

文章编号: 2095-2163(2020)06-0218-04

中图分类号: TP305

文献标志码: A

面向豆瓣网影评的水军识别研究

王 亚

(许昌学院, 河南 许昌 461000)

摘要: 豆瓣网影评中存在很多恶意刷高分或恶意贬低电影的水军用户, 这些用户的存在严重干扰了正常用户使用豆瓣网进行信息的获取与交流。本文通过分析豆瓣网水军用户与正常用户的区别, 提取用户静态行为特征和动态行为特征, 采用逻辑回归的二分类方法对这两类特征进行模型训练, 并将这两类特征融合进行模型训练, 以最大程度的提高识别影评水军用户的有效性。实验表明, 本文抽取的这两类特征对水军用户识别的准确率最高可达近 67.8%。

关键词: 水军用户; 特征提取; 逻辑回归

Research on the identification of water army facing douban.com film critics

WANG Ya

(College of Information Engineering, Xuchang University, Xuchang Henan 461000, China)

[Abstract] There are many spam users who maliciously brush high scores or maliciously degrade movies in Douban website. The existence of these users seriously interferes with normal users' use of Douban for information acquisition and communication. This paper analyzes the differences between Douban spam users and normal users, extracts user static behavior characteristics and dynamic behavior characteristics, uses logistic regression binary classification to perform model training on these two types of features, and fuses the two types of features for model training, which maximize the effectiveness of identifying film review spam users. Experiments show that the accuracy of the two types of features extracted in this paper for spam users is up to 67.8%.

[Key words] Spam users; Feature extraction; Logistic regression

0 引言

随着互联网社交网络的发展, 用户在虚拟社区中分享或表达自己的观点已成常态, 如在微博、博客和各类相关网站中, 用户可以撰写自己对某一事件的看法, 也可以对产品或服务进行评论。用户在虚拟社区中的言论自由是用户的基本权利, 但与此同时, 也存在由恶意众包下的网络水军组成的众多垃圾用户, 他们受雇于不法商家或不良组织或个人, 在社交网络中进行恶意交互活动, 如发布垃圾贴、广告贴, 从事话题炒作、事件营销、发布虚假评论等活动, 这些网络水军的存在干扰了网络舆情传播, 严重影响了虚拟社区正常的互动分享与生态平衡。因此, 识别和检测虚拟社区的垃圾用户已经成为了一个重要的议题。

豆瓣网是一个提供书影音推荐、线下同城活动、小组话题交流等多种服务功能于一体的创新网络服务社区网站。而豆瓣电影是中国最大与最权威的电影分享与评论社区, 收录了百万条影片与影人的资料, 有 2500 多家电影院加盟, 更汇聚了数千万热爱电影的人。而影评在一定程度上影响着电影的票

房, 很多人在选择一部电影之前, 会习惯性的打开评分网站, 查看该电影的评分或影评, 评分过低或影评太差可能就不在观影范围之内, 对于已经存在的高评分电影, 用户会呈现出更高的观影倾向。在这种利益的推动下, 当前有很多即将上映的电影为了获取更高的票房, 雇佣大量的网络水军在豆瓣影评中进行虚假评论, 恶意抬高电影评分。同时, 还有购买虚假评论对同期上映的电影进行恶意差评的行为。所有这些发布的虚假的、不符合实际的评论信息被称为垃圾评论信息, 发布这些虚假评论的用户被称为水军用户。虚假的影评信息会打破电影生态环境, 引起电影票房的不公平竞争。

对于垃圾用户识别和检测的研究, 目前国内外学者已经取得了一定的成果。其中, 对于电子邮件^[1]、电话网络^[2]、网站论坛^[3]等类型的研究开展的较早, 而针对 Facebook、Twitter、微博等新型社交网络虚拟社区中的垃圾检测则是近几年兴起的研究热点。在虚拟社区中, 垃圾用户的检测通常从多种信息源出发, 例如: 用户资料、微博内容、图结构等。而对于此类垃圾用户的检测, 如何找到好的用户特

基金项目: 许昌学院校级科研项目(2020YB017); 河南省高等院校重点科研项目(19A520036)。

作者简介: 王 亚(1986-), 女, 硕士, 讲师, 主要研究方向: web 数据挖掘。

收稿日期: 2020-03-26

征无疑是研究的重点之一。当前已经存在的工作主要是基于垃圾用户发布的内容评论进行特征的抽取,如复制的评论、评论词汇量的大小、评论的长度,是否进行了多次评论、评分、用户的联系方式(如URL或电话)、评论情感分等^[4-7],另外一类方法则基于用户的行为特征,如用户的活跃天数和活跃的平均时间段^[8],用户注册日期^[7]等。除了文本特征和行为特征外,还有用户肖像特征,如粉丝数和关注数^[8]、发布的评论数或发表的博文数^[7]等也会被用于垃圾用户的检测。相比于文本特征,用户肖像特征更容易得到。但以上对于垃圾用户的检测多用于对微博垃圾用户的判断,而对于豆瓣网影评垃圾用户的检测目前还未发现有文章涉及,微博用户和豆瓣网影评用户在用户信息、行为特征方面并不完全相同,因而本文旨在对豆瓣网影评数据进行分析,抽取特征,从而完成豆瓣网影评水军的识别,其主要贡献如下:

