

doi: 10.12052/gdutxb.170036

面向汽车评论的细粒度情感分析方法研究

陈炳丰¹, 郝志峰^{1,2}, 蔡瑞初¹, 温 雯¹, 王丽娟¹, 黄 浩¹, 蔡晓凤¹

(1. 广东工业大学 计算机学院, 广东 广州 510006; 2. 佛山科学技术学院 数学与大数据学院, 广东 佛山 528000)

摘要: 情感分析方法能够在海量的汽车评论信息中挖掘出有价值的信息, 在汽车产品设计、品牌营销等方面具有较大的应用价值. 针对汽车评论分析的细粒度分析要求, 本文提出了基于实体的细粒度情感分析方法. 首先, 对汽车评论数据进行文本细粒度处理, 然后采用Linear-chain CRF模型对评论数据进行情感实体识别和情感倾向分类; 再对Linear-chain CRF模型进行改进, 提出了一种构造双层结构的CRF模型的方法, 解决2个任务间的关联问题. 实验结果表明, 双层结构CRF模型的情感分析效果优于Linear-chain CRF模型, 能够满足汽车评论在情感实体识别与情感倾向分类的需求.

关键词: 汽车评论; 情感分析; 情感词典; 细粒度; 条件随机场

中图分类号: TP181 文献标志码: A 文章编号: 1007-7162(2017)03-0008-07

A Fine-grained Sentiment Analysis Algorithm for Automotive Reviews

Chen Bing-feng¹, Hao Zhi-feng^{1,2}, Cai Rui-chu¹, Wen Wen¹, Wang Li-juan¹, Huang Hao¹, Cai Xiao-feng¹

(1. School of Computers, Guangdong University of Technology, Guangzhou 510006, China; 2. School of Mathematics and Big Data, Foshan University, Foshan 528000, China)

Abstract: Sentiment analysis method can mine valuable information from a mass of automotive reviews, which has great application value in automotive product design and brand marketing. For the requirements of fine-grained analysis, a fine-grained sentiment analysis algorithm is put forward based on the entity. Firstly, the automotive reviews are preprocessed, then the model of Linear-chain CRF is used to do sentiment entity recognition and sentiment classification. Secondly, in order to relate the entity recognition with sentiment classification, the model of Linear-chain CRF is improved, and a method of two-level CRF proposed. Experimental results show that two-level CRF is better than Linear-chain CRF in sentiment analysis, which can meet the demand of fine-grained sentiment analysis of automotive reviews.

Key words: automotive reviews; sentiment analysis; sentiment lexicon; fine-grained; conditional random field

随着互联网的高速发展, APP、论坛和留言板等媒体网络上产生海量的用户评论信息, 对这些海量信息进行分析整理与价值挖掘, 成为当前文本情感分析技术的研究热点^[1-2]. 文本情感分析是指对带有情感色彩的文本信息进行分析、处理、归纳和推理的过程^[3]. 文本情感分析涉及数据挖掘、信息检索、机器学习等专业知识, 应用范围也相当广泛.

随着汽车论坛中的评论信息越来越多, 仅依靠人工的方法已经无法应对海量信息的处理, 目前的

文本分析技术在处理这些汽车评论信息方面仍存在许多难点^[4]. 一是分析粒度太大, 对整篇文章或句子的情感倾向性做出判断, 忽略了句子中更细粒度的信息, 造成文本中有价值的信息丢失; 例如: “雷凌双擎油耗较低, 但是电池寿命较短.” 这句话就无法从整句上来判断评论人的意图究竟是褒义还是贬义, 对这句话进行细粒度分析后可知评论者对“油耗”持褒义, 对“电池”持贬义; 二是将汽车实体识别和情感倾向分类分别单独进行, 忽略了汽车实体识别与情

收稿日期: 2017-02-20

基金项目: 国家自然科学基金资助项目(U1501254, 61472089, 61572143); 广东省自然科学基金资助项目(2014A030308008); 广东省自然科学基金杰出青年基金资助项目(2014A030306004); 广东省科技计划项目(2015B010108006); 广东省教育厅项目(2015KQNCX027)

作者简介: 陈炳丰(1983—), 男, 助理研究员, 博士研究生, 主要研究方向为数据挖掘、自然语言处理. E-mail: 735180@qq.com.

