

# 在线健康社区活跃用户识别 及其交互类型分析\*

董伟<sup>1</sup> 李建红<sup>2</sup> 陶金虎<sup>1</sup>

(1. 天津大学教育学院, 天津 300350;

2. 天津大学管理与经济学部, 天津 300072)

**摘要:** [目的/意义] 在线健康社区作为用户进行医疗健康相关信息共享和问题解决的平台, 其作用日渐显著, 研究在线健康社区中用户交互行为的特征, 对进一步发现在线社区用户交互规律和相关平台建设工作有一定的参考价值。[方法/过程] 利用数据挖掘相关工具获取在线健康社区用户交互行为信息, 其中以百度贴吧鼻炎社区为例, 并从用户自身和交互两个维度的相关指标出发, 对活跃用户进行识别和分析。在此基础上, 利用内容分析法对活跃用户评论内容进行编码, 进一步划分用户交互行为类型。[结果/结论] 研究发现, 用户交互行为可细分为预防、求助、症状、治疗、反馈、提醒、情感、离题 8 种类型。治疗类、求助类交互行为是百度鼻炎吧中主要的交互行为类型。鉴于以上研究结果, 在线健康社区在今后的管理与运营中应该结合不同类型用户交互行为的特点, 关注不同疾病用户的差异化健康信息需求, 以适当方式为用户提供精准的、优质的信息服务。

**关键词:** 在线健康社区 活跃用户 交互行为类型

**分类号:** G350

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2020.01.09

## 0 引言

截至 2019 年 6 月, 我国网民规模达 8.54 亿, 较 2018 年底增长 2598 万, 互联网普及率达 61.2%, 较 2018 年底提升 1.6 个百分点; 我国手机网民规模达 8.47 亿, 较 2018 年底增长 2984 万<sup>[1]</sup>。有调查表明, 在美国使用互联网的成年人中, 有 80% 的网民通过互联网获取健康信息, 34% 的用户会阅读他人发布在网站或博客上的治疗经历, 18% 的网民会在网上寻找与自己有相似病症的人, 16%

\* 本文系国家社会科学基金青年项目“在线健康社区用户交互行为及其对用户健康效用影响研究”(项目编号: 16CTQ029)的研究成果之一。

[作者简介] 董伟 (ORCID: 0000-0001-7820-292X), 男, 副教授, 博士, 研究方向为数据挖掘与分析、在线社区用户行为研究, Email: dongwei83@tju.edu.cn; 李建红 (ORCID: 0000-0002-0053-1771, 通讯作者), 女, 硕士研究生, 研究方向为在线社区用户行为, Email: ljh519@tju.edu.cn; 陶金虎 (ORCID: 0000-0002-5456-3955), 男, 硕士研究生, 研究方向为数据挖掘、教育技术学, Email: 1254331347@qq.com。

的用户会上网查找对医生的评论或者评分<sup>[2]</sup>。在线社区是基于互联网技术由一群具有共同兴趣爱好或目的的用户组成,通过信息生产及消费来完成在线协作、知识分享及在线交易等活动<sup>[3]</sup>。在线社区有诸多类型:交易型社区、关系型社区、知识型社区和兴趣型社区<sup>[4]</sup>。在线健康社区作为知识型社区的代表,借助于网络技术的应用,可以为不同主题的健康信息或知识的交流与协作提供服务平台<sup>[5]</sup>,也为患者用户获取相关健康知识提供了便利。一个在线社区的发展不能仅仅依靠网络技术的发展,更需要用户的积极参与,活跃用户作为在线社区中最核心最有价值的用户,他们的参与行为关系着网站的生存和发展,并且在线健康社区作为一种提供相关健康信息的特殊在线平台,用户的参与行为更为主题化。鉴于此,识别在线健康社区中活跃用户,细化用户的行为类型,深入了解在线健康社区中用户需求进而提高用户的参与度对于在线健康社区的运营和发展有重要意义。

因此,本文选取典型在线健康社区作为研究对象,对其中的相关社区用户进行活跃度划分从而识别出活跃用户,并在此基础上细化在线健康社区中用户交互行为类型,识别用户需求,从而帮助我们理解和把握在线健康社区中用户的行为特征,发现在线健康社区中用户交互规律,这对于在线健康社区今后的精细化管理和运营有重要启示。

