

APP软件的用户评论模式分析方法

冉 猛 姜 瑛

(云南省计算机技术应用重点实验室 昆明 650500) (昆明理工大学信息工程与自动化学院 昆明 650500)

摘 要 面对海量的APP软件,不同用户对其评论的侧重点、表达方式以及情感倾向程度等都不相同,这给APP软件的用户行为分析和质量评价带来了困难。提出一种APP软件用户评论模式分析方法,首先综合分析用户评论信息与APP软件信息之间的关系,根据用户对APP软件的评论特征将用户评论信息进行分类;接着分析每类用户评论信息的词性组合;然后计算用户评论信息的情感倾向程度,以分析出该APP软件用户的评论模式;最后通过实验验证了该方法的有效性。

关键词 APP软件,用户评论信息,特征分类,词性组合,评论模式

中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.11.027

Analytical Method for APP Software's User Comment Patterns

RAN Meng JIANG Ying

(Yunnan Key Lab of Computer Technology Application, Kunming 650500, China)

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract Faced with the massive user's comment of APP software, there are different results of comment focus, expression style and emotional tendency for different users. These differences bring difficulties to the user behavior analysis and the quality evaluation of the APP software. This paper presented a method to analyze APP software's user comment patterns. At first, the relationships between user's comment information and APP software information are analyzed. The user's comment information is classified based on the comment feature of APP software. Then the combination of part-of-speech for each type of user's comment information is analyzed. After calculating the emotional tendency degree of user's comment information, the user's comment patterns are analyzed. Finally, the experimental results show that the method is effective.

Keywords APP software, User's comment information, Feature classify, Combination of part-of-speech, Comment patterns

1 引言

面对海量的APP软件,如何评价其质量成为了一个亟需解决的问题。Jiang等^[1]首先对APP软件用户进行调研,然后通过监督学习方法分析各个APP软件属性的重要性,并建立APP软件质量模型,以评价APP软件的质量。Ma等^[2]通过半监督学习方法分析现有APP软件质量之间的差异性,以预测出具有恶意行为的APP软件。上述研究主要通过分析APP软件本身的属性特征来评价APP软件的质量。在使用APP软件之后,用户通常会发表相应的评论信息和评分星级,这些信息中往往隐含了用户对APP软件特定属性的关注程度、情感倾向程度等潜在行为信息,有助于研究人员对APP软件的质量进行评价。然而,不同用户的行为是不相同的,即用户对APP软件进行评论的侧重点、表达方式和情感

倾向程度存在差异,这给APP软件的用户行为分析和质量评价带来了困难。

2 相关工作

目前,用户行为分析在微博领域和商品领域日趋成熟^[3-4]。李敏等^[3]对微博用户的签到时间、签到地点以及发表评论的情感倾向等行为进行分析,以获取用户的需求信息。王伟等^[4]通过分析用户评论中对产品的关注特征和满意度行为来建立用户需求和偏好模型。然而,APP软件的用户评论与普通的用户评论不同。通过分析两万余条APP软件的用户评论信息,我们发现它具有如下特点:1)相对于微博评论,APP软件用户评论信息的字数较少,最短的评论只有一个字^[5],例如“好...”;2)用户评论信息中含有大量的网络词汇,例如“就是个坑爹的软件”和“聊天的神器”等;3)用户在选择

到稿日期:2016-10-15 返修日期:2016-12-09 本文受国家自然科学基金(61462049,61063006,60703116)资助。

冉 猛(1989—),男,硕士生,主要研究方向为软件工程、软件质量保证与测试,E-mail:374085783@qq.com;姜 瑛(1974—),女,博士,教授,主要研究方向为软件质量保证与测试、云计算、软件工程、大数据分析,E-mail:jy_910@163.com(通信作者)。

APP 软件前无法通过广告或品牌获得该软件的质量信息^[6], 并不了解该软件的真实质量, 因此用户评论信息中对 APP 软件的情感倾向程度可以作为评价该 APP 软件的重要依据。

目前, APP 软件用户行为分析和质量评价也取得了一系列的研究成果^[7-10]。Pagano 等^[7]通过分析用户评论的时间趋势和评论信息中特征之间的差异性行为对用户社区的影响, 来帮助开发团队对 APP 软件进行需求分析。Guzman 等^[8]通过自然语言处理技术对用户评论信息中的特征进行聚类, 并分析用户对 APP 软件的正反态度, 帮助开发人员分析用户对 APP 软件的需求。Anam 等^[9]通过分析用户评论信息中评论特征之间的差异性以及对 APP 软件的正反态度, 以建立一个用户对 APP 软件特征的偏好排名系统。Khalid 等^[10]通过人工标注方法筛选出评论信息中用户对 APP 软件功能的抱怨行为, 并进行分类, 以帮助开发者从用户的角度来提高 APP 软件的质量。