感倾向分类之间的联系;三是缺乏反馈机制以及自动学习机制,缺少对错误的结果进行再次学习的过程。

为解决上述问题,本文提出一种面向汽车评论信息的细粒度情感分析方法;使用自定义汽车评论爬虫爬取汽车论坛中带星级标注的信息,减少人工标注;以汽车实体为最小分析粒度,构造一种双层结构的条件随机场模型,用于同时进行情感实体识别和情感倾向分类两个任务;引入反馈机制,使模型能够对错误的样本数据进行学习再更新。

1 相关工作

目前,国内外有很多关于情感分析的研究,但是大部分偏向于社交网络方面的文本情感分析,专门针对汽车行业的评论信息进行研究的还比较少。情感分析技术从研究任务的角度,可分为情感词库构建、情感实体抽取、情感倾向分类等;从文本粒度的角度可分为多篇章、单篇章、句子、短语、单词等几个研究级别;从文本类别的角度,可分为产品评论和社会舆情评论两种情感分析。

1.1 汽车评论分析

汽车评论分析是对论坛中海量的评论信息进行价值挖掘,结合多种技术对评论者的意图进行分析,得到有价值的分析结果。张晶等^[5]针对汽车评论文本中具有多方面内容等问题,提出了一种基于多标记学习的汽车评论文本多方面性能识别方法。廖健等^[6]基于观点袋模型对汽车评论信息进行情感极性分类。目前,国内采用条件随机场对汽车评论信息进行细粒度情感实体识别和情感倾向分类的研究仍处于起步阶段。

1.2 情感词典构建

情感分析研究工作的首要任务是构建一个高质量与大数量的情感词库。构建情感词典主要有基于已有极性词词典或本体库的方法和基于机器学习的方法^[7]。Gatti等^[8]基于SentiWordNet词典库比较了最常用技术和新的混合学习框架,手动建立情感极性词典库。Devaraj等^[9]结合4种不同情感词库以不同的方式来确定不同文本数据的情感极性,并试图在机器学习分类过程中融入情感词典知识。

1.3 情感实体识别

情感实体识别是识别文本信息中有价值的情感信息,提取与情感倾向有关实体,如评价对象、主题、评价者等。实体识别中对隐藏情感对象进行识别是

一个难题,Chen Bingfeng等^[10]通过在条件随机场模型中添加两个全局变量来代表主题及评论者来对隐藏情感对象进行识别。Zhang Shaowu等^[11]构造启发式规则,结合机器学习来识别关键观点的句子来进行交叉领域情感分类。郑敏洁等^[12]提出了一种基于层叠条件随机场的评价对象抽取方法,通过低层条件随机场获得候选评价对象集,对复合短语评价对象进行合并,最后由高层模型抽取评价对象。

1.4 情感倾向分类

情感倾向分类是对文本信息中的情感对象进行极性分类。Wu Fangzhao等^[13]提出一个在特定领域融合多源情感分类器的分类方法,多源情感包含情感对象与极性、情感分类器、无标号数据域、标签数据目标域等。郝志峰等^[14]针对微博社交网络图单节点多标签特性,设计多标签节点树的频繁子树挖掘算法,使用MLTreeMiner挖掘信息传播模式。温雯等^[15]提出一种基于多类别语义词簇的新闻读者情绪分类方法,使用word2vec模型对文本进行初始的语义表达,将传统的文本词向量表达改进为语义词簇上的向量表达。陈培文等^[16]对情感特征进行选取和加权,使用SVM分类器对文本进行情感识别及分类,在Spark分布式计算平台上执行分类模型,取得较好的分类精度和时间代价。

2 情感分析过程设计

汽车评论的情感分析是一个复杂的文本处理过程,需要经过一系列步骤才能完成。在进行情感分析之前要先收集汽车评论训练数据、构建汽车评论情感词典和进行模型训练,然后再对用户输入的汽车评论信息进行分析。分析的过程主要有原始数据处理、分句、分词、词性标注、加载词典、特征抽取,再将向量化的文本信息输入已训练好的情感分析模型中,输出情感分析结果;最后用户在得到分析结果后可以对有误差的结果进行修正,将修正结果反馈给训练样本数据库进行词库更新。目前反馈机制中的误差修正和词库更新等工作内容主要依靠人工的形式进行处理。情感分析过程如图1所示。

