参考价值。但是,强解释性和弱解释性在概念

外延上的界限比较模糊,相应的强弱解释性意见句识别问题也更具挑战性。

2.2 解释性意见标注语料库构建

为了训练解释性意见分类词向量特征和分类模型,我们从手机和汽车领域的在线评价构建了两个解释性意见标注语料库。其中,手机评价原始语料来自数据堂网络^①,约有 25000 句;汽车评价是我们从网络上收集的,约有 10000 句。我们定义 3 种解释性意见类别,即非解释性句、弱解释性意见句和强解释性意见句,分别用 0, 1 和 2 表示。为了保证数据的可靠性,防止偶然因素影响实验结果,采取两人同时对语料进行人工标注,并利用 Kappa 值^[16]来评价人工标注的一致性。Kappa 检验计算公式为

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \quad (1)$$

其中, $\Pr(a)$ 是评分员间相对观察到的吻合; $\Pr(e)$ 是偶然吻合的假想概率,是用观察到的数据计算出的每个观察者随机选择各个种类的概率。如果评分员完全吻合,则 $k=1$; 如果评分员间除了期望的偶然发生的吻合外没有吻合,则 $k=0$ 。一般情况下, $k>0.75$ 说明语料标注已经取得相当满意的一致程度;当 $k<0.4$ 时,说明语料标注一致程度不够理想,也说明所研究的问题具有很大难度。

计算我们标注的语料,如果只区分是否解释性意见句, $k=0.683$; 当考虑解释性的强弱,相应的 $k=0.604$ 。可见,相对于解释性与非解释性意见句,强/弱解释性意见句的区分更加困难。

如表 1 所示,采用上述标准,将原始语料中一致性较好的评价句作为最终实验的数据,手机领域 5670 句,汽车领域 2340 句。

表 1 解释性意见标注语料库中各类别意见句的分布情况
Table 1 Distributions of different types of opinion sentences in the explanatory opinion annotated corpora

类别	手机语料		汽车语料	
	语料规模	比例/%	语料规模	比例/%
2-强解释性	2060	36.4	1407	60.1
1-弱解释性	2158	38.0	555	23.7
0-非解释性	1452	25.6	378	16.2
总数	5670	100	2340	100

① 数据堂. <http://www.datatang.com/data/44156>

3 方法

3.1 任务描述

本文将解释性意见句识别看成是一个二分类问题。考虑到强解释性和弱解释性界定的困难,在本文中暂不区分解释性的强弱。因此,给定一个含有 m 个词语的意见句 $S=w_1w_2\dots w_m$, 解释性意见句子识别的任务就是根据句子内部特点判断该句子是解释性的还是非解释性的。为了发掘输入的意见句 S 蕴含的潜在语义,采用词向量作为解释性意见句分类特征,并采用 SVMs 分类器完成解释性意见句分类任务。如图 1 所示,本文系统主要包含以下 3 个模块。

1) 词向量学习。为了获取解释性意见句分类的词向量,我们采用循环神经网络语言模型^[3,14]。该方法可以无指导地学习词向量。显然,语料规模越大,所获取的词向量特征表达的语义越准确。因此,我们在构建自动编码的词向量库时,采用网络中大规模的未标注产品语料(约 100 万句)作为词向量的训练语料。通过自动编码后,意见句子中的每个词都表示为一种 n 维向量的形式。

2) 基于词向量的解释性意见句分类特征训练。研究表明,词向量的每一维都蕴含含有隐含语义,而且在同一文本训练的词向量中,每一维所表达的隐含语义一般是相同的^[3,14]。基于这一点,需要进一步将意见句中每个词的词向量合成,从而将句子也表示为 n 维向量的形式。我们用这种含有语义的向量形式,替换了传统文本分类广泛使用的向量空间模型(vector space models, VSMs),并进一步融合 SVMs 训练方法,从 3.2 节的解释性意见标注语料中构建基于词向量的解释性意见句分类模型。

3) 解释意见句分类。采用 SVMs^[4]作为分类框架。通过在 n 维特征空间中找到间隔最大的超平面作为判别模型,将待识别的意见句经过分词、转换及合成,同样映射到该空间进行判别是否为解释性意见句,从而完成分类的过程。

下面重点介绍词向量学习和解释性意见句分类特征训练两个模块。

3.2 词向量学习

为了获取解释性意见句识别的语义信息词向量,同时考虑到参数空间和参数学习效率,本文采用词向量表示解释性/非解释性意见句分类特征,同时采用循环神经网络语言模型 RNNLM^[3,14]从意见文本中学习相应的词向量。

