

学校代码: 11482

学 号: 160506010002



浙江财经大学

ZHEJIANG UNIVERSITY OF FINANCE AND ECONOMICS

# 硕士学位论文

MASTER THESIS

论文题目 社交网络水军识别及其在营销信息

传播模型中的应用研究

作者姓名 李岩

专 业 管理统计学

所在学院 信息管理与工程学院

指导教师 邓胜春

完成日期 2019 年 1 月

# 硕士学位论文

## 社交网络水军识别及其在营销信息传播模型中的应用研究

2019 年 1 月

# **MASTER THESIS**

## **Spammer Detection and its Application Research in Marketing Information Dissemination Model on Social Network**

**January 2019**



## 浙江财经大学硕士研究生学位论文原创性声明

本人郑重声明：

此处所提交的硕士学位论文是本人在导师指导下，在浙江财经大学攻读硕士学位期间独立进行研究所取得的成果。据本人所知，论文中除已注明部分外，不包含他人已发表或撰写过的研究成果，对本文的研究工作做出重要贡献的个人和集体均已注明。本声明的法律结果将完全由本人承担。

作者签名：李岩

日期：2019年1月17日

## 浙江财经大学硕士研究生学位论文使用授权书


本论文系本人在浙江财经大学攻读硕士学位期间在导师指导下完成的硕士学位论文。本论文的研究成果归浙江财经大学所有，本论文的研究内容不得以其他单位的名义发表。本人完全了解浙江财经大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权浙江财经大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

作者签名：李岩

日期：2019年1月17日

经审查，确认该论文已符合浙江财经大学硕士学位论文的要求。

导师签名：



日期：2019年1月17日

## 摘要

随着社交网络的不断发展与渗透，使之拥有数以亿计的用户群体。企业经常利用如此巨大规模用户群体的优势，在社交网络上进行产品推广、品牌维护、建立客户关系等营销活动。然而，看似平静的社交网络中却存在着一股“暗流”，即水军用户。它们在利益的驱使下，有目的地大规模地产生点赞、评论、转发等虚假行为，从而达到舆论引导、产品推广以及控制口碑等目的，甚至有些会诋毁竞争对手企业以及相关产品。水军用户的存在不仅严重地影响了公平的社交网络营销环境，而且有些甚至违背商业道德，造成恶劣的影响。

传统研究着重利用静态行为特征来识别水军用户，随着水军策略的不断演变，越来越趋近于正常用户，此类方法已无法有效地检测出愈发隐蔽的水军用户。因此，观察和分析水军用户的动态行为愈发重要；此外，传统模型大多只针对在离线场景下的水军用户检测，由于社交网络数据的持续性使得传统模型难以应用其中，无法适应社交网络在线检测的场景。

由于水军用户在社交网络上产生大量有目的性的行为，干扰营销信息传播的过程，扰乱公平的社交网络营销环境。然而前人在社交网络营销信息传播模型的工作中，较少细粒度地考虑到水军用户作用的影响因素。只有细粒度地研究水军用户对营销信息传播过程的影响，发现水军用户内在的作用机理，才能有效治理与应对水军带来的危害，从而维护公平、有序的社交网络营销环境。

本文的主要工作与创新性如下：

1) 针对水军策略的不断演变、愈发隐蔽以及传统模型难以应用于社交网络在线检测场景的问题，本文使用计算波动率的方式构造了社交网络用户的动态行为特征，构造的特征计算量小、便于实际运用；在离线环境中对比验证了水军用户与正常用户动态行为特征的差异性与有效性；再结合用户的动静态性行为特征与半监督模型 Tri-Training 构建了在线水军检测模型。因此，本文不仅把水军检测领域中对静态行为特征的研究推广到了对动态行为特征的研究，并且弥补了在线检测场景下的研究不足。

2) 针对水军用户会干扰营销信息传播过程从而扰乱社交网络营销环境的问题，由于在社交网络营销信息传播模型中较少有工作细粒度地考虑水军用户的影响，因此本文以基于主体的 IC 模型为基础，结合社交网络营销理论构建了营销信息传播模型；并且在模型中纳入水军的影响因素，构建了考虑水军用户的营销信息传播模型。

3) 最后通过抓取新浪微博用户的数据，在离线环境中对比分析了用户的动态行为特征的差异性与有效性，并且验证了在线水军检测模型的有效性：F1 值为 93.33%，平均训练时间约为 2 分钟；通过抓取新浪微博营销信息的数据，验证了考虑水军用户营销信息传播模型的有效性：热度与曝光量更符合实际传播结果，并且分析了水军用户对营销信息传播过程的影响，以及结合模型的参数，为个人、企业及有关部门提供相关治理意见。

**关键词：**社交网络；水军识别；动态行为；Tri-Training 模型；在线检测；IC 模型；营销信息传播模型

## ABSTRACT

With the continuous development and penetration of social network, it has owned hundreds of millions of users. Enterprises often take advantages of such large-scale users to promote products, maintain brands, and build customer relationships on social network. However, there exists a kind of users, namely spammers. Driven by interests, they purposefully produce larger-scale false behaviors such as praise, comment and forward, in order to guide public opinions, promote products and control word-of-mouths. Some even vilify competitors and related products. The existence of spammer not only seriously affects the fair marketing environment of social network, but also some even violates business morality, which can lead to extremely bad influence.

Traditional research focuses on the user static behavior features to detect spammers. As strategy of spammers evolves, spammers become more and more close to normal users, such method has been unable to effectively detect increasingly hidden spammers. Therefore, it is increasingly important to observe and analyze the dynamic behavior features of spammers. Moreover, traditional model detects spammers mostly only in offline environment. Due to the persistence of social network data, traditional model is difficult to apply and cannot adapt to the environment of online detection of social network.

Spammers generate a large number of purposeful behaviors on social network, which interfere with the process of marketing information dissemination and disrupt the fair marketing environment of social network. However, in the work of marketing information dissemination model of social network, predecessors have less considered the influencing factor of spammer in a fine-grained manner. Only by researching the impact of spammer on the process of marketing information dissemination in a fine-grained manner, finding the internal mechanism of the spammers, can we effectively govern and deal with the harm brought by spammers and maintain a fair and orderly social network marketing environment.

The main work and innovation of this paper are as follows:

- 1) Because spammers strategy is becoming more constantly changing and concealed, and traditional model is difficult to apply to the online detection environment of social network, this paper used calculating-volatility method to construct the dynamic behavior features of the social network users. It has low time complexity and convenient for

practical use; the difference and effectiveness of the dynamic behavior features between spammers and normal users were compared and verified in offline environment; then the online spammer detection model was constructed by combining the user's dynamic and static behavior features with the semi-supervised model Tri-Training. Therefore, this paper not only extended the study of behavior features in spammer detection field: from static behavior features to dynamic behavior features, but also made up for the research deficiency in online detection environment.

2) In view of the problem that spammers will interfere with the process of marketing information dissemination and disturb the social network marketing environment, there are few works to take into account the influence of the spammer in a fine-grained manner in the marketing information dissemination model of social network. So, based on the Agent-Based IC model, this paper constructed a marketing information dissemination model with social network marketing theory; a marketing information dissemination model considering spammer was also constructed by incorporating influencing factor of the spammer into the model.

3) Finally, by crawling the user data of Sina Weibo, the differences and effectiveness of the user's dynamic behavior features were compared and analyzed in the offline environment, and the effectiveness of the online spammer detection model was verified: the F1 value is 93.33%, average training time is about 2 minutes; by crawling the marketing information data of Sina Weibo, the effectiveness of the marketing information dissemination model considering spammer was verified: the heat and exposure are more closer with the actual dissemination results, in addition, the impact of spammer on the process of marketing information dissemination was analyzed and relevant governance advice for individuals, enterprises and relevant departments was provided with the parameters of the model.

**Keywords:** social network; spammer detection ; dynamic behavior; Tri-Training model; online detection; IC model; marketing information dissemination model



# 目录

|                                |    |
|--------------------------------|----|
| 第一章 绪论.....                    | 1  |
| 第一节 研究背景和意义.....               | 1  |
| 第二节 国内外研究现状.....               | 2  |
| 第三节 研究目标与创新点.....              | 6  |
| 第四节 主要内容与结构安排.....             | 6  |
| 第二章 基础理论与模型.....               | 8  |
| 第一节 数据挖掘基础理论.....              | 8  |
| 第二节 社交网络营销信息传播模型基础理论.....      | 12 |
| 第三章 社交网络水军用户的动态行为分析与在线检测.....  | 15 |
| 第一节 社交网络水军用户检测.....            | 15 |
| 第二节 动态行为特征定义与分析.....           | 16 |
| 第三节 在线水军检测模型.....              | 18 |
| 第四节 实验分析.....                  | 21 |
| 第五节 本章小结.....                  | 27 |
| 第四章 社交网络水军识别在营销信息传播模型中的应用..... | 29 |
| 第一节 营销信息传播过程分析.....            | 29 |
| 第二节 只存在正常用户的营销信息传播模型.....      | 31 |
| 第三节 考虑水军用户的营销信息传播模型.....       | 33 |
| 第四节 实验分析.....                  | 35 |
| 第五节 本章小结.....                  | 43 |
| 总结与展望 .....                    | 44 |
| 参考文献 .....                     | 46 |
| 作者在读期间发表的学术论文 .....            | 51 |
| 致谢 .....                       | 52 |

# 第一章 绪论

社交网络为企业带来了新的商业机会，与传统营销媒介相比，社交网络由于其互动性、分享性、放大性等特点，企业在其上发布的营销信息能够高速、广泛的进行传播，因此企业能够以更低的成本、更好的效果在社交网络上完成品牌建立、产品推广等营销活动。然而社交网络上也存在水军用户，它们或被企业自身利用、或被竞争对手利用，极大地影响了社交网络正常的营销信息传播过程，扰乱了营销环境。因此如何有效地识别水军用户以及水军用户如何影响了营销信息传播过程从而扰乱营销环境是摆在现实面前需迫切解决和研究的问题。

本章首先阐述本文的研究背景及研究意义，综述现有的研究成果，定位本文的研究目标，即对社交网络水军用户进行识别及其在营销信息传播模型中的应用研究，最后给出本文的主要内容及结构安排。

## 第一节 研究背景和意义

### 一、研究背景

随着互联网技术、内容的不断发展，互联网用户群体也变得越来越庞大，截止至 2018 年 6 月，我国的网民规模达到了 8.02 亿，互联网普及率为 57.7%（中共中央网络安全和信息化委员会办公室，2018）。其中，社交网络作为互联网最重要的平台，在内容分享、娱乐交友、电子商务等各方面发挥不可替代的作用。例如，“社交+”的商业模式的应用：电商平台通过与社交网络平台合作或利用电商本身的社交功能，使商品通过用户的好友链以分享的方式进行传递，从而吸引更多的用户购买商品，降低引流成本，提高销量；新型营销方式网络营销的应用：企业在社交网络平台上发布产品、活动等营销信息，通过社交网络信息裂变的传播方式，使得营销信息以指数级的方式传播，从而高效地达到产品宣传、品牌维护等目的。

社交网络已经渗透于我们生活的方方面面，主要包括以下几类：音视频音乐社交网络、论坛社交网络、知识问答社交网络、新闻阅读社交网络、直播社交网络、微博社交网络等。社交网络种类繁多、内容丰富、用户基数大，为人们提供了良好的平台环境，极大便利了人们的生活。但也产生了一些利益驱动的水军用户，它们在利益的驱使下，在社交网络上有目的地大规模产生点赞、评论、关注、转发等行为，达到舆论引导、产品推广以及控制口碑等目的。不仅如此，它们还利用社交网络环境，通过提高点击量、关注量、制造虚假评论等手段，扭曲诋毁企业产品，损

害名誉，或人为制造公众话题，给相应的企业及其产品在口碑及声誉等方面造成负面的影响，直接导致产品销售的急剧下滑。社交网络上的水军行为严重地影响了民众的辨识力、扰乱了社交网络秩序，急需识别与治理。

目前，社交网络水军识别技术已成为数据挖掘领域较为活跃的研究方向之一。如何从海量的社交网络用户信息中挖掘出潜藏的水军特征与行为模式，从而识别出水军，已成为一项非常重要且具有挑战性的工作（莫倩和杨珂，2014）。现阶段社交网络水军识别的研究工作着重围绕社交网络用户的静态行为特征，使用的方法侧重于传统的统计学、数据挖掘方法，上述方法已无法有效地检测出社交网络平台上水军用户日益隐蔽的行为模式并且难以应用于在线检测场景中。此外，社交网络水军通过制造或传播虚假信息、炒作话题、提高用户影响力等恶意行为，对社交网络上的新闻内容传播、营销信息传播等众多方面产生重大的影响，尤其在营销信息传播领域。已有研究表明，正常用户的网络购买行为严重受到网络水军的影响（郑春东等，2015）。因为社交网络水军的存在，干扰了营销信息传播的过程，为社交网络营销的正常良性发展带来了极大的隐患，引起越来越多的误导与纠纷，给社交网络用户和平台都带来了难以估计的损失。

## 二、研究意义

由于水军用户严重影响视听，扰乱社交网络正常的营销环境，不仅影响正常用户对社交网络平台的使用意愿和信用程度，更会影响整个社交网络行业的健康发展。因此，社交网络平台急需有效的水军检测方法，不仅可以准确地检测出行为日益复杂、隐蔽性愈来愈高的水军用户，还需要能够应用于社交网络在线检测的场景中；此外，由于营销信息传播过程中存在水军用户的作用，严重干扰了营销信息传播的过程，因此，需细粒度地研究水军用户如何影响正常的营销信息传播过程。只有充分了解水军作用的内在机理，才能有效应对，防范自如。

## 第二节 国内外研究现状

### 一、社交网络水军识别

水军从最早的邮件领域、发展到电子商务领域，再到社交网络领域，可见，水军的发展牢牢紧跟着互联网的发展，它们的出现有一定的时代背景。在邮件领域、电子商务领域对水军识别的研究较多，而在拥有庞大用户规模的社交网络领域对水军识别的研究却不足，并且面对社交网络数据规模性与实时性的特点，其他领域的水军识别方法难以应用其中。因此，本文的研究定位于社交网络领域中的水军识别。

国内的文献把社交网络水军识别定位在社交网络学科下（李立耀等，2015），主要的研究思路为：依托新浪微博平台的数据，通过观测水军行为，构造出识别水军的特征，把水军识别看作二分类问题，使用监督学习、半监督学习、无监督学习等方法对其进行识别。张良（2015）通过构造用户账号属性、行为属性并使用逻辑回归的方法识别社交网络水军；张艳梅等（2017）通过构造水军新的特征指标并使用基于遗传算法的贝叶斯模型对社交网络水军进行识别；程晓涛等（2015）通过构造社交网络关系图特征并使用朴素贝叶斯、贝叶斯网络以及决策树对社交网络水军进行识别，通过利用监督学习模型对水军进行识别，能够达到较为不错的精确率和召回率。韩忠明等（2013）通过使用概率图将用户属性和行为属性连接起来，视水军为隐变量的方法识别社交网络水军，该方法能够有效地发现社交网络上的水军。陈侃等（2015）挖掘了不同传播主体间的交互模式并使用决策树建模，可以有效地对社交网络水军进行识别。谭侃等（2017）利用主动学习方法对未标记样本进行价值评估，使用排序和聚类相结合的双层采样算法选出最有价值的样本给专家标注，使得模型在只有少量有标记样本的条件下能达到较好的识别效果。孙卫强（2015）增加了社交网络用户的时间维度并使用深度信念网络识别社交网络水军，使得模型不仅具有较高的准确率，而且具有较好的鲁棒性。对于水军团体的检测，叶施仁等（2017）使用水军用户高度聚集性以及与实际用户之间关系稀疏性的特点对水军进行搜索，并使用 Fast Unfolding 进行社区检测，能够较好的发现水军团体。

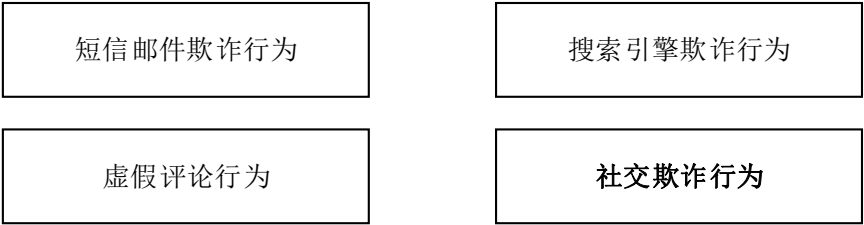


图 1-1 国外社交网络水军识别定位于异常用户行为检测中

国外的文献把社交网络水军识别定位在异常用户行为检测中(Jiang 等, 2016)，如图 1-1，异常用户行为包括：短信邮件欺诈行为、搜索引擎欺诈行为、虚假评论行为、社交欺诈行为，其中社交欺诈行为包含了社交网络水军行为。对其主要的研究思路为：依托 Twitter、Facebook 等社交网络平台的数据，通过观测水军行为，构造出识别水军的特征，使用监督学习、无监督学习、半监督学习、离群点分析等方法对水军进行识别。Benevenuto 等（2010），Shen 等（2016）通过构造用户属性特征、用户内容特征、用户行为特征，使用 SVM 等监督学习方法对社交网络正常用户和水军用户进行分类，从而识别出水军用户。也可以使用无监督聚类的方式识别水军团体（Hu 等，2010）。这两种方法代表了目前主流的研究方法，它们最大的

缺点是在社交网络拥有如此大规模数据的情况下,相比于无标记数据,有标记的数据很少,未能充分利用大量无标记的数据对学习器进行训练。使用半监督学习可以有效地利用无标记数据提高数据利用率从而提高学习器的精度和效率。Li 等(2015)提出一种半监督学习框架,结合社交网络理论中的 Trust Rank 和 Anti-Trust Rank 不停的从未标记样本中选取样本标记,从而有效地利用大量未标记样本数据提高训练效率。Zhang 等(2016)使用 K-Means 和 Co-Training 模型充分利用社交网络大量的无标记样本数据对水军识别,达到了 94.72%的准确率和 75%的 F1 值。Yu 等(2017)利用 NMF 矩阵加入社交信息正则化的半监督学习框架对社交网络水军进行识别,达到了 86%的 F1 值。以上方法虽然充分利用了大量无标记数据提高了模型的准确率,然而却不能有效地应对社交网络中日益隐蔽、行为不断变化的水军用户。Miller 等(2014)把社交网络水军视为离群点,将流式聚类 Den Stream 与 Stream KM++ 两种方法相结合,达到了 100%召回率和 2.8%的假阳性率。Fu 等(2017)抽取社交网络用户动态的行为特征,使用 SVM 作为分类器,对水军用户检测达到了 91%的召回率和 92%的 F1 值。以上文献虽然构造并分析了水军用户的动态行为特征,但是计算过于复杂且基于离线场景,难以应用于在线检测的环境中。Sedhai 和 Sun(2017)使用四个模块:黑名单检测模块,重复信息检测模块,可信信息检测模块以及分类检测模块对社交网络进行在线欺诈检测,能够达到不错的检测效果,但需要依赖大量的人工配置,模块构造繁琐,难以实际运用。

还有一些使用自然语言处理等深度学习技术的方法识别社交网络水军用户。Liu 等(2016)通过 LDA 主题模型,构建了全局异常值和局部异常值特征,使用随机森林模型识别社交网络水军,达到了 97%的 F1 值。Wu 等(2017)使用 Word2Vec 模型构建用户发布信息的特征向量,使用多层感知器作为分类器,对微博信息是否欺诈的检测达到了 95%的精确度和召回率。

由于水军用户隐藏策略的不断演变,它们行为愈发趋向于正常用户(程晓涛等,2015),传统的统计学、数据挖掘方法使用静态行为特征已无法有效地识别社交网络平台中行为不断变换的水军用户。因此,观察和分析水军用户的动态行为愈发重要。虽然 Fu 等(2017)构造了社交网络用户动态行为特征,但计算复杂且基于离线环境,无法应用于在线检测场景中。