2.1 汽车评论数据采集

目前我国知名的汽车网站有新浪汽车、搜狐汽车、汽车之家、易车网和太平洋汽车等,这些网站都建有汽车论坛,而且会对汽车的车系、讨论的主题、车主所在地区等进行分类,成为一个规模庞大的论坛体系。如此大量的汽车评论信息,无法以人工的方

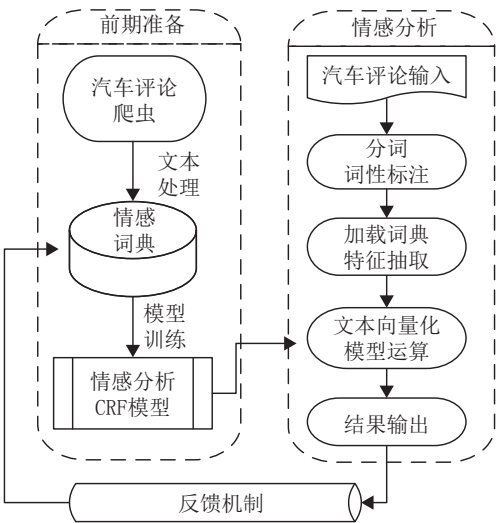


图 1 情感分析流程图

Fig.1 Flow diagram of sentiment analysis

式来获取,须借助网络爬虫技术来解决数据采集问题.为防止他人恶意下载论坛数据,论坛一般会采取反爬虫的措施,如某IP在短时间内不断下载数据则对该IP进行下载限制等.

针对反爬虫问题,本文基于已有爬虫技术,进行反爬虫技术升级,采取不定时更换用户代理、多爬虫并行执行等策略来获取论坛数据.论坛上的评论数据一般都带有好评、差评和星级标注等情感倾向信息,爬虫利用该标注抽取其中的评论信息并对正负面情感信息进行整理,通过数据接口与汽车评论训练数据样本情感词典建立联接,将处理好的格式化数据存入训练样本数据库中.

2.2 文本预处理

通过网络爬虫从汽车论坛上采集到的信息是不能够直接进行情感分析的,为达到更好的情感分析效果,必须经过一系列的文本处理工作.文本处理过程有过滤无效词、分词、词性标注、句法依赖解析等,分词、词性标注和句法依赖解析使用斯坦福大学提供的自然语言处理工具进行处理.

中文句子中的“单词”是含有语义的基本单元,也是细粒度情感分析的基本元素,汽车评论信息在处理过程中先解析成一个个单词后才能进行结构化处理.为便于后续研究,采用斯坦福大学提供的开源Java中文文本分析工具Word Segmenter对汽车评论信息进行分词.

评论帖子进行分词后,还得进行词性标注,也就是对分出来的每个单词标注其词性(如名词、形容词等),通过对单词进行词性标注后可以获取更多

的文本结构信息,本文采用Stanford POS Tagger对单词进行标注.不同的标签代表不同的词性,如“PN”表示代名词(Pronoun),“VV”表示其他动词(Other verbs),“NN”表示其他名词(Other Noun),“PU”表示标点符号(Punctuation).

经过分词和词性标注后,就可以对句子进行句法解析了.句法依赖解析是分析句子中的单词之间的语法关系,得到词与词的依赖关系和逻辑关系,在较深层次有助于计算机理解文本内容.采用句法解析工具Stanford Parser对文本进行句法依赖解析.“nsubj”是依赖关系标签,表示标称主题关系(nominal subject),图2示例中“小明”依赖于“喜欢”并构成标称主题依赖关系;“ccomp”表示从句补语,一般由两个动词构成,中心语引导后一个动词所在的从句;“dobj”表示直接宾语;单词上面的数字表示该词在句中的位置.