同一般的神经网络语言模型(neural network language model, NNLM)^[12]类似,RNNLM 使用一个三层的神经网络来构建语言模型<输入层,隐含层,输出层>,并通过输入前 $n-1$ 个词的向量形式($C(w_{k-n+1}, w_{k-n+2}, \dots, w_{k-1})$)来预测下一个词 w_k 。但不同的是,在构建 RNNLM 时需要将上一个隐含层参数作为下一次计算的输入,以进一步计算下一个隐含层参数。因此,RNNLM 语言模型可以很好地利用上下文中所隐含的信息。

实际上,词向量的每一维都包含有丰富的上下文信息,可以很好地表现出该词在当前文本中的语义规则。从前面的例子分析可以看出,解释性意见句并不能通过简单的关键词来进行识别,很大程度上需要在语义层次进行判断。因此,我们将意见句中的词表示为具有隐含语义的词向量,并以此作为解释性意见句分类特征。为了保证学习得到的向量的有效性,我们还进行了部分规则性测试,发现在

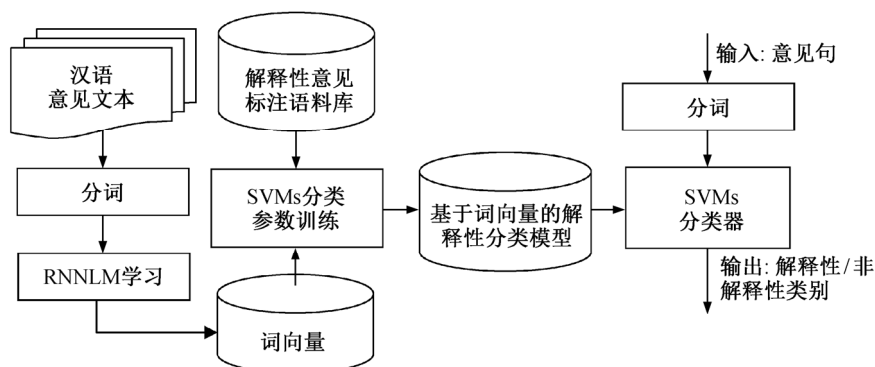


图 1 基于自动编码技术的解释性意见句分类系统框图

Fig. 1 Overall diagram of the explanatory sentence classification system based on word embeddings

意见文本上学习得到的向量同样保证了语义的规则性^[3]，如“安全带”-“安全性”+“舒适性”=“座椅”，“apple”-“ios”+“安卓”=“google”等。例 3 给出从语料中学习到的几个典型词语的词向量。

例 3 屏幕 = [0.267494 0.008962 0.104337
0.341712 -0.230426...]

软件 = [0.110570 0.113541 -0.150757 -
0.075154 0.068729...]

游戏 = [0.241467 0.083909 -0.091359
0.093864 -0.124306...]

3.4 解释性意见句分类特征

在基于向量空间模型(vector space model, VSM)的传统文本分类系统中，待分类的文本的每个词语被看做是一个潜在的离散的特征，而分类文档可表示为一组特征向量。特征值一般为布尔权重或者词频。为了减少分类特征空间，同时去掉噪声，往往会采用 χ^2 或者 TF-IDF 等方法来进行特征选择。但是，传统的基于主题的主题分类特征选择方法无法表示句子中每个词语隐含的语义信息，而这些信息对于正确的解释性意见句分类十分重要。

为了弥补这一点，在学习得到的词向量基础上，通过将句子中词语的词向量融合，得到隐含语义的句子向量表示。在此基础上，采用 SVMs 训练方法从解释性意见标注语料库中得到相应的解释性意见分类特征。

假设 $S=w_1w_2\dots w_m$ 表示一个含有 m 个词语的意见句， $C(w_i)$ 表示 S 中第 i ($1 \leq i \leq m$) 个词 w_i 的 N 维向量形式，则句子 S 的 N 维向量

$$C(S) = C(w_1) + C(w_2) + \dots + C(w_m), \quad (2)$$

通过式(2)的词向量的合成，可将每个意见句 S 也映射到一个 N 维特征空间。这样，不仅可以将句子表示为一个语义句子集合，而且比传统的 VSM 模型具有更低的维度，从而加快训练速度。不同于情感分析，这里我们假设解释性信息与词的位置没有关系，直觉上也确实如此。图 2 给出从词向量到句子向量合成的示例，其中，实线箭头即句子“屏幕分辨率高”根据词向量合成后的结果。

