深圳大学硕士

数据仓库与数据挖掘

|  |  |  |
| --- | --- | --- |
| 姓名、学号 | ： | 王浩2070276004 |
| 专业名称 | ： | 电子信息 |
| 调研方向 | ： | 用户行为分析调研 |
| 任课教师姓名 | ： | 陈小军 |

# 1 用户行为分析介绍

从1953年行为分析的概念被提出到现在为止，研究领域已经从刚开始的心理研究逐渐发展为使用机器学习等方法进行研究。用户行为分析已经被证明了在一些领域的重要性。例如推荐产品的潜在客户，分析用户的偏好信息，异常用户的识别等例子都有用户行为分析的身影。

到目前为止，用户行为分析领域已经有许多人进行了重要的研究。这些研究主要是集中于研究如何对用户过去、现在以及未来的行为进行理解、建模和预测。目前用户行为分析已经涉及多个领域。例如有关于社交网络、网络安全、安全健康以及服务改善等领域都被涉及。本文的调查内容主要针对社交网络中的异常行为以及服务改善两方面进行调查。

对于研究数据来说，在用户分析领域涉及到的数据往往是结构化数据和非结构化数据(语音图像)以及半结构化数据(数据之间的关联)。这些数据在获取程度上又分为了可显式获取到的数据和隐式反馈数据，针对两种形式的数据，目前学术界研究也取得了客观的成果。数据可以从点击流数据以及常见日志数据等多方面进行获取计算。针对于研究方法，在该领域中目前有多种研究方法，其中以机器学习算法模型为主，但不仅限于机器学习算法。另外还有基于基本统计指标，关联规则、协同过滤、神经网络等一系列研究方法。本文首先简单介绍一下针对异常行为的分析，然后再分别针面向营销服务和面向web服务进行调查。本文调查内容通过相关文献获取。

# 2 调查方式与文章结构

本文基于文献与相关博客展开调查。调查的具体步骤为：首先将调查方向分为社交网络与服务改善，然后逐个入手；在社交网络领域中主要对异常行为分析和用户识别分类展开调查，涉及到的主要模型是基于点击流和基于深度学习的分析方法；在服务质量改善领域，主要对Web导航服务以及用户推荐进行调查。通过对具体方向的调查来了解基于统计指标，协同过滤、关联规则，点击流分析等方式进行分析的研究方法。

在第1节主要介绍了用户行为分析的背景及现状；本节主要说明了具体的调查方式与领域。而在本文剩余部分的组织如下所示：第3节主要介绍一下用户行为分析用于社交网络和服务交付改善的相关模型和方法。最后，在第4节从调查结果中得出一些总结以及对未来研究的一些想法。

# 3 相关模型与方法

在这一小节，主要介绍目前通过结合文献所调查到的一些模型方法。主要涉及社交网络中的异常行为检测、异常用户分类和服务改善领域中的推荐系统等方面。这些方法主要分为基于统计指标分析、基于点击流的机器学习算法、基于神经网络的研究方法等。

## **3.1 面向社交网络的异常行为检测**

社交网络是人们用来分享、交流的平台或工具。在社交网络上，理解用户行为是很重要的，也是有挑战性的。通过研究分析用户之间的联系或者对用户进行更细粒度的分类能够更好的区分某一用户的意外行为和具有异常行为的用户。（社交网络中代表用户行为的计算模型：向量表示、用层次概念表示、本体表示、用图形表示、点击流表示。这里主要调查了基于点击流的计算模型）这里调研了两种方式来分析用户的异常行为以及对用户群体进行分类。这两种方式分别为基于用户点击流的分析和基于深度学习的用户异常行为分析。具体介绍如下所示。

### 3.1.1 基于点击流的分析模型(点击事件、基于时间、两者混合模型)

点击流模型主要分为：1、点击序列模型；点击序列模型是将每个用户的点击流当作一系列点击事件，然后按照到达顺序进行排序；然后通过算法进行用户行为的分析。2、基于时间的模型；基于时间的模型是关注事件之间的间隔分布，每个用户的点击流由间隔时间列表表示[t1，t2，t3，...，tn]，其中n是用户点击流中的点击次数。3、两者混合模型；混合模型结合了点击类型和点击间隔时间。每个用户的点击流被表示为有序的点击序列以及点击之间的间隔【3】。图3.1中说明了混合模式下将两个样本点击流离散化为事件序列的详情。其中A,B,C,D表示事件类型t表示第i个点击事件和第i+1个点击事件之间的间隔时间。这三种通过比较点击流的相似度来对用户进行分类。