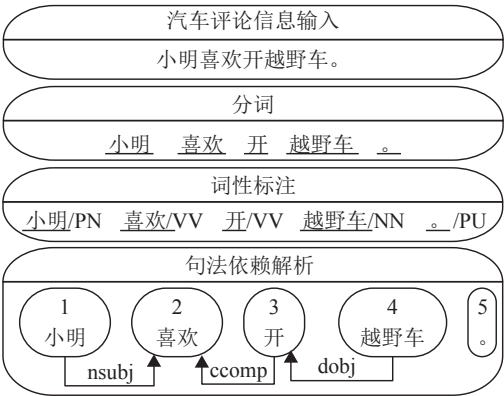


图 2 文本预处理示例图

Fig.2 Example of text preprocessing

2.3 模型选择

在情感分析领域,以往的工作在进行情感实体识别和情感倾向判断时都分别单独采用条件随机场模型,这种方法忽略了这两个任务间的联系,针对该问题,本文对该方法进行优化,构造两层结构的条件随机场模型,将汽车评论的情感实体识别与情感倾向分类统一到一个模型中进行计算,实现了两个任务间的信息相互联系.

汽车评论信息进行文本预处理后得到细粒度的文本数据,将文本数据转化为适合模型训练的数据格式;使用转化后的训练数据调整模型参数,对情感分析模型进行训练,输出训练好的情感分析模型;将向量化的文本信息输入已训练好的情感分析模型中进行运算,输出情感分析结果.

本文采用经典的Linear-chain CRF模型对细粒度

的汽车评论数据分别进行实体识别与情感倾向分类;采用由Linear-chain CRF优化后的双层结构CRF模型对评论数据同时进行实体识别与情感倾向分类。

3 情感分析方法优化

情感分析方法优化是面向汽车评价细粒度情感分析方法的核心部分,重点介绍如何对线性链条件随机场模型进行优化,通过加入势函数来联结两个子任务,形成双层结构的条件随机场模型。

3.1 线性链条件随机场模型

条件随机场(CRF)是一种无向图模型,结合了最大熵模型和隐马尔可夫模型的特点,用于序列标注的机器学习模型^[17]。条件随机场模型用于文本信息的序列标注、数据分割等自然语言处理任务;中文分词、词性标注、情感实体识别、情感倾向分类等任务也常常使用CRF模型。汽车评论数据属于文本信息,因此采用条件随机模型来处理其分析任务。条件随机场属于典型的判别式模型,最常用的是线性链条件随机场(Linear-chain CRF)。若观测序列为 $X=(x_1, x_2, \dots, x_n)$,而标注序列为 $Y=(y_1, y_2, \dots, y_n)$,则在给定输入序列的情况下,Linear-chain CRF模型可定义为

$$P(Y|X) = \frac{1}{Z(X)} \prod_{i=1}^n \psi_i(y_i, y_{i-1}, X). \quad (1)$$

$$Z(X) = \sum_y \prod_{i=1}^n \psi_i(y_i, y_{i-1}, X). \quad (2)$$

$$\psi_i(y_i, y_{i-1}, X) = \exp\left(\sum_k \lambda_k f_k(y_i, y_{i-1}, X, i)\right). \quad (3)$$

其中, $Z(X)$ 是正则化因子, ψ_i 是势函数, f_k 是一个任意的特征函数, λ_k 是对应的特征函数的权值。

对Linear-chain CRF模型的训练,采用最大似然法则获得定义特征的权值,采用Viterbi算法来寻找在观测序列 X 条件下最大条件概率的标注序列,即

$$Y = \arg \max_Y P(Y|X). \quad (4)$$

3.2 双层条件随机场模型

Linear-chain CRF模型只能分别适用于汽车评论信息情感实体识别任务与情感倾向分类任务,为了将这两个任务联系起来,提出了一种构造双层结构的CRF模型的方法。

双层结构CRF模型可以看作是两个Linear-CRF的结合,它的结构包含两条标记的线性链以及观察序列,同时在相同时间点的不同层次的标记节

点中相互连接。在细粒度情感分析任务中,第1层标记序列对评论句中的汽车实体以及对应的情感词进行识别,即对应有3种标记T(实体名)、S(情感词)和O(其它词);第2层标记序列中对汽车实体和情感词的情感倾向进行分类,对应3种标记P(正面情感)、N(负面情感)和O(没有情感)。