## 1 研究现状

对于一个社交网络网站而言,用户越活跃越能给网站增加价值<sup>[6]</sup>,用户活跃度的高低反映了用户对网站的忠诚度和回头率,侧面上也反映了网站的质量水平,用户活跃度关系着网站的生存和发展,并且用户越是活跃的网站,其生命特征就越明显,其生命周期就越长久<sup>[7]</sup>。在线社区活跃用户相关研究方面,其中有学者通过分析用户影响力对活跃用户进行分析,Yaguang Shi认为用户影响力在一定程度上代表了用户的活跃度,研究了如何抽取独立级联模型下的影响力结点,提出了一种基于贪婪算法的获取影响力结点的近似解的算法,从而对活跃用户进行分析<sup>[8]</sup>。Yan X等人探讨了基于用户喜好的社交网络中顶级k影响力结点集的识别问题<sup>[9]</sup>。还有学者对用户参与的活跃度指标进行了设计和分析,黄炜和刘璇等人分别通过主成分分析法和积分方法建立综合分析模型和指标,进而对用户活跃度进行分析<sup>[10][11]</sup>。林鑫等人结合标签相关性判断的特点对活跃度指数的计算方法进行改造,提出一种基于多次活跃度指数迭代的标签相关性判断策略<sup>[12]</sup>。曹磊在比较新浪微博与腾讯微博上用户的活跃度时,综合使用“收听数”“听众数”“广播数”“广播频率”这四个指标来反映用户的活跃度<sup>[13]</sup>。邓岩和霍英男从状态数、日志数、好友数以及行为频次等来判断用户的活跃度<sup>[14][6]</sup>。

在线健康社区用户的行为类型主要包括健康信息浏览行为、信息搜寻行为、信息共享行为和利用行为<sup>[15]</sup>。Huang和周军杰均基于社会支持行为分别分析了乳腺癌和前列腺癌社区和以癌症为主题的QQ群中的用户行为类型,将其分别分为需求支持、信息支持、情感支持、尊重支持和网络支持五种支持行为及知识共享、情感支持、实物支持、网络支持、尊重支持、离题六种行为<sup>[16][4]</sup>。赵栋祥从信息科学视角指出国内在线健康社区用户健康信息行为的研究主要涉及信息披露行为、信息获取与搜寻行为、信息共享行为、信息服务使用及持续使用行为、社会支持行为等行为类型<sup>[17]</sup>。以往研究多数将在线健康社区中用户的信息行为作为整体进行研究,并没有

针对在线健康社区中用户信息行为的具体内容进行细分, 用户往往是基于自己的需求参与到在线健康社区中, 例如有关疾病的预防、诊断、用药等, 将用户信息行为作为整体研究不能深入了解在线健康社区用户的需求内容。

综上所述, 现有学者对于在线健康社区用户行为研究已有了一定成果, 但是对于在线健康社区用户活跃度的划分尚未形成一致的结论, 并且已有研究鲜有涉及针对在线健康社区这一特殊平台用户交互规律的挖掘以及基于用户的评论内容对用户交互行为类型进行精细化的划分, 因此本文将从以下几个方面对在线健康社区中用户的行为进行研究: (1) 本文利用数据爬虫工具获取在线健康社区用户交互行为信息, 以百度贴吧鼻炎社区为例, 从用户自身和交互两个维度的相关指标出发, 通过对用户类型的划分识别活跃用户。(2) 选取活跃用户, 基于活跃用户评论内容采用内容分析法进行编码, 针对在线健康社区对用户交互行为进行精细化的划分, 识别在线健康社区中用户交互类型, 进而了解在线健康社区用户的需求。在此研究基础上为在线健康社区的运营和发展提供可行性建议。