因此,本文重新构建了社交网络用户的动态行为特征,构造的动态行为特征容易理解、计算量小、便于实际应用;进一步在离线环境中对比分析了水军用户与正常用户动态行为特征的差异性与有效性;再结合动静行为特征及半监督模型建立了在线检测模型,以适应社交网络在线检测的场景。

## 二、社交网络营销信息传播模型

早期的网络营销包括电子邮件营销、论坛（BBS）营销、搜索引擎营销等，随着社交网络平台的发展与成熟，并且拥有了庞大的用户群体，已成为企业开展营销活动的天然优质场所。目前，不仅大型企业使用社交网络平台进行营销，越来越多的中小型企业也开始利用社交网络平台开展营销活动（Ahani 等，2017）。因此，本文着眼于对社交网络平台上营销信息传播模型的研究。

国内对于社交网络信息传播模型的研究，代表性的综述有（海沫和郭庆，2016；陈慧娟等，2014），其中较为常用的信息传播模型为 SIR 动力学模型。通过对社交网络信息建立 SIR 动力学模型，可以较好的模拟社交网络信息传播的过程，但其最大的缺点是未考虑到社交网络中“人”的特征。因此王彦本和蔡皖东（2016）拓展了 SIR 转移状态增加了个体的遗忘机制，使之更符合信息传播的实际情况；郭东伟等（2014）在 SIR 模型中引入博弈论，从理性博弈和非理性博弈的角度出发对个体进行建模，更好的体现了信息传播机制；程晓涛（2015）考虑了每个传播节点影响力的差异，将其引入到 SIR 模型中，较细粒度的描绘了信息传播的过程。

社交网络最重要的特点是拥有信息传播的高速性以及广泛性，因此能够放大消费者之间的口碑（Word of Mouth）传播，故国外的文献着重于对于营销信息口碑传播的研究，也称作口碑营销或病毒式营销。利用口碑营销策略得到的客户比传统营销策略得到的客户有着更久的客户价值（Villanueva 等，2008）；对于客户熟悉的品牌，正口碑产生的影响通常大于负口碑，对负口碑抵抗的程度越大，客户购买该品牌产品的概率越大，对正口碑抵抗的程度越大，客户购买该品牌产品的概率越小（East 等，2008）；口碑营销策略也有较强的延滞效应（Trusov 等，2009），即经过一段时间之后才能显示出相应的营销效果。对口碑营销建立的模型称为营销信息传播模型或创新扩散模型，主要利用基于主体的模型构建（Renana 等，2010）。Goldenberg 等（2001）利用首次利用基于主体的模型（细胞自动机）对口碑营销进行建模，解决了个体数据缺失、无法验证主要假设的问题，这一工作开创了利用基于主体的模型来研究复杂的口碑营销问题；之后，利用主体的模型被用来研究口碑营销中面对的各种难题，例如网络外部性（Goldenberg 等，2010）、异质性（Kumar 等，2007）等问题。目前的研究主要集中于使用基于主体的模型（小世界网络以及无标度网络）对口碑营销的不同方面进行模拟分析。Jun 等（2006）在小世界网络上模拟了通过客户推荐的社交网络口碑营销传播过程，发现当客户互相认识的概率增大时，产品利润随之缓慢增长，并非预期的急速增长；Li 和 Du（2017）在小世界网络以及无标度网络上模拟了口碑营销传播的过程，结果显示客户背景越不相同、越有自己的观点，口碑营销的效果越好；Samuel（2018）在小世界网络上模



拟了口碑营销中营销信息知识抽取问题，发现人群中掌握知识的比例越大，营销信息在知识抽取方面的效率越低。

综上所述，各个学者从不同方面、不同角度对社交网络营销信息传播过程建立相应的模型，从而探究影响营销信息传播过程中的不同因素。由于现实社交网络中存在水军用户，它们会对社交网络营销信息传播过程产生干扰（朱湘等，2015），对于此实际现象，目前文献中并未有较多细粒度的研究。因此，本文基于此视角，以基于主体的模型为基础，应用社交网络及营销相关的理论知识构建营销信息传播模型；并且在其中纳入水军的影响因素，构建考虑水军用户的营销信息传播模型，据比较细粒度地探讨水军对营销信息传播过程的影响。

### 第三节 研究目标与创新点

针对水军策略的不断演变、愈发隐蔽以及传统模型难以应用于社交网络在线检测场景的问题，本文使用计算波动率的方式构建了社交网络用户的动态行为特征，构造的动态行为特征计算量小、便于实际应用；进一步在离线环境中对比分析了水军用户与正常用户动态行为的差异性与有效性；再以半监督模型 Tri-Training 为基础，结合动静行为特征的方式构建在线检测模型；最后使用抓取的新浪微博数据，验证了该模型的有效性。

针对目前社交网络营销信息传播模型中未考虑到水军用户的影响因素，为了较细粒度地分析水军用户对营销信息传播过程的影响，本文以基于主体的 IC 模型为基础，结合社交网络以及营销的相关理论知识构建营销信息传播模型；并且在其中纳入水军的影响因素，构建考虑水军用户的营销信息传播模型；使用新浪微博营销信息的数据验证了模型的有效性，分析了水军用户对营销信息传播过程的影响，最后结合模型的信息传播参数，为个人、企业和相关部门提供了治理意见。

### 第四节 主要内容与结构安排

#### 一、主要内容

为了检测水军用户愈发隐蔽的伪装策略，首先使用计算波动率的方式构造社交网络用户的 9 个动态行为特征，将社交网络用户的静态行为特征拓展到动态行为特征，接着在离线环境下对比分析了正常用户和水军用户这 9 个动态行为特征的差异性与有效性；进一步地，以半监督 Tri-Training 算法为基础，结合动静行为特征构建在线检测模型；最后使用抓取的新浪微博的数据对该模型进行精度与效率的检验。

为了探讨营销信息传播过程中水军的影响因素，首先使用社交网络、营销相关的理论以及基于主体的 IC 模型构建营销信息传播模型；接着在其中纳入水军的影响因素，构建考虑水军用户的营销信息传播模型；使用新浪微博营销信息的数据，验证了模型的有效性，分析了水军用户对营销信息传播过程的影响，并且结合模型的参数进行灵敏度分析，给出了合理的治理意见。

## 二、结构安排

本文的结构安排如图 1-2 所示，各个章节的主要内容如下：

**第一章：**介绍本文的相关背景与研究意义，指出本文研究的目标与创新点，说明本文的研究内容与结构安排。

**第二章：**介绍本文使用的相关模型，包括数据挖掘模型及基于主体的模型。重点介绍基于分歧的 Tri-Training 半监督学习模型以及 IC 模型。

**第三章：**首先构造社交网络用户的动态行为特征，再结合用户的动静态行为特征构造基于 Tri-Training 的在线水军检测模型；抓取新浪微博的数据，验证动态行为特征的差异性与有效性以及在线水军检测模型的精度与效率。

**第四章：**结合社交网络以及营销理论，以 IC 模型为基础建立营销信息传播模型；并且在模型中加入水军的影响因素，构建考虑水军用户的营销信息传播模型。通过抓取新浪营销信息数据，验证模型的有效性，分析水军用户对营销信息传播过程的影响，结合模型的信息传播参数，为相关方提供治理意见。

最后，总结本文的工作过程与结论，分析本文的不足，指出未来的研究方向。

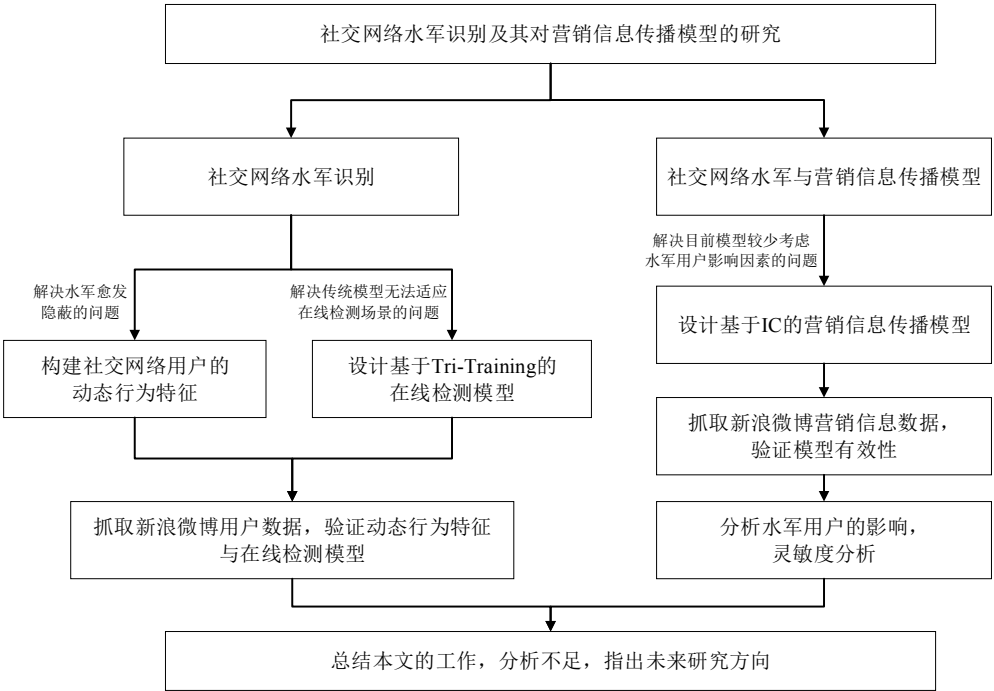


图 1-2 本文结构安排

## 第二章 基础理论与模型

社交网络每天产生数以亿计的用户行为数据，包括网民的搜索数据、购物数据、交互数据等等。如何从这些丰富的数据中寻找有用的规律和知识是一项极具富有挑战性的工作。为此数据挖掘提供了强有力的工具。常用的数据挖掘模型为有监督学习模型，应用效果较好却忽略了社交网络大量的无标记数据，造成数据浪费。使用半监督学习模型与无监督学习模型可以充分地利用大量无标记数据提高模型的性能，并且由于在线社交网络的数据具有大规模性与实时性的特点，因此需要模型的增量学习/训练方法。另外，社交网络的营销信息传播过程遵循创新扩散过程，且成裂变式传播，主要利用基于主体的模型对其建模，使用其中的 IC 模型可细粒度地对营销信息传播过程进行建模。

本章较为详细地介绍了有监督学习模型 NN、半监督学习模型 Tri-Training、无监督学习模型 K-means 以及模型的增量学习/训练方法；介绍了社交网络营销信息传播的基础理论与模型，包括创新扩散理论与基于主体的 IC 模型。

### 第一节 数据挖掘基础理论

#### 一、有监督学习模型 NN

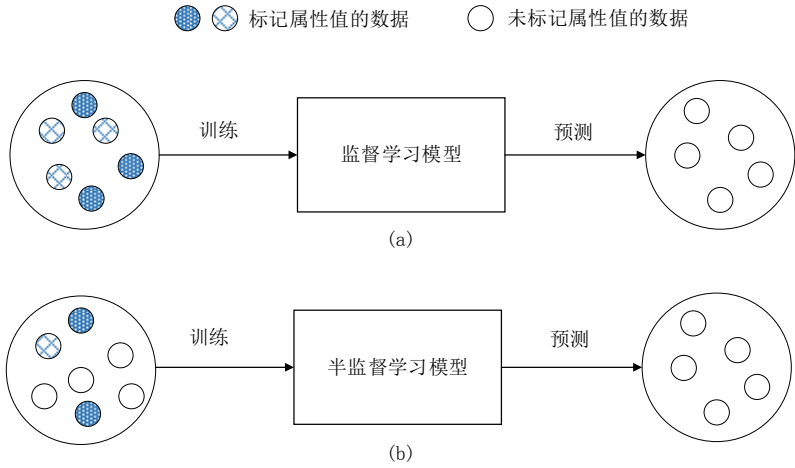


图 2-1 有监督学习与半监督学习示意图

有监督学习是指使用已标记属性值的数据进行学习从而预测未标记数据的属性值（如图 2-1,a），用  $(x, y)$  表示已标记属性值的数据，其中  $x \in R^n, y \in Z | y \in R$ ，那么直接或间接学习到的决策函数  $y = f(x)$  或条件概率分布  $p(y|x)$  称为有监督学习模型。当  $y$  的属性值为连续属性时，即  $y \in R$ ，学习到的模型称为回归模型；当

$y$  的属性值为离散属性时，即  $y \in Z$ ，学习到的模型称为分类模型。社交网络面临的大部分问题为分类问题，例如预测某个用户是否会购买某样产品、是否会点击某个广告、是否为某种用户类型等；也存在一些回归问题，例如预测信息传播的最高转发量、最终转发量等。

常见的分类模型包括逻辑回归、SVM、NN 等，NN（神经网络）由于其强大的非线性映射能力，近年来受到广泛的关注。从理论上讲，神经网络能够以任意精度逼近任何复杂的非线性函数，这一优良性能使其能够作为多维非线性函数的通用数学模型。其中应用最广泛的神经网络为多层前馈神经网络，每层神经元与下一层神经元互连，神经元之间不存在同层连接，也不存在跨层链接，如图 2-2 所示，其中  $W_{ij}$  表示神经元之间相连的权重。

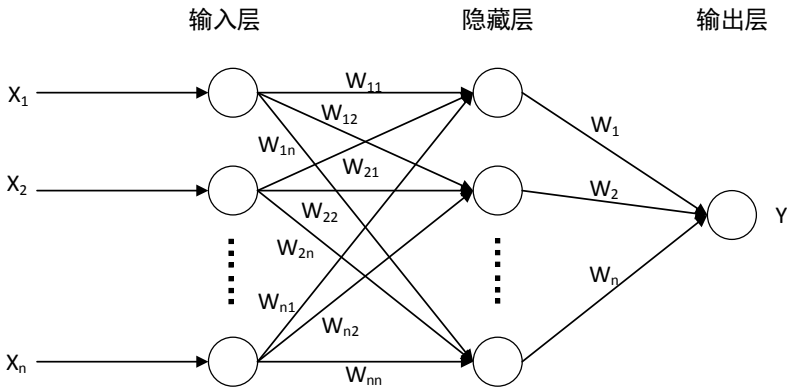


图 2-2 多层前馈神经网络

对于这样的多层前馈神经网络，常使用反向传播算法对其进行训练。核心的思想非常简单：对于训练数据  $(x, y)$ ，若当前神经网络的输出为  $Y$ ，则神经网络的权重调整为  $W = W + \Delta W$ ， $\Delta W = \eta(y - Y)x$ ，其中  $\eta$  为学习率。当假设训练结果正确时，则权重  $W$  不会发生改变；假如训练结果错误，则会对  $W$  进行一定调整。多层前馈神经网络通过从输出层往输入层逆方向传播误差进行权重调整形式的训练，即可拥有强大的函数映射功能。

## 二、半监督学习模型 Tri-Training

有监督学习模型的最大特点是需要大量的标记数据，然而社交网络中的数据大都为无标记数据，标记大量的数据又需要人力与物力，代价十分昂贵。利用半监督学习模型可以充分利用大量无标记数据，减小标记数据的代价，提高数据利用率从而提高模型的预测准确率。

半监督学习是指利用大量无标记属性值的数据与有限少量有标记属性值的数据一起训练模型进行学习（如图 2-1,b），此训练方法不仅有效地利用了数据，避免数据的浪费；同时解决了有监督学习中模型泛化能力不足的问题（刘建伟等，2015）。

使用  $L(x_1, x_2, \dots, x_l)$  表示已标记数据， $U(x_{l+1}, x_{l+2}, \dots, x_{l+u})$  表示无标记数据，将其结合同时训练模型，弥补了只存在标记数据  $L$  的不足，可获得性能更优的模型。半监督学习模型在分类任务中主要使用的是基于分歧的方法（周志华，2013），其中最为核心的为 Tri-Training 模型。

Tri-Training 从 Co-Training 发展而来，Co-Training 使用条件要求数据视图一定要满足充分冗余和独立性，否则因学习器的差异不足而降低学习效率。Tri-Training 克服了 Co-Training 数据视图一定要满足充分冗余性和独立性的条件（Zhou 和 Li, 2005）：它不需要充分冗余的视图，不对基学习器有特定要求，因此该算法便于实际应用（Xu 等，2017）。Tri-Training 模型的流程训练如下，训练过程如图 2-3 所示：

|  |
|--|
| 输入：训练数据 $T$ ，其中已标记数据为 $L$ ，未标记数据为 $U$ ；  |
| 输出：Tri-Training 模型；  |
| <p>Step 1 利用 bootstrap 采样从标记数据 <math>L</math> 中抽取三份训练数据 <math>T_i</math>，训练三个不同的基分类器 <math>C_i</math>；</p> <p>Step 2 从 <math>U</math> 中取出若干未标记数据 <math>U_i</math>，当 <math>U_i</math> 非空时；</p> <p>Step 2.1 分别在三个分类器 <math>C_i</math> 上预测 <math>U_i</math>，若有两个分类器的分类结果相同且与第三个分类器不一致时，则将 <math>U_i</math> 以及“同意”的分类结果放入第三个分类器的训练数据 <math>T_i</math> 中，即以“少数服从多数”的原则选择数据；</p> <p>Step 2.2 在新的训练数据 <math>T_i</math> 下重新训练三个分类器，继续迭代；</p> <p>Step 3 对三个基分类器 <math>C_i</math> 进行投票集成，获得最终模型；</p> |

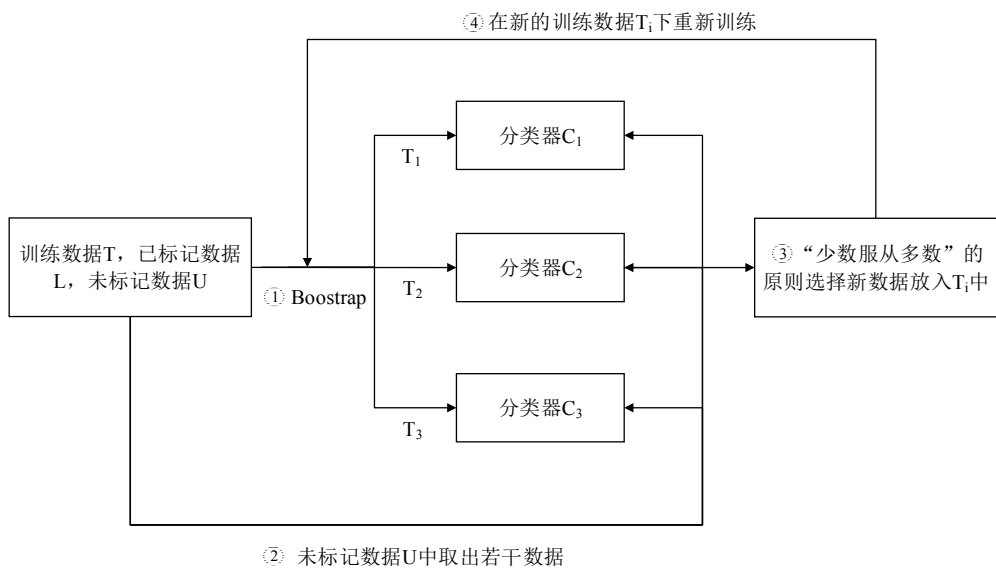


图 2-3 Tri-Training 模型训练流程

### 三、无监督学习模型 K-means

半监督学习同时利用了少量的标记数据以及大量的无标记数据对模型进行训练从而提高其性能,然而当每一例数据的标记都无法获取时,半监督学习进而失效。此时,无监督学习应运而生。

无监督学习的核心之一为聚类。聚类的目的是找到相似对象的群组,组内之间相似性远大于组间之间的相似性。在社交网络场景下,寻找社区、发现相似爱好的用户等都属于聚类的应用场景。聚类模型发展 40 余年,种类繁多(Berkhin, 2006),应用广泛。最具代表性的莫过于划分聚类 K-means, K-means 是一种简单、高效的聚类方法,具有较好的可伸缩性,广泛运用于各类无监督应用场景中。