4 实验结果与分析

为了评价本文方法，我们面向产品评价构建一

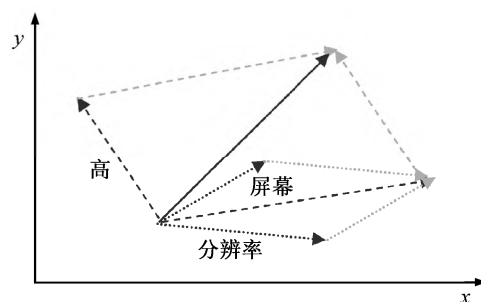


图 2 由词向量合成句子向量

Fig. 2 From word embeddings to sentence embeddings

个汉语解释意见分类系统，并采用 3.2 节所构建的语料对系统进行测试。

4.1 实验设置

实验数据来自表 1 的解释性意见标注语料，并按 3:1 的比例将表 1 的两个语料分为训练和测试两部分，如表 2 所示。

为了评价解释性意见句分类性能，采用精确度 (precision, P)、召回率 (recall, R) 和 F 值 (F -score, F) 3 个指标。此外，实验系统的词向量采用 Google 的开源工具包 Word2Vec^① 网络中真实的产品评价语料 (约 100 万句) 中学习获得，并采用 libsvm^② 作为识别解释性意见句的分类器。

本文主要从两个方面进行验证：1) 首先通过不同维度的词向量的分类结果对比，验证词向量大小对解释性意见识别性能的影响，从而确定解释性分类最佳的词向量维度；2) 前文分析表明词向量可以发掘并准确表示意见句所蕴含的潜在语义，因而有利于提高解释性意见句的识别性能，为了验证这一点，我们将词向量特征与其他常规的特征选择方法进行比较，以进一步证明本文所采用的词向量特征的有效性。

表 2 训练和测试语料库基本统计信息
Table 2 Basic statistics of the training and test corpora

类别	手机产品评价		汽车产品评价	
	训练语料	测试语料	训练语料	测试语料
解释性	3171	1050	1473	490
非解释性	1089	360	282	95
总数	4260	1410	1755	585

① word2vec. <https://code.google.com/p/word2vec/>

② libsvm. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

4.2 实验结果

4.2.1 词向量维度的选择

词向量维度选择是自动编码方法的一个重要问题。前人研究表明,一般将词语表示为 50 或者 100 维的向量形式效果较好^[3,14-15]。然而针对解释性意见识别任务,词向量维度为多大时,才能很好地表达解释性意见,还需要进一步的实验验证。为此,我们通过分别将每个词语表示为 10 维、50 维、100 维、200 维及 500 维向量,并通过考察相应的解释性意见句识别结果来确定面向解释性意见句识别的词向量最佳的维度,实验结果如表 3 所示。

从表 3 可以看出:随着词向量维度从 10 维逐渐增加到 500 维,系统在手机和汽车两个领域上,精确率 P 和 F 值都有所增加,而召回率 R 有所降低,并且训练时间也会增加。当维度继续增加时,解释性意见分类性能变化趋于平缓,甚至有所下降,其原因可能是当维度过高时,产生过拟合现象。我们在开发集的参数交叉验证实验中也出现了类似现象,即随着特征维度的增多,开发集的分类精确率不断上升,但是对测试集中数据的泛化能力却有所降低。我们在效率与性能方面进行折中考虑,最终采用基于 200 维的词向量表示形式,因为其不仅具有很好的泛化能力,而且在效率方面也不会受影响。

4.2.2 与常规特征选择方法的比较

为了进一步验证词向量方法的有效性,选取卡方(χ^2)、信息增益(IG)和 TF-IDF 值等常见的特征选择方法及他们之间的互相组合作为实验的基线系统,同本文提出的词向量方法(即 WE_200)进行对比。实验结果如表 4 所示,其中,粗体数字为该评测指标的最好结果。

表 3 手机和汽车评论领域不同词向量维度的解释性意见句分类结果

Table 3 Results of explanatory opinion sentence recognition for word embeddings with different dimensions

维度	手机评论领域			汽车评论领域		
	P	R	F	P	R	F
10	0.792	0.936	0.858	0.867	0.961	0.912
50	0.813	0.944	0.874	0.892	0.943	0.917
100	0.832	0.926	0.877	0.888	0.951	0.918
200	0.842	0.925	0.881	0.889	0.951	0.919
500	0.846	0.923	0.883	0.888	0.959	0.922

表 4 词向量方法与常规特征选择方法的对比实验结果
Table 4 Results for comparing word embeddings to some normal feature selection methods

特征选择方法	手机评论领域				汽车评论领域			
	维度	P	R	F	维度	P	R	F
χ^2	2845	0.873	0.863	0.868	2815	0.897	0.886	0.891
IG	1874	0.668	0.855	0.750	3108	0.843	0.908	0.874
TF-IDF	3000	0.860	0.852	0.856	3000	0.894	0.873	0.883
IG+ χ^2	900	0.752	0.926	0.830	992	0.839	0.939	0.886
IG+TF-IDF	883	0.756	0.863	0.806	1593	0.840	0.943	0.888
χ^2 +TF-IDF	2464	0.878	0.861	0.869	2154	0.895	0.865	0.880
WE_200	200	0.842	0.925	0.881	200	0.889	0.951	0.919