## 2 研究设计

为识别在线健康社区中活跃用户以及对用户交互行为类型的细分, 本文基于百度鼻炎吧中用户的相关数据进行了以下流程设计, 结果如图 1 所示。

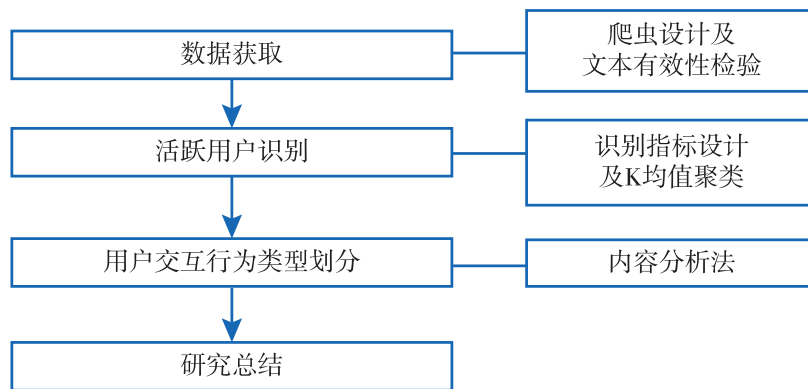


图 1 研究设计

首先是数据获取。本文数据来源为百度鼻炎吧, 通过使用 Python 语言设计爬虫工具获取用户信息, 包括用户昵称、用户评论内容、发帖量、回帖量等相关数据, 在此基础上, 通过对用户数据进行数据清洗和处理以及用户评论内容的有效性检验获取有效的用户信息。

其次是活跃用户识别。聚类分析是一种数据归约技术, 它可以把大量的观测值归约为若干个类, 使得类内观测值的相似度比类间的相似度高。K 均值聚类分析是最常见的划分方法, K-Means 聚类也称快速聚类, 从反复寻找类质心角度设计算法。本文利用 R 语言中的 `wssplot` 函数确定聚类分析中的类的最佳数目, 采用统计分析方法中的 K 均值算法, 通过 RapidMiner 工具对鼻炎贴吧中活跃用户进行识别。

最后为用户交互行为类型划分。内容分析法作为质性研究的重要工具之一，主要是通过系统性的编码、主题或模式的分类过程等途径，来实现对文本数据内容的主观解读。本文采用内容分析法对用户评论内容进行手工编码，针对用户评论内容对用户交互行为类型进行划分。

### 3 在线健康社区活跃用户识别与分析

#### 3.1 实验数据

百度鼻炎吧是一个比较典型的在线健康社区，目前有超过13万个用户关注，在鼻炎吧里，用户可以通过发帖、回帖、浏览等形式来进行交流和互动，从而获取有效的信息。因此本文以百度鼻炎吧为数据来源，以鼻炎吧中的用户为研究对象，通过使用Python语言设计爬虫工具抽取鼻炎吧中用户的吧龄、访客数、粉丝数、发帖量、回帖量、关注吧数、头衔等级、经验值、留存时间和评论内容，具体数据收集时间为2019年3月1日至2019年3月15日，共获取原始数据8691条，在数据清洗过程中，本文首先删除了网页爬取的出错数据、缺项数据和重复数据，为了保证用户评论内容的质量，本文对用户评论内容进行了审核，主要使用百度ai开放平台的文本审核功能，该系统通过训练海量相关数据构建了多种语义模型，识别率较高，能够对复杂违规文本进行识别，根据检测结果对存在恶意推广和低质灌水行为的用户评论内容进行了排除，经过筛选最终获取了2481个用户和5287条有效用户评论内容。

#### 3.2 用户活跃度指标设计

既往研究中对在线社区用户进行分类时采用的指标包括微博数、评论数、发帖量、回帖量、用户留存时间、经验值、关注的人数，可以看出之前的研究指标在一定程度上反映了用户的活跃程度，但是在进行用户活跃程度分类时采用的指标不够全面，本研究认为用户活跃度在一定程度上反映了用户对于在线社区的忠诚度，一方面体现在用户自身对于社区内容的贡献方面，另一方面体现在与其他用户的交互方面。因此本研究在之前研究的基础上将用户行为指标进行进一步的综合整理，在划分用户活跃类型时选取了基于两个维度的指标进行用户活跃度分析：

(1) 基于用户本身活跃度：发帖量、用户留存时间（用户第一条发帖和最后一条发帖的时间间隔）、经验值、头衔等级。其中，在线社区中用户的经验值与头衔等级是根据用户发帖回帖，以及签到来计算的，因此在一定程度上体现了用户的活跃程度。