假设原始数据为  $x, x \in R^n$ , 在给定预设的群组数  $k$  ( $k \leq n$ ) 的条件下, K-means 的目标是将  $x$  划分为  $k$  类 ( $k$  个群组), 用  $S\{S_1, S_2, \dots, S_k\}$  表示, 其中组内误差最小, 也就是优化式(2.1), 其中  $u_i$  表示群组  $S_i$  的中心。

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - u_i\|^2 \quad (2.1)$$

K-means 模型的训练流程如下:

输入: 预设的群组数  $k$ ;

输出: 划分完毕的  $k$  个群组  $S\{S_1, S_2, \dots, S_k\}$ ;

Step 1 从原始数据  $x$  中随机选取  $k$  个元素, 作为  $k$  个群组的中心;

Step 2 当聚类中发生变化;

Step 2.1 在  $x$  中选择除  $k$  个中心以外的数据, 计算它们与  $k$  个中心之间的距离, 将它们划分到距离最近的那个群组;

Step 2.2 划分完毕后, 重新计算  $k$  个群组各自的中心;

Step 3 输出  $k$  个群组;

从最优化角度来看, K-means 的迭代过程即是优化目标函数(2.1)的过程, 将式(2.1)改写为式(2.2), 即得到每一次迭代优化的具体目标函数  $J$ , 其中  $N$  代表数据总个数,  $K$  代表需要划分的群组个数,  $r_{nk}$  表示数据点  $x_n$  是否被划分为群组  $u_k$ , 是为 1, 不是为 0。当采取两步迭代策略进行优化时, 即得到上述的训练过程: 首先固定  $u_k$ , 寻找最优的  $r_{nk}$ , 当把  $x_n$  划分为离它最近群组的中心时,  $J$  最小, 此时的  $r_{nk}$  最优; 再固定  $r_{nk}$ , 寻找最优的  $u_k$ , 对  $J$  求导即可得到此时最优的  $u_k$  (式(2.3)), 即当前所划分的群组中数据点的平均值。由于  $J$  在每一步的优化过程中非递增, 因此最终可得到最优或次优的  $k$  个群组。

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - u_k\|^2 \quad (2.2)$$



$$u_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \quad (2.3)$$

## 四、增量学习

上文介绍了有监督学习模型 NN、半监督学习模型 Tri-Training 以及无监督学习模型 K-means，常规的学习（训练）模型的方式为批量学习（训练），即模型学习的过程中每一次迭代需要使用全部的数据。但当数据规模增大时，例如有上百万条数据，上万维特征，不仅会导致模型学习的速度非常慢，更棘手的是计算机由于没有足够的内存，无法一次性载入如此大规模的数据量。对于此问题，增量学习（训练）可以有效解决。

增量学习（Incremental learning）的思想是先用部分数据初始化模型的参数，当线上来一个数据就更新一次参数。例如，梯度下降常用来训练有监督学习模型，梯度下降每次使用全部的数据更新模型的参数，随机梯度下降是对梯度下降的一个近似，它每次只使用一个数据来更新模型的参数。使用随机梯度下降后，模型的参数更新更频繁，收敛更快也更容易跳出局部最小值（Raschka, 2014），因而效率会更高。

基于增量学习的原理和优势，可将其应用于在线学习中，即模型被实时的学习。通过在线学习，模型可以大大减小每一次的训练时间、且不用考虑内存不足数据无法载入的问题；且更新完模型后可将数据丢弃，以减小存储成本。由于社交网络的数据具有大规模性和实时性，增量学习方法为在线社交网络场景下的模型学习提供了良好的解决方案。

有监督学习模型 NN 的反向传播算法本质上就是梯度下降法，因此可通过随机梯度下降法使其具备在线学习的能力；半监督学习模型 Tri-Training 依赖于有监督学习器，因此可利用有监督学习器的增量学习达到 Tri-Training 模型的增量学习（Tang 和 Han, 2010），使 Tri-Training 模型具备在线学习的能力；无监督学习模型 K-means，对目标函数  $J$  的优化不仅可以使用两步迭代的方法，也可使用随机梯度下降法（Bottou 和 Bengio, 1995）对其进行优化，进而也使其具备在线学习的能力。因此，有监督学习模型 NN、半监督学习模型 Tri-Training 与无监督学习模型 K-means 都可应用于在线检测的场景中。

## 第二节 社交网络营销信息传播模型基础理论

### 一、社交网络营销信息传播模型

企业的产品从开始发布到被消费者全面接纳购买的过程称之为产品的创新扩散过程。创新扩散过程是信息传播领域较为经典和重要的理论之一，它描述了产品从引入期到成熟期的过程，该过程成 S 型曲线。其中消费者的采纳过程被分为五个阶段：了解阶段、兴趣阶段、评估阶段、试验阶段及采纳阶段，即是消费者获知产品、对产品产生兴趣、考虑购买产品、尝试购买产品到全面接纳产品的五个阶段。社交网络上的营销信息在用户节点之间的传播与产品在消费者之间的创新扩散传播类似（Zafarani 等，2014）。由于社交网络的关系网建立在现实关系基础之上，不仅弥补了传统的社交关系，而且使得用户之间的关系更密集，因此营销信息在其上传播更高效，并且多成星形或裂变式传播（易成岐等，2013）。此外，若发布该营销信息的节点中心度越大、网络中用户的关系越紧密，该营销信息传播的速度和范围会越大。

对社交网络营销信息传播过程建立的模型称为创新扩散模型或营销信息传播模型，常用基于主体的（Agent-Based）模型对其进行建模，基于主体的模型由于从个体层面出发，可以去探究信息传播过程中各种复杂的问题（Kiesling 等，2012）。有三种较为常见的模型（Chang 等，2018）：传染病（SIR）模型、线性阈值（LT）模型以及独立级联（IC）模型。传染病模型将信息比作成病原体，病原体的感染过程比作为信息传播的过程，它假设人与人之间的关系是不可见的，因此它只能研究信息传播全局的属性特点，不能够细粒度的展示信息传播的过程；线性阈值模型与独立级联模型都可以较细粒度展现信息的传播过程，但线性阈值模型中的阈值难以量化确定，为其应用带来了限制；而独立级联模型没有阈值的限制，因此被广泛应用。下面单独介绍独立级联模型。

## 二、独立级联（IC）模型

在社交网络中，人们通常转发他人发布的信息，尤其是好友发布或转发的信息。信息在好友之间的传播过程称为信息级联（IC）。信息级联模型是为信息级联过程建立的模型，其中以独立级联模型最具代表性，模型的假设（Kempe 等，2003）如下：

- 网络可表示为有向图；
- 网络中的每个节点只有激活和待激活两个状态，激活代表节点采用了某种行为动作，例如点赞、评论、转发信息；
- 当一个节点被激活后，便可以激活它的邻居节点；
- 节点状态转移不可逆，只能从待激活状态转移到激活状态。

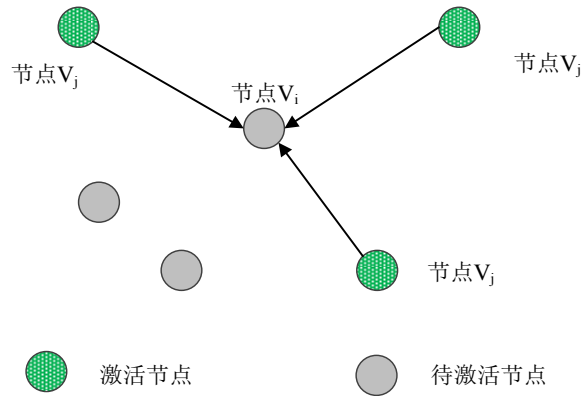


图 2-4 IC 模型

将激活节点视为信息的传播者，待激活节点视为信息的接收者，那么在独立级联模型中，发布者会尝试激活接收者，如图 2-4 所示：若  $V_i$  的邻居节点  $V_j$  在  $t$  时刻被激活，那么在  $t+1$  时刻， $V_j$  会变为激活状态，并以独立的以概率  $p(V_j, V_i)$  激活节点  $V_i$ 。在时刻  $t$  被激活的节点只能在时刻  $t+1$  仅有一次尝试激活其邻居节点的机会。

然而，朴素的 IC 模型并不足以模拟社交网络信息传播的实际情况，因此可以从各种角度对其进行拓展，例如：增加信息特征（Zhou 等，2015）、考虑主题感知（Barbieri 等，2013）等。由于本文是对营销信息传播过程的研究，因此利用社交网络以及营销相关的理论知识拓展 IC 模型，构建以 IC 模型为基础的营销信息传播模型，简称 IC 营销信息传播模型。

### 第三章 社交网络水军用户的动态行为分析与在线检测

针对水军策略的不断演变、愈发隐蔽以及传统模型难以应用于社交网络在线检测场景的问题，本章首先利用计算波动率的方式构建了社交网络用户的动态行为特征，经过离线环境下的对比分析，验证了构造的动态行为特征有较平衡的准确性和计算时间，构造简单便于实际使用；再结合社交网络用户的动静态行为特征以及半监督学习 Tri-Training 模型，构建了社交网络在线水军检测模型；通过抓取新浪微博平台上的数据，验证了该模型的有效性。

#### 第一节 社交网络水军用户检测

社交网络可抽象成图  $G(V, E)$ ，它包括两个集合：结点  $V$  和连接节点的边  $E$ ，边  $E$  可分为有向边和无向边，无向边构成的图为无向图，例如人人网、Facebook 等，有向边构成的图为有向图，例如新浪微博、Twitter 等（由于本文基于新浪微博平台的数据，所以隐指有向图）。图上的每个节点代表一个用户  $V_i$ ，每个用户  $V_i$  拥有  $n$  个特征，记为  $Feature(i, j), j = 1 \dots n$ ，根据每个用户的  $n$  个特征可将其分为的不同类别  $y, y \subseteq Z$ ；水军检测模型的目的是构建一个判别函数  $f(x)$ ，当输入用户特征  $Feature(i)$  时，函数  $f(Feature(i))$  输出其对应类别  $y, y \subseteq \{0, 1\}$ ，0 代表正常用户，1 代表水军用户。本章首先在传统研究定义的静态行为特征上进行拓展，构造社交网络用户的 9 个动态行为特征，基于这 9 个动态行为特征在离线检测环境下对比验证其有效性，再结合动静态行为特征与半监督学习 Tri-Training 模型构建在线水军检测模型，模型结构如图 3-1 所示。虽然本章使用的是新浪微博平台上的数据，但构造的动态行为特征以及设计的在线检测模型仍可拓展应用于其他社交网络中。

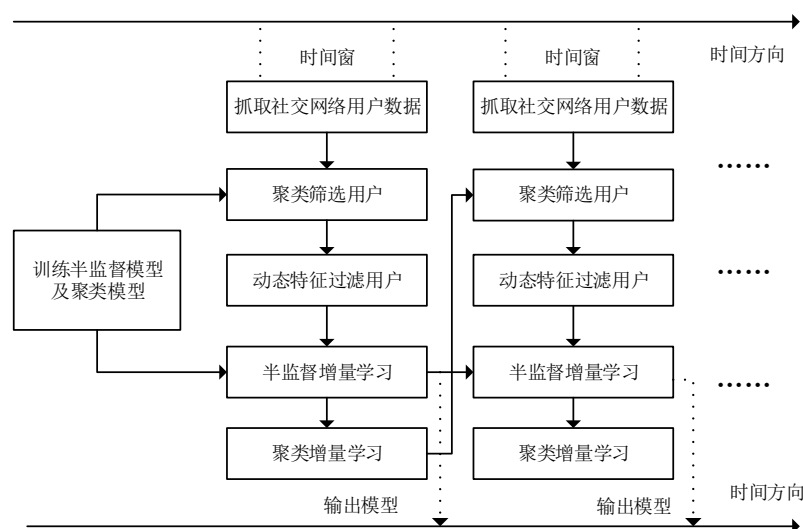


图 3-1 社交网络在线水军检测模型

## 第二节 动态行为特征定义与分析

水军用户与普通用户有着不同的行为特征，目前的文献已经定义了较多的静态行为特征，其中较为重要的包括粉丝关注比，聚类系数以及度数。本文在此基础上拓展，定义如下的社交网络用户的动态行为特征。

### 定义 1 动态粉丝关注比 (Dynamic Fans Follower Ratio: DFFR):

静态粉丝关注比的定义如式(3.1)，其中  $Fans(t)$  表示  $t$  时刻用户的粉丝数， $Followers(t)$  表示  $t$  时刻用户的关注数。正常用户与水军用户在  $t$  时刻有着不同的粉丝/关注模式：正常用户通常关注熟知的用户，如亲戚朋友，这些用户通常也会成为他们的粉丝，从而粉丝关注比不会太小；而水军用户为了提高利益用户的影响力或提高自己的中心性以逃避检测，通常会关注大量与其不相识的用户，导致其关注数较高但粉丝数较小。因此，在  $t$  时刻，用户的粉丝关注比越小，越有可能是水军。

$$FF(t) = \frac{Fans(t)}{Followers(t)} \quad (3.1)$$

在  $t, t+1, \dots, t+n$  时间段内，动态粉丝关注比的定义如式(3.2)，其中  $DFF_{diff}$  的计算方式如式(3.3)、式(3.4)、式(3.5)。动态粉丝关注比衡量了在  $t-n \sim t$  时刻的粉丝关注比的波动率：正常用户由于较强的社交性，粉丝/关注数量变化较大，导致其波动率较高；水军用户由于在社交网络中大部分时期处于潜伏状态，当受到利益驱动时才会产生关注利益用户等行为，导致其波动率较小。因此，水军用户的动态粉丝关注比与正常用户相比较小。

$$DFFR = \frac{\text{std}(DFF_{diff})}{\text{mean}(DFF_{diff}) + 1} \quad (3.2)$$

$$DFF_{diff} = \ln(FF(t+i)) - \ln(FF(t+i-1)), i = 1, 2, \dots, n-1, n \quad (3.3)$$

$$\text{mean}(DFF_{diff}) = \frac{\sum_{i=1}^N DFF_{diff}}{N} \quad (3.4)$$

$$\text{std}(DFF_{diff}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (DFF_{diff} - \text{mean}(DFF_{diff}))^2} \quad (3.5)$$

### 定义 2 动态聚类系数 (Dynamic Clustering Coefficients: DCC):

静态聚类系数定义如式(3.6)，其中  $k(t)$  表示  $t$  时刻用户的邻居节点数， $E(t)$  表示  $t$  时刻用户的邻居节点之间相连的边数。聚类系数用于衡量  $t$  时刻用户的传递行为，表示用户的邻居之间互相认识的实际数量占可能数量的比例。正常用户节点的邻

居都是熟人，他们之间相互认识，彼此连接，所以聚类系数相对较大；而水军用户节点的邻居大都互不相识，邻居之间相连较少，所以聚类系数相对较小。

$$CC(t) = \frac{E(t)}{\frac{k(t)(k(t)-1)}{2}} \quad (3.6)$$

在  $t, t+1, \dots, t+n$  时间段内，动态聚类系数定义如式(3.7)，其中  $DC_{diff}$ ， $\text{mean}(DC_{diff})$ ， $\text{std}(DC_{diff})$  计算方式与上文相同。动态聚类系数衡量了在  $t-n \sim t$  时刻聚类系数的波动率：正常用户由于较强的社交性，社交关系变动频繁，故动态聚类系数波动较大；而水军用户长期潜伏社交性较弱，当受到利益驱动时才会产生关注利益用户等行为，且利益用户互相之间并无关联，致其动态聚类系数波动较小。因此水军用户的动态聚类系数与正常用户相比较小。

$$DCC = \frac{\text{std}(DC_{diff})}{\text{mean}(DC_{diff}) + 1} \quad (3.7)$$

### 定义3 动态 PageRank 值 (Dynamic PageRank: DPR):

在衡量社交网络用户的中心性时，度是目前文献最常用的度量指标。度中心性表示为  $C(t) = d$ ， $d$  指在  $t$  时刻用户节点邻居的数量。如果一个用户在  $t$  时刻拥有的邻居数量越多，那么中心性越高，这显然不是最合理的。因为并不是认识的朋友越多他越重要，而是认识重要的朋友越多他才越重要。PageRank 值可以衡量这一特征：社交网络一个用户的中心性由所有指向该用户的节点中心性决定，因此采用 PageRank 值替代度衡量用户的中心性。

静态 PageRank 值定义如式(3.8)，式中  $A(t)$  表示  $t$  时刻的邻接矩阵， $D(t) = \text{diag}(d_1^{out}, d_2^{out}, \dots, d_n^{out})$  表示  $t$  时刻关于度的对角矩阵。在  $t$  时刻，大部分正常用户的 PageRank 值分布在低值区间内，少部分高威望用户则拥有较大的 PageRank 值，成幂律分布。水军用户由于粉丝关注比的极其不对称性以及较低的粉丝影响力，PageRank 值比正常用户小。

$$PR(t) = \beta \left( I - \alpha A(t)^T D(t)^{-1} \right)^{-1} \quad (3.8)$$

在  $t, t+1, \dots, t+n$  时间段内，动态 PageRank 值定义如式(3.9)，其中  $DP_{diff}$ ， $\text{mean}(DP_{diff})$ ， $\text{std}(DP_{diff})$  计算方式与上文相同。动态 PageRank 值衡量了在  $t-n \sim t$  时刻的 PageRank 值的波动率：由于水军用户潜伏在社交网络中，很少与正常用户建立社交联系，加上粉丝的影响力较低，导致其 PageRank 值一直较小，波动也较小；正常用户由于社交性较强，中心性变化较大，PageRank 值波动较大。故水军用户的动态 PageRank 值与正常用户相比较小。



$$DPR = \frac{\text{std}(DP_{diff})}{\text{mean}(DP_{diff}) + 1} \quad (3.9)$$

**定义 4 动态微博相关特征 (Dynamic Tweet Related Features: DTRF):**

水军用户或是被机器人操控,或是被人为操控,发布的微博特征与普通用户相比而言不同,包括:微博的数量(Tweet Number: TN),微博字符的平均长度(Tweet Length: TL),微博的平均评论数(Tweet Comment: TC)、平均点赞数(Tweet Attitude: TA)、平均转发数(Tweet Repost: TR)以及微博转发占比(Retweet Ratio: RR)。正常用户由于较强的交流、分享特性,微博相关特征的波动较大;水军用户长期潜伏并短期内以提高话题热度、操作话题等目的才发布微博,微博相关特征的波动较小。以上 6 个微博相关的特征都可作为区分水军用户和正常用户的动态行为特征,计算波动值的方法与上文相同。

以上定义的社交网络用户的 9 个动态行为特征差异明显,计算简单,便于实际应用。把静态行为特征拓展到动态行为特征,发现水军用户与正常用户的动态行为模式,不仅可以提高离线和在线水军检测模型的准确率,且对水军检测领域更具有有一定的借鉴意义。

### 第三节 在线水军检测模型

#### 一、Tri-Training 半监督学习模型与动态行为特征过滤

表 3-1 模型的符号意义

| 符号                                   | 说明  |
|--------------------------------------|---|
| $T_s$                                | 起始训练时间  |
| $T$                                  | 训练时间间隔, 时间窗                                   |
| $N_i$                                | $N$ 个用户中第 $i$ 个用户                             |
| $Feature(i)$                         | 第 $i$ 个用户的行为特征向量                              |
| $Feature(i)_{T_s}$                   | 第 $i$ 个用户在 $T_s$ 时刻的静态行为特征向量                  |
| $Feature(i)_{T_s+T}$                 | 第 $i$ 个用户在 $T_s+T$ 时刻的静态行为特征向量                |
| $Feature(i)_{T_s \sim T_s+T}$        | 第 $i$ 个用户在 $T_s \sim T_s+T$ 时间间隔的动态行为特征向量     |
| $Total\_Feature(i)_{T_s \sim T_s+T}$ | 第 $i$ 个用户在 $T_s \sim T_s+T$ 时间间隔的动态行为特征向量的加权和 |

通过上文分析，我们发现社交网络水军用户与正常用户的动态行为特征存在明显不同。本文以半监督模型 Tri-Training 为基础，利用静态行为特征聚类以及动态行为特征筛选出最有价值的用户数据增量训练 Tri-Training，以此构建在线检测模型，模型的符号意义如表 3-1 所示。

