

# 基于用户异常行为数据的个性化推荐模型研究

章群英, 房书豪, 章嘉愉, 曾莹, 王申奥

(嘉兴学院, 浙江 嘉兴 314001)

**摘要** 作为缓解信息过载、信息迷航等问题的一种重要方式, 个性化推荐系统可以帮助用户从海量信息中快速找到有价值的信息, 但现有的推荐系统却普遍存在推荐不精准的现象。本文首先讨论了个性化推荐系统推荐不精准的原因, 并引出用户异常行为的发生是系统难以精准推荐用户兴趣的结论, 而后基于此结论, 详细讨论了用户异常行为的定义与特征, 并提出了两种可以用于系统数据筛选的用户异常行为的价值判断方法。其次, 基于此方法构建了一个基于用户异常行为数据的个性化推荐模型, 以期可以对优化个性化推荐系统的推荐服务有所裨益。

**关键词** 个性化推荐; 用户异常行为; 兴趣动态变化

**中图分类号**: TP3

**文献标识码**: A

**文章编号**: 1007-0745(2023)03-0010-04

数字经济时代, 信息的冗余和过载现象日益凸显, 精准定位用户喜好并进行个性化推荐的运营方式成为大多数C端企业的商业共识, 可以有效地对信息进行过滤和筛选, 帮助用户以个性化的方式来检索符合其需求的信息资源, 缓解信息过载的个性化推荐系统<sup>[1]</sup>, 也成为当前的研究热点。

个性化推荐系统是根据用户对信息的反馈互动行为来判断用户和信息之间的匹配程度, 从而向用户进行信息推荐的一类数据分析系统, 它通过获取用户的历史行为数据, 如网页的浏览数据、购买记录、社交网络信息、用户地理位置等, 来推断用户偏好, 并据此进行推荐<sup>[2]</sup>。由此可知, 个性化推荐系统往往与用户本身的兴趣强相关, 且兴趣可以通过特定的行为来表现。这意味着用户网络行为产生的数据是推荐系统的重要数据源, 换言之, 推荐系统的推荐不精确问题, 常常是因为无意义、偶发性且不能代表用户偏好的数据所导致。

基于上述原因, 本文通过分析偶发性用户数据的产生原理, 认为此类数据的产生原因是用户在体验系统服务的过程中发生的自身兴趣的动态变化。并利用此结论, 以用户的异常行为数据为基础, 架构了一个可以精确筛选异常数据、辅助系统优化的理论模型。该模型可以在一定程度上提高传统个性化推荐系统的推荐准确程度, 提升用户在使用相关服务时的体验感。

## 1 模型原理概述

偶发性用户数据也称之为异常行为数据, 是系统所捕捉到的用户在使用系统时发生的用户异常行为。

这种行为往往不符合用户在使用系统时的通常习惯, 可以作为用户兴趣动态变化的标识, 并由此进一步优化系统对用户的精确推荐。本节将主要阐述兴趣动态变化与用户异常行为之间的内在联系, 讨论哪些行为的捕捉对于系统分析用户兴趣更具有价值, 并尝试总结其行为规律。

### 1.1 用户异常行为

个性化推荐系统的实质是通过理解用户的兴趣和偏好帮助用户过滤大量无效信息并获取感兴趣的信息或者物品的信息过滤系统<sup>[3]</sup>, 但由于影响兴趣变化的因素众多, 且兴趣动态变化规律相对复杂, 导致用户异常行为也相应地出现难预知性、突发性和无周期性等特征。合理分析用户异常行为的特征, 是模型筛选有价值信息的基础。

#### 1.1.1 定义与特征

用户异常行为是用户个体做出的不符合自身网络社交活动习惯的特殊行为, 此类行为往往不符合用户原有的喜好偏向, 并表现为与原有喜好有较大差异甚至对立。譬如用户点赞以往不喜欢的视频类型, 这一现象就会被视为异常现象。