(2) 基于用户对虚拟社区贡献的交互角度：关注贴吧数、关注人数、粉丝数、最近访客数、回帖量。

本文的具体指标设计如表1所示：

#### 3.3 用户活跃程度分析

在线健康社区可以视为一个包含信息、用户和社区三个要素相互影响、相互依存的复杂系统，信息是用户参与交互行为和社区运行情况的记录；用户是在线健康社区的参与者、贡献者和管理者，用户产生、传播、获取、评价和使用健康信息；社区是用户线上活动的场所和用户间的信息交流空间<sup>[17]</sup>。在线健康社区中用户之间的交流和互动对于患者和社区均至关重要，在线社区由于用户之间的互动会提供给用户有效的健康信息，患者可以从其他用户得到生理和心理的支

持和帮助。但是在线健康社区中不同用户的活跃程度差别较大, 活跃度的高低影响着用户之间的交互行为, 因此本文通过以上的数据指标对鼻炎贴吧的用户数据进行活跃用户的识别。首先利用 wssplot 函数来确定聚类分析中的类的最佳数目, 得到组内平方和随着聚类个数变化图。如图 2 所示, 从一类到三类变化时, 组内的平方和有明显的下降趋势, 三类以后, 下降的速度减弱, 表明聚成三类对于数据来说是一个很好的拟合, 所以本文中取  $k=3$ 。

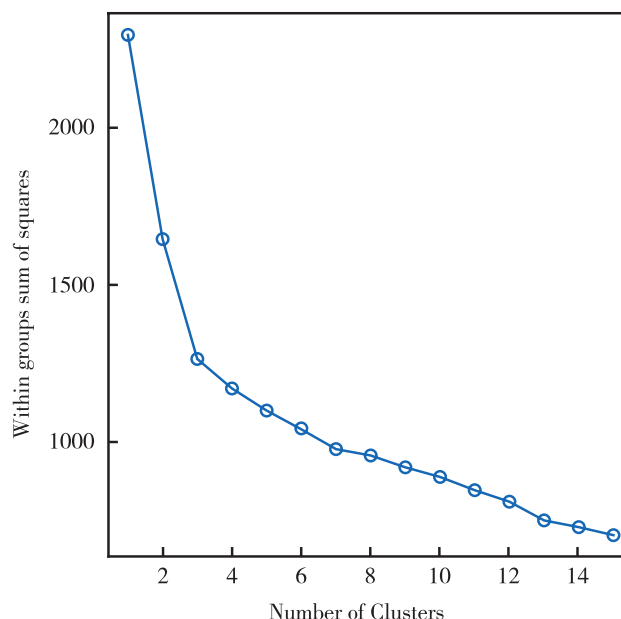


图 2 组内平方和随聚类数变化

表 1 用户活跃度分类指标

	变量名称	变量含义
自身角度	发帖量	用户发的总的帖子数量
	留存时间	用户第一条发帖与最后一条发帖的时间差
	经验值	贴吧里的经验值是根据用户行为计算的, 包括用户的签到、发帖、回帖、顶帖等行为
	头衔等级	贴吧的头衔等级与用户的在线时长、参与讨论、促进用户交流和互动程度有关
交互角度	回帖量	用户总的回帖量、评论数
	关注贴吧数	用户关注的贴吧数量
	关注人数	用户关注的其他用户数量
	粉丝数	该用户被其他用户关注的数量
	最近访客数	最近访问该用户主页的总人数

本文使用 RapidMiner 工具中的 K-Means 算法对处理后的用户相关指标数据进行聚类分析, 最终得到的用户活跃度聚类结果如下: 第一类用户 90 人, 第二类用户 1298 人, 第三类用户 1417 人。结果如图 3 所示:

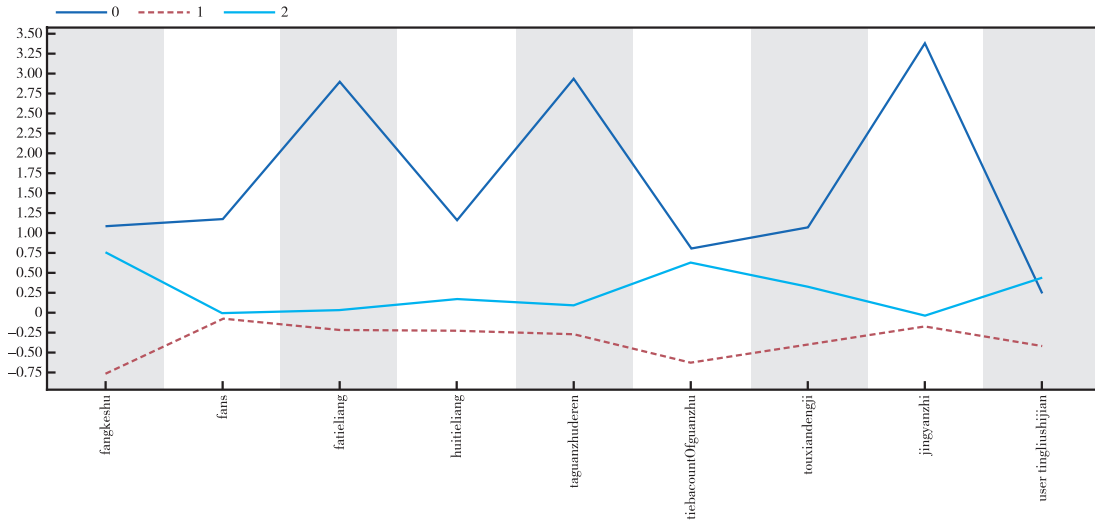


图3 三类用户相关数据

注：折线由上至下依次代表第一类用户、第二类用户和第三类用户。

K-Means 用户指标质心结果如表 2 所示：

结合图 3 和表 2，我们可以看出，第一类用户相关数据指标均处于较高的水平，明显高于其他两类用户，发帖量和关注的人数、经验值明显处于较高的水平，贴吧里面的经验值是根据用户的签到数、发帖回帖量、被别人回复的帖子数和上传的照片数计算的，由此可见，第一类用户不仅局限在自身发帖方面，更热衷于回复其他用户、与其他用户进行积极的互动和交流。

表 2 各类用户指标质心

指标	第一类用户	第二类用户	第三类用户
访客数（人）	1.092	0.757	-0.763
粉丝数（人）	1.173	0.001	-0.001
发帖量（条）	2.914	0.039	-0.039
回帖量（条）	1.147	0.171	-0.171
关注人数（人）	2.934	0.092	-0.092
关注贴吧数（个）	0.796	0.630	-0.630
头衔等级	1.064	0.319	-0.319
经验值	3.373	-0.041	-0.041
用户留存时间（天）	0.231	0.440	-0.440

为进一步了解第一类用户的特征，本研究提取出部分第一类用户相关指标，如表 3 所示。可以看出，第一类用户的发帖量远大于回帖量，该类用户更倾向于主动分享，他们对于鼻炎贴吧的发帖量和发帖内容起着至关重要的作用，成为百度鼻炎吧中主要的信息提供者。第一类用户不仅局限在自身发帖方面，更热衷于回复其他用户，与其他用户进行积极的互动和交流。如表 3 中序号为 2 的用户免费提供自用治愈的鼻炎偏方，包括配方、使用方法、用法用量、注意事项以及对

患者的安慰和鼓励, 并且他的帖子得到很多用户的积极回应, 有些用户进行了回帖, 从而可以看出该用户的发帖促进了贴吧的活跃度, 并且带动了其他用户的交流。相较于第一类用户, 第二类和第三类用户在指标数据处于较低水平, 参与水平较低。张昱认为活跃度是用户使用某种功能或者服务的频繁程度<sup>[18]</sup>, 周静认为活跃度用来衡量用户活动的频繁程度<sup>[19]</sup>, 并且已有研究表明非活跃用户会聚集在活跃用户的周围<sup>[20]</sup>, 因此本文选取鼻炎贴吧中表现最为活跃的第一类用户作为活跃用户代表, 进行活跃用户交互行为的分析。