Tri-Training 模型通过 bootstrap 采样从带标记的训练样本集中产生 3 个基分类器，当用无标记数据进行训练时，若两个基学习器预测结果相同且与另一个基学习器不一致，则放入另一个基学习器中学习，即以“少数服从多数”的原则学习，以此充分利用无标记数据，有效避免了数据浪费，最终分类器由三个基学习器组合而成。由于社交网络数据的大规模性，每次使用全部的无标记数据训练 Tri-Training 定不可取，鉴于水军用户和正常用户动态特征的差异性，以用户的动态行为特征筛选数据训练 Tri-Training 成为可行方案。

一个时间窗  $T_s \sim T_s + T$  训练的具体作法为：分别在  $T_s$ 、 $T_s + T$  时刻抓取用户数据  $N$ ，为每个用户  $N_i$  计算其  $Feature(i)_{T_s}$ ， $Feature(i)_{T_s+T}$ ， $Feature(i)_{T_s \sim T_s+T}$ ， $Total\_Feature(i)_{T_s \sim T_s+T}$ ； $Total\_Feature(i)_{T_s \sim T_s+T}$  为每个用户动态行为特征的等权加权和，即式(3.10)，选取  $Total\_Feature(i)_{T_s \sim T_s+T}$  最大的  $K_1$  个代表正常用户，最小的  $K_2$  个代表水军用户，即式(3.11)与式(3.12)。

$$Total\_Feature(i)_{T_s \sim T_s+T} = \frac{\sum_{j=1}^9 Feature(ij)_{T_s \sim T_s+T}}{9} \quad (3.10)$$

$$\overset{Max}{K_1} \left( Sort \left( Total\_Feature(i)_{T_s \sim T_s+T} \right) \right) \quad (3.11)$$

$$\overset{Min}{K_2} \left( Sort \left( Total\_Feature(i)_{T_s \sim T_s+T} \right) \right) \quad (3.12)$$

通过以上过滤，得到  $K_1 + K_2$  个用户；将  $K_1 + K_2$  个用户的  $Feature(i)_{T_s+T}$  增量训练 Tri-Training 模型。通过动态行为特征的筛选，使得 Tri-Training 模型能充分利用最有价值的无标记用户数据学习，且基于增量训练可以适应在线检测。

## 二、K-means 聚类模型与在线水军检测模型

Tri-Training 最大的缺点是容易受到噪音的影响。社交网络数据由于其数据的多样性，抓取程序的不稳定性，或是网络的不稳定性，容易导致产生离群点或噪声数据，从而影响动态行为特征过滤后的数据质量和 Tri-Training 的学习效果。为了过滤噪声数据，本文选取 K-means 为代表的聚类模型，K-means 为最典型的一种划分聚类模型，方便实践且效果良好，并且适应增量学习可将其应用于在线检测场景中。

在一个时间窗  $T_s \sim T_s + T$  的训练过程中，首先对抓取的用户  $N$  使用  $Feature(i)_{T_s \sim T_s+T}$  聚类（簇数  $k = 2$ ），从中选取离中心点最近的  $K$  个用户，即

$$\min_K Sort(Dist(N_i - O)) \quad (3.13)$$

其中 $O$ 为中心点。这样不仅可以过滤掉离群点或噪声，提高了数据的可靠性而且因为过滤掉不可靠数据为下层动态特征筛选减少了部分计算。一个时间窗 $T$ 的训练步骤如图 3-2 所示，整个模型的流程如下：

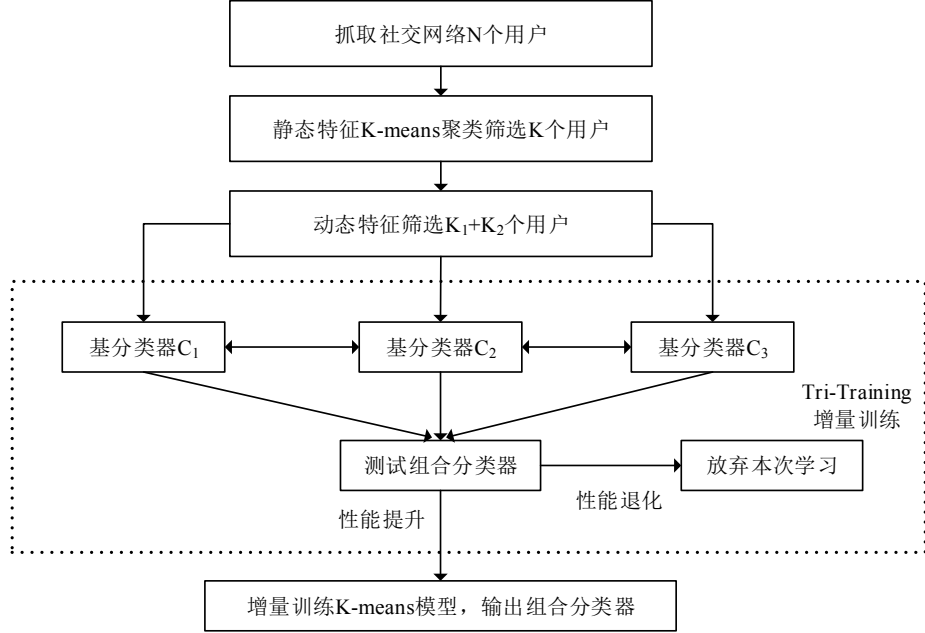


图 3-2 一个时间窗  $T$  的训练步骤

输入：起始训练时间  $T_s$  以及在此时刻带标签的用户数据集  $L$ ，训练时间间隔  $T$ ，抓取用户数  $N$ ，聚类过滤用户数  $K$ ，动态行为特征筛选最大的  $K_1$  个用户和最小的  $K_2$  个用户；

输出：每个训练时刻三个基学习器  $C_1, C_2, C_3$  的组合模型；

Step 1 在  $T_s$  时刻使用 Hold Out 方式对  $L$  划分为训练集和测试集，训练集使用 Bootstrap 采样划分三份训练集  $L_1, L_2, L_3$ ，使用它们的  $Feature(i)_{T_s}$  训练三个基分类器  $C_1, C_2, C_3$  以及聚类模型  $C$ ；

Step 2 一直循环，直到手动退出；

Step 2.1 分别在  $T_s$ 、 $T_s + T$  时刻抓取社交网络用户数据  $N$ ；

Step 2.2 使用所有用户  $N$  的  $Feature(i)_{T_s+T}$  利用式(3.10)、式(3.11)、式(3.12)筛选得到  $K_1 + K_2$  个用户；

Step 2.3 计算  $K$  个用户的  $Feature(i)_{T_s}$  与  $Feature(i)_{T_s+T}$ ，利用式(3.10)、式(3.11)、式(3.12)筛选得到  $K_1 + K_2$  个用户；

Step 2.4 将  $K_1 + K_2$  个用户的  $Feature(i)_{T_s+T}$  放入基学习器  $C_1, C_2, C_3$  中学习，当两个基学习器学习到标签相同且与另一个学习器不一致时，将学习到的标签  $y$  以及  $Feature(i)_{T_s+T}$  放入另一个学习器增量学习，同时保存标签  $y$  和  $Feature(i)_{T_s+T}$ ；

Step 2.5 使用测试集测试组合学习器，若组合学习器性能退化，则放弃本次

学习；若组合学习器性能提升，将保存的标签  $y$  和  $Feature(i)_{T_s+T}$  放入 K-means 聚类模型中进行增量学习；输出此时的  $C_1, C_2, C_3$  组合模型；

Step 2.6  $T_s$  被更新为  $T_s+T$ ；

---

在每个时间窗的训练步骤中，忽略抓取数据的时间，且计算都基于向量化计算，所以基本操作包括  $N$  个用户 K-means 聚类、 $K_1+K_2$  个用户的 Tri-Training 增量训练以及  $K_1+K_2$  个用户的 K-means 增量训练。K-means 聚类及增量训练的时间复杂度分别为  $O(N)$ 、 $O(K_1+K_2)$ ，Tri-Training 增量训练的时间复杂度取决于具体的基分类器，本文使用神经网络分类模型，因一个神经网络分类模型的时间复杂度为  $O((K_1+K_2)M^2)$ ，其中， $M$  为隐藏层的节点数，故 Tri-Training 的时间复杂度  $O((K_1+K_2)^3M^6)$ ，所以每个时间窗模型训练的总时间复杂度为  $O(N(K_1+K_2)^4M^6) \approx O(N)$ ，因  $K_1, K_2, M$  在模型中为常数。空间复杂度包括存储抓取的数据、K-means 的参数和 Tri-Training 的参数，它们的空间复杂度分别为  $O(2N)$ 、 $O(N)$ 、 $O(6M)$ ， $M$  为隐藏层的节点数，所以每个时间窗模型训练的空间复杂度为  $O(3N+6M) \approx O(N)$ 。

## 第四节 实验分析

### 一、数据抓取与统计

实验采用的硬件环境为 4GB 内存，i5 处理器的单台笔记本电脑，软件环境为 Windows 10 操作系统，Python 作为程序设计语言。采用 Scrapy 数据采集框架从新浪微博定时采集数据，新浪微博每个用户都有唯一标识 ID。首先构造三个文件，分别存入正常用户 ID，水军用户 ID 以及未知标签用户 ID，再每隔 2 小时根据 ID 抓取其用户、微博、关注及粉丝相关数据存入 MongoDB 数据库中。常规做法是人工标记水军用户与正常用户，但因标记繁琐以及存在错标等情况，因此本文按如下方式构造用户 ID：水军用户 ID 从网络公关公司买入，正常用户 ID 根据人工挑选的公众用户及一般普通用户作为种子用户，对其关注列表使用宽度优先搜索抓取 ID 存入文件，只对关注列表抓取保证正常用户的可靠性，未知标签用户 ID 挑选上文若干水军用户和正常用户 ID 作为种子用户，对它们的关注粉丝列表使用宽度优先搜索抓取 ID 存入文件。考虑到单台笔记本的处理能力，以及剔除 ID 被封、数据跟踪失败以及网络断线丢失的数据，最终得到 184 个水军用户、332 个正常用户和 274 个未标记用户在 2017 年 12 月 13 日 12:00:00~2017 年 12 月 21 日 22:03:34 每隔两小时时间段的数据。用户微博不仅包含纯文本数据，有些还包含了表情、图片等多媒体数据，由于本文构造的微博相关特征只统计微博的平均字符个数，对于包含多媒体信息的微博，只统计其纯文本字符个数，没有纯文本的微博不纳入统计。

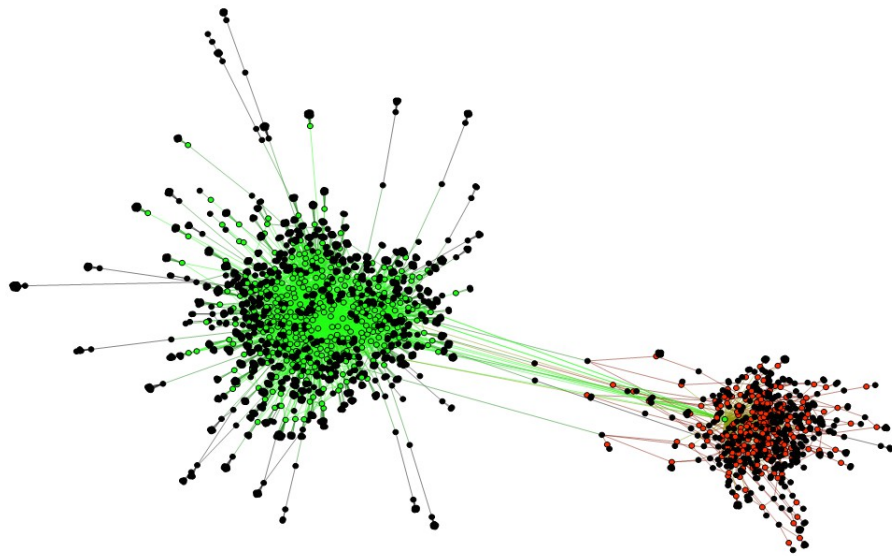


图 3-3 初始时刻用户构成的部分网络图

表 3-2 初始时刻水军用户与正常用户的静态行为特征平均值对比

| 静态行为特征 \ 用户类型   | 水军用户       | 正常用户         |
|-----------------|------------|--------------|
| 粉丝关注比 (FF)      | 0.073110   | 54404.744567 |
| 聚类系数 (CC)       | 0.000075   | 0.006654     |
| PageRank 值 (PR) | 0.000079   | 0.000193     |
| 微博发布数量 (TN)     | 21.663043  | 7447.719879  |
| 微博字符平均长度 (TL)   | 123.504997 | 273.350338   |
| 微博平均评论数 (TC)    | 0.196887   | 3578.487170  |
| 微博平均点赞数 (TA)    | 0.031319   | 19356.170168 |
| 微博平均被转发数 (TR)   | 0.185372   | 57277.957331 |
| 微博转发占比 (RR)     | 0.211405   | 0.356502     |

图 3-3 展示了在初始抓取时刻 2017 年 12 月 13 日 12:00:00 时，用户构成的部分网络图，采用 Gephi 软件中 ForceAtlas2 布局算法。图中可以观察到，正常用户大部分聚集在左边的簇（绿色），水军用户则大部分聚集在右边的簇（红色）。它们的平均值对比如表 3-2 所示，由于正常用户包含一些公众用户，他们的影响力和社交需求较大，行为特征值较大，因此特征对比较明显。

表 3-3 统计了在 3 个不同时间点全部用户构成网络图的参数。可以观察到，用户关联不紧密，平均聚类系数较小，这是由于本文根据三个用户 ID 文件抓取用户数据，它们之间的联系较少，并且只抓取了用户的一层社交圈，进一步导致网络稀疏。随着时间的增加以及用户社交活动的变化，用户关联缓慢增加，因此，平均聚类系数也相应缓慢增加，平均最短路径缓慢减少。

表 3-3 不同时间点的网络参数统计

| 时间点                       | 节点个数  | 节点边数   | 平均度数  | 平均聚类系数 | 平均最短路径 |
|---------------------------|-------|--------|-------|--------|--------|
| 2017 年 12 月 13 日 12:00:00 | 71216 | 140628 | 3.949 | 0.0188 | 2.389  |
| 2017 年 12 月 17 日 12:00:00 | 71288 | 140380 | 3.938 | 0.0186 | 2.334  |
| 2017 年 12 月 21 日 12:00:00 | 67463 | 126636 | 3.754 | 0.0197 | 2.193  |

下文的实验步骤如下：

- 1) 分析 184 个水军用户及 332 个正常用户的动态行为特征，对比两者差异；
- 2) 利用动态行为特征的差异性构造离线检测模型，使用静态行为特征、动态行为特征、对比模型特征以及动静态行为特征相结合的四中特征计算方法，比较它们的准确率和效率；
- 3) 利用动静态特征结合方式训练在线检测模型，分析在线检测模型的准确率和效率。

## 二、动态行为特征分析

表 3-4 水军用户与正常用户动态行为特征的平均值对比

|                    |      | 粉丝关注<br>比       | 聚类<br>系数            | PageRank<br>值     | 微博发布<br>数量      | 微博字符<br>平均长度    | 微博平均<br>评论数     | 微博平均<br>点赞数     | 微博平均<br>被转发数    | 微博转发<br>占比      |
|--------------------|------|-----------------|---------------------|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $T = 1$<br>(2 小时)  | 水军用户 | 0.00042         | 0.000018            | 330.375525        | 0.000017        | 0.008343        | 0.003035        | 0.000380        | 0.003046        | 0.001890        |
|                    | 正常用户 | 0.001134        | 0.0139562268        | 251.502           | 0.000285        | 0.029428        | 0.049885        | 0.045491        | 0.066501        | 0.007314        |
|                    | 差值   | <b>0.000714</b> | <b>0.0139381937</b> | <b>875.977</b>    | <b>0.000268</b> | <b>0.021085</b> | <b>0.046850</b> | <b>0.045111</b> | <b>0.063455</b> | <b>0.005424</b> |
| $T = 2$<br>(4 小时)  | 水军用户 | 0.000303        | 0.000018            | 802.578332        | 0.000024        | 0.005621        | 0.003048        | 0.000292        | 0.003034        | 0.001548        |
|                    | 正常用户 | 0.001473        | 0.0121631664        | 363.715           | 0.000397        | 0.028230        | 0.047434        | 0.043340        | 0.062837        | 0.007229        |
|                    | 差值   | <b>0.001170</b> | <b>0.012145</b>     | <b>861.785383</b> | <b>0.000373</b> | <b>0.022609</b> | <b>0.044386</b> | <b>0.043048</b> | <b>0.059803</b> | <b>0.005681</b> |
| $T = 3$<br>(6 小时)  | 水军用户 | 0.000322        | 0.000016            | 565.172998        | 0.00003         | 0.003934        | 0.002203        | 0.000417        | 0.002755        | 0.00068         |
|                    | 正常用户 | 0.001767        | 0.011089            | 505.475624        | 0.000482        | 0.033616        | 0.056134        | 0.051135        | 0.076305        | 0.008541        |
|                    | 差值   | <b>0.001445</b> | <b>0.011073</b>     | <b>-59.697374</b> | <b>0.000452</b> | <b>0.029682</b> | <b>0.053931</b> | <b>0.050718</b> | <b>0.073550</b> | <b>0.007861</b> |
| $T = 4$<br>(8 小时)  | 水军用户 | 0.000102        | 0.000007            | 122.632696        | 0.000034        | 0.003782        | 0.003546        | 0.000243        | 0.003843        | 0.000972        |
|                    | 正常用户 | 0.001955        | 0.010589            | 517.019733        | 0.000551        | 0.031284        | 0.053554        | 0.051028        | 0.069196        | 0.008238        |
|                    | 差值   | <b>0.001853</b> | <b>0.010582</b>     | <b>394.387037</b> | <b>0.000517</b> | <b>0.027502</b> | <b>0.050008</b> | <b>0.050785</b> | <b>0.065353</b> | <b>0.007266</b> |
| $T = 5$<br>(10 小时) | 水军用户 | 0.000502        | 0.000019            | 83.602259         | 0.000038        | 0.006142        | 0.002091        | 0.000282        | 0.002758        | 0.001037        |
|                    | 正常用户 | 0.002023        | 0.012409            | 204.767017        | 0.000599        | 0.033703        | 0.057937        | 0.051257        | 0.074353        | 0.008669        |
|                    | 差值   | <b>0.001521</b> | <b>0.012390121</b>  | <b>164.7577</b>   | <b>0.000561</b> | <b>0.027561</b> | <b>0.055846</b> | <b>0.050975</b> | <b>0.071595</b> | <b>0.007632</b> |

表 3-4 对比了 184 个水军用户和 332 个正常用户在 2017 年 12 月 13 日 12:00:00 ~ 2017 年 12 月 21 日 22:03:34 每隔  $T$  时间段 9 个动态行为特征的平均值，其中，1 单位  $T$  等于 2 小时。对于每个时间间隔  $T$ ，水军用户由于潜伏性的特点，行为波动较小，动态行为特征与正常用户相比较小，与文中定义相符。随着时间间隔  $T$  的



增加，两者的动态 PageRank 值差值迅速降低，意味着动态 PageRank 值的区分能力迅速变弱；动态聚类系数的差值保持平稳，区分能力保持稳定；而动态粉丝关注比以及动态微博相关特征的差值随着时间间隔  $T$  的增加而微弱增加，区分能力缓慢增强。因此，可以认为在  $T=1$  时，动态行为差异与其他时间间隔  $T$  相比较明显，区分水军用户和正常用户的能力较强。

进一步地，动态 PageRank 在  $T=1$  时的分布如图 3-4 所示：横轴为动态 PageRank 值（In 对数），纵轴为用户数量。水军用户的动态 PageRank 值分布分散且较小，37.5%的用户大于 4；正常用户的动态 PageRank 值分布成正态分布且 96% 的动态 PageRank 值大于 4。由此可见水军用户由于中心性的稳定性，导致中心性波动率较小，动态 PageRank 值较小；正常用户由于有着较强的社交性，导致中心性波动率较大，动态 PageRank 值较大。