用户异常行为具有一些重要的特征: 在内容上, 表现为较强的难预知性。由于用户异常行为的定义是用户对“反常规偏好”做出的“反常规反馈”的反馈, 因此在内容上会出现与原有偏好的较大差异, 内容跳跃性强且无法被系统为用户已建立的规律所感知。在时间上, 表现为突发性、短时性、低频性和无周期性。

★基金项目: 2021-2022年浙江省大学生科技创新活动计划(新苗人才计划)项目“基于数据分析的个体信息需求匹配研究”(编号: 2022R417A015)研究成果。

用户异常行为的发生通常不会持续太长时间,既不会持续发生,也不会定期发生。

### 1.1.2 价值判断推论

用户异常行为可以具体细分为有价值和无价值两种,异常行为的价值判断取决于该行为数据对用户自身的喜好分析是否具有实际意义。举例而言,某用户在短期内重复性浏览具有相同内容标签但并不属于该用户传统兴趣偏好内容的信息,这有可能是因为该用户的兴趣偏好的确出现变化,也有可能是因为用户错误操作(如忘记关闭 APP、系统卡顿造成误触等)所导致。前者对于系统分析用户兴趣有优化作用,故可称为有价值的用户异常行为,后者则称为无价值的用户异常行为。

推荐系统为用户所推荐的内容之所以时常出现某种偏差,核心就是因为对异常行为的错误的价值衡量,从而导致在筛选数据时忽略了部分有价值行为数据或提取了无价值行为数据。

## 1.2 用户异常行为的价值判断法

有价值的用户异常行为有助于系统对用户的内容偏好做出更精确的判断,因而在系统筛选行为数据时,应当具有一个可靠稳定的筛选机制来获取这些有价值的用户异常行为数据并加以处理。对于价值判断而言,其实质是判断用户的用户异常行为是否满足兴趣动态变化规律。兴趣动态变化规律分为兴趣演化与兴趣漂移两种,相应的,用户异常行为的价值判断同样具有两种:用户精力变化判断和用户经验累积判断。

### 1.2.1 用户精力变化判断

用户精力变化判断是基于兴趣漂移规律的异常行为为价值判断,其目的在于筛选出关于兴趣漂移的异常行为。兴趣漂移规律被定义为用户随时间的改变而不断发生兴趣的动态变化,其主要方式是通过计算用户随时间的推移在其本身已有的不同偏好中分配的精力来推测用户的主要兴趣。基于此,用户精力变化判断的基本假设是:若用户在其发生用户异常行为前存在精力分配的变化趋势,比如逐渐增加了导致用户发生异常行为的偏好内容的关注,则可以判定用户的异常行为是基于兴趣动态变化现象所导致,故而可以判定其异常行为对于系统优化有利。

根据兴趣漂移规律,这类行为的判定一般基于时间的长短。用户分配在新偏好上的时间越多,则用户的主要偏好朝着新偏好发展的可能性就越大。因此,对于精力判断而言,时间序列分析方法将是必要的,并需要考虑到用户对新内容的遗忘程度。

### 1.2.2 用户经验累积判断

经验判断则是与兴趣演化相关。对于兴趣演化规律,目前学界的基本假设是:用户的兴趣演化是伴随一个可累积的因素的变化而产生的<sup>[4]</sup>。我们将这个可累积的因素总结为“经验”。举例而言,就是摄影爱好者随着摄影经验的的增长而减少对非专业相机的偏好,转而追求专业或专家级相机的情况。由此可以看出,精力变化判断与经验累积判断的不同在于,前者考虑不同偏好的转移,而后者考虑单个偏好内的深化。

这一判断的核心在于用户是否在发生异常行为前存在单一领域内经验的增长进而达到某种兴趣的质变,比如由于用户接触的知识水平的提高,用户发生了对其偏好中更高层次的事物的喜好。同精力判断一致,若确有此行为,则判定对系统优化有利。依从兴趣演化理论的假设来看,这类行为的判定可以基于某种可累积因素的变化而产生。从兴趣演化的现象来看,其本质应该是由用户的学习能力所影响。