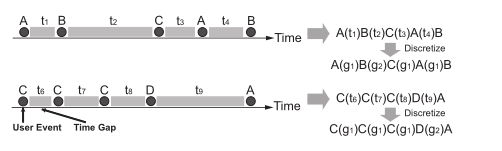


图3.1 混合模式点击流模型

通过点击流方式来对用户进行分析是指从点击流数据中捕获用户行为，并且容易将检测到的行为可视化出来。采用点击流分析的方式时，可以使用多种算法来对用户进行分类并预测用户的行为。例如：在Deepak Pai等人提出了采用了有监督学习算法来对用户进行分类【4】；虽然有监督学习方法表现得并不差，但是这类模型是需要大量的真实数据进行训练的，一旦模型被训练，他们就只能识别训练数据中所出现的异常行为，而不能识别新的异常。而在Gang Wang等人提出了使用无监督聚类方法来实现识别意外的用户行为，预测用户未来行为【1】。这里以后者进行说明。

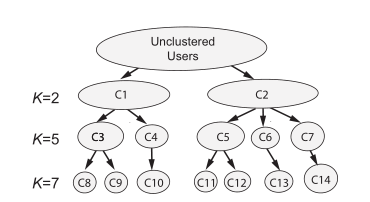


图3.2 用户分类树

后者通过利用点击流之间的相似性来构建相似图，以此来按照不同的粒度来对用户进行分类并识别异常用户。将各类用户集群利用树状结构表现出来。如图所示。然后利用迭代特征剪枝算法通过卡方统计移除父聚类的主要特征，按照剩余的次要特征进一步划分父聚类，直到聚类质量达到最小阈值算法停止。最后通过构建可视化工具，来查看具体分类用户的行为及特征。

### 3.1.2 基于深度学习异常行为检测方法

随着数据的增加，近年来基于深度学习的异常检测算法越来越流行，并已经应用于各种任务。基于深度学习的异常检测方法往往会优于大部分传统的机器学习方法。并且基于深度学习对社交网络中异常进行检测的方法在近几年也有较多的研究。

目前基于深度学习的异常检测方法主要包括三种：1、基于社交网络的图形分析的结构化方法；2、基于用户活动的提取和分析的行为方法；3、以及结合上述两种方法的混合方法。【7】其中第一类侧重于处理用户在社交网络上活动内容的模型。它主要考虑到用户之间是相互独立的。此外，由于用户行为的演变依赖于用户本身的频繁出现，这可能是这种方法的一个严重限制。所以，这里检测异常行为是基于用户本身来看的；第二类强调对社交网络图的属性进行分析，社交网络图对不同用户之间的关系进行建模。该方法主要考虑用户之间的交互性，每个用户都可以与其他用户通过提及、分享等建立联系。这种方法可以通过静态图或者动态图来完成对关系的分析。可以利用该方法识别异常用户；第三类分析社交网络图结构中的用户活动。该方法主要对用户本身的网络结构和用户的活动关系进行分析。

其中在基于神经网络研究方法的调查中有发现到，王晓东等人提出一种基于图神经网络（GNN）和循环神经网络（RNN）的深度学习模型来预测用户的异常行为。通过将GNN和RNN相结合，利用GNN捕捉用户局部行为之间潜在的逻辑关系，同时使用RNN来捕捉事件序列上的信息来构建异常检测分析模型；【9】Zhang和Ding等人利用深度学习自动提取行为模式特征的优势并尝试应用于用户异常行为检测；【8，10】Yun Pan等人结合网络流量分析技术，提出了一种基于深度神经网络的用户异常行为检测模型。实现了对网络流量的细粒度分析使用户异常检测能够满足特定网络环境的要求。【6】

## 3.2 面向服务改善的模型与分析方法

在这个领域的重点主要是为了去提升服务本身的质量。在面向服务改善的领域中，用户行为分析可以帮助理解并服务用户，以便于提供更准确的分析和预测。例如，在电子商务的环境下，基于用户偏好进行开发推荐系统可以提高其销量。并且如果根据用户的共同特征来对用户进行划分，这样有助于确定用户的准确需求并实现个性化定制。在调查中主要涉及面向营销服务和web导航服务两方面。