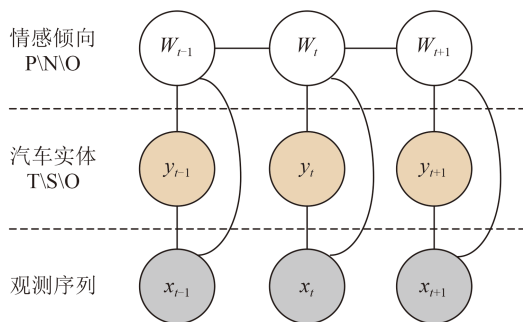


图3 双层结构CRF模型

Fig.3 Two-level model of CRF

双层结构CRF模型可定义为

$$P(y|x) = \frac{1}{Z(x)} \left(\prod_{t=1}^{n-1} \prod_{l=1}^L \psi_l(y_{l,t}, y_{l,t+1}, x, t) \right) \left(\prod_{t=1}^n \prod_{l=1}^{L-1} \phi_l(y_{l,t}, y_{l+1,t}, x, t) \right). \quad (5)$$

$$\psi_l(y_{l,t}, y_{l,t+1}, x, t) = \exp\left(\sum_k \lambda_k f_k(y_{l,t}, y_{l,t+1}, x, t)\right). \quad (6)$$

$$\phi_l(y_{l,t}, y_{l+1,t}, x, t) = \exp\left(\sum_k \lambda_k f_k(y_{l,t}, y_{l+1,t}, x, t)\right). \quad (7)$$

其中, ψ_l 是同一标记序列上的势函数; ϕ_l 是两条标记序列间的势函数; n 为同一标记序列上的结点数, L 为标记序列数,本文取值为2; $f_k(y_{l,t}, y_{l,t+1}, x, t)$ 和 $f_k(y_{l,t}, y_{l+1,t}, x, t)$ 分别为定义在同一标记序列和不同标记序列间的特征函数。双层结构CRF模型的参数估计和模型推理过程可参考Charles Sutton等^[18]的研究成果。

对比Linear-CRF模型和双层CRF模型的定义可以得出这两个模型之间的区别在于双层CRF模型加入了势函数 $\phi_l(y_{l,t}, y_{l+1,t}, x, t)$ 。势函数 ϕ_l 是对汽车实体识别和情感倾向分类两个任务间联系的形式化描述。双层结构CRF模型在Linear-CRF模型上,对不同层中相同时间节点位置的标记做马尔科夫假设,融合不同标记序列上的不同标记间的联系信息,根据最大熵准则,将联系信息形式化描述为势函数 ϕ_l 。通过势函数 ϕ_l 建立起不同标记序列间的信息交互,将相互独立、没有信息交互的两条标记序列联系到一起,因此,双层CRF模型不但具有Linear-CRF的优点,而且还在其基础上融入更加丰富的不同标记序列间的特征,在序列标记问题上得到比Linear-CRF更好

的效果。

4 实验结果与分析

通过爬虫采集汽车之家和易车网论坛中的数据做为实验训练数据和测试数据,通过对比Linear-chain CRF模型和双层结构CRF模型在卡罗拉、雷凌和速腾3个品牌型号的汽车评论数据中的情感分析效果,验证细粒度情感分析方法在海量汽车评论中的作用。

4.1 实验环境

实验的硬件环境为DELL PowerEdge T630服务器: Intel E5-2620 2.4 GHz六核处理器、64 G内存、1.5 T存储;实验操作系统为64位的Linux CentOS 6.4;实验软件环境为myeclipse 2015,实验工具为UMASS(马萨诸塞大学)的开源GRMM工具包,GRMM是一个概率图模型软件工具包,专门用于实现CRFs模型,可实现多种推理算法和参数估计算法。

4.2 实验数据集

实验的训练数据和测试数据均来自于国内知名汽车网站上的论坛,本实验的训练数据通过自定义爬虫从汽车之家、易车网等论坛上获取50 965条数据,包含日期、车型、帖类型、点击量、回复量、标题、正文等字段。汽车实体专业词库从训练数据中的正文字段中抽取,共抽取汽车实体数1 426个,实体评价词5 525个。情感词典由情感实体与其评价词和所对应的情感倾向构成,通过将汽车实体与评价词相结合后标注情感倾向,得到35 231个情感实体。测试数据共有37 056条数据,包含12 851条卡罗拉汽车评论、8 252条雷凌汽车评论和15 953条速腾汽车评论。