上述研究一方面通过分析 APP 软件用户评论信息中特征之间的差异性并对其进行分类, 来了解用户对 APP 软件的需求; 另一方面通过结合用户评论信息中的特征和用户对 APP 软件的正反态度并进行综合分析, 来帮助开发者从用户的角度提高 APP 软件的质量。然而, 传统的人工标注方法很难适应目前海量的 APP 软件评论信息。APP 软件用户评论信息是一种具有自由性和随意性的典型的网络评论, 不同用户的评论模式是不相同的, 这给对 APP 软件的用户评论信息的分析带来了一定困难。因此, 本文针对如何分析 APP 软件的用户评论模式展开研究。首先对 APP 软件的用户评论模式进行定义; 接着将用户评论信息进行分类; 然后分析每类用户评论信息的词性组合; 最后计算用户评论信息的情感倾向程度, 以分析出 APP 软件的用户评论模式。

3 APP 软件的用户评论模式

由于不同的用户对 APP 软件的需求存在差异, 因此用户对 APP 软件评论的侧重点也就不同, 即用户评论信息中的评论特征不同。通过对 APP 软件信息和用户评论信息进行综合分析, 我们将用户评论特征分为以下 3 类。

(1) 属性特征: 表示评论信息中针对 APP 软件信息的功能、所占空间、版本号等特征;

(2) 综合特征: 表示评论信息中用户对 APP 软件进行的综合评论特征;

(3) 其它类特征: 除属性特征和综合特征以外的其它特征。

例如, 有 3 条关于“微信”APP 软件的评论: “为什么发不了消息”、“微信很好”和“下载不了”。其中, “为什么发不了消息”针对的是该 APP 软件的“发消息”属性特征; “微信很好”针对的是该 APP 软件的综合特征; 而“下载不了”针对的是该 APP 软件的“下载”特征, 属于其它类特征。

此外, 由于网络的开放性和自由性, 用户的表达方式具有随意性, 有的用户习惯用“{副词+形容词}”的词性组合进行评论, 有的用户习惯用“{动词+副词+形容词}”的词性组合进行评论, 而有的用户通常使用网络情感词汇进行评论。不同的用户对 APP 软件的评判标准不同, 表达的情感倾向也就不同, 例如, 不同的用户对“QQ”APP 软件的评论有: “QQ 是

个不错的软件”和“QQ 就是个垃圾”, 二者表达的情感倾向截然相反。因此, 本文对词性组合和情感倾向程度分别做如下定义。

定义 1 词性组合 $S = \{s_x + [s_x] + \dots s_x | s_x\}$ 。其中, s_x 的取值为 n(名词)、v(动词)、vn(动名词)、d(副词)、a(形容词)和 a_n(网络情感词汇), “+”表示“和”, “[]”表示该项为“可选项”, “|”表示“或”。

定义 2 情感倾向程度 $E = \{e_x\}$ 。其中, e_x 的取值为 1(很差)、2(差)、3(一般)、4(好)和 5(很好)。

通过分析 APP 软件的用户评论发现, 每个用户发表评论时都有自己的侧重点、表达方式和评判标准。因此, 本文将用户评论信息的特征类别、词性组合和情感倾向程度称为 APP 软件用户评论的评论模式, 并对其做如下定义。

定义 3 评论模式 $P = \{\{F\}, S, E\}$ 。其中, F 表示特征类别, 取值为综合特征、属性特征和其它类特征。

例如, 某用户对“QQ”APP 软件的评论信息为“下载很慢”。通过分析, 特征类别 $F = \{\text{其它类特征}\}$, 该用户评论信息的词性组合 $S = \{v + d + a\}$, 用户对该 APP 软件的情感倾向程度 $E = \{1\}$, 评论模式 $P = \{\{\text{其它类特征}\}, \{v + d + a\}, \{1\}\}$, 具体分析见第 4 节。