## 2 基于用户异常行为数据的个性化推荐模型

### 2.1 模型结构

模型分为数据筛选模块和个性化推荐模块两部分。其中,模型将异常行为检测机制与兴趣动态变化量化结合,从而提高数据筛选的准确度。模型既考虑到了兴趣的动态变化也考虑到了时间效应对于用户兴趣的影响,提高了模型的预测精度。

在推荐系统中考虑一系列因子,包括项目信息、用户信息和用户行为信息。项目信息包括项目的说明、标签等表征项目特征的信息;用户信息包括用户性别、语言偏好等;用户行为信息包括用户点赞、收藏、分享、评论等行为信息。

### 2.2 数据筛选模块架构

数据筛选细分为用户异常行为检测和兴趣动态变化的判断。异常行为检测模块,利用项目信息、用户信息和用户行为信息的集成数据即预数据作为该模块的输入,最后输出模型数据作为个性化推荐系统的输入数据。

本文在模型训练阶段通过引入目标项目来训练异常行为检测模块以做出自主判断的能力。通过将用户的瞬时兴趣  $k_i$  与目标项目  $t_i$  串联输入到一个多层感知器 MLP 中,多层感知器 MLP 据此建模并输出预测向量  $p_i$ ,若向量  $p_i$  趋向于 0,则用户的瞬时兴趣受到污染,即为异常行为数据,反之为正常行为数据。然后将异常行为数据作为兴趣动态变化判断模块的输入数据。

表1 模型模块功能简述

系统相关模块名称	模块功能
异常行为检测模块	检测异常行为并将异常行为数据从用户行为数据中抽离出来。
数据筛选	对抽离出的异常行为分别进行经验和精力判断，通过对比分析，筛选有价值的的数据；将正常数据和有价值的的数据进行集成，得到模型数据。
自然语言处理模块 NLP	对评论内容进行正负向判断，以便于获得更加准确的经验判断。
个性化推荐模块	基于时间效应和用户兴趣变化的模型
	提高预测精度；减少新用户或用户的新兴趣造成的冷启动问题的影响。