4.3 实验结果分析

汽车评论数据经过文本预处理后,得到的是一个有标注信息的序列数据,可采用经典的隐马尔科夫模型(HMM)进行实体识别,但是,HMM模型的当前状态仅依赖于前一个状态且与其他状态相互独立的特点,导致HMM模型在识别过程中丢弃较多信息,在存在多个复杂相关的特征或者长距离依赖的情况下,HMM模型无法利用更多的上下文信息来进行推导,使得模型跟数据的拟合度较差。

HMM模型虽然能够较好地解决序列表示问题,但是条件独立性假设带来方便与高效率的同时也存在一些不足。为更好利用文本预处理后的上下文信息,实验采用条件随机场模型进行序列表示。采用Linear-chain CRF模型对测试数据进行情感分析后的

实验结果如表1所示,测试的数据按卡罗拉、雷凌和速腾3个汽车品牌进行分类,分别识别其评论中含有正向、中性、负向的情感实体。

表1 Linear-chain CRF模型情感实体识别效果

Tab.1 Recognition results of Linear-chain CRF

汽车品牌型号	正向	中性	负向	汇总
卡罗拉	4 748	243	1 307	6 298
雷凌	2 797	163	833	3 793
速腾	3 997	227	1 081	5 305

采用双层结构CRF模型对测试数据进行情感分析后的实验结果如表2所示。

表2 双层结构CRF模型情感实体识别效果

Tab.2 Recognition results of two-level CRF

汽车品牌型号	正向	中性	负向	汇总
卡罗拉	6 094	189	1 398	7 681
雷凌	3 553	143	884	4 580
速腾	4 876	159	1 142	6 177

为了比较两种模型在汽车评论情感分析任务中的效果,采用情感实体识别率来进行分析评价。情感实体识别率可定义为

$$RR = \frac{SE}{Num} \times 100\%.$$

(8)

其中,RR是识别率,SE是识别出来的情感实体数量,Num是本平台测试数据总量。

从图4可以看出,双层结构CRF的识别率均高于Linear-chain CRF,在卡罗拉品牌是高11%、在雷凌品牌是高10%、在速腾品牌是高6%。两种模型在不同品牌间识别率差不一样的主要原因在于汽车情感词典数量有限,不同品牌的评论倾向不一样,使用同一词典对不同品牌进行情感分析效果就会有差异;另一个原因是语义模糊性,不同人对同一评论的理解可能出现截然相反的结果,如“车架较高”对于身高较高的人是优点,对于身高较矮的人是缺点。

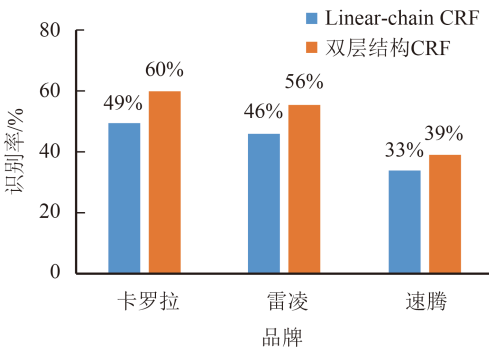


图4 情感实体识别率

Fig.4 Recognition rates of sentiment entity

图5、图6是采用双层结构CRF模型对汽车评论测试数据进行情感分析后的系统运行结果的部分截图。

外观(923)	动力(801)	空间(798)	价格(767)	座椅(717)
发动机(374)	配置(363)	提车(333)	车型(269)	性价比(213)
功能(168)	后视镜(164)	大灯(162)	前脸(149)	脚垫(134)
后排(102)	变速箱(101)	车膜(100)	门(100)	设计(96)
中控台(75)	颜色(72)	舒适性(72)	屁股(71)	质量(71)
屏幕(61)	天窗(58)	总体(58)	高速(57)	中控(57)
服务(51)	省油(50)	灯(49)	油门(48)	钥匙(46)
口碑(35)	隔音(34)	舒适度(34)	胎噪(32)	黑色(32)
态度(27)	地板(27)	订车(26)	样子(25)	多功能方向盘(25)

图5 情感实体统计

Fig.5 Statistics of sentiment entity

15

4. **大空间**爱好者:坐上雷凌,完全感觉不到他是一台紧凑级:简约的设计理念,良好的空间利用率,让双擎的空间大而有趣。

102

【空间】空间感觉前面**一般**,后排空间大!