4 APP 软件的用户评论模式分析方法

本节基于 APP 软件用户评论模式的定义, 分析用户评论信息中的特征类别、词性组合以及用户对 APP 软件的情感倾向程度。

4.1 用户评论信息分类

用户发表的评论通常是对该 APP 软件某类特征是否满足自己的需求所进行的情感表达。不同用户对同一个 APP 软件进行评论时所表达的情感倾向程度必然是不同的, 原因在于有的用户侧重于 APP 软件的综合特征, 有的则更关注 APP 软件的属性特征等。这一现象说明, APP 软件的用户评论信息中的评论特征及其情感倾向程度在评价 APP 软件质量中有着不可忽视的作用。因此, 接下来首先提取用户评论信息的特征情感词对; 再通过分析特征情感词对的关系, 结合 APP 软件信息和评论信息分析所针对的特征, 将用户评论信息分为 3 类。

4.1.1 提取 APP 软件的特征情感词对

由于 APP 软件的用户评论信息是自然语言, 因此需要对其进行结构化或半结构化的处理。目前, 大多数研究通过提取用户评论中的特征词和情感词来获得用户对产品特征进行的描述或评论的数据。然而, APP 软件的用户评论信息中的特征词和情感词往往是成对出现的。例如, 用户对“微信”APP 软件评论: “下载很麻烦”, 其中用情感词“麻烦”修饰了特征词“下载”, 用副词“很”修饰了情感词“麻烦”。因此, 本文称其为 APP 软件特征情感词对, 并对其做如下定义。

定义 4 APP 软件特征情感词对 $f_i = \{Wh, Wd, Wa\}$ ($i = 1, 2, \dots, n$)。

其中, Wh 代表用户评论信息中用户关注 APP 软件的特征词, 如“下载”、“画面”等; Wa 代表修饰特征 Wh 的情感词,

如“麻烦”、“流畅”等,其中也包括网络情感词,其表达了用户对 APP 软件特征的主观感受,如“神器”等; Wd 代表修饰情感词 Wa 的副词,表达了用户对特征情感的程度,如“很”、“非常”等。

本文在文献[11]提取特征词、情感词的基础上,对特征词、副词、情感词同时进行提取,即提取 APP 软件的特征情感词对 f 。

4.1.2 分类方法

APP 软件信息中的软件名称、软件类别、软件简介、软件来源等均由开发者官方发布,这些信息通常描述的是该 APP 软件特有的功能、属性等特征。为了更好地对用户评论信息进行分类,本文将 APP 软件信息中的名词、动词以及名词词组进行提取,建立 APP 软件信息特征词库,并将这些词归为 APP 软件的属性特征。

此外,部分用户对 APP 软件进行评论的信息中往往省略了特征词 Wh ,或者 Wh 为该 APP 软件的软件名称。例如,用户对“微信”APP 软件的评论信息“很好,很满意”和“微信很不错”中评论对象是一致的,都是对“微信”APP 软件的综合评论。因此,本文将这类用户评论信息归为针对 APP 软件综合特征进行的评论。

另有部分用户的评论信息中的特征词 Wh 是用户对 APP 软件评论的其它类特征。例如,同样是对“微信”APP 软件进行的评论信息“下载不了”和“界面太差”,特征词 Wh 分别为“下载”和“界面”,二者都不属于该 APP 软件的属性特征和综合特征。因此,本文将这类信息归为用户针对 APP 软件的其它类特征进行的评论。

用户评论信息的分类方法流程图如图 1 所示。

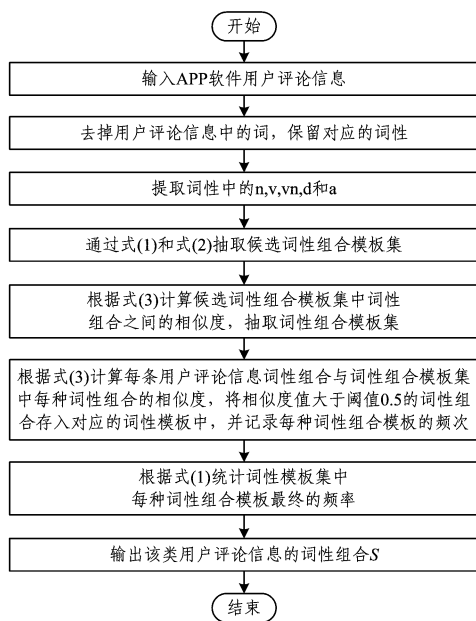


图 1 用户评论信息的分类方法流程图

图 1 中, $count_{WL}$, $count_{AL}$ 和 $count_{OL}$ 分别表示 APP 软件综合特征、属性特征和其它类特征的用户评论信息条数。

4.2 词性组合分析

不同类别的用户评论信息的语法结构是不同的,即用户评论信息对应的词性组合是不同的。通常,用户针对 APP 软件属性特征和其它类特征进行评论的信息比较复杂;而针对 APP 软件综合特征进行评论的信息相对比较简单,往往用