### 3.2.1 面向营销服务

营销服务是为了使产品或服务对消费者具有吸引力，来提高产品的销售额。在面向营销服务领域的方法的主要目标是利用用户行为分析获得的用户行为模式来实现用户个性化定制和个性化营销。面向营销服务主要倾向于推荐系统和消费者划分，来对特定用户推荐产品以及对特定的一类用户发起营销服务和个性化定制。

对于推荐系统来说，它的最终目的是为了帮助消费者找到他们可能感兴趣的内容或者产品。通常利用用户在电商中对产品的购买行为和评价行为的历史信息来进行开发。使用推荐系统往往会对用户的购买决策产生较大影响。为了完成对用户的分析，了解用户的内容偏好和评级模式是很重要的。这个可以基于网络日志来挖掘用户偏好来实现。推荐系统在设计中往往会从基于内容，协同过滤以及关联分析等方式进行较为广泛的使用。其中协同过滤算法作为推荐系统中的核心技术一般可以分为基于用户的协同过滤算法和基于物品的协同过滤算法。协同过滤最常用的就是矩阵因子分解技术，它将用户和物品嵌入到一个共享的低维空间内构造一个二维矩阵，其中每一行代表一个用户，每一列代表一个物品。通过将矩阵分解为两个低秩稠密的矩阵P和Q，其中P为用户的隐因子矩阵，Q为物品的隐因子矩阵，通过这两个矩阵来预测用户对物品的评分。

在学术界这些技术有着不同的变体和改进方法。例如由Laishram A等人提出的利用遗传算法来改进协同过滤【11】，改进了协同过滤中的有效预测评分。由S Feng等人提出的将用户社交网络融合到推荐推荐商品的模型中来加强模型可以更好执行群体推荐【12】。推荐系统不仅可以用在电商中的产品推荐，还可以在实体商家中来对产品位置进行预测。例如：由W Liang等人提出的基于室内客户的跟踪数据来检测用户到达的热点区域，来找到最佳的广告牌位置【13】。H Shi等人利用聚类算法和因子分解分析用户移动信息，来完成对用户推荐位置的任务【14】。对于消费者划分来说，在调查过程中发现，在消费者划分中通常使用点击流日志来分析用户在购买前后的行为，通过用户对添加到购物车到删除购物车产品等一系列行为的分析，可以对消费者在一定程度上进行划分，并有助于针对特定用户展开营销服务的各类活动。

### 3.2.2 面向Web导航服务

在这一领域，通常要考虑用户在网络中执行的不同类别的操作。例如用户在浏览器中的搜索以及用户在网络中执行的一系列点击事件等与互联网进行交互的操作。通过对浏览器、搜索引擎等web服务进行用户行为分析可以获取用户对网络推荐或者个性化配置兴趣的倾向，以便更好地改进它们的服务。在该领域进行用户行为分析大多是基于用户在web上的浏览信息数据或点击流数据来进行分析。

通过对用户在web上的浏览信息可以进一步转换为用户点击网站的顺序，通过进一步分析就可以来预测用户下一次将访问的网站、来提取用户的兴趣倾向、进行网页推荐、识别用户以及其他个性化网络应用等。在面向web导航服务的领域中，已经有了很多优秀的成果。在该领域的探索中，通过基于机器学习或深度学习的方法已经在多方面产生成果。例如：在用户识别方面，Haidar R等人基于网站中用户实际的导航数据针对特定的网站构建分类器【16】。分类器使用增强决策树分类模型会随着机器人的发展定期进行重新训练来学习新的假设。除了特定针对于某一个网站的机器人，它通常可以检测出来任何类型的机器人。Y Yang提出通过用户的网页浏览行为建立用户档案来捕获用户行为模式的强度以便更好地识别用户【17】。在搜索引擎中进行搜索实体推荐方面来说，主要分为实体发现和实体排序两方面。实体发现主要有：基于知识图谱发现实体的方法、基于搜索日志的发现方法以及基于网页文档的发现方法。实体排序主要有：基于查询的排序方法、基于查询和用户偏好的排序方法以及基于查询和上下文信息的排序方法。黄际洲在面向搜索引擎的实体推荐方面从基于排序学习与信息新颖性增强方法、基于深度多任务学习的上下文相关实体推荐等方面进行了研究分析【18】。