102

【空间】**空间**感觉前面**一般**,后排空间大!

146

这是销售顾问讲的,我就抄袭了一下,我可没那么专业,哈哈屁股挺好看的,很大气,看着就高端。**空间还不错**回家在跟家里人哈哈,我也要努力,说不定我也快成为有车一族了!

163

油耗:混合动力,油耗非常不错外观:霸气的前脸,流线的**排空间非常充裕**,后排中央位置同样很充裕动力:动力模式下动力精准配置:导航、定速巡航等科技配置齐全性价比:最实惠的混动地板平整,舒适度很不错。

175

当时店里赠送的仪表台垫喜欢这样的**平整**后台, **空间更大**。

175

当时店里赠送的仪表台垫喜欢这样的平整后台, **空间更大**。

图6 情感实体识别明细

Fig.6 Detail of sentiment recognition

图5是对情感实体进行识别并统计每个实体在测试数据中出现的次数,表示该评论测试数据对该实体的关注程度. 点击图5中的实体名称,系统将会把测试数据中所有含有该实体的句子抽取出来,并用红色字体把实体和评价词在句中标注出来. 如点击图5中的“空间(798)”,则可得到图6的结果。

5 结束语

本文针对汽车评论的信息量越来越大,采用细粒度的情感分析方法对该问题进行研究. 通过自定义爬虫在知名汽车论坛上爬取汽车评论数据,对数据进行过滤、分词、词性依赖标注等文本预处理工作,根据已有专业词库构建汽车情感词典,通过已经训练好的模型对输入数据进行情感实体识别和情感倾向分类. 通过研究条件随机场模型对文本情感分析的作用及效果,基于Linear-chain CRF模型,对该模

型进行优化并构建双层结构CRF模型,双层结构CRF模型在引入上下文信息以及依赖信息后,在情感实体识别和情感倾向分类任务上,通过实验验证了双层结构CRF模型的情感分析效果优于Linear-chain CRF模型. 最后将双层结构CRF模型应用于汽车评论测试数据并以网页形式展示情感分析效果,实验结果表明,本文提出的情感分析方法能够满足汽车评论在情感实体识别和情感倾向分类的需求。

参考文献:

[1] RANA T A, CHEAH Y. Aspect extraction in sentiment analysis: comparative analysis and survey [J]. Artificial Intelligence Review, 2016, 46(4): 459-483.

[2] RAVI K, RAVI V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications [J]. Knowledge-based Systems, 2015, 89: 14-46.

[3] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.

ZHAO Y Y, QIN B, LIU T. Sentiment analysis [J]. Journal of Software, 2010, 21(8): 1834-1848.

[4] 李纲, 程明结, 寇广增. 基于情感倾向识别的汽车评论挖掘系统构建[J]. 情报学报, 2011, 30(2): 204-211.

LI G, CHENG M J, KOU G Z. The construction of car comments mining system based on sentiment analysis [J]. Journal of the China Society for Scientific and Technical Information, 2011, 30(2): 204-211.

[5] 张晶, 李德玉, 王素格. 基于多标记学习的汽车评论文本多性能识别[J]. 计算机工程与科学, 2016, 38(1): 188-194.

ZHANG J, LI D Y, WANG S G. Multiple performances identification for car review texts based on multi-label learning [J]. Computer Engineering and Science, 2016, 38(1): 188-194.

[6] 廖健, 王素格, 李德玉, 等. 基于观点袋模型的汽车评论情感极性分类[J]. 中文信息学报, 2015, 29(03): 113-120.

LIAO J, WANG S G, LI D Y, et al. The Bag-of-Opinions method for car review sentiment polarity classification [J]. Journal of Chinese Information Processing, 2015, 29(03): 113-120.

[7] 王山雨. 面向产品领域的细粒度情感分析技术[D]. 哈尔滨: 哈尔滨工业大学 计算机科学与技术学院, 2011.

[8] GATTI L, GUERINI M, TURCHI M. Sentiwords: Deriving a high precision and high coverage lexicon for sentiment analysis [J]. Ieee Transactions on Affective Computing, 2016, 7(4): 409-421.