“ $\{v+d+a\}$ ”或“ $\{a\}$ ”等方式发表评论。因此,本小节先对属性特征和其它类特征的用户评论信息所对应的词性组合进行分析,再对综合特征类的用户评论信息的词性组合进行分析。

4.2.1 属性特征和其它类特征的词性组合分析

针对 APP 软件属性特征和其它类特征的用户评论信息比较复杂,例如,“QQ”APP 软件的评论信息“额……为什么发不了消息了”。然而,用户评论信息中许多词是没有实际意义的,例如“的”“地”“了”等,而用户评论信息中词性为 n , v , vn , d 和 a 才是该条用户评论的关键信息。因此,需要提取每条用户评论信息中的 n , v , vn , d 和 a 。

此外,为了降低词性组合分析的复杂度,本文使用“词频”(Term Frequency, TF)统计的方法对词性组合进行过滤,抽取候选词性组合模板集。TF 统计公式如下:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_K n_{K,j}} \quad (1)$$

其中, $tf_{i,j}$ 表示第 i 种词性组合出现的频率, $n_{i,j}$ 表示 i 种词性组合在该类用户评论信息中出现的频数, K 表示该类用户评论信息中的词性组合的种类, $\sum_K n_{K,j}$ 表示该类用户评论信息中词性组合的总数。候选词性组合模板的 $tf_{i,j}$ 值越高,它越可能是真正的词性组合模板。这里需要设定一个阈值,只有当词性组合的 $tf_{i,j}$ 值大于阈值时,该词性组合才被存入到候选词性组合模板集中。抽取候选词性组合模板集的公式如下:

$$isAttr(tf) = \begin{cases} 1, & tf_{i,j} > \frac{1}{K} \\ 0, & tf_{i,j} \leq \frac{1}{K} \end{cases} \quad (2)$$

由于该类用户评论信息中的词性组合有 K 类,即每种词性组合出现的平均频率为 $\frac{1}{K}$,因此只有频率 $tf_{i,j}$ 值大于平均频率 $\frac{1}{K}$ 的词性组合才具有代表性,将其作为候选词性组合模板。

通常,候选词性组合模板集中的词性组合之间存在彼此相似的情况,例如某候选词性组合模板集为 $A = \{\{v+d+a\}, \{v+a\}\}$,其中,“ $\{v+d+a\}$ ”与“ $\{v+a\}$ ”相似。因此,需要将候选词性组合模板集中相似的词性组合进行合并,以抽取词性组合模板集。本文利用 Levenshtein 相似度计算公式^[12]计算词性组合之间的相似度,计算公式如下:

$$Sim_{i,j} = 1 - \frac{ld}{\max(m, n)} \quad (3)$$

其中, $Sim_{i,j}$ 表示词性组合 i 和词性组合 j 之间的相似度, ld 表示词性组合 i 和词性组合 j 之间的距离, m 和 n 分别为两个词性组合中字符的长度。为了判断两个词性组合是否相似,需要设定一个阈值。 $Sim_{i,j}$ 的取值范围为 $[0, 1]$,在该取值区间中通常将 0.5 作为阈值,当两个词性组合的相似度值 $Sim_{i,j}$ 大于 0.5 时,认为两个词性组合是相似的。本文通过实验发现,当阈值取 0.5 时,效果达到最优。

接下来需要记录每种词性组合的频次,以计算该类特征中用户评论信息中最终词性组合的频率,从而分析出该类评论特征中用户评论信息的最终词性组合。

属性特征和其它类特征的词性组合分析流程如图 2 所示。

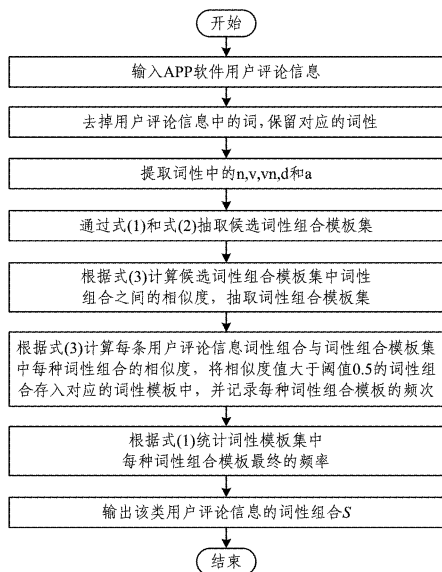


图2 属性特征和其它类特征的词性组合分析流程图

4.2.2 综合特征的词性组合分析

针对本文第3节中APP软件用户评论的前两个特点,将对APP软件综合特征的用户评论信息对应的词性组合分为两种情况:

1)部分用户用“{[d]+a}”, “[d]+v}”和“{n+d+a|v}”等方式发表评论;

2)部分用户用网络情感词汇对该类特征进行评论,如用户评论信息“QQ就是个神器”用网络情感词汇“神器”对“QQ”APP软件的综合特征进行评论。

本文将以上两种情况的词性组合分别定义为: “[n]+[d]+a|v}”和“{[n]+[d]+a_n}”。

此外,目前的情感词词典不能识别网络情感词汇,为了对APP软件综合特征用户评论信息的词性组合进行更好的分析,本文建立了网络情感词汇库,收集了诸如“TMD”, “SB”和“神器”等高频度网络情感词汇,共计138个。

APP软件综合特征类用户评论信息的词性组合分析流程如图3所示。

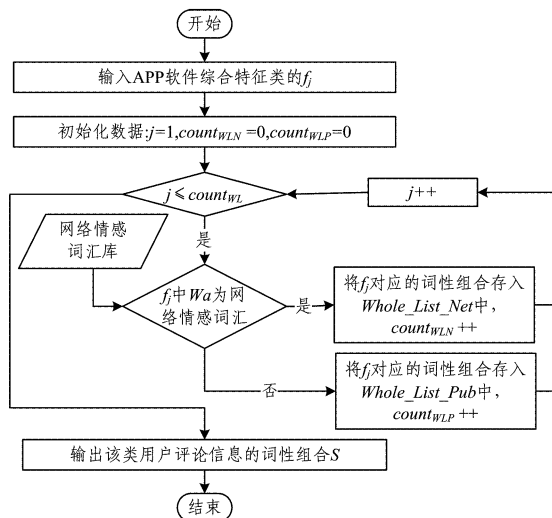


图3 APP软件综合特征词性组合分析流程图

图3中, f_j 表示综合特征中第 j 条用户评论信息的特征情感词对; $count_{WLN}$ 和 $count_{WLP}$ 分别表示该类用户评论信息中使用网络情感词汇和未使用网络情感词汇的评论的条数。

4.3 量化情感倾向程度

针对第3节中APP软件用户评论信息的第3个特点,目前对APP软件的情感倾向程度成为了评价该APP软件的重要依据。然而,在APP软件的用户评论信息中,不同用户对APP软件的评判标准不一样,这导致不同的用户对APP软件所表达的情感倾向程度是有差异的,从而给APP软件的用户评论模式分析带来了一定的影响,也给APP软件的质量评价带来了影响,因此需要对APP软件用户评论信息中的情感倾向程度进行量化。

目前,量化用户的情感倾向程度的方法主要分为基于词典^[13]的方法和基于语料库的方法^[14]。然而,APP软件的用户评论是一种典型的网络评论,用户在针对APP软件的属性特征进行评论时往往会使用许多网络情感词汇,这给量化APP软件的情感倾向程度带来了一定的影响。因此,本文采用文献[15]中的量化方法,首先从情感词的极性出发,分析APP软件特征情感词对 f 中 Wh , Wd 和 Wa 三者的语法关系;然后结合《知网》情感词语料词典和网络情感词汇库,对情感倾向程度进行逐级量化计算;最后再计算每条评论信息的综合评分,将APP软件用户评论信息的情感倾向程度划分为5个等级:5,4,3,2和1。

通过以上方法,可以分析出用户对APP软件评论的侧重点和表达方式,针对特征的评论进行量化,从而得到评论的情感倾向程度,以分析出用户的评论模式 $P = \{\{\text{其它类特征}\}, \{v+d+a\}, \{1\}\}$ 。

5 实验结果及分析

5.1 数据来源及处理

为了验证本文方法的有效性,开发了一个爬虫工具,使用该工具从安卓电子市场¹⁾随机爬取了38款APP软件的用户评论(共计41293条,评论时间为2014年3月到2016年7月),并据此建立了APP软件信息库和APP软件用户评论库。

本文将ICTCLAS 2016作为数据预处理的工具,完成了评论分词及词性标注,并提取了APP软件特征情感词对,部分数据的处理结果如表1所列。

表1 数据处理结果(部分)

分词前	分词后	特征情感词对
QQ太差了	QQ/n 太/d 差/a 了/y	{QQ/n, 太/d, 差/a}
下载很慢	下载/v 很/d 慢/a	{下载/v, 很/d, 慢/a}