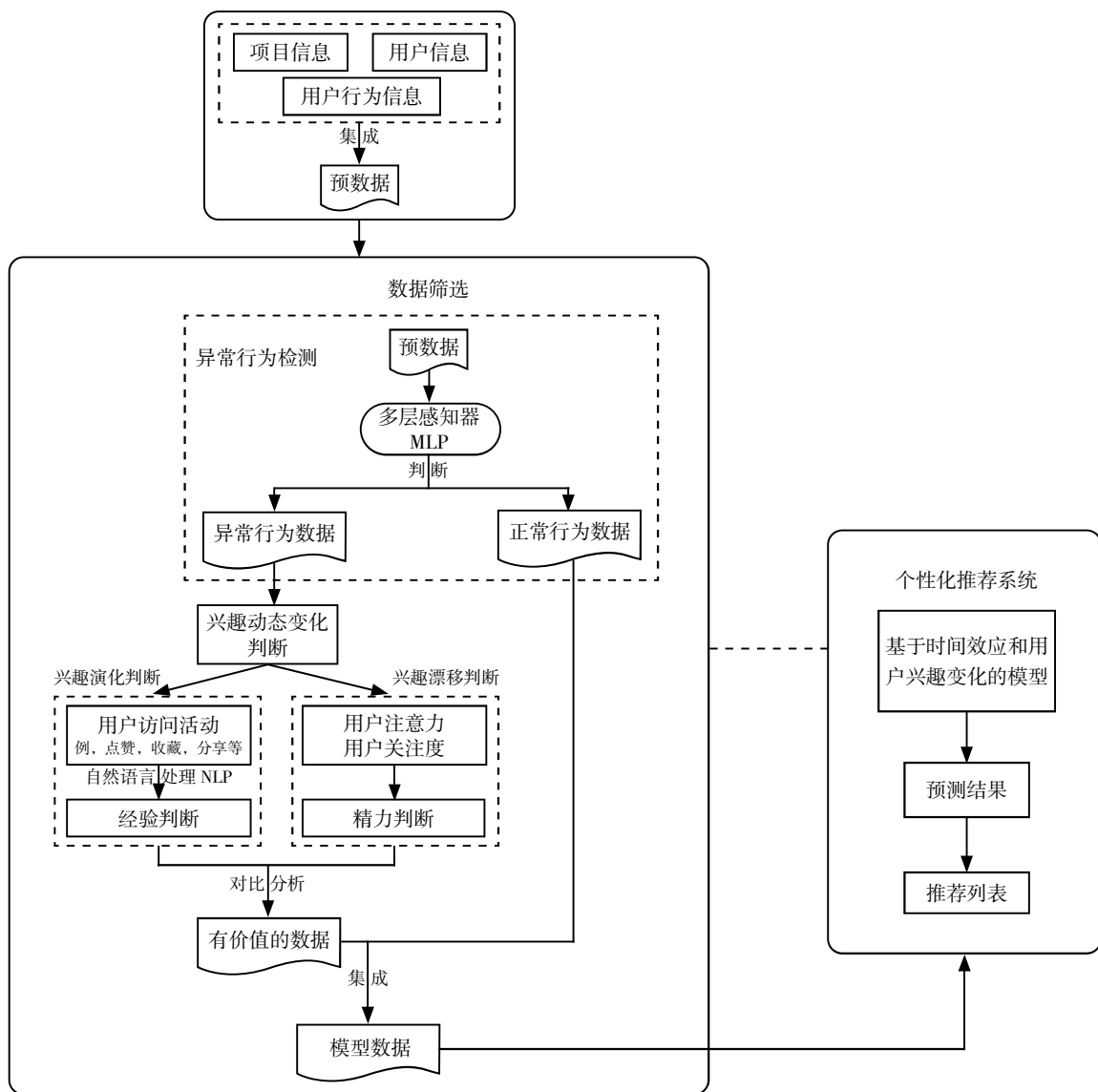


图1 基于用户异常行为的个性化推荐模型

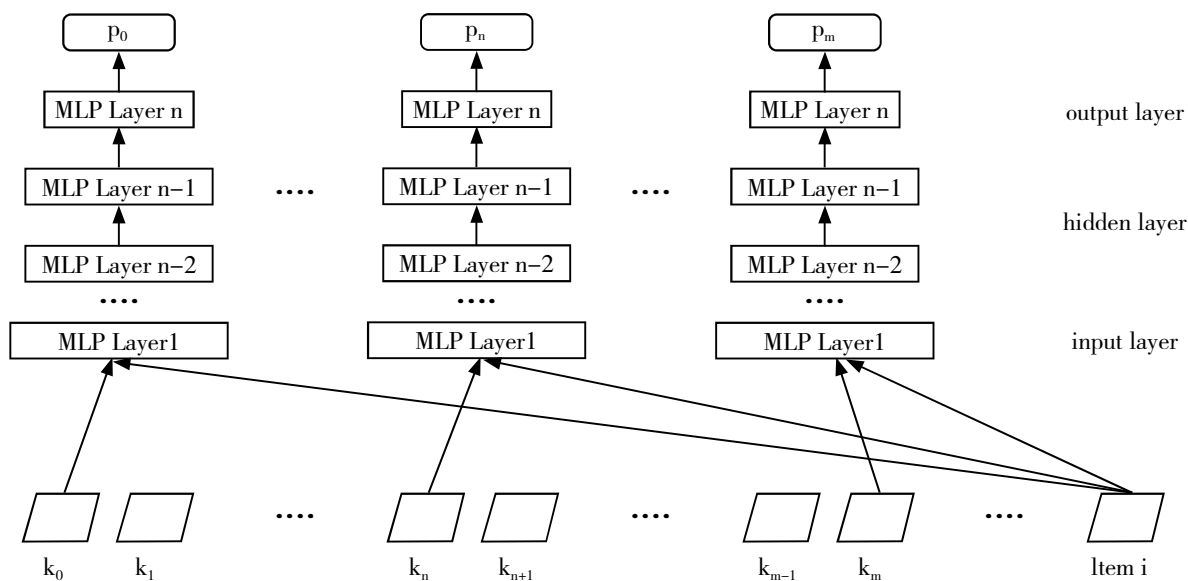


图 2 用户异常行为监测示意图

在兴趣动态变化判断中，分别对用户具体的学习状态数据、具体的历史数据以及访问频率进行经验和精力判断。因用户的评论行为可在多场景触发，若不对用户评论的内容进行控制，则无法精准地进行经验判断。因此，本文采用自然语言处理方法（NLP）的意见挖掘进行情感判断，利用训练好的情感分类器对用户评论的内容进行判断。通过选取情感词作为特征词并统计各特征词的词频，再利用情感分类器对文本进行积极（正向）和消极（负向）的二分类。例如，“华为手机非常好”表达的是正向情感，“华为手机使用起来很卡顿，不好用！”表达的是负向情感。

通过对比分析用户的经验和精力，从而区分有价值数据和无价值数据。若用户在该领域上既无经验也无精力，则为无价值数据，反之则为有价值数据且需召回到正常行为的数据中。

### 2.3 个性化推荐模块

通过研究用户的点赞、分享等访问行为实现用户兴趣点的检测和推荐，点赞、分享等行为具有序列性，这对预测用户的行为和兴趣具有重要意义。但基于用户兴趣的特征，使得推荐系统要满足动态和适应性强的特点，且能考虑项目的时效性和用户兴趣的动态变化，从而做出相应的响应。