(1) 本文通过对获得的豆瓣网影评数据进行用户信息和行为分析,提取了用户静态行为特征和动态行为特征作为用户的基本特征。

(2) 本文首先人工判别出部分水军用户和正常用户,以逻辑回归作为二分类方法对这些数据进行模型训练,并以这些作为训练数据以逻辑回归方法进行模型训练。同时对多个特征进行结合以查看模型的训练效果,从而最大程度的提高豆瓣网影评水军的识别效率。

实验表明采用本文提取的相关特征进行训练和学习,对于豆瓣网影评水军的识别准确率能达到将近67.8%,这足以证明本文提取特征的可行性和有效性。

1 豆瓣网用户特征提取

对豆瓣网数据进行采集,获得了指定的电影信息、电影评论信息、评论用户基本信息和评论用户关系信息等构成的数据集。通过对数据集中用户的影评进行分析发现,豆瓣网影评水军用户一般具有如下静态行为特征和动态行为特征:

(1) 影评水军用户一般受雇于不良组织,在社交网络中发布虚假评论,此类用户要么是为专门发表虚假影评评论而注册的用户,或者是一些专业的水军用户,已广泛的发表虚假评论,这些用户一般具有较少的粉丝,其关注的用户一般也较少,另外,其基本信息资料也会存在一定程度的不完善性,基于这些静态的特征,将提取相关的特征,构建基于静态行为特征的水军识别模型,以进行水军用户的识别。

(2) 影评水军用户为了进行电影营销,恶意抬高或贬低电影。为了影响网络民意,制造较强的舆论导向,水军用户所发布的评论信息的情感度一般较为激烈。另外,其以利益为目的,评论频率、评论时间、评论打分、评论内容都明显起到了引导舆论倾向的目的。基于这些动态行为特征,构建相应的动态行为特征模型,以进行水军用户的识别。

1.1 豆瓣网用户静态行为特征

用户静态行为特征主要是指用户的基本信息,如:用户的自我介绍、常居地、个性签名、粉丝数、关注数等静态特征。一般来说,正常用户只有一个账号,而水军用户则存在多个账号,主要区别在于基本信息的完整度。基于此,建立三个特征 *desc*, *local_name*, *signature* 分别表示用户是否填写了自我介绍、常居地与个性签名,另外建立了三个特征 *desclen*, *locnamelen*, *signaturelen* 分别表示自我介绍、常居地、个性签名的长度。这三个长度特征由于阈值与其他特征有区别,因此采用 *max_min* 线性函数使其归一化为特征 *desclen**, *locnamelen**, *sighlen**, 见公式(1)、(2)和(3)。

$$desclen^* = \frac{desclen - min}{max - min}, \quad (1)$$

$$locnamelen^* = \frac{locnamelen - min}{max - min}, \quad (2)$$

$$sighlen^* = \frac{signlen - min}{max - min}. \quad (3)$$

通过分析数据集,为用户ID与备用ID建立一个特征 *id_uid*,表示该用户的用户ID与备用ID是否相等,一般地,普通用户可能会不一致,而水军用户为降低成本,则会采用默认形式保持一致。

抽取 *flowerCount* 和 *friendCount* 作为粉丝数量和关注人数量的特征,并采用 *min_max* 线性函数将其归一化为特征 *flowercount**, *friendcount**, 见公式(4)和(5)。

$$flowercount^* = \frac{flowerCount - min}{max - min}, \quad (4)$$

$$friendcount^* = \frac{friendCount - min}{max - min}. \quad (5)$$

1.2 豆瓣网用户动态行为特征

用户的动态行为特征主要是用户对影评的评论特征,包括:影评时间、影评部数、影评分数等。

将用户发表影评时间与用户注册账号时间的时间差、以及电影上映时间与电影评论时间的时间差分别作为一个特征,从而建立一个特征值 T_{c-u} , 用

来表示用户发表评论时间与用户注册账号时间的时
间差,如果该用户发表了多条电影评论,则 T_{c-u} 为
时间差的平均值,见公式(6)。建立另一个特征值
 T_{c-r} ,用来表示用户发表影评时间与电影上映时
间的时间差,如果该用户发表了多条电影评论,则 T_{c-r}
为时间差的平均值,具体见公式(7)。

$$T_{c-u} = \frac{\sum_{i=1}^m | \text{commenttime}_{film_i} - \text{time}_{register} |}{m}. \quad (6)$$