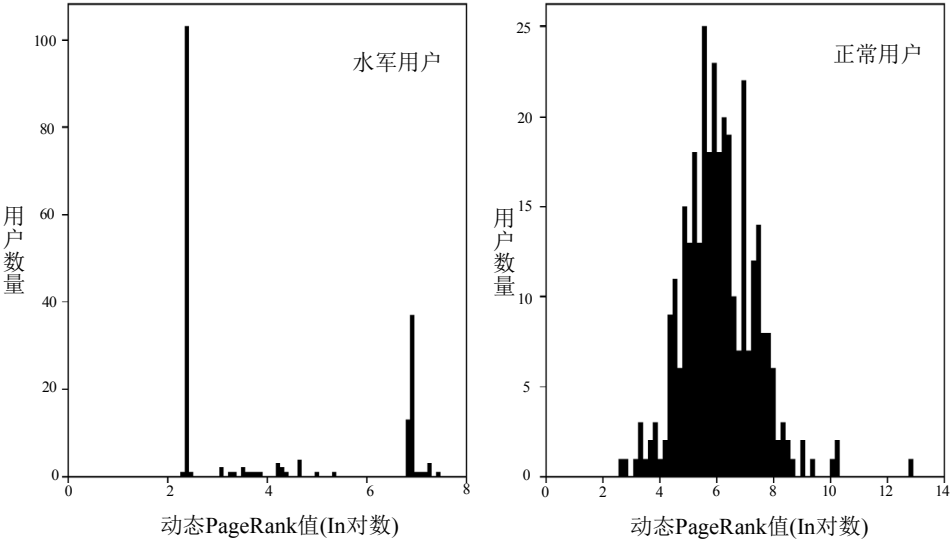


图 3-4 水军用户（左）和正常用户（右）动态 PageRank 值分布

### 三、离线水军检测模型检验

由于水军用户与正常用户的动态行为特征具有明显的差异性，可据此构造离线水军检测模型，步骤如下：首先计算全部用户每隔 2 小时时间间隔（即  $T=1$ ）的动态行为特征作为训练特征，再使用 10 折交叉验证（10-Fold Cross-Validation）把模型分为 10 次进行训练，每一次训练把数据集划分为 10 份，9 份用来训练模型，1 份用来测试模型，模型的评估指标取 10 次训练结果的平均值。评估指标使用查准率（Precision）、查全率（Recall）和 F1 值（F1 Score）。

对比模型采用（Fu 等，2017）构造的离线水军检测模型：使用滑动窗计算每个用户特征的变化情况，计算相应的均值、标准差和离群点数作为用户动态行为特

征，再通过用户之间层次聚类，得到每个用户所属的簇特征，结合用户动态行为特征和聚类特征得到用户总的特征放入监督模型中训练。

本文采用神经网络（Neural Network）监督模型，比较静态特征、动态特征、对比模型特征以及动静特征结合的特征计算时间、查准率、查全率和 F1 值。首先通过 10 折交叉验证得到 4 种特征计算方法在单隐藏层 30 个节点下各自 F1 值达到最大值并且保持稳定，在此相同的神经网络配置下通过 10 折交叉验证得到表 3-5 的对比结果：

表 3-5 单隐藏层 30 个节点的神经网络配置下 4 种特征计算方法对比

|        | 特征计算时间 | 查准率    | 查全率    | F1 值   |
|--------|--------|--------|--------|--------|
| 静态特征   | 49 秒   | 94.99% | 85.75% | 89.90% |
| 动态特征   | 26 分钟  | 84.35% | 90.30% | 87.00% |
| 对比模型特征 | 3 天    | 99.05% | 98.33% | 98.60% |
| 动静特征结合 | 27 分钟  | 98.22% | 93.01% | 95.41% |

表 3-5 所示：使用静态特征检测水军可达到 89.90% 的 F1 值，虽然动态特征查准率没有静态特征高，但查全率与之相比高，达到 90.30%；对比模型构造的特征有着最高的查准率、查全率和 F1 值，但其特征计算时间高达 3 天；当把动静特征相结合，查准率、查全率和 F1 值都与对比模型特征接近且计算时间与之相比大幅减少。

由于本文提出的动态特征计算时间较少且能有效地体现水军用户和正常用户的行为，但其查准率与静态特征相比较低，故融合静态和动态行为特征构建在线水军检测模型，以满足社交网络在线检测场景。

#### 四、在线水军检测模型检验

在初始训练时间  $T_s$ ，将用户数据采用保持（Hold Out）方法划分为训练集和测试集，即 2/3 为训练集，1/3 为测试集；训练集使用 Bootstrap 采样方法训练基学习器和聚类模型，基学习器采用单隐藏层 30 个节点的神经网络，K-means 聚类模型中  $k=2$ ；再每隔时间段  $T$  从数据库中读入未标记用户数据  $N$  训练在线检测模型，每次训练完后使用测试集数据对 Tri-Training 模型进行测试；结果得到各个时间段 Tri-Training 模型的查准率（Precision）、查全率（Recall）和 F1 值（F1 Score）。

图 3-5 显示了模型在默认参数下的训练曲线，横坐标为训练时间点，纵坐标为模型的查准率、查全率和 F1 值。在初始训练阶段，Tri-Training 模型的查准率为 97.55%，查全率为 72.72%，F1 值为 83.30%。随着每个时刻对无标记数据的在线学习，查准率一直保持 97% 左右，查全率缓慢增加且在训练结束时稳定在 89.09%，

F1 值缓慢增加且在训练结束时稳定在 93.33%。说明 Tri-Training 模型在各个训练时间点充分利用了最有价值的无标记用户数据在线学习，因使用聚类剔除社交网络中的噪声数据及利用用户的动态特征筛选数据，使得模型在保持查准率平稳的情况下，大大提高了查全率和 F1 值。

在离线环境中计算用户的动态特征由于计算了整个时间段的图特征，故计算时间较长，而在在线检测环境中，只需计算每个训练时间窗内的图特征，因此所需计算量明显减少。图 3-6 显示了模型在默认参数下每个训练时刻模型所花费的时间，因模型的时间复杂度为  $O(N)$ ，所以在每个时刻模型的训练时间较平稳，均值为 118 秒，说明模型可以较好地适应在线检测环境。

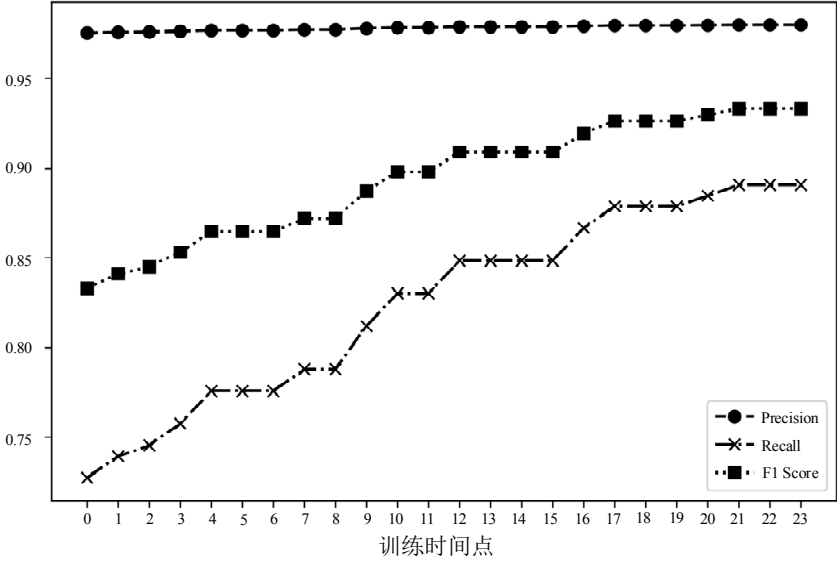


图 3-5 默认参数下，模型的训练曲线

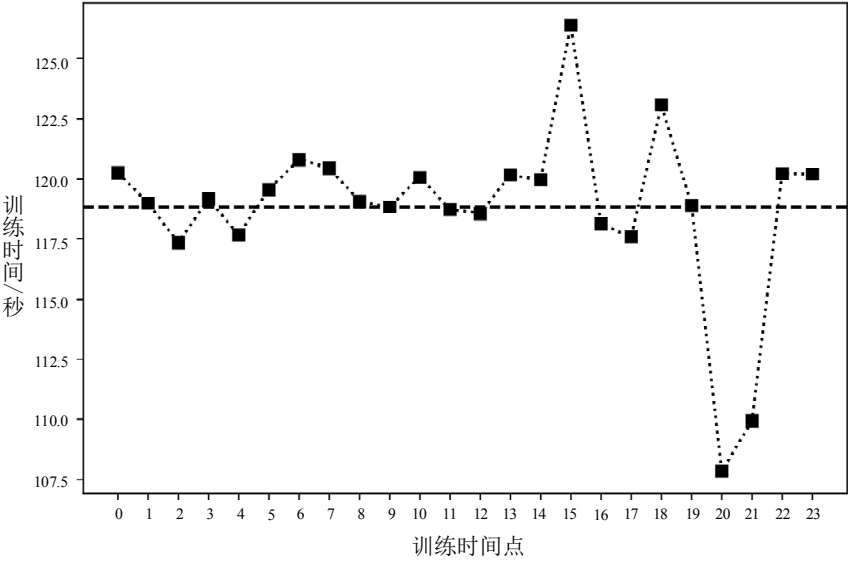


图 3-6 默认参数下，模型在每个时间点训练所花费的时间

为了验证聚类 and 动态特征筛选机制的有效性，对比了在噪声环境下，经过聚类 and 动态特征筛选和未经过聚类 and 动态特征筛选的迭代曲线，仍在默认参数下对比，结果如图 3-7 所示：开始时，两者的 F1 值相差很小，分别为 83.30% 和 81.34%。但随着每一次的在线训练，未经过聚类 and 动态特征筛选的模型因其内部的噪声积累，F1 值较快达到最大值且稳定在稳定 85.89%，而经过聚类 and 动态特征筛选的模型因每次挑选最有价值的数据进行在线学习，极大减小了噪音和无价值的样本数据对学习器的影响，F1 值不断增加且最终达到 93.33% 并保持稳定。

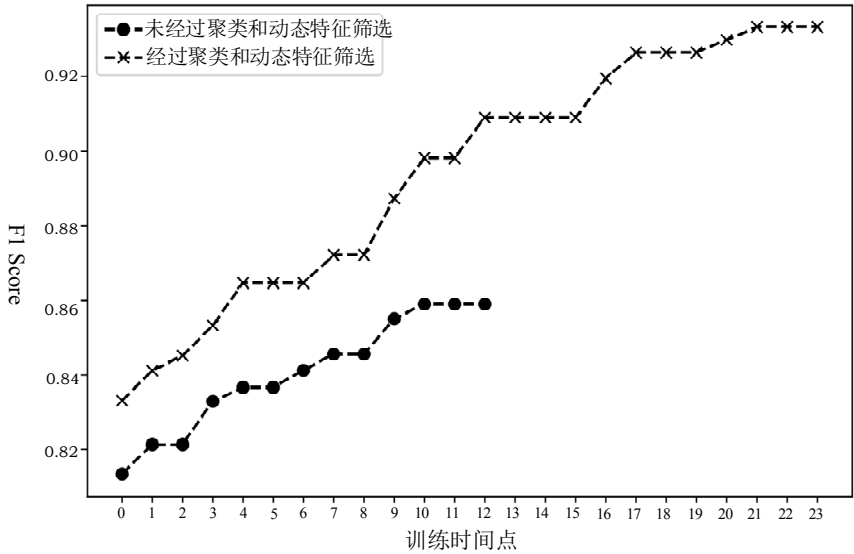


图 3-7 在噪声环境下，是否经过聚类 and 动态特征筛选的对比（默认参数下）

本文构造的在线水军检测模型的核心基于 Tri-Training 半监督模型，它对初始的标记数据敏感，初始标记数据的质量高低决定了 Tri-Training 模型的基础性能，从而影响了模型整体的最终性能。因此，本文模型的可调参数  $T, N, K, K_1, K_2$  对模型最终性能的影响与初始标记数据的质量相比甚微，但参数  $T$  会影响模型的训练次数以及参数  $N, K, K_1, K_2$  会影响模型在每个时间点的训练时间： $T$  越小，模型的训练时间点越多，训练次数越多； $N, K, K_1, K_2$  越大，模型在每个时间点训练时选择的用户数越多，计算时间会相应增大。因此，在实际使用时，为了提高模型的准确率，可尝试增大  $N, K, K_1, K_2$  的参数值，选取更多的用户训练；当模型的准确率达到业务要求时，可减小参数  $N, K, K_1, K_2$  的值降低模型每次训练的训练时间。

### 第五节 本章小结

本章通过对社交网络正常用户和水军用户的行为分析，构造简单、易于使用的动态行为特征，在离线环境中分析并验证了构造特征的差异性和有效性，据此提出

了动静特征相结合的在线水军检测模型，此模型不仅充分利用筛选后最有价值的未标记用户的数据，又基于增量学习，可适应在线检测。把静态行为特征推广到动态行为特征对水军检测的研究方向有一定的借鉴意义；且构造的在线水军检测模型填补了这一领域的研究不足。

## 第四章 社交网络水军识别在营销信息传播模型中的应用

社交网络与传统媒介相比,信息传播更加快速与高效,企业在社交网络平台上更由于直接面对消费人群,投入少、回报高,已成为企业不可忽视的营销利器之一。但由于水军用户的存在,干扰了正常营销信息的传播过程,扰乱了有序公平的营销环境。因此,本章首先对社交网络营销信息传播过程进行抽象的分析,并将其分为两个阶段:初始传播阶段与热门传播阶段,定位建模的部分为直接粉丝链的部分;再分别以只存在正常用户、考虑水军用户的两种场景下进行建立以 IC 模型为基础的营销信息传播模型;通过抓取新浪微博营销信息的数据,验证了模型的有效性;在考虑水军用户的营销信息传播模型中,分析了水军用户的影响,以及结合模型的参数,给出了合理的治理意见。

### 第一节 营销信息传播过程分析

企业在社交网络平台上开展营销活动主要包括投放广告以及推广两种形式,广告投放由平台操作,水军较难干预;而推广形式的营销活动例如发布企业新商品的信息、商品的折扣促销信息、企业的相关新闻信息等,就为水军提供了乘虚而入的机会:正常用户浏览到企业相关的营销信息若产生兴趣,便会产生点赞、评论或转发行为,通过社交网络的粉丝关注、推送机制以及热门发现功能不断迭代给其他用户产生点赞、评论、转发行为,经过如此反复迭代,该营销信息的热度既可提升,曝光量也相应增加,即可达到一定的营销效果;水军用户在营销信息传播过程中,它们因利益驱动,对其产生大量点赞、评论、转发行为,借机提高营销信息的热度及曝光量,扰乱社交网络正常的营销信息传播过程,通过此手段,产生“不正当的竞争”。

首先定义两个概念:直接粉丝链与间接粉丝链。直接粉丝链是指直接关注企业的用户以及它们的粉丝递归构成的路径;间接粉丝链是指非直接关注企业的用户以及它们的粉丝递归构成的路径,如图 4-1 所示。当企业发布一条营销信息时,传播过程会分为两个阶段,第一阶段经过该企业直接粉丝链上的用户进行传播,直接粉丝链上的用户包括影响力较高的用户,也包括影响力较低的用户,影响力较高的用户能带动更多的下一级传播,也可以通过间接粉丝链直接由企业指派影响力较高的用户传播,通过这种影响力较高的用户也可以带动更多的下一级传播。在第一阶段的传播过程中,营销信息的热度和曝光量会逐渐升高;当此营销信息的热度和曝光量升高到一定程度时,可到达第二传播阶段:变成“热门”营销信息。通过热门发现、推送、订阅等功能,可被其他非企业指派的间接粉丝链上的用户浏览到(此

阶段直接粉丝链的传播部分归纳在初始传播阶段中，因为直接粉丝链上的用户并不是因为此营销信息“热门”而传播，而是本身就关注该企业而传播)，进而可带动更多的传播，当此营销信息的热度及传播量趋于饱和时，传播逐渐停止。由于本文以 IC 模型为基础，研究的是营销信息如何在邻居用户之间的传播过程，因此忽略了非直接粉丝链上的用户，只考虑直接粉丝链上的用户。

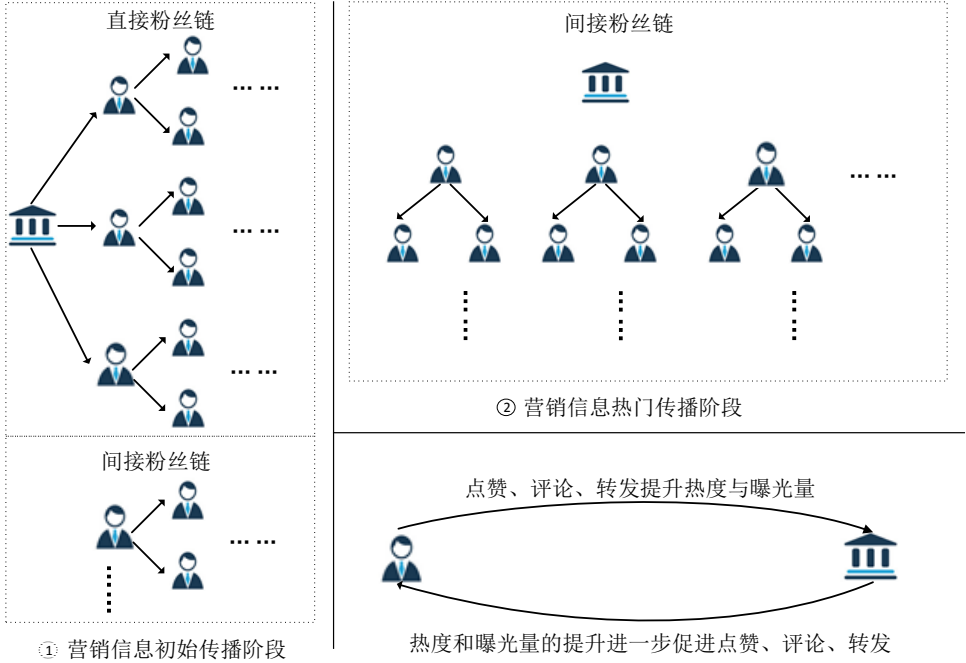


图 4-1 营销信息传播过程

在营销信息传播过程中，用户的点赞、评论、转发行为能提升信息的热度与曝光量，热度与曝光量的提升进一步促进更多用户产生点赞、评论、转发行为，通过此循环交替的过程，营销信息的热度与曝光量才得以不断提升，从而能够达到相应的营销效果。本文抽象出社交网络用户直接、共性的行为，只考虑了营销信息被点赞、评论、转发的三种行为，忽略了一些社交网络平台上用户非直接、具体的行为，例如：知乎、果壳、微博平台上的用户可以对营销信息进行搜索以提升其热度，微博平台上的用户还可以发表以#开头的相关营销信息的微博以提升其热度等，但是用户都可以在其上产生点赞、评论、转发基本社交行为。

根据 IC 模型理论，在  $t$  时刻当前节点  $V_i$  的邻居节点中，已被激活的邻居节点  $V_j$  会以独立的概率  $P(V_j, V_i)$  尝试激活当前节点  $V_i$ ，若当前节点  $V_i$  被其中一个邻居节点激活，那么在  $t+1$  时刻它会变成激活状态，将会尝试激活以它作为邻居的节点。这里节点被激活是指用户节点对营销信息产生了点赞、评论、转发任意一种的行为。