5.2 实验结果及分析

5.2.1 实验一

从APP软件信息库和APP软件用户评论库中随机选取了12款APP软件,共计11358条用户评论,并筛选出了9846条与该APP软件相关的用户评论信息。然后,根据本文第4节中的方法对用户评论信息进行分类、词性组合分析和情感倾向程度量化,得出每条用户评论信息的评论模式。相关实验结果数据如表2和表3所列。

¹⁾ <http://apk.hiapk.com/apps>

表 2 APP 软件用户评论信息中各类特征所占比例/%

APP 软件	QQ	微信	快手	百度视频	土豆视频	今日头条	QQ 音乐	美图秀秀	墨迹天气	安居客	搜狗输入法	讯飞输入法
属性特征	5.12	8.17	2.83	12.85	14.80	4.96	3.55	1.60	2.27	17.61	9.02	11.51
综合特征	27.50	29.35	58.48	25.71	17.91	43.69	36.03	52.40	45.90	18.87	27.44	26.66
其它类特征	67.38	62.48	38.69	61.44	67.29	51.35	60.42	46.00	51.83	63.52	63.54	61.83

表 3 APP 软件的用户评论模式 1(部分)

序号	APP 软件	用户 ID	用户评论信息	F	S	E	P
1	微信	597734266083872802	发不了消息	{属性特征}	{v+d+n}	{2}	{{属性特征},{v+d+n},{2}}
2	微信	601724559133118478	听不到语音	{属性特征}	{v+d+v+n}	{2}	{{属性特征},{v+d+v+n},{2}}
3	微信	596908575402627077	非常好	{综合特征}	{d+a}	{5}	{{综合特征},{d+a},{5}}
4	微信	590391633615392782	坑爹的微信	{综合特征}	{a_n+n}	{1}	{{综合特征},{a_n+n},{1}}
5	微信	602401519802912792	牛逼啊	{综合特征}	{a_n}	{5}	{{综合特征},{a_n},{1}}
6	微信	543218047436066818	下载费劲啊	{其它类特征}	{v+a}	{2}	{{其它类特征},{v+a},{2}}
7	QQ 音乐	561204814248284165	太赞了	{综合特征}	{d+a_n}	{5}	{{综合特征},{d+a_n},{5}}
8	QQ 音乐	566844475419267073	主题不好	{属性特征}	{n+d+a}	{2}	{{属性特征},{n+d+a},{2}}
9	QQ 音乐	558852571821903883	背景不好!	{其它类特征}	{n+d+a}	{2}	{{其它类特征},{n+d+a},{2}}
10	美图秀秀	588920691202592800	安装快快滴	{其它类特征}	{v+d+v}	{4}	{{其它类特征},{v+d+v},{4}}
11	美图秀秀	586415433419198472	空间就不能小点吗	{属性特征}	{n+d+v+a}	{2}	{{属性特征},{n+d+v+a},{2}}
12	美图秀秀	599079379372478471	太棒了	{综合特征}	{d+a_n}	{5}	{{综合特征},{d+a_n},{5}}

由表 2 可以看出,即使针对同一款 APP 软件进行评论,不同用户评论的侧重点也是不同的。大多数用户主要针对 APP 软件的综合特征或其它类特征进行评论,但也有部分用户对 APP 软件的属性特征进行评论。

其次,不同用户发表的评论信息的词性组合也有差异,由表 3 可以看出,每条用户评论信息对应的词性组合均不相同。其中,针对综合特征评论的词性组合相对比较简单,并且存在使用网络情感词汇进行评论的情况,如第 4 条和第 5 条用户评论信息。

此外,不同用户对 APP 软件的情感倾向程度不同,即使评论特征相同,最后的情感倾向程度也存在差异性。例如,表 3 中第 3 条和第 4 条用户评论信息虽然都是针对“{综合特征}”的,但是它们表达出来的情感倾向程度截然相反,分别为“{5}”和“{1}”。

5.2.2 实验二

为了验证 APP 软件用户评论模式识别方法的有效性,本文邀请了 11 位用户根据自己的实际使用情况对实验一中的 12 款 APP 软件进行了多次评论,然后运用本文的方法对 APP 软件的用户评论模式进行分析。实验结果如图 4(其中柱状上方数字表示评论条数)和表 4 所示。

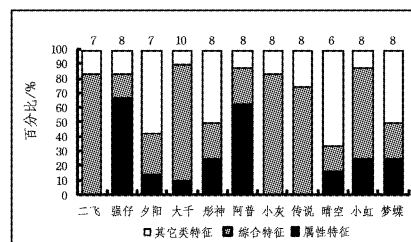


图 4 APP 软件用户评论信息的特征类别分布情况

表 4 APP 软件的用户评论模式 2(部分)

