

评论信息抽取技术在网络贸易系统平台上的应用

赵俊芹 杨丹 余森 陈林

(重庆大学软件工程学院 重庆 400044)

摘要 准确挖掘购物网站中的用户评论对商家和顾客进行有效的推荐具有重要作用。本文研究了现有的网络贸易系统平台的功能模块,给出了用户评论信息抽取的具体方法,并指出用户评论信息抽取模块应用到网络贸易系统平台的优势和实际意义。

关键词 网络贸易,评论抽取,句法分析

Technique of Reviews Extraction Apply to Online Shopping System

ZHAO Jun-Qin YANG Dan YU MIAO CHEN LIN

(School of Software Engineering, Chongqing University, Chongqing 400044)

Abstract Mining the customer reviews accurately in commercial websites has significant meaning in effective recommendation for trade company and customers. This article studies function modules of the existing online shopping system platform, proposes a technique to customer reviews extraction, and points out that the customer reviews extraction module apply to online shopping system platform is highly effective and has practical significance.

Keywords Online shopping, Reviews extraction, Syntactical parser

随着互联网的发展,网络贸易开始在中国兴起,给传统的贸易方式带来了巨大的冲击,被人们喻为“一场商贸方式的革命”。网络贸易,是消费者通过网络(购物网站)向供应商购买商品或享受服务。消费者可浏览网上商品目录,比较、选择满意的商品或服务,通过网络下订单,进行网上付款或离线付款;而供应商处理订单、网上送货或离线送货,从而完成整个网上购物过程。这里所指的购物网站,可以是自身提供产品销售服务的纯粹经销商,或者是网络交易平台,也可以是两者的混合体。这里所指的供应商,可以是企业,也可能是个体户或者个人。在价格制定方式上,可以由供应商制定价格,也可以是采用消费者和供应商通过竞价、议价的方式最终确定。

中国网络贸易市场潜力巨大。据统计,截止 2006 年底,中国网上贸易交易额为 312 亿元,用户市场规模达到 4310 万,预计到 2010 年市场交易额将达到 1460 亿元^[1]。网上贸易如此迅猛发展,网络贸易平台在满足顾客快速找到满意产品的设计方面显得尤为重要。但是,现有的网络贸易平台并没有很好地实现这一重要方面,通过对目前没有充分利用的用户评论的抽取可以解决这一问题。

本文主要研究了现有的网络贸易系统平台的功能模块,给出了产品用户评论信息抽取的具体方法,并指出产品用户评论信息抽取模块应用到网络贸易系统平台的优势和实际意义。

1 网上购物的主要特点

网上商店与传统商店在门类结构和功能上没有区别,不同点在于其实现这些功能和结构的方法以及商务运作的方式,与传统商店相比,网上商店主要有六大特点:1)成本低廉;2)“零库存”商店;3)经营规模不受场地限制;4)便于与顾客交互;5)365 天 24 小时经营;6)无国界、区域限制。网上购物从前台看是一种特殊的 Web 服务器,现代 Web 网站的多媒体支持和良好的交互性功能成为网上购物的基础,使得顾客可以像在真实的超级市场推着购物车挑选商品,并最后在收银台结账。

2 产品用户评论信息抽取

网上购物网站一般都有用户对已购产品的评论区域,随着

网上购物的迅猛发展,已购用户的评论数量也在迅速地增加。大量的用户评论有褒有贬,这使得打算购买产品的顾客在是否选择购买的时候难以下结论,也使得商家对产品受欢迎程度比较难以把握。但是,通过挖掘和总结某一产品的用户评论,可以使顾客和商家准确做出决策。挖掘产品评论,分为四个步骤:1)抽取评论涉及的产品特征词;2)识别评论中修饰已抽取的产品特征的极性词;3)识别极性词极性褒贬;4)统计某一产品特征褒贬两类评论数,在用户界面可视化显示数据。

2.1 抽取评论涉及的产品特征词

特征词是句子评论的对象,在实际应用中,用户往往关注某类产品的特征,一个特征词需要满足以下三个条件之一^[2]:1)是给定主题的一部分;2)是给定主题的一个属性;3)是给定主题的部分的一个属性。