表 3 部分活跃用户指标数据

序号	访客数 (人)	粉丝数 (人)	发帖量 (条)	回帖量 (条)	关注的人数 (人)	关注的贴吧数 (个)	头衔等级	经验值	留存时间
1	11	103	6862	285	88	18	9	1075	43 days, 0:00:00
2	7	166	4268	347	75	20	12	7684	856 days, 0:00:00
3	4	13	515	299	26	5	7	454	13 days, 0:00:00
4	12	24	826	363	40	2	7	253	52 days, 0:00:00
5	9	193	8624	294	32	20	6	124	275 days, 0:00:00
6	11	336	3.3 万	212	349	20	4	30	469 days, 0:00:00
7	12	914	2.7 万	216	4	20	2	10	335 days, 0:00:00
8	11	721	1.8 万	450	730	20	9	1828	113 days, 0:00:00
9	12	610	1.7 万	846	484	17	8	868	499 days, 0:00:00
10	10	158	1.7 万	181	72	6	4	46	147 days, 0:00:00
11	12	786	1.3 万	79	63	20	5	51	73 days, 0:00:00
12	11	272	1.2 万	1921	27	20	5	90	237 days, 0:00:00
13	10	288	1.2 万	926	11	20	7	224	910 days, 0:00:00
14	6	59	7299	36	41	20	3	25	23 days, 0:00:00
15	12	368	3.6 万	347	621	20	7	233	504 days, 0:00:00
16	11	721	1.8 万	450	256	20	4	48	386 days, 0:00:00
17	11	132	7696	80	599	20	5	60	1449 days, 0:00:00
18	11	336	3.3 万	212	79	20	9	1367	193 days, 0:00:00
19	10	158	1.7 万	181	54	20	7	442	96 days, 0:00:00
20	11	57	7707	1574	328	13	6	112	0:00:00

## 4 用户交互类型研究

本部分将在用户分类的基础上以第一类用户作为活跃用户进行研究, 采用内容分析法对百度鼻炎吧中的用户交互行为进行深入研究。通过已有研究可以发现: 第一, 活跃用户在百度贴吧里面发帖回帖积极性均很高, 并且得到了其他用户的积极回应; 第二, 活跃用户参与的评论内容占据了百度鼻炎吧中用户评论的 47%, 本文中这些评论内容能够很好地代表鼻炎贴吧中用户交互的主要内容, 并且已有相关研究中所分析发现的在线社区中核心成员和活跃成员贡献绝大部分信息交流内容<sup>[21]</sup>。因此本文选取了 90 个活跃用户, 通过爬取与 90 个活跃用户相关的发帖回帖评

论内容对在线健康社区用户的交互行为进行深入研究。

在线医疗健康社区主要提供两种功能：一是提供医疗健康类的信息；二是提供社会支持<sup>[22]</sup>，包括信息支持、情感支持、陪伴。信息支持指社区成员之间的信息、建议或指导的传递，在线健康社区中通常是与疾病和健康相关的转诊信息、建议和个人经历等信息。情感支持指有关理解、鼓励、同情、关心等的情绪表达，能够帮助降压和减轻焦虑。陪伴指参与者之间的聊天、开玩笑、分享日常生活等行为，传达的内容中不包含与疾病相关的信息，它能够加强社区成员之间的联系和社区认同感<sup>[4]</sup>。本文将结合社会支持行为编码和健康社区的特点对用户评论内容进行编码分类，为保证编码的准确性，本文采用内容分析法对评论内容进行手工编码。

#### 4.1 评论内容处理及编码

本文首先对爬取的活跃用户的评论内容进行预处理，共得到 2805 条活跃用户交互评论内容，根据用户之间评论内容的主题进行编码，为了避免个人主观性，保证编码的可靠性，本文遵循两个人编码原则，两个人各自独立对用户评论内容进行编码，一致性为 92%，并对不一致的编码经过研讨后，形成一致的结果。根据编码结果，本文将在线健康社区用户交互行为分为预防类、求助类、症状类、治疗类、反馈类、提醒类、情感类、离题类八种交互类型。部分编码示例如表 4 所示：