用户名	APP 软件	用户评论信息	F	S	E	P
二飞	QQ	太卡了	{综合特征}	{d+v}	{1}	
	微信	很好用,	{综合特征}	{d+v}	{5}	
	今日头条	牛逼	{综合特征}	{a_n}	{5}	{{综合特征},{[n]+d+a v},{5}}
	畅读	蛮好的...	{综合特征}	{d+a}	{5}	
	搜狗输入法	下载很慢!!!	{其它类特征}	{v+d+a}	{1}	
强仔	快手	快手很不错哟	{综合特征}	{n+d+a}	{5}	
	微信	发不了消息	{属性特征}	{v+d+n}	{2}	
	QQ	看不了图片了!!	{属性特征}	{v+d+n}	{2}	
	畅读	很不错	{综合特征}	{d+a}	{5}	{{属性特征},{v+d+v n},{2}}
	今日头条	图片不清晰...	{属性特征}	{n+d+a}	{2}	
夕阳	安居客	搜索不行	{属性特征}	{v+d+v}	{2}	
	墨迹天气	安装不了	{其它类特征}	{v+d}	{2}	
	QQ 音乐	安装很慢啊!	{其它类特征}	{v+d+a}	{1}	
	百度视频	反应迟钝	{其它类特征}	{v+a}	{2}	
	土豆视频	浪费流量	{其它类特征}	{v+n}	{2}	
梦蝶	QQ	界面挺美观的	{其它类特征}	{n+d+a}	{4}	{{其它类特征},{v+[d]+a n},{2}}
	微信	超赞	{综合特征}	{a_n}	{5}	
	今日头条	赞一个!	{综合特征}	{a_n}	{5}	
	搜狗输入法	键盘不能使用!	{属性特征}	{n+d+v}	{2}	

由图 4 和表 4 可以看出,每个用户都有自己的评论侧重点,并且表达方式和评判标准都是有差异的。例如,表 4 中的“二飞”用户的评论信息中有其它类特征和综合特征,但是他更侧重于 APP 软件的综合特征,其情感倾向程度普遍偏高,

评论所用的词性组合虽然存在一些差异,但大多用“{[n]+d+a|v}”词性组合发表评论信息。用户“强仔”则更关心 APP 软件的属性特征,并且对 APP 软件的要求较高,所表达出来的情感倾向大多是负面的,通常用“{v+d+v|n}”词性组

合发表评论信息。而用户“夕阳”更侧重于针对其它类特征进行评论,虽然其表达方式的词性组合比较多样化,但大多使用“{v+[d]+a|n}”的词性组合,其情感倾向程度波动比较大。

通过实验一和实验二可以看出,本文提出的用户评论模式分析方法是有效的,可以较为准确地分析出用户评论的特征类别、词性组合以及情感倾向程度。

结束语 本文提出了一种 APP 软件的用户评论模式分析方法,首先通过分析用户评论信息和 APP 软件信息之间的关系,将用户评论信息分为 3 类;然后分析每类用户的评论信息的词性组合;最后量化每条用户评论信息的情感倾向程度,以分析出 APP 软件的用户评论模式。通过实验证明了本文方法的有效性。但该方法也存在不足,即当用户对 APP 软件的评论特征类别比较分散时只做了简单的频率计算,并没有做深入分析。因此,下一步将对评论特征类别的分析、网络情感词汇的收集以及 APP 软件的用户可信度计算等方面进行深入研究。