# 4 结论

在经过对面向社交网络的用户异常行为检测和面向服务改善用户行为分析的调查之后，发现在对用户行为进行分析时，往往离不开机器学习或深度学习技术。在此基础之上提出针对各种领域的研究模型和方法。基于学习的用户异常行为分析研究在建模算法和行为模式提取等方面都已经做了很多工作。但是就建模而言并没有达到非常好的效果，仍需要在原有研究的基础之上提出更成熟更有效的解决方法。基于点击流数据、日志数据和浏览信息等数据和对用户一些行为的统计数据的分析已经在推荐和识别用户等方面有所成果，不仅有基于电商，网络的推荐系统也有基于浏览器信息的web导航服务。但是目前基于搜索引擎的用户数据具有不完整性，因此需要在此基础之上提出更准确的方法。目前大多数是基于网络中的显式数据来对用户行为进行分析，而在基于隐式反馈信息方面的研究并不太多，基于隐式数据来对用户进行分析可提高分析的准确性及价值。

# 5 参考文献

[1] Martín A G, Fernández-Isabel A, de Diego I M, et al. A survey for user behavior analysis based on machine learning techniques: current models and applications[J]. Applied Intelligence, 2021: 1-27.

[2] Wang G, Zhang X, Tang S, et al. Unsupervised clickstream clustering for user behavior analysis[C]//Proceedings of the 2016 CHI conference on human factors in computing systems. 2016: 225-236.

[3] Wang G, Zhang X, Tang S, et al. Clickstream user behavior models[J]. ACM Transactions on the Web (TWEB), 2017, 11(4): 1-37.

[4] Pai D, Sharang A, Yadagiri M M, et al. Modelling visit similarity using click-stream data: A supervised approach[C]//International Conference on Web Information Systems Engineering. Springer, Cham, 2014: 135-145.

[5] Chalapathy R, Chawla S. Deep learning for anomaly detection: A survey[J]. arXiv preprint arXiv:1901.03407, 2019.

[6] Pan Y. Network security and user abnormal behavior detection by using deep neural network[J]. Internet Technology Letters, 2021, 4(3): e260.

[7] Chaabene N E H B, Bouzeghoub A, Guetari R, et al. Deep learning methods for anomalies detection in social networks using multidimensional networks and multimodal data: a survey[J]. Multimedia Systems, 2021: 1-11.

[8] 王晓东,赵一宁,肖海力,王小宁,迟学斌.使用GNN与RNN实现用户行为分析[J].计算机科学与探索,2021,15(05):838-847.

[9] Zhang J, Ou J, Ding C, et al. An abnormal behavior detection based on deep learning[C]//2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2018: 61-65.

[10] 丁珊. 基于深度学习的入侵检测关键技术研究[D].北京交通大学,2018.

[11] Laishram A , Padmanabhan V . Discovery of user-item subgroups via genetic algorithm for effective prediction of ratings in collaborative filtering[J]. Applied Intelligence, 2019, 49(11):3990-4006.

[12] Feng S , Zhang H , Cao J , et al. Merging user social network into the random walk model for better group recommendation[J]. Applied Intelligence, 2018.

[13] Liang W , Wu Z , Jie C , et al. Understanding Customer Behavior in Shopping Mall from Indoor Tracking Data[C]// 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2018.

[14] Shi H , Chen L , Xu Z , et al. Personalized location recommendation using mobile phone usage information[J]. Applied Intelligence, 2019, 49(2).

[15] 刘华真,王巍,谷壬倩,张屹晗,郝亚奇.基于用户浏览行为的个性化推荐研究综述[J/OL].计算机应用研究:1-12[2021-05-25].https://doi.org/10.19734/j.issn.1001-3695.2020.10.0347.

[16] Haidar R , Elbassuoni S . Website Navigation Behavior Analysis for Bot Detection[C]// 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2017.

[17] Yang Y . Web user behavioral profiling for user identification[J]. Decision Support Systems, 2010, 49(3):261-271.

[18] 黄际洲. 搜索引擎中的实体推荐关键技术研究[D].哈尔滨工业大学,2019.