下文以 IC 模型为基础，在直接粉丝链上分别考虑两种场景建立具体的营销信息传播模型。

## 第二节 只存在正常用户的营销信息传播模型

根据社交网络与营销理论，正常用户对营销信息产生的点赞、评论、转发行为（激活）受到较多因素的影响，图 4-2 展示了一条营销信息在只包含正常用户的直接粉丝链上传播的过程图：企业用户发布一条营销信息，首先经过企业用户直接粉丝链上第一层的用户进行传播，每个用户节点受到若干影响因素决定是否被激活（点赞、评论或转发该营销信息），若某个用户被激活，那么经过此激活用户的粉丝进行第二层传播；若不被激活，则不经过此激活用户的粉丝进行第二层传播。经过层层迭代激活，此营销信息就被传播出去，当企业用户的直接粉丝链上不再有用户被激活，该营销信息的传播即达到稳定，传播过程结束。

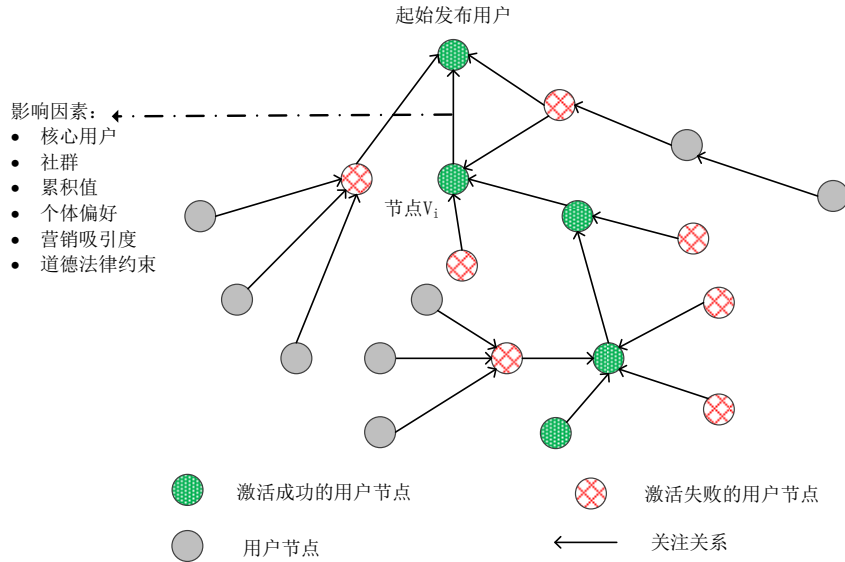


图 4-2 只存在正常用户的营销信息传播模型

影响正常用户激活的因素较多，本文根据现有理论归纳如下的 6 个影响因素：核心用户、社群、累积值、个体偏好、营销吸引度及道德法律约束，下文详细说明：

1) 社交网络的核心用户由于他们的社交地位、产品经验及知识背景可以影响其他用户，进而可以显著的影响其他用户对于产品的购买（Bao 和 Chang, 2014），因此正常用户对营销信息产生的行为会受到核心用户的影响，此影响以概率  $PCore(V_j, V_i)$  表示，按式(4.1)计算。在当前用户  $V_i$  的关注列表  $FollowUsers(V_i)$  中，每个已被激活的用户  $V_j$  拥有的粉丝数为  $Fans(V_j)$ ，粉丝数越大，说明该关注的用户  $V_j$  社交地位越高，对当前用户  $V_i$  的影响越大，激活当前用户  $V_i$  的概率越大；

$$PCore(V_j, V_i) = \frac{\text{Max}(Fans(V_j))}{\text{Sum}(Fans(FollowUsers(V_i)))} \quad (4.1)$$

$$V_j \in FollowUsers(V_i) \& V_j \in IsActivated$$



若当前用户的关注列表中只有一个用户被激活，那么  $V_j$  为当前用户的直接驱用户，否则  $V_j$  为影响力最大的那一个用户。

2) 研究表明，社交网络用户容易形成高度相关的紧密社区，因而能够对企业的网络营销产生正向或反向的作用（Java 等，2007）。因此正常用户对营销信息产生的行为会受到社群内用户的影响，此影响以概率  $PCommunity(V_i)$  表示，按式(4.2)计算。把当前用户  $V_i$  的关注列表  $FollowUsers(V_i)$  及粉丝列表  $FansUsers(V_i)$  组成一个社群，社群内已经激活的用户数越多，激活当前用户  $V_i$  的概率越大，即社群效应越明显；

$$PCommunity(V_i) = \frac{Count(V_k)}{Count(FansUsers(V_i) + FollowUsers(V_i))} \quad (4.2)$$

$$V_k \in FansUsers(V_i) \cup FollowUsers(V_i) \& V_k \in IsActivated$$

3) 在营销信息传播过程中，已激活的用户数量越多，说明该营销信息的热度与曝光量越高，因此，越容易激发当前用户  $V_i$  的兴趣，即当前用户  $V_i$  越易被激活，此影响以概率  $PAccumulate(V_i)$  表示，按式(4.3)计算：

$$PAccumulate(V_i) = \frac{Count(V_k)}{Count(AllUsers)} \quad (4.3)$$

$$V_k \in AllUsers \& V_k \in IsActivated$$

4) 此外，信息的配置以及客户的动机能够显著影响营销信息的传播（Je 等，2009），因此营销信息的吸引度、个体的偏好也可以影响用户对营销信息产生的行为。个体偏好的概率记为  $PReference(V_i)$ ，值为[0,1]范围的正态分布；营销信息吸引度的概率记为  $PAttractiveness(V_i)$ ，值为[0,1]范围的可变参数。当前用户  $V_i$  对企业的营销信息偏好越强、营销信息越富有吸引力，则当前用户  $V_i$  被激活的概率越大；

5) 最后，正常用户虽然出现在社交网络平台上，但仍属于“社会人”，仍受到一定程度的约束，其中包括现实社会的约束，也包括社交网络平台的约束。当面对不符合道德、违规的营销信息时，若约束越强，用户越会冷静对待，越不会跟风产生盲目的营销行为。此影响以概率  $PConstraint(V_i)$  表示，值为[0,1]范围的可变参数，概率值越大，则当前用户  $V_i$  越不容易被激活。

因此，营销信息传播过程在只包含正常用户的情况下，当前用户  $V_i$  被邻居节点  $V_j$  激活的概率  $P(V_j, V_i)$  按式(4.4)计算：

$$P(V_j, V_i) = PCore(V_j, V_i) + PCommunity(V_i) + PAccumulate(V_i) + PReference(V_i) + PAttractiveness(V_i) - PConstraint(V_i) \quad (4.4)$$

最后，使用热度（Heat）及曝光量（Exposure）评价营销信息传播的效果。营

销信息的热度（Heat）按式(4.5)计算，即当前已激活的节点用户个数：

$$Heat = Count(User), User \in AllUsers \& User \in IsActivated \quad (4.5)$$

营销信息的曝光量（Exposure）按式(4.6)计算，即当前已激活节点用户的粉丝数之和：

$$Exposure = Sum(Fans(User)), User \in AllUsers \& User \in IsActivated \quad (4.6)$$

因此，当只存在正常用户的情况下，营销信息传播模型如下：

Step 1 选择社交网络平台上一条营销信息，以发布的用户为起点抓取其粉丝链用户，构造直接粉丝链，全部用户都视为正常用户；

Step 2 激活起始节点用户  $V_j$ ；

Step 3  $V_j$  加入可激活节点列表  $L_1$  中；

Step 4 已激活的节点列表  $L_2$  赋值为空；

Step 5  $t=0$ ；

Step 6 初始化用户的 PReference、PAttractiveness、PConstraint；

Step 7 一直循环；

Step 7.1  $L_1$  追加给  $L_2$ ；

Step 7.2 使用  $L_2$  计算式(4.5)  $Heat$ 、式(4.6)  $Exposure$ ；

Step 7.3 对于  $L_1$  中每一个可激活节点用户  $V_j$ ；

Step 7.3.1 对于  $V_j$  的每一个粉丝节点用户  $V_i$ ；

Step 7.3.1.1 使用式(4.1)、式(4.2)、式(4.3)、式(4.4)计算  $PCore(V_j, V_i)$ 、 $PCommunity(V_i)$ 、 $PAccumulate(V_i)$ 、 $P(V_j, V_i)$ ；

Step 7.3.1.2 若  $P(V_j, V_i) > 1.0$ ，令  $P(V_j, V_i) = 1.0$ ；

Step 7.3.1.3 若  $rand[0,1] < P(V_j, V_i)$ ，则粉丝节点用户  $V_i$  标记为激活；

Step 7.4 清空  $L_1$ ；若标记激活的节点用户  $V_i \notin L_2$ ，则加入  $L_1$  中；

Step 7.5 若  $L_1$  不为空，进入下一次迭代， $t=t+1$ ；否则，退出循环；

### 第三节 考虑水军用户的营销信息传播模型

上文仅考虑只存在正常用户的场景，然而实际社交网络上存在着大量的水军用户。水军用户受雇于相关利益方，通过对营销信息有目的地大规模点赞、评论、转发，暗中操作营销信息传播的过程。水军用户的行为受利益驱动，利益驱动越大，水军用户数量越多，即意味着对营销信息点赞、评论、转发的数量就越多。以  $Interest$  代表利益， $PInterest(V_i)$  代表当前水军用户  $V_i$  被激活的概率，以此间接的表示水军用户的数量，按式(4.7)计算： $Interest$  越大， $PInterest(V_i)$  越大，当前水军

用户就越容易被激活，即代表雇佣的水军用户数量越多。

$$PInterest(V_i) = \frac{Interest}{Count(AllSpammerUser)} \quad (4.7)$$

$$Interest \subset [0, Count(AllSpammerUser)]$$

根据观察与分析，水军用户主要在以下两个方面产生作用，如图 4-3 所示：

1) 水军用户在直接粉丝链上被激活后，可以带动下一层级的粉丝用户激活。水军用户为了提高自己的中心性逃避检测，会关注一些正常用户。正常用户由于出于礼貌等心里因素，也会反向关注它们。因此，水军用户能够影响正常用户的  $PCore(V_j, V_i)$  和  $PCommunity(V_i)$ ，从而影响正常用户的激活概率，但是此影响作用较弱；

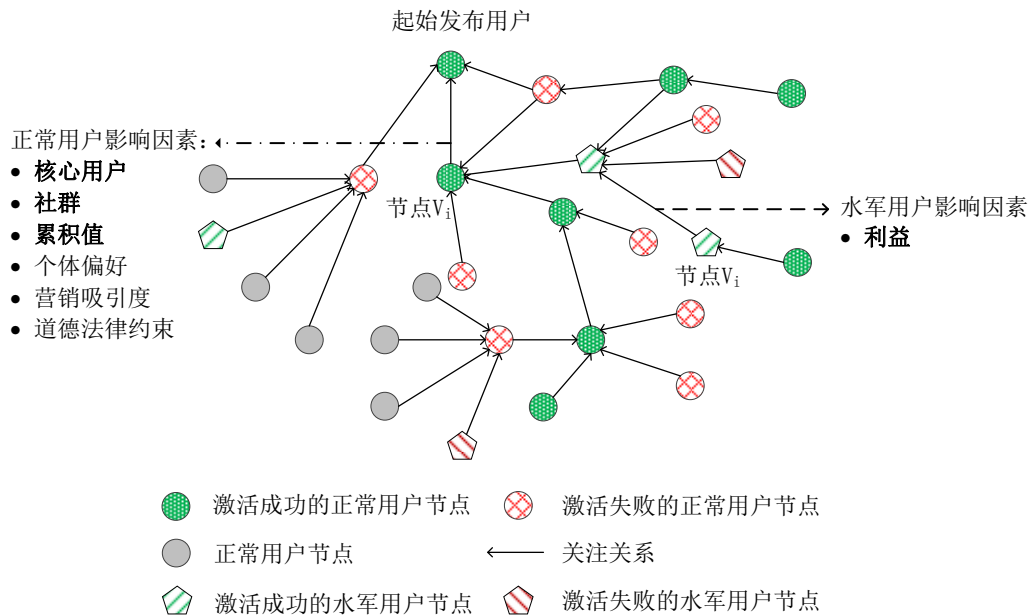


图 4-3 考虑水军用户的营销信息传播模型

2) 水军用户在直接粉丝链上被激活后，影响正常用户的累计激活概率。水军用户被激活后意味着对营销信息产生了点赞、评论或转发行为，那么营销信息的热度与传播量便会升高。营销信息的热度与传播量越高，越会有更多的正常用户参与其中产生营销行为（激活）。因此，水军用户主要影响了信息传播过程中已激活的用户数量，从而影响了正常用户的激活概率，即影响了正常用户的  $PAccumulate(V_i)$ ，此概率值比只存在正常用户的场景下要大。

因此，在考虑水军用户的场景下，水军用户既可以带动下一层级的正常用户传播，也可通过提升整体营销信息的热度与传播量从而带动更多的正常用户传播，前者影响较弱，后者为影响较强。对于水军用户自身的激活概率  $PInterest(V_i)$ ，按式(4.7)计算。

最后，在考虑水军用户的营销信息传播模型中，使用第三章训练好的水军检测模型来区分水军用户与正常用户。具体地，使用模型序列化技术导入训练好的模型。

因此，在考虑水军用户的场景下，营销信息传播模型如下：

Step 1 选择社交网络平台上的一条营销信息，以发布的用户为起点抓取其粉丝链用户，构造直接粉丝链；

Step 2 使用序列化技术导入第三章训练好的在线水军检测模型，将直接粉丝链上的用户区分为两类用户，正常用户与水军用户；

Step 3 激活起始节点用户  $V_j$ ；

Step 4  $V_j$  加入可激活节点列表  $L_1$  中；

Step 5 已激活的节点列表  $L_2$  赋值为空；

Step 6  $t=0$ ；

Step 7 初始化正常用户的 PReference、PAttractiveness、PConstraint；

Step 8 初始化水军用户的 PInterest；

Step 9 一直循环；

Step 9.1  $L_1$  追加给  $L_2$ ；

Step 9.2 使用  $L_2$  计算式(4.5) *Heat*、式(4.6) *Exposure*；

Step 9.3 对于  $L_1$  中每一个可激活节点用户  $V_j$ ；

Step 9.3.1 对于  $V_j$  的每一个粉丝节点用户  $V_i$ ；

当  $V_i$  属于正常用户：

Step 9.3.1.1 使用式(4.1)、式(4.2)、式(4.3)、式(4.4)计算  $\text{PCore}(V_j, V_i)$ 、 $\text{PCommunity}(V_i)$ 、 $\text{PAccumulate}(V_i)$ 、 $\text{P}(V_j, V_i)$ ；

Step 9.3.1.2 若  $\text{P}(V_j, V_i) > 1.0$ ，令  $\text{P}(V_j, V_i) = 1.0$ ；

Step 9.3.1.3 若  $\text{rand}[0,1] < \text{P}(V_j, V_i)$ ，则粉丝节点用户  $V_i$  标记为激活；

当  $V_i$  属于水军用户：

Step 9.3.1.4 若  $\text{rand}[0,1] < \text{PInterest}(V_i)$ ，则粉丝节点用户  $V_i$  标记为激活；

Step 9.4 清空  $L_1$ ；若标记激活的节点用户  $V_i \notin L_2$ ，则加入  $L_1$  中；

Step 9.5 若  $L_1$  不为空，进入下一次迭代， $t=t+1$ ；否则，退出循环；

## 第四节 实验分析

### 一、数据抓取与分析

首先跟踪抓取社交网络平台上企业有关的营销信息，由于当前正是互联网企业百花齐放的时期，因此本文选取有关互联网产品的营销信息，如图 4-4：选取滴

滴滴出行公司在新浪微博平台上推广紧急联系人功能的产品营销信息，只抓取该营销信息的转发关系数据以及滴滴出行用户的直接粉丝链数据。

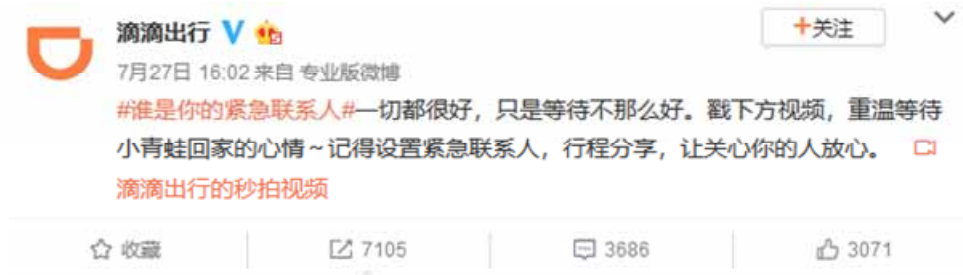


图 4-4 新浪微博平台上滴滴出行发布的产品营销信息

实验的软硬件环境配置与第三章一致，仍使用 Python 作为程序设计语言。首先获取此营销信息的 mid 号，为 4266425340068297，接着获取该营销信息转发关系的用户 ID，根据用户 ID 号抓取相应信息，包括基本信息、粉丝关注信息以及微博信息存入 MongoDB 数据库中，此为转发关系数据。再以滴滴出行用户（ID 号为 2838754010）起点，跟据其粉丝列表使用宽度优先搜索策略抓取粉丝用户的相关信息，包括基本信息、粉丝关注信息以及微博信息存入 MongoDB 数据库中，此为直接粉丝链数据。

图 4-5（左）为此营销信息的转发关系图，滴滴出行发布此营销信息后不仅由直接粉丝链上的用户转发，还包括一些由企业指定的非直接粉丝链上的用户转发，如“Youtube 精选”、“大哥王振华”等高影响力用户，通过这些高影响力用户更能有效地带动营销信息的传播。由于接口权限不足及数据过大的限制，并不能抓取全部的转发关系及用户，而是抓取一部分代表性的数据，通过去除转发链断掉及不完整的数据，共得到 1515 条转发关系，1477 个参与转发的用户（若有足够抓取权限，能够抓取 7105 个转发关系）。

图 4-6 为滴滴出行的直接粉丝链图，即直接关注滴滴出行的用户以及它们的粉丝递归构成的图，滴滴出行影响力较大，粉丝较多，由于用户权限的限制，只能抓取前 5 页的粉丝数据，宽度优先搜索策略的层数为 2，最终共得到 2060 个粉丝关系，1981 个粉丝用户。

最后，为了得到转发关系图中直接粉丝链转发的部分，剔除转发关系图中不在直接粉丝链上的用户，并且剔除重复的转发关系，最终得到 954 个用户节点，954 个转发关系，如图 4-5（右）所示。由于只能抓取到直接粉丝链上一部分的用户，这里也只能提取一部分的用户，因此导致迭代层数较少，即直接关注企业并转发此营销信息的用户并没有递归的构成传播路径。所以对营销信息真正的传播过程不易检验，只能对营销信息最终传播的结果进行检验。

下文的实验流程如下：

- 1) 在滴滴出行的直接粉丝链上对模型进行检验，与转发关系图中直接粉丝链转发的部分作对比，验证本文构建的营销信息传播模型的有效性。将朴素的 IC 营销信息传播模型、只存在正常用户的 IC 营销信息传播模型、考虑水军用户的 IC 营销信息传播模型作对比分析；
- 2) 对考虑水军用户的 IC 营销信息传播模型进行分析，分析水军用户对信息传播过程的影响，并结合模型的参数，给出个人、企业及相关单位对应的治理意见。