参 考 文 献

- [1] JIANG H, MA H, REN Z, et al. What makes a good app description? [C]//Proceedings of the 6th Asia-Pacific Symposium on Internetware on Internetware. ACM, 2014:45-53.
- [2] MA S, WANG S, LO D, et al. Active Semi-Supervised Approach for Checking App Behavior Against Its Description[C]// 2015 IEEE 39th Annual Computer Software and Applications Conference (COMPSAC). IEEE, 2015:179-184.
- [3] LI M, WANG X C, ZHANG J, et al. Study on Check-in and Related Behaviors of Location-based Social Network Users [J]. Computer Science, 2013, 40(10):72-76. (in Chinese)
李敏, 王晓聪, 张军, 等. 基于位置的社交网络用户签到及相关行为研究[J]. 计算机科学, 2013, 40(10):72-76.
- [4] WANG W, WANG H W, MENG Y. The collaborative filtering recommendation based on sentiment analysis of online reviews [J]. Systems Engineering-Theory & Practice, 2014, 34(12):3238-3249. (in Chinese)
王伟, 王洪伟, 孟园. 协同过滤推荐算法研究:考虑在线评论情感倾向[J]. 系统工程理论与实践, 2014, 34(12):3238-3249.
- [5] ZHANG L, QIAN G Q, FAN W G, et al. Sentiment Analysis Based on Light Reviews [J]. Journal of Software, 2014, 25(12):2790-2807. (in Chinese)
张林, 钱冠群, 樊卫国, 等. 轻型评论的情感分析研究[J]. 软件学报, 2014, 25(12):2790-2807.
- [6] LIU Y, LIAO X W, LIU Y. The impact of online review on software and platform's pricing strategies [J]. Journal of Systems Engineering, 2014, 29(4):560-570. (in Chinese)
刘洋, 廖貅武, 刘莹. 在线评论对应用软件及平台定价策略的影响[J]. 系统工程学报, 2014, 29(4):560-570.
- [7] PAGANO D, MAALEJ W. User feedback in the appstore: An empirical study[C]//International Conference on Requirements Engineering. 2013:125-134.
- [8] GUZMAN E, MAALEJ W. How do users like this feature? a fine grained sentiment analysis of app reviews[C]//2014 IEEE 22nd International Requirements Engineering Conference (RE). IEEE, 2014:153-162.
- [9] ANAM A S M I, YEASIN M. Accessibility in smartphone applications: what do we learn from reviews? [C]//Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, 2013:35.
- [10] KHALID H. On identifying user complaints of iOS apps[C]//International Conference on Software Engineering. 2013:1474-1476.
- [11] HU Z K, ZHENG X L, WU Y F, et al. Product recommendation algorithm based on users' reviews mining [J]. Journal of Zhejiang University (Engineering Science), 2013, 47(8):1475-1485. (in Chinese)
扈中凯, 郑小林, 吴亚峰, 等. 基于用户评论挖掘的产品推荐算法[J]. 浙江大学学报(工学版), 2013, 47(8):1475-1485.
- [12] LEVENSHTAIN V I. Binary codes capable of correcting deletions, insertions, and reversals [J]. Soviet Physics Doklady, 1966, 10(8):707-710.
- [13] LI X N, ZHANG S W, YANG L, et al. ARES: Autoregressive Emotion-Sensitive Model for Predicting Sales Performance [J]. Journal of Computer Research and Development, 2013, 50(8):1722-1727. (in Chinese)
李雪妮, 张绍武, 杨亮, 等. ARES:用于预测的情感感知自回归模型[J]. 计算机研究与发展, 2013, 50(8):1722-1727.
- [14] LIN Q H. Design and Implementation of the Product Reviews Analysis System Based on Affective Computing [D]. Shanghai: Fudan University, 2013. (in Chinese)
林钦和. 基于情感计算的商品评价分析系统设计与实现[D]. 上海:复旦大学, 2013.
- [15] RAN M, JIANG Y, XIANG Q, et al. Method of Consistency Judgment for App Software's User Comments[C]//International Conference of Young Computer Scientists, Engineers and Educators. Springer Singapore, 2016:470-483.
- [11] WANG Y H. Improvement and Application of Random Forest Algorithm in recommender systems [D]. Hangzhou: Zhejiang University, 2016. (in Chinese)
王宇恒. 推荐系统中随机森林算法的优化与应用[D]. 杭州:浙江大学, 2016.
- [12] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1):5-32.
- [13] ZHANG Y, GAO Q Q. Water quality evaluation of Chaohu Lake based on random forest method [J]. Chinese Journal of Environmental Engineering, 2016, 10(2):992-998. (in Chinese)
张颖, 高倩倩. 基于随机森林分类算法的巢湖水质评价[J]. 环境工程学报, 2016, 10(2):992-998.

(上接第 167 页)

- [9] FANG K, ZHANG Q, ZHUANG Z, et al. Making Recommendations Better: The Role of User Online Purchase Intention Identification[C]// 2016 International Conference on Software Networking (ICSN). IEEE, 2016:1-4.
- [10] LIU N, CHEN Y T, YU H Q, et al. Traffic Flow Forecasting Method Based on Elman Neural Network [J]. Journal of East China University of Science and Technology, 2011, 37(2):204-209. (in Chinese)
刘宁, 陈昱颀, 虞慧群, 等. 基于 Elman 神经网络的交通流量预测方法[J]. 华东理工大学学报, 2011, 37(2):204-209.