[9] DEVARAJ M, PIRYANI R, SINGH V K. Lexicon ensemble and lexicon pooling for sentiment polarity detection [J]. IETE Technical Review, 2016, 33(3): 332-340.

[10] CHEN B, HAO Z, CAI R, et al. Sentiment target extraction based on CRFs with multi-features for Chinese microblog [C]// MORISHIMA A. Web Technologies and Applications-

APWeb 2016. Heidelberg: Springer, 2016: 29-41.

[11] ZHANG S, LIU H, YANG L, *et al.* A cross-domain sentiment classification method based on extraction of key sentiment sentence [C]//LI J, JI H, ZHAO D, *et al.* Natural Language Processing and Chinese Computing. Nanchang: LNAI, 2015: 9362, 90-101.

[12] 郑敏洁, 雷志城, 廖祥文, 等. 基于层叠CRFs的中文句子评价对象抽取[J]. 中文信息学报, 2013, 27(3): 69-76.

ZHENG M J, LEI Z C, LIAO X W, *et al.* Identify sentiment-objects from chinese sentences based on cascaded conditional random fields [J]. Journal of Chinese Information Processing, 2013, 27(3): 69-76.

[13] WU F, HUANG Y, YUAN Z. Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources [J]. Information Fusion, 2017, 35: 26-37.

[14] 郝志峰, 黄灿锦, 蔡瑞初, 等. 结合用户兴趣的微博信息传播模式挖掘[J]. 模式识别与人工智能, 2016, 29(10): 924-935.

HAO Z F, HUANG C J, CAI R C, *et al.* User interest related information diffusion pattern mining in microblog [J]. Pattern Recognition and Artificial Intelligence, 2016, 29(10): 924-935.

[15] 温雯, 吴彪, 蔡瑞初, 等. 基于多类别语义词簇的新闻读者情绪分类[J]. 计算机应用, 2016, 36(8): 2076-2081.

WEN W, WU B, CAI R C, *et al.* Emotion classification for news readers based on multi-category semantic word clusters [J]. Journal of Computer Applications, 2016, 36(8): 2076-2081.

[16] 陈培文, 傅秀芬. 采用SVM方法的文本情感极性分类研究[J]. 广东工业大学学报, 2014(03): 95-101.

CHEN P W, FU X F. Research on sentiment classification of texts based on SVM [J]. Journal of Guangdong University of Technology, 2014(03): 95-101.

[17] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data [C]//Proceedings of ICML. [S.l.:s.n.], 2001: 282-289.

[18] SUTTON C, MCCALLUM A, ROHANIMANESH K. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data [J]. Journal of Machine Learning Research, 2007, 8: 693-723.

广东工业大学荣获8项2016年广东省科学技术奖

2017年2月7日，广东省委、省政府在广州召开广东省创新发展大会，深入学习贯彻习近平总书记系列重要讲话精神，贯彻落实中央经济工作会议和全国科技创新大会精神，部署加快推动广东省创新驱动发展战略。会议表彰了获得2016年度广东省科学技术奖的先进单位和个人。省委书记胡春华出席会议并讲话，强调要牢牢抓住国家科技产业创新中心这个核心定位，加快实施创新驱动发展核心战略，不断开创新发展新局面。省长马兴瑞主持会议，省人大常委会主任李玉妹、省政协主席王荣出席会议。

会上颁发了2016年度广东省科学技术奖。省委常委、省人大常委会副主任徐少华宣读《广东省人民政府关于颁发2016年度广东省科学技术奖的通报》。2016年共颁发省科学技术奖项237项，广东工业大学陈新教授团队的成果“面向电子装备点位操作的高速精密运动规划与测量技术及应用”获得了2016年度广东省技术发明一等奖，戴青云教授团队荣获2016年度广东省科技进步二等奖。另外，作为参与完成单位获得6项奖励，其中参与完成的项目“东莞横沥镇模具产业协同创新体系的建设与实践”荣获特等奖。

此次获奖，是广东工业大学连续4年获得了广东省科学技术一等奖，这进一步为实现高水平大学建设国家奖目标打下了基础。在此对获奖团队表示热烈祝贺！

（广东工业大学科技处、人文社科处 供稿）