本文首先利用 DeParser 对评论里的每个句子进行预处理。DeParser 是由哈尔滨工业大学信息检索实验室开发的汉语语法分析器。其输出结果中,用三行表示一个句子,第一行是输入的原句;第二行对句子进行了分词和词性标注,并在句子中每个词及词性的前面加上序号,句子的末尾增加一个句尾标志“<EOS>”,由其支配全句的核心词;第三行是句子中词与词之间的依存关系。依存关系中,每个关系以一个依存对表示,依存对中的第一个词是核心词,支配第二个词。

这里,我们根据第二行输出,选择词性是名词[n]的词语作为产品特征词。实际用户评论中也有动词或形容词作为特征词,本文不作讨论,在后面的工作中再作研究。

2.2 识别评论中修饰已抽取的产品特征的极性词

极性词是指句子中带有感情色彩的词语,极性词是判断句子极性和文档极性的前提条件和根本依据。我们可以利用已抽取的产品特征词识别评论句子中的极性词。例如下面的句子:1)“观后视镜设计得很可爱。”2)“耗油量太大。”

在 1)中,特征词是“观后视镜”,与它邻近的极性词是形容词“可爱”。

在 2)中,特征词是“耗油量”,与它邻近的极性词是形容词“大”。

由此可见,对评论中的每个句子,假如它含有特征词,根

据 DeParser 第二行和第三行的输出,抽取特征词附近的形容词,即修饰产品特征的特征词,该形容词可被看作特征词的极性词,且此处的附近限定为 3 个词语的距离^[5]。

2.3 识别极性词极性褒贬

一个词的极性分为两种情况,一种是原极性,这种极性可以利用 WordNet 得到;另一种是上下文极性^[4]。后者可以出现在两种语境中:一是句子含有否定前缀,例如,“不漂亮”;二是句子含有强调前缀,例如,“非常漂亮”。这个问题,可以定义两个字典:否定字典和强调字典解决^[3]。对一个句子进行句法分析后,通过研究词与词之间的依存关系,可以找到否定前缀和强调前缀,从而得到极性词的极性褒贬情况。

2.4 统计某一产品特征褒贬两类评论数,在用户界面可视化显示数据

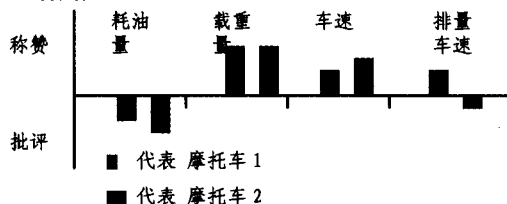


图 1 两种类型摩托车的用户评论信息形象化比较

如图 1, X 轴代表用户评论抽取出来的产品特征, Y 值是由公式 $Y = \text{该特征称赞得评论数} - \text{该特征批评得评论数}$ 得出。在 X 轴的上方,表示该产品特征称赞得更多些,在 X 轴下方,表示该产品特征批评得更多些,其代表的 Y 值是由公式 $Y = \text{该特征称赞的评论数} - \text{该特征批评的评论数}$ 。这样的可视化表示,可使顾客一目了然地比较产品的各个特征。图 1 给出了摩托车 1 和摩托车 2 的四类特征比较,在载重量和车速两个特征,两种类型的摩托车的用户都是好评多些,但是摩托车 2 的好评多于摩托车 1 的好评。在耗油量方面,两类车型都是差评,但摩托车 2 的差评稍微多了一些。在排量方面,两类车型的比较很明显,摩托车 1 总体是好评,摩托车 2 总体是差评。根据这些比较,顾客可以方便地选择自己喜爱的产品。

3 目前网络贸易平台功能模块介绍

与传统购物过程相似,网上购物需要找到自己所需商品,放进购物车,然后付账来完成整个购物过程。除了这些必要的购买程序之外,目前购物网站一般设有用户评论区,但是它不是用户必须登录使用的板块。

目前的网上购物系统平台购物流程如图 2 所示:

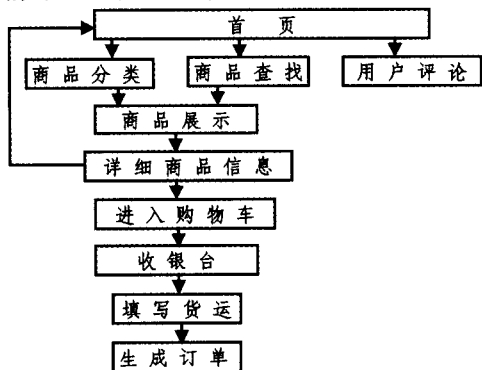


图 2 目前网上购物系统平台购物流程

现有的网络贸易平台功能模块一般包括:会员专区、商品展示、在线查询、信息发布、购物车的显示、订单的生成以及用户评论模块等^[6]。

会员专区模块:是指会员所享受的权利,它包括会员注册、登录、修改或删除自己的资料,发布供应或求购信息,查

看自己购物车的内容及自己的历史订单等,为会员提供方便。

商品展示模块:动态地从数据库中取出商品信息,包括商品名字、图片、价格、联系人等各种信息,还包括特价商品的展示和分类商品的展示等。

在线查询模块:提供客户网上查询的功能,主要有:①商品查询。根据商品名称,商品类型等查询该商品的详细信息,包括外观、规格、性能、价格、产地以及已售出的数量等。②订购信息查询。包括订购日期,已订购商品的价格、数量以及目前的状态(未发送、途中)等。

商品信息发布模块:提供会员商品信息发布,注册取得授权后,可在系统提供的产品管理里随时增添、修改、更新自己的销售商品信息。

购物车设计模块:当顾客看中某商品时,就启动购物车功能。包括:在每种商品信息中,都有购买按钮;选购完商品后,系统给出购物清单,在清单中列出所购物品及数量、每件物品的价格和购物总价格;整个购物过程有完整的确认系统。

订单生成模块:是顾客在填完货运资料和付款方式后,为用户生成订单。

用户评论模块:顾客在购买商品后,发表使用评论的区域。

4 产品用户评论信息抽取模块应用到网络贸易平台

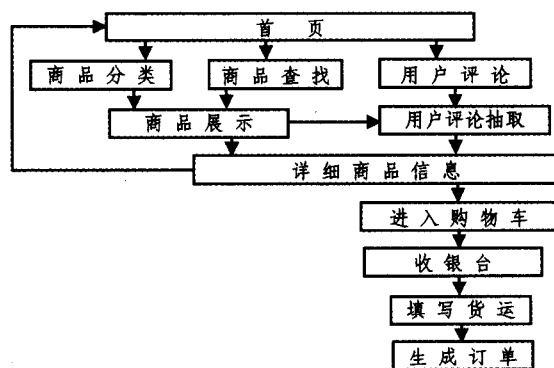


图 3 用户评论抽取模块应用到网络贸易平台后的购物流程

如图 3, 在产品用户评论信息抽取模块应用到网络贸易平台上以后,先前的没有利用的用户评论模块得以充分利用。在商品展示后,顾客选择看中的商品,然后再参看用户评论的信息,得到已购用户对选中产品的评价,就可以决定是否购买。这样顾客就免去了自己比较商品的麻烦,且节省了时间。当然,有时间且爱逛店铺的顾客也可以选择先前的方式购买商品。

结语 在目前的网络贸易平台上,用户评论信息不能充分利用。商家只能动用大量的人力、物力做产品满意度调查;顾客面对大量的褒贬不一的评论也很难做出购买决定。通过把产品用户评论信息抽取模块应用到网络贸易平台,商家和顾客都可把握产品的各个特征的性能及用户反馈情况,使网络贸易更受欢迎。

参考文献

- 1 iResearch Consulting Group. China online shopping Research Report [R]. 2006
- 2 Yi J, Niblack W. Sentiment mining in WebFountain [A]. In: Proceedings of the 21st International Conference on Data Engineering (ICDE 2005) [C]. Washing, DC, USA: IEEE Computer Society Press, 2005. 1073~1083
- 3 姜得成, 姚天防. 汉语句子语义极性分析和观点抽取方法的研究 [J]. 计算机应用, 2006
- 4 LIU B, HU M. Opinion observer. analyzing and comparing opinions on the web [A]. In: International World Wide Web Conference Committee (IW3C2) [C]. Chiba, Japan, 2005
- 5 Shi Bin, Chang Kuiyu. Mining Chinese Reviews. International Conference on Data Mining-Workshops (ICDMW'06) [C]. IEEE, 2006
- 6 谯志群, 王大全. 网上购物系统若干关键技术研究 [J]. 杭州电子工业学院学报, 2003