其中: m 代表评论的电影总数, $\text{commenttime}_{film_i}$
表示对第 i 部电影的评论时间, $\text{time}_{register}$ 表示该用
户的注册时间。

$$T_{c-r} = \frac{\sum_{i=1}^m | \text{commenttime}_{film_i} - \text{releasetime}_{film_i} |}{m}. \quad (7)$$

其中: m 代表评论的电影总数, $\text{commenttime}_{film_i}$
表示对第 i 部电影的评论时间, $\text{releasetime}_{film_i}$ 表示
第 i 部电影的上映时间。

影评信息里的星级评价同影评平均分的差值也
能作为一个特征,建立一个特征值 commentrating ,
用来表示用户对影评信息中的星级评价同电影平均
分的偏差,如果一个账号评论了多部电影,则
 commentrating 表示每部电影的评分与该电影平均
分的差值,求和再进行平均,所得的值,具体见公式
(8)。

$$\text{commentrating} = \frac{\sum_{i=1}^m | \text{commentrating}_{film_i} - \text{avgrating}_{film_i} |}{m}. \quad (8)$$

其中: m 代表用户评分的电影总数,
 $\text{commentrating}_{film_i}$ 表示用户对第 i 部电影的评分,
 $\text{avgrating}_{film_i}$ 表示第 i 部电影的平均评分。

本文提取的豆瓣影评数据中有七部电影,因而
对于每个用户建立一个 7 维向量用以表示用户对哪
一部电影做出了影评, $\text{film} = (f_1, f_2, \dots, f_i)$ 表示电
影合集,对每个用户建立一个向量 $v_{\text{film}} = (vf_1, vf_2,$
 $\dots, vf_i)$,其中 vf_i 表示该用户是否对第 i 部电
影做出了评论,由于电影数选择为 7 部,因此, v_{film} 是
一个 7 维向量。另外建立特征 $\text{total}_{\text{film}}$ 表示用户
评论的电影的总数,该值小于等于 7。

建立了一个特征值 commentvote ,用来表示用
户发表的影评所获得的赞数,如果用户发表了多条
影评,则 commentvote 表示每部影评所获得的平均赞

数。见公式(9)。

$$\text{commentvote} = \frac{\sum_{i=1}^m | \text{commentvote}_{film_i} |}{m}. \quad (9)$$

其中: m 代表该账号下发表的评论被赞的电影
总数, $\text{commentvote}_{film_i}$ 表示用户评价的第 i 部电
影的影评获得的赞数。对该投票数进行线性函数归
一化,获得特征 commentvote^* ,见公式(10)。

$$\text{commentvote}^* = \frac{\text{commentvote} - \text{min}}{\text{max} - \text{min}}. \quad (10)$$

2 豆瓣网数据采集与数据标记、实验结果

(1)数据采集与数据标记。本文采集了豆瓣七
部电影的影评信息,获取了 126880 个用户的基本信
息,结合用户的基本信息、影评信息以及社交网信
息,人工标注出一些垃圾用户样本和普通用户样本
作为研究对象。标注出的 2458 个样本信息,其中垃
圾用户 897 个,正常用户 1561 个。在实验中,为了
保持样本的均衡性,选取了 897 个垃圾用户和 897
个正常用户参与训练。实验采用五折交叉验证。

(2)实验结果。本实验根据提取的用户特征,
采用逻辑回归的分类方法进行模型的训练,回归参
数采用默认值,使用准确率、精确率、召回率、F1 值
这四种指标对模型进行评价。对不同的特征进行融
合,进行模型的训练,所得的实验结果如表 1 所示。

表 1 对不同特征融合进行模型训练所得的实验结果

Tab. 1 Experimental results of model training on different feature fusions

特征	Accuracy	Precision	Recall	F1
用户静态行为特征	0.617 08	0.592 06	0.762 51	0.665 37
用户动态行为特征	0.670 57	0.659 12	0.744 7	0.692 99
用户静态行为特征+ 用户动态行为特征	0.677 84	0.659 13	0.741 37	0.697 02

从实验结果可以看出,融合用户的不同特征进
行训练将得到不同的模型结果。仅用户的静态行为
特征参与进行模型的训练,识别水军的准确率可以
达到 61.7%。而仅用户的动态行为特征参与进行模
型训练,识别水军的准确率可达到 67.1%。如果将
用户动态行为特征和静态行为特征融合进行模型的
训练,识别水军的准确率可提高到 67.8%,而模型精
确率和 F1 值也都有些许提高。从实验结果可以看
出,用户的动态行为特征在水军用户方面有比较显
明的优势,而单纯的采用静态行为特征效果并不是
太好,将这两类特征进行融合检测水军的准确率将
提高到 67.8%,具有较好的效果。(下转第 224 页)