从表 4 的对比实验结果可以看出: 1) 相对于传统的分类特征选择方法,词向量方法的参数空间虽然大大减少,但解释性意见句识别的总体性能有所提高; 2) 虽然将传统特征融合后,在特征空间减少的情况下性能并没有下降,但是无论是性能还是维度,距离自动编码特征还是有一定差距。造成这种错误的原因,我们认为有以下几个方面。1) 特征模糊。比如“屏幕清晰”和“屏幕好”,前一个是解释性句,而后一个是非解释性句,而通过传统方法选择的特征,基本都有“屏幕”这个特征,因而容易造成错误分类。2) 语义模糊。一些比较句,既可以认为是解释性句,也可认为是非解释性句。比如“电池能比过 iphone 吗?!",既可认为其提供了解释信息(电池没 iphone 耐用),也可认为没有任何信息(假设大家都不知道 iphone),因此这种问题给任务本身带来难度。3) 泛化能力差。传统的特征彼此是孤立的,“屏幕”与“触摸屏”的差别和“手机”与“汽车”的差别一样,这显然是不合理的。因此需要发掘特征之间内在的联系。相反地,词向量特征方法没有受语料不平衡的影响,对于解释性意见句有很好的识别能力,其原因是,在词向量特征方法的系统中,所有的句子都有同样的维度,每一维度都有数值,特征非常稠密,在降低维度的同时还保留了隐含语义信息,因而对不同的意见句均可取得较好的分类效果。

5 结论及未来研究展望

在本文研究中,我们将汉语解释性意见句识别形式化为一个二分类问题,并引入词向量特征,在 SVMs 框架下实现了一个面向产品评价的解释性意

见句识别系统。同时, 通过在手机和汽车两个领域上的实验考察了词向量维度对解释性意见分类的影响, 并与常规的文本分类特征选择方法进行对比。实验结果表明, 当词向量维度为 500 时, 系统可以取得最佳的 F 值, 在手机和汽车领域上的解释性意见句识别的 F 值分别可达 88.2% 和 92.2%。另外, 本文提出的基于词向量的解释性意见分类方法均优于相应的传统的特征选择方法及相应的特征组合, 包括卡方、信息增益和 TF-IDF 等。

在未来的工作中, 我们将继续探讨如何将识别出来的解释性意见句, 根据其强弱进行排序。另外, 我们的实验结果显示, 词向量特征不需要过多的人工参与, 即不需要经过去除停用词、特征选择等过程, 其每一维所代表的隐含语义, 可以很准确地理解句子与类别的关系。因此, 我们计划将词向量方法拓展到其他二分类甚至多分类任务中, 以进一步探讨词向量特征在不同任务中的应用效果。

参考文献

- [1] B. Liu. Sentiment analysis and subjectivity // The Handbook of Natural Language Processing. 2nd ed. New York: Chapman & Hall, 2010: 627–666
- [2] Kim H D, Castellanos M G, Hsu M, et al. Ranking explanatory sentences for opinion summarization // Proceedings of the 36th annual international ACM SIGIR conference (SIGIR). Dublin, 2013: 1069–1072
- [3] Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations // Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). Atlanta, 2013: 746–751
- [4] Chang C C, Lin C J. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2011, 27(2): 1–27
- [5] Hu M, Liu B. Mining and summarizing customer reviews // Proceedings of the 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD). Seattle, 2004: 168–177
- [6] Titov I, McDonald R. A joint model of text and aspect ratings for sentiment summarization // Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT). Columbus, 2008: 308–316
- [7] Ganesan K, Zhai C X, Han J. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions // Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, 2010: 340–348
- [8] Paul M J, Zhai C X, Girju R. Summarizing contrastive view point in opinionated text // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). MIT Stata Center, 2010: 66–76
- [9] Coates A, Ng A Y, Lee H. An analysis of single-layer networks in unsupervised feature learning // Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, 2011: 215–223
- [10] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders // Proceedings of the 26th Annual International Conference on Machine Learning (ICML). Montreal, 2008: 1096–1103
- [11] Le Q V. Building high-level features using large scale unsupervised learning // Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Vancouver, 2013: 8595–8598
- [12] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model. Journal of Machine Learning Research, 2003, 3: 1137–1155
- [13] Mnih A, Hinton G. A scalable hierarchical distributed language model // Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS). Vancouver, 2008: 1081–1088
- [14] Mikolov T. Statistical language models based on neural networks [D]. Brno: Brno University of Technology, 2012
- [15] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. The Journal of Machine Learning Research, 2011, 12: 2493–2537
- [16] Carletta J. Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 1996, 22(2): 249–254