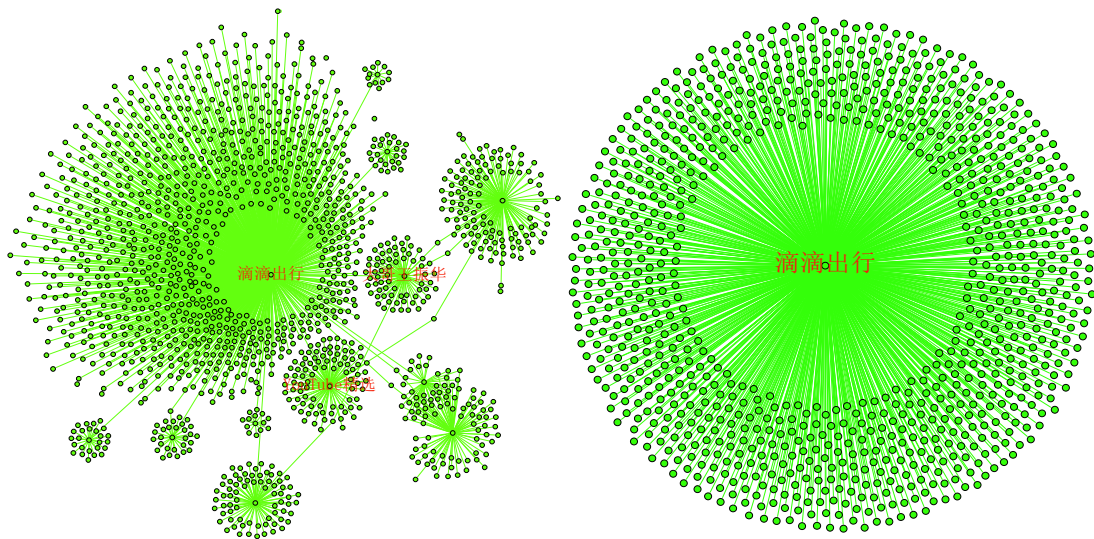


图 4-5 滴滴出行产品营销信息的转发关系图（左）  
与其上的直接粉丝链转发的部分（右）

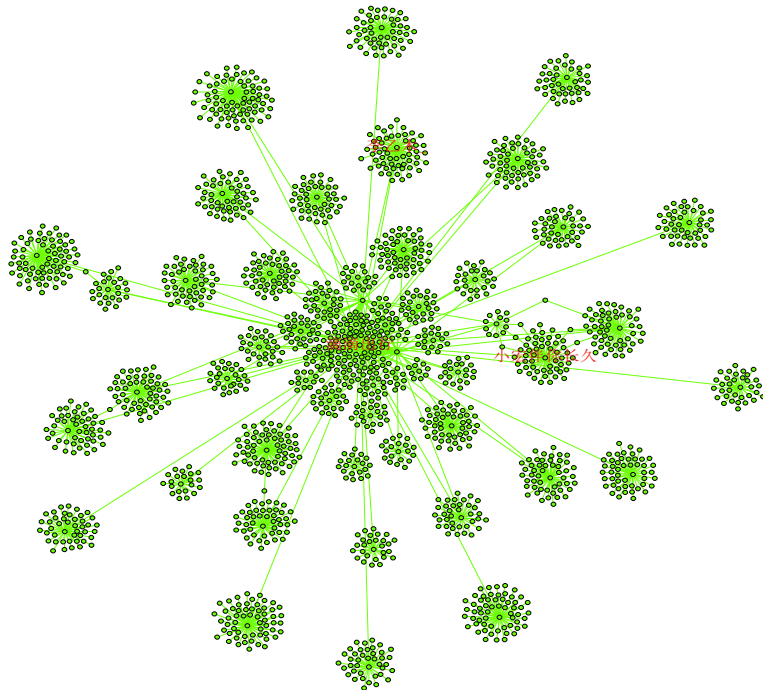


图 4-6 滴滴出行的直接粉丝链图

## 二、营销信息传播模型检验

在滴滴出行的直接粉丝链上对以下三种营销信息传播模型进行检验，模型配置如下：

1) 朴素的 IC 营销信息传播模型：在此模型中用户被其邻居节点激活的概率  $P(V_j, V_i)$  设为固定值，代表新浪微博平台上用户点赞、评论或转发一条营销信息的概率，设置为 0.3；

2) 只存在正常用户的 IC 营销信息传播模型：PReference 初始化为均值 0 方差为 0.3 的正态分布、PAttractiveness 初始化为 0.1、PConstraint 初始化为 0；

3) 考虑水军用户的 IC 营销信息传播模型：使用第三章的水军检测模型将用户划分为正常用户与水军用户：在 1981 个用户中，水军用户个数为 661，占比 33%，正常用户个数为 1320，占比 67%，说明此营销信息有一定的水军用户参与。模型的参数如下配置，PInterest 设置为 0.6、PReference 初始化为均值 0 为方差为 0.3 的正态分布、PAttractiveness 初始化为 0.1、PConstraint 初始化为 0。

表 4-1 营销信息传播模型的检验

| 营销信息传播模型             | 热度（准确率）     | 热度提升<br>百分比     | 曝光量（准确率）      | 曝光量提升<br>百分比    |
|----------------------|-------------|-----------------|---------------|-----------------|
| 真实传播                 | 954         |                 | 17973         |                 |
| 朴素的 IC 营销信息传播模型      | 331(34.70%) |                 | 7710(42.90%)  |                 |
| 只存在正常用户的 IC 营销信息传播模型 | 424(44.44%) | ↑ 9.74%         | 10074(56.05%) | ↑ 13.15%        |
| 考虑水军用户的 IC 营销信息传播模型  | 638(66.88%) | ↑ <b>32.18%</b> | 14962(83.25%) | ↑ <b>40.35%</b> |

三个模型的传播结果与实际传播的结果（转发关系图中直接粉丝链转发的部分）作对比。评价指标采用热度式(4.5)（激活的节点个数，即转发的规模（海沫和郭庆，2016））与曝光量式(4.6)（激活节点的粉丝数，为实际产品中采用的评价指标<sup>①</sup>）。表 4-1 显示了模型传播结果的对比，可以看出在符合实际意义的参数配置下，在只存在正常用户的 IC 营销信息传播模型中，由于加入了影响力、社群、累积值等因素，更符合实际营销信息传播过程，比朴素的 IC 营销信息传播模型效果更好，热度提升了 9.74%，曝光量提升了 13.15%；进一步地，在考虑水军用户的 IC 营销信息传播模型中更为准确的拟合出了真实传播的热度与曝光量，与朴素的 IC 营销信息传播模型相比，热度提升了 32.18%，曝光量提升了 40.35%。因此，在 IC 营销信息传播模型中纳入水军因素更符合此营销信息实际传播的情况，拟合结

<sup>①</sup>参见 <http://www.weiboreach.com/>



果更加准确。

图 4-7 显示了在上述参数的配置下三个模型的迭代过程，左图为营销信息热度，右图为营销信息曝光量。由于抓取的滴滴出行直接粉丝链的最大层数为 2，因此迭代的次数不多。图中可以看出，朴素的 IC 营销信息传播模型迭代速度较慢，热度和曝光量缓慢达到饱和，只存在正常用户的 IC 营销信息传播模型迭代速度适中，热度和曝光量的饱和值相比小幅提升，考虑水军用户的 IC 营销信息传播模型迭代更为快速，热度和曝光量的饱和值更趋近此营销信息的实际值，对此营销信息的传播过程拟合效果更好。

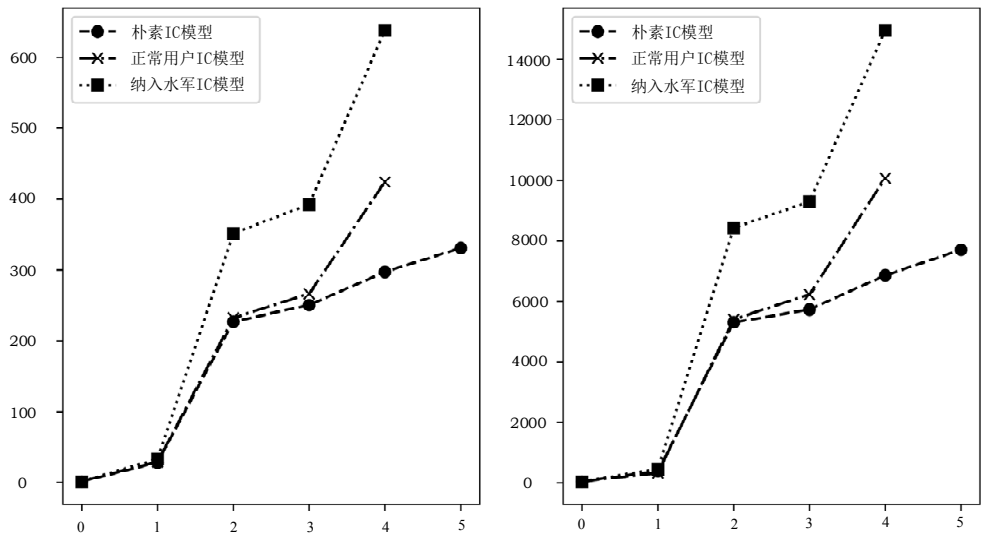


图 4-7 营销信息传播模型的迭代过程（左为热度，右为曝光量）

### 三、考虑水军用户营销信息传播模型的水军用户影响分析

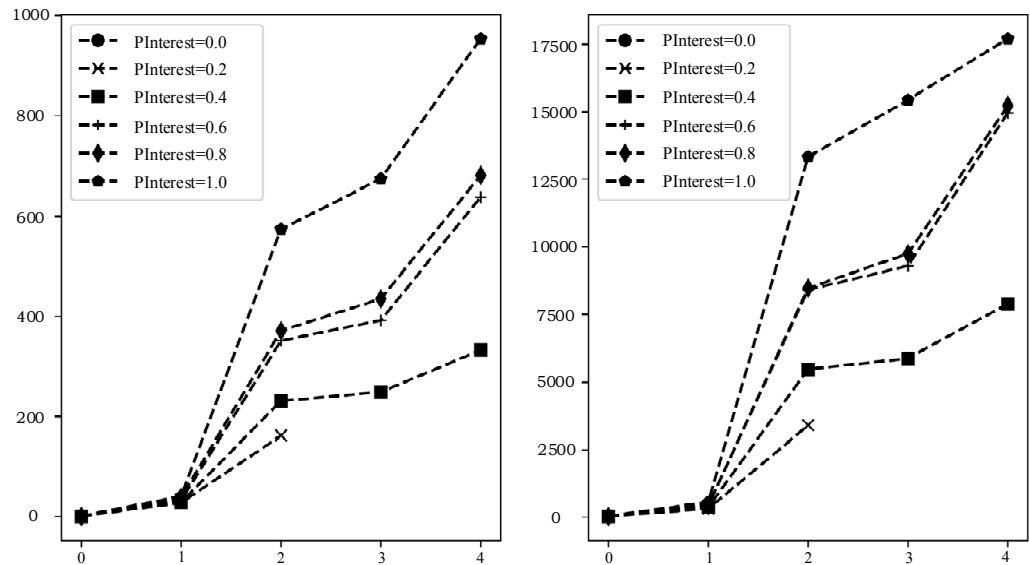


图 4-8 水军受到利益的影响（左为热度，右为曝光量）



图 4-8 展示了水军受到的利益与营销信息传播的关系，参数配置如下： $P_{Reference}$  初始化为均值 0 为方差为 0.3 的正态分布、 $P_{Attractiveness}$  初始化为 0.1、 $P_{Constraint}$  初始化为 0。图中可以看出，随着利益驱动  $P_{Interest}$  的增大，营销信息传播的热度与曝光量增长速度变快，饱和值增大，即代表水军用户受到的利益驱动越大，数量越多，那么对营销信息传播过程的影响就越明显，与现实情况相吻合。

从营销信息传播模型或过程的角度来看，水军用户的具体影响不仅表现在激活的水军用户可以直接带动它的粉丝用户激活，更明显的是激活的水军用户增加了信息传播过程中已激活的用户数量，进一步促进更多的正常用户激活。图 4-9 显示了在第二节模型参数配置的下，在只存在正常用户和考虑水军用户的模型中，正常用户累计激活概率  $P_{Accumulate}$  的迭代过程对比。从中可以看出，与只存在正常用户的模型相比，考虑水军用户模型的正常用户累计激活概率  $P_{Accumulate}$  迭代增长更快，概率值更大，因为概率值的变大，所以激活的正常用户数会相应增多。因此，水军用户通过此作用，“暗中”提升了营销信息的最终传播效果。

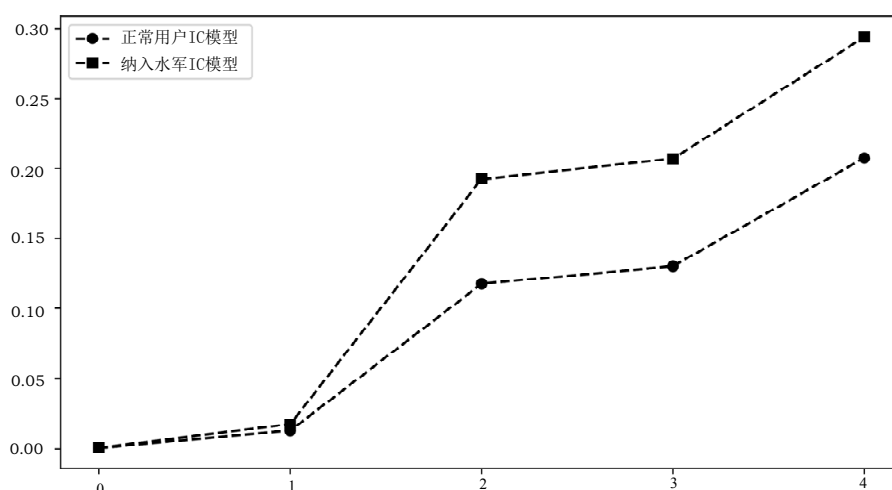


图 4-9 营销信息传播模型中正常用户累计激活概率的迭代过程对比

从单个企业来看，通过雇佣水军来提升营销信息效果也许是有利的，但从整个市场来看，这将无疑会产生极其有害的结果。如果一家企业在社交网络上通过雇佣水军为其营销信息大量炒作，倘若取得了较好的效果，那么其他企业便会纷纷效仿，当此举成为行业暗藏的“潜规则”，企业在社交网络上开展每一次营销活动，都要雇佣水军为其炒作，那么便会造成企业之间严重的不公平竞争，造成失衡的社交网络营销环境。不仅如此，如果企业雇佣水军恶意攻击其他企业声誉或其产品，例如“蒙牛陷害门”，更是违背道德法律，进一步恶化社交网络营销环境。更进一步地看，当社交网络充斥着大量的水军用户，任其肆意的炒作，使得传播的营销信息难辨真假，那么社交平台自身的公信力便会慢慢丢失，当用户把社交网络平台只当作虚假信息传递的场所，便对其慢慢丧失兴趣，最终导致用户的流失与平台的消亡。

#### 四、考虑水军用户营销信息传播模型的灵敏度分析

因此,为了维护公平、有序的社交网络营销环境,保障社交网络平台能够稳定、长远的发展,需有效治理水军。以下将从3个方面进行探讨:

1) 对于企业来说,无需强行雇佣水军为其营销信息大肆推销炒作,因为企业声誉、产品好坏实质由企业本身决定,即使企业通过水军强行推广其产品或声誉,也会经不住市场的考验,更是白白浪费了相应财力。企业若想通过社交网络平台开展效果较好的营销活动,那么可以从发布的营销信息内容考虑。图4-10展示了营销信息吸引度的灵敏度分析,参数配置如下:  $PReference$  初始化为均值0 为方差为0.3的正态分布、 $PConstraint$  初始化为0、 $PInterest$  初始化为0。图中可以看到随着营销信息吸引度  $PAttractiveness$  的增大,企业发布的营销信息热度与曝光量增长速度变快,饱和值增大。因此,企业可以考虑如何将营销信息内容变的更引人入胜以此提高营销效果。例如,可以从趣味性、互动性、利益性、故事性等方面考虑:对于新产品的发布与推广,可以发布一些内含有奖竞猜、限量免费赠送等营销信息,使用图文配套、甚至可以为新发布的产品拍摄小短片,以此增加用户的参与行为,提高营销信息的热度与曝光量;对于企业的品牌建设,可以发布一些关于企业自身的经营理念、重要事件,也可以发布关于企业故事、品牌故事等营销信息,获得用户的心里认可,从而在不经意间提升营销效果。

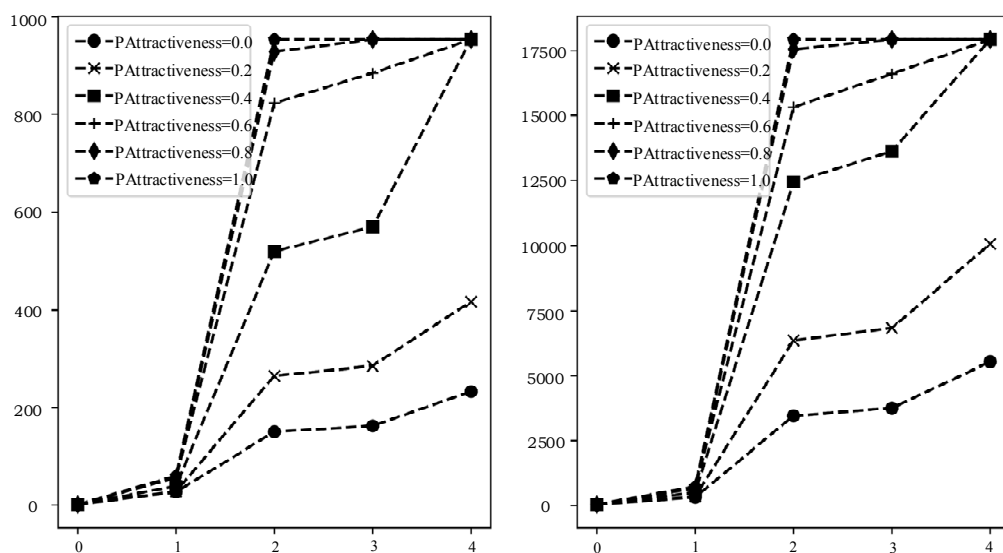


图 4-10 营销信息吸引度的影响 (左为热度, 右为曝光量)

2) 由于水军用户扰乱了营销信息传播过程,因此需有相关单位针对性的进行治理。图4-11展示了约束的灵敏度分析,参数配置如下:  $PAttractiveness$  初始化为0.1,  $PReference$  初始化为均值0 为方差为0.3的正态分布,  $PInterest$  初始化为0.6。图中可以看出随着约束  $PConstraint$  的增大,企业发布的营销信息热度与曝光量增

长速度减小，饱和值减小。说明随着治理的加大，可有效地减少水军对营销信息传播过程的干扰。可从社交网络平台与政府治理两方面入手。第一，针对社交网络平台，应不断提高网络技术能力，从技术上修补漏洞，避免被利益驱动者轻易使用。例如，对于社交网站的 API 接口，应使用多次混淆加密的方式，增加破解的难度；对于访问者的 IP 地址，应加大限制，避免被使用代理后不限频率的访问；对于访问者的账号，增加验证码的识别难度，使用人体工程学的方式设计验证码，避免被机器学习等新型识别技术轻松破解。当然，更应利用数据挖掘、机器学习等技术开发新型的、有效的水军检测模型，例如本文第三章所示的模型。由于利益驱动者的网络技术能力不断提升，这就需要社交平台的技术能够与时俱进，以维护可靠的社交网络营销平台。第二，针对政府管治，国家已经制定了《关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释》，对水军的若干行为做出了一些规定，但远远不够，利益的巨大诱惑依旧使得利益驱动者大肆地产生水军，游走于法律的灰色地带。因此，政府需制定明确的、全面的互联网法律，加大对水军的约束，以促进建立良好、有序、公平的社交网络营销环境。