因此，在个性化推荐模块采用基于时间效应和用户兴趣变化的模型，该模型分为信息时效性过滤和预测推荐两部分。在信息时效性过滤部分，将艾宾浩斯遗忘曲线与传统的协同过滤算法结合，对相似用户集

进行时效性划分以避免冗余用户和项目对推荐结果的影响。在预测推荐部分，将通过项目时效性及用户的动态兴趣变化计算得到的概率矩阵和随机游走算法结合，从而实现个性化推荐系统的优化。

### 3 结语

用户网络行为作为最能反映用户兴趣偏好的数据，一直是推荐系统进行推荐的主要依据。本文阐述了用户异常行为带来的数据偏差是推荐系统推荐不精准的原因，并对兴趣动态变化与用户异常行为进行了概念梳理，提出了异常行为检测与价值判断串联的方法。基于此，本文通过整合兴趣的动态变化与时间效应对于用户兴趣的影响，提出了一种基于兴趣动态变化的个性化推荐模型。该模型在理论上可以在一定程度上优化个性化推荐系统的数据筛选能力，进而提升系统对于用户的精准推荐能力。

### 参考文献：

- [1] 于蒙, 蔡利平, 周绪川, 等. 融合注意力机制的深度学习推荐模型 [J]. 西南民族大学学报 (自然科学版), 2022, 48(05): 550-560.
- [2] 刘君良, 李晓光, 个性化推荐系统技术进展 [J]. 计算机科学, 2020, 47(07): 47-55.
- [3] 赵梦媛, 黄晓雯, 桑基韬, 等. 对话推荐算法研究综述 [J]. 软件学报, 2022, 33(12): 4616-4643.
- [4] 冯浩源. 动态用户兴趣模型构建及推荐算法研究 [D]. 天津: 天津大学, 2017.