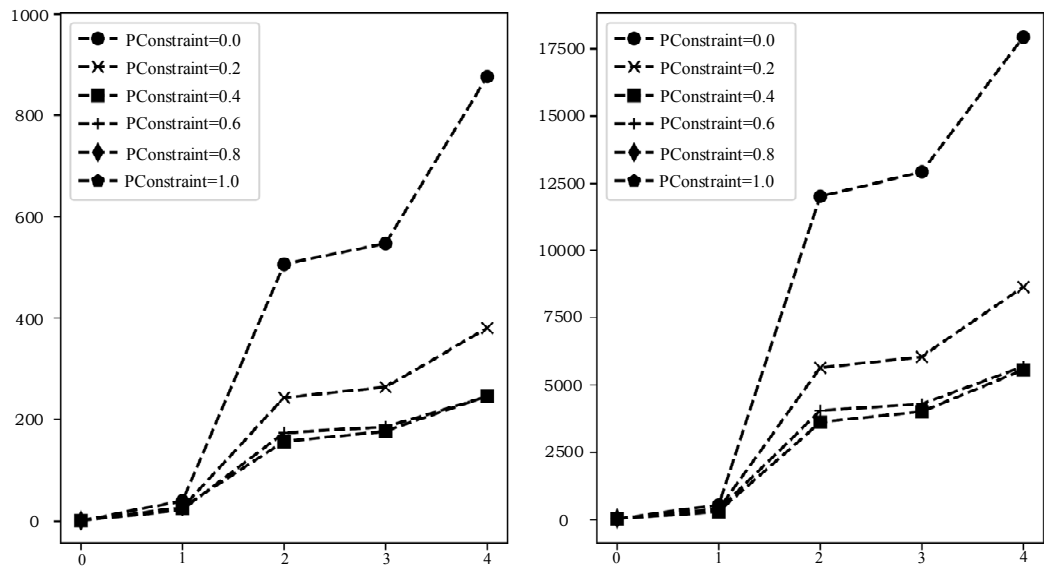


图 4-11 约束的影响（左为热度，右为曝光量）

3) 除了企业自身、社交网络平台和政府以外，水军用户之所以有效，还有一个较大的影响因素就是个体。图 4-12 展示了个体偏好的灵敏度分析，参数配置如下：PAttractiveness 初始化为 0.1、PInterest 初始化为 0.6、PConstraint 初始化为 0。图中显示了随着个体偏好 PReference 的增大，企业发布的营销信息热度与曝光量增长速度变快，饱和值增大。个体的媒介素质越低，越容易产生盲从、轻信的行为，越容易扩大水军对其的作用力。例如最近深陷造假事件的“科大讯飞”，虽后来澄清

事件，但在 2018 年 9 月 25 日，股价跌近 4%，报收价 28.93 元/股<sup>①</sup>。因此，每个人都应培养明辨是非的能力，不盲从也不去造谣，为营造良好的社交网络营销环境做出一份贡献。

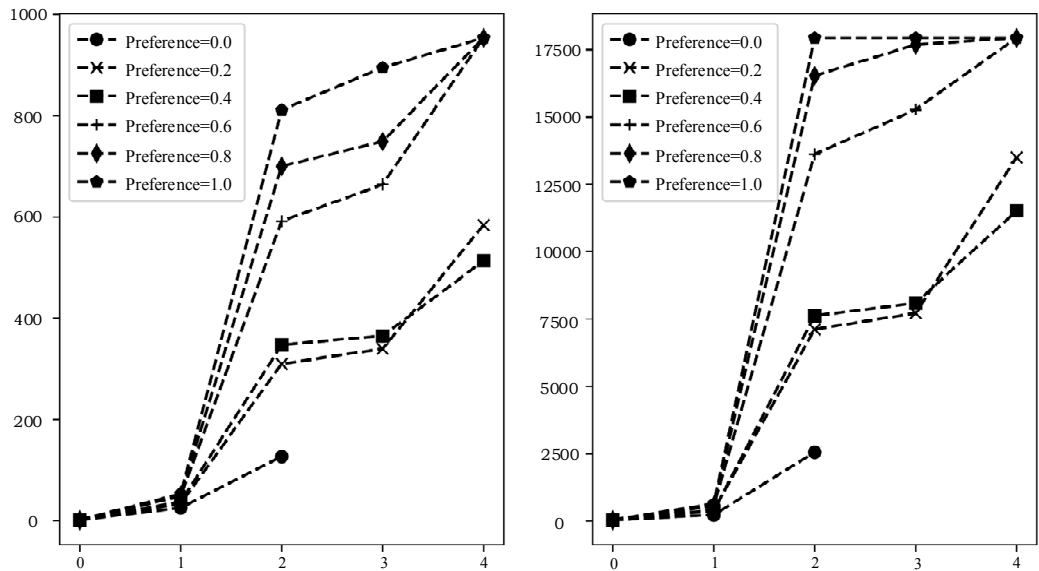


图 4-12 个体偏好的影响（左为热度，右为曝光量）

### 第五节 本章小结

本章首先对社交网络营销信息的传播过程做出了简单的分析，针对营销信息传播过程中直接粉丝链的部分，以 IC 模型为基础，分别以只包含正常用户、考虑水军用户的两种场景下建立营销信息传播模型；通过抓取新浪微博营销信息的数据，验证了构建模型的有效性；并且在考虑水军用户的营销信息传播模型中，分析了水军用户的干扰过程，并且结合了模型的参数，给出了个人、企业、平台与政府相关治理意见。

<sup>①</sup>参见 <http://tech.sina.com.cn/it/2018-09-26/doc-ihkmwytp1444814.shtml>

# 总结与展望

## 一、本文工作总结

随着社交网络的蓬勃发展，用户规模不断增长，每天数以万计的信息在其上流动，已成为人们获取信息、分享交流的首选之地。企业也利用如此庞大规模的直接消费用户群体，在社交网络上开展产品推广、品牌建立、客户关系维护等营销活动，从而提高产品销量、维护企业形象。然而，社交网络上也存在着此类用户——水军用户，它们看不见、摸不着，我们却能感知它们的存在。它们在社交网络上有目的地产生大规模的关注、点赞、评论、转发等行为，干扰了正常营销信息的传播过程，扰乱了公平有序的社交网络营销环境。因此，本文基于以下两个视角进行了研究：第一，对社交网络愈发隐蔽的水军用户进行分析与识别；第二，水军用户对营销信息传播过程的影响进行建模与分析。

对于水军用户策略的不断提升、愈发趋向于正常用户，传统研究使用静态行为特征已无法有效地检测出愈发隐蔽的水军用户的问题。本文构造了简单、易于使用的动态行为特征，在离线环境下分析了水军用户与正常用户动态行为特征的差异性与有效性；为了弥补传统研究在在线水军检测模型方面的不足，结合动静态行为特征以及 Tri-Training 模型构建了在线水军检测模型；通过抓取的新浪微博数据，验证了该模型的有效性。

对于水军用户影响了营销信息传播的过程从而扰乱社交网络营销环境的问题。本文结合社交网络营销理论与 IC 模型建立了在两种场景下的营销信息传播模型：只存在正常用户的营销信息传播模型与考虑水军用户的营销信息传播模型；通过抓取的新浪微博营销信息的数据，验证了考虑水军用户营销信息传播模型的有效性，分析了水军用户对营销信息传播过程的影响，并且根据模型的参数给出了相关方治理意见。

本文的主要创新点如下：

- 1) 在社交网络水军识别方面，不仅把水军检测领域从对静态行为特征的研究推广到了对动态行为特征的研究，并且弥补了在在线检测场景下模型的研究不足；
- 2) 在社交网络营销信息传播模型方面，考虑到水军用户的影响因素，构建了考虑水军用户的营销信息传播模型，较细粒度地分析了水军用户对营销信息传播过程的影响；
- 3) 使用抓取的新浪微博平台上的数据，验证了动态行为特征、在线水军检测模型以及考虑水军用户营销信息传播模型的有效性；并且结合模型的参数，为个人、企业和相关部门提供了治理意见。

## 二、未来研究展望

由于实验条件与本人时间有限，对问题的认识以及所做的工作仍存在不足。在未来的研究工作中，应从以下几个方面进一步深入研究：

1) 在社交网络水军识别方面，由于实验条件有限，本文抓取的数据不足，未能充分的模拟实际社交网络的场景，未能充分展现模型的动态特征筛选与在线检测的能力。因此，在未来的工作中，应搭建分布式的集群计算环境，抓取足量的数据，能够全方位的测试模型在线检测的能力；此外，由于社交网络中存在大量的非结构化数据，不仅局限于本文所用的数值、文本型数据，还存在大量的图片、声音等多媒体数据，如何充分利用这些数据对水军检测领域是个挑战；最后，如何构造更高效的模型也是个难点，因在线检测属于流数据计算的范畴，如何利用流数据算法构造在线检测模型也是待挖掘的方向之一。

2) 在社交网络营销信息传播模型方面，本文只考虑了对主体影响较为常见的因素，然而对于其他存在的影响因素以及这些影响因素之间复杂的交互关系考虑不足；此外，本文只模拟了单条营销信息的传播过程以及单方水军的影响因素，当存在正反两方水军或多条营销信息的情况都未纳入考虑。因此，在未来的研究工作中，应在模型中纳入更为全面的影响因素，并且考虑到存在正反两方水军以及多条营销信息的场景，探究水军更复杂的影响过程，提供更全面的决策意见。

## 参考文献

- [1] 陈慧娟, 郑啸, 陈欣. 微博网络信息传播研究综述[J]. 计算机应用研究, 2014, 31(2):333-338.
- [2] 陈侃, 陈亮, 朱培栋, 熊岳山. 基于交互行为的在线社会网络水军检测方法[J]. 通信学报, 2015, 36(7):120-128.
- [3] 程晓涛, 刘彩霞, 刘树新. 基于关系图特征的微博水军发现方法[J]. 自动化学报, 2015, 41(9):1533-1541.
- [4] 程晓涛. 微博网络水军识别技术研究[D]. 郑州: 解放军信息工程大学, 2015.
- [5] 郭东伟, 乌云娜, 邹蕴, 孟祥燕. 基于非理性博弈的舆情传播仿真建模研究[J]. 自动化学报, 2014, 40(8):1721-1732.
- [6] 海沫, 郭庆. 在线社交网络信息传播模型研究[J]. 小型微型计算机系统, 2016, 37(8):1672-1679.
- [7] 韩忠明, 许峰敏, 段大高. 面向微博的概率图水军识别模型[J]. 计算机研究与发展, 2013, 50(s2):180-186.
- [8] 李立耀, 孙鲁敬, 杨家海. 社交网络研究综述[J]. 计算机科学, 2015, 42(11):8-21.
- [9] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法[J]. 计算机学报, 2015, 38(8):1592-1617.
- [10] 莫倩, 杨珂. 网络水军识别研究[J]. 软件学报, 2014, 25(7):1505-1526.
- [11] 孙卫强. 基于深度信念网络的网络水军识别研究[D]. 长沙: 湘潭大学, 2015.
- [12] 谭侃, 高旻, 李文涛, 田仁丽, 文俊浩, 熊庆宇. 基于双层采样主动学习的社交网络虚假用户检测方法[J]. 自动化学报, 2017, 43(3):448-461.
- [13] 王彦本, 蔡皖东. 社交网络中考虑遗忘机制的谣言传播[J]. 西北工业大学学报, 2016, 34(2):349-355.
- [14] 叶施仁, 叶仁明, 朱明峰. 基于网络关系的微博水军集团发现方法[J]. 计算机工程与应用, 2017, 53(6):96-100.
- [15] 易成岐, 鲍媛媛, 薛一波, 姜京池. 新浪微博的大规模信息传播规律研究[J]. 计算机科学与探索, 2013, 7(6):551-561.
- [16] 张良. 面向新浪微博的水军识别技术的研究与实现[D]. 长沙: 国防科学技术大学, 2015.
- [17] 张艳梅, 黄莹莹, 甘世杰, 丁熠, 马志龙. 基于贝叶斯模型的微博网络水军识别算法研究[J]. 通信学报, 2017, 38(1):44-53.
- [18] 郑春东, 韩晴, 王寒. 网络水军言论如何左右你的购买意愿[J]. 南开管理评论,

- 2015, 18(1):89-97.
- [19]中共中央网络安全和信息化委员会办公室. 第 42 次《中国互联网络发展状况统计报告》[EB/OL]. [http://www.cac.gov.cn/2018-08/20/c\\_1123296882.htm](http://www.cac.gov.cn/2018-08/20/c_1123296882.htm), 2018-08-20.
- [20]周志华. 基于分歧的半监督学习[J]. 自动化学报, 2013, 39(11):1871-1878.
- [21]朱湘, 贾焰, 聂原平,等. 基于微博的事件传播分析[J]. 计算机研究与发展, 2015, 52(2):437-444.
- [22]Ahani A., Rahim N. Z. A., Nilashi M. Forecasting social CRM adoption in SMEs: A combined SEM-neural network method[J]. *Computers in Human Behavior*, 2017, 75:560-578.
- [23]Bao T., Chang T. L. S. Finding disseminators via electronic word of mouth message for effective marketing communications[J]. *Decision Support Systems*, 2014, 67(C):21-29.
- [24]Barbieri N., Bonchi F., Manco G. Topic-aware social influence propagation models[J]. *Knowledge & Information Systems*, 2013, 37(3):555-584.
- [25]Benevenuto F., Magno G., Rodrigues T., Almeida V. Detecting spammers on twitter[C]. *In Proceedings of 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, Redmond, Washington, US, July 13-14, 2010: 75-83.
- [26]Berkhin P. A survey of clustering data mining techniques[J]. *Grouping Multidimensional Data*, 2006, 43(1):25-71.
- [27]Bottou L., Bengio Y. Convergence properties of the K-means algorithms[C]. *In Proceedings of the 7th International Conference on Neural Information Processing Systems*, Denver, Colorado, USA, 1995:585-592.
- [28]Chang B., Xu T., Liu Q., Chen E. H. Study on information diffusion analysis in social networks and its applications[J]. *International Journal of Automation and Computing*, 2018, 15(4):1-26.
- [29]East R., Hammond K., Lomax W. Measuring the impact of positive and negative word of mouth on brand purchase probability[J]. *International Journal of Research in Marketing*, 2008, 25(3):215-224.
- [30]Fu Q., Feng B., Guo D., Li Q. Combating the evolving spammers in online social networks[J]. *Computers & Security*, 2018, 72:60-73.
- [31]Goldenberg J., Libai B., Muller E. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata[J]. *Academy of Marketing Science*



*Review*, 2001, 9(3):1-18.

- [32] Goldenberg J., Libai B., Muller E. The chilling effects of network externalities[J]. *International Journal of Research in Marketing*, 2010, 27(1):4-15.
- [33] Gao H., Hu J., Wilson C., Li Z., Chen Y., Zhao B.Y. Detecting and characterizing social spam campaigns[C]. *In Proceedings of the 17th ACM Conference on Computer and Communications Security*, Chicago, Illinois, USA, October 4-8, 2010: 681-683.
- [34] Java A., Song X., Finin T., Tseng B. Why we twitter: understanding microblogging usage and communities[C]. *In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, California, USA, August 12-12, 2007: 56-65.
- [35] Jiang M, Cui P, Faloutsos C. Suspicious behavior detection: current trends and future directions[J]. *IEEE Intelligent Systems*, 2016, 31(1):31-39.
- [36] Jun T., Kim J. Y., Kim B. J., Choi M.Y. Consumer referral in a small world network[J]. *Social Networks*, 2006, 28(3):232-246.
- [37] Kempe D., Kleinberg J., Tardos É. Maximizing the spread of influence through a social network[C]. *In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, August 24-27, 2003:137-146.
- [38] Kiesling E., Günther M., Stummer C., Wakolbinger L. M. Agent-based simulation of innovation diffusion: a review[J]. *Central European Journal of Operations Research*, 2012, 20(2):183-230.
- [39] Kumar V., Petersen J. A., Leone R. P. How valuable is word of mouth?[J]. *Harvard Business Review*, 2007, 85(10):139-144.
- [40] Li F., Du T. C. The effectiveness of word of mouth in offline and online social networks[J]. *Expert Systems with Applications*, 2017, 88:338-351.
- [41] Li Z., Zhang X., Shen H., Liang W., He Z. A semi-supervised framework for social spammer detection[C]. *In Proceedings of Advances in Knowledge Discovery and Data Mining, PAKDD 2015*, Ho Chi Minh City, Vietnam, May 19-22, 2015: 177-188.
- [42] Liu L., Lu Y., Luo Y., Zhang R., Itti L., Lu J. Detecting" smart" spammers on social network: a topic model approach[J]. arXiv preprint arXiv:1604.08504, 2016.
- [43] Miller Z., Dickinson B., Deitrick W., Hu W., Wang A. H. Twitter spammer detection using data stream clustering[J]. *Information Sciences*, 2014, 260(1):64-73.
- [44] Peres R., Muller E., Mahajan V. Innovation diffusion and new product growth models: A critical review and research directions[J]. *International journal of research in*

*marketing*, 2010, 27(2):91-106.

- [45] Phelps J., Lewis R., Mobilio L., Perry D., Raman N. Viral marketing or electronic word-of-mouth advertising: examining consumer responses and motivations to pass along email[J]. *Journal of Advertising Research*, 2004, 44(44):333-348.
- [46] Raschka S. Python machine learning[M]. Birmingham, UK: Packt Publishing Ltd, 2015:42-46.
- [47] Sedhai S., Sun A. Semi-supervised spam detection in twitter stream[J]. *IEEE Transactions on Computational Social Systems*. 2018, 5(1):169-175.
- [48] Shen H., Ma F., Zhang X., Zong L., Liu X., Liang W. Discovering social spammers from multiple views[J]. *Neurocomputing*, 2016, 225: 49-57.
- [49] Tang X. L., Han M. Ternary reversible extreme learning machines: the incremental tri-training method for semi-supervised classification[J]. *Knowledge & Information Systems*, 2010, 23(3):345-372.
- [50] Thiriot S. Word-of-mouth dynamics with information seeking: Information is not (only) epidemics[J]. *Physica A Statistical Mechanics & Its Applications*, 2018, 492:418-430.
- [51] Trusov M., Bucklin R. E., Pauwels K. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site[J]. *Journal of Marketing*, 2009, 73(5):90-102.
- [52] Villanueva J., Yoo S., Hanssens D. M. The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth[J]. *Journal of Marketing Research*, 2008, 45(1):48-59.
- [53] Wu T., Wen S., Liu S., Zhang J., Xiang Y., Alrubaian M., Hassan M. M. Detecting spamming activities in twitter based on deep - learning technique[J]. *Concurrency & Computation Practice & Experience*, 2017, 29(19):e4209.
- [54] Xu G., Zhao J., Huang D. An improved social spammer detection based on tri-training[C]. *In Proceedings of 2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, December 5-8, 2016:4040-4042.
- [55] Yu D., Chen N., Jiang F., Fu B., Qin A. Constrained NMF-based semi-supervised learning for social media spammer detection[J]. *Knowledge-Based Systems*, 2017, 125:64-73.
- [56] Zafarani R., Abbasi M. A., Liu H. Social media mining: an introduction[M]. Cambridge, UK: Cambridge University Press, 2014:232-233.
- [57] Zhang X., Bai H., Liang W. A social spam detection framework via semi-supervised

- learning[C]. *In Proceedings of Trends and Applications in Knowledge Discovery and Data Mining, PAKDD 2016*, Auckland, New Zealand, April 19, 2016: 214-226.
- [58] Zhou D., Han W., Wang Y. A fine-grained information diffusion model based on node attributes and content features[J]. *Journal of Computer Research & Development*, 2015, 52(1):156-166.
- [59] Zhou Z. H., Li M. Tri-training: exploiting unlabeled data using three classifiers[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2005, 17(11):1529-1541.

## 作者在读期间发表的学术论文

- [1] 李岩, 邓胜春, 林剑. 社交网络水军用户的动态行为分析及在线检测[J]. 计算机工程. (Online)

## 致谢

2.5 年学制的研究生生涯即将结束，在毕业论文撰写完成之际，我首先想感谢我的导师邓胜春导师。导师为人谦和、平易近人、笑容可掬，平日里导师的工作十分繁忙，但仍不忘为我的论文给予及时的指导。不仅需要感谢导师对我论文的指导，也需要感谢他对我为人处世、做人做事的指导，这些都指引着我今后的人生道路。在此，向邓胜春导师表示崇高的敬意和衷心的祝福。

此外，感谢深深关心、为我默默付出的家人及同学。家人为我解决了后顾之忧，同学之间相互互助、共同进步。尤其感谢吴昕扬同学，读研路上一路相伴，一起成长。

最后，对在百忙之中评审我硕士学位论文的各位专家和学者表示衷心的感谢。



**Zhejiang University of Finance and Economics**

**浙江财经大学**

---

地 址：中国杭州市下沙高教园区学源街 18 号

邮 编：310018

电 话：0571-86754515

网 址：<http://www.yjs.zufe.edu.cn>