表 4 编码示例

核心观点	交互类型分类
大气污染重，戴好口罩	预防
煮菜放入姜葱蒜以达到平衡	
不想堵塞鼻孔，还得继续研究	
喷剂能使用吗？	求助
怎么治好的？可以私信不？	
楼主你鼻涕倒流吗？鼻子塞吗？	
晚上睡觉鼻塞	症状
鼻涕倒流、白色结痂、鼻毛脱落	
鼻塞，交替性不通	
停用布地奈德	治疗
针灸艾灸、鼻塞	
使用中药	
好了很多	反馈
金霉素眼膏……未见效	
鼻炎膏不行的	
别耽搁时间	提醒
不要胡乱用药	
这种药，不能用	
好烦人啊！	情感
看到你回复心情好！	
心情又有点不太好，不敢哭，加油	
这报纸是你自己打印的	离题
我也 40 岁了，大姐	
VX 号……	



(1) 预防类交互行为指用户之间进行讨论关于避免鼻炎发生所涉及的内容行为, 比如用户在饮食、出行方面的注意事项, 关于鼻炎是否遗传的探讨, 该类交互行为的用户大多是鼻炎患者, 他们在患病之后告诫其他用户提前预防鼻炎的发生, 提供一些预防方法和建议。

(2) 咨询与询问交互行为指用户通过提问具体问题以求得其他用户的解答, 涉及的内容包括咨询鼻炎的症状、治疗的方法、治疗效果等, 该类交互行为是鼻炎贴吧的主要行为类型之一。在线健康社区作为用户获取医疗健康类的信息的主要来源之一, 可见大多数鼻炎患者或者患者家属渴望通过在线健康社区得到有关疾病信息的解答。

(3) 症状类交互类型指的是用户通过描述自己的鼻炎症状来与其他用户进行交流, 为社区用户提供参考, 通过对用户交互内容的编码和分析, 发现用户提到鼻炎症状更多的是鼻塞、鼻干, 一部分用户提到鼻炎难受导致睡眠不足, 可见患者用户深受鼻炎疾病的困扰, 该类交互行为的用户少数是之前患者, 大多数用户现在遭受鼻炎的困扰, 希望通过症状的描述来得到其他用户的帮助。

(4) 治疗类交互行为的表现是用户之间的交流内容包括鼻炎治疗过程的讨论, 用户评论文本中涉及的内容有中药、西药、中医治疗、手术等治疗方法, 也有少数用户提及偏方治疗和通过饮食治疗。用户通过分享自己的治疗方法和治疗过程与其他用户进行交互, 该类交互行为是鼻炎社区中最重要的交互行为, 涉及的用户最多, 由此可见用户参与在线健康社区主要目的是寻求治疗方法。

(5) 反馈类交互行为中用户通过对自己之前治疗方法效果的反馈给其他用户提供了参考, 例如, 用户评论某种偏方使用之后并没有任何效果、通过坚持使用某种药物可以治疗鼻炎, 用户的反馈内容能够帮助其他用户更准更快地找到有效的治疗方法, 避免采用无效甚至有害的治疗方式。

(6) 提醒类交互行为涉及的主要内容是用户对于其他用户关于社区中一些欺骗行为的善意提醒, 鼻炎贴吧中有少数用户存在打广告卖药的行为, 用户在交互过程中会对该类用户的行为提出怀疑并提醒其他用户防止上当受骗。提醒类交互行为还涉及对用户治疗时间的提醒, 鼻炎早期治疗效果更好, 因此用户会建议有鼻炎症状的用户尽早治疗。

(7) 情感类交互行为表现为用户情绪的表达, 鼻炎贴吧中通过对用户交互内容的分析涉及的情绪表达有鼻炎困扰的难过与迷茫、用药无效之后的失望、对于治疗方法的怀疑、患者之间的鼓励与祝福等, 在线健康社区为具有相同经历的患者提供了情感表达的平台, 通过用户之间的互动, 他们可以得到安慰和支持。

(8) 离题类交互行为指用户的交互内容与鼻炎贴吧主题无关的行为, 主要表现为用户之间的闲谈, 虽然用户之间的离题交互并没有涉及有关鼻炎的内容, 但是有研究表明在线社区中用户之间的闲谈大多涉及的是一些比较轻松的话题<sup>[4]</sup>, 这种交互行为有助于缓解由于鼻炎痛苦形成的压抑的氛围。

#### 4.2 编码结果与交互行为统计

根据以上编码方式和交互类型分类, 得到的活跃用户评论内容的统计结果如表 5 所示:

结果表明, 治疗类交互类型共计 1226 条, 占比 43.7%, 求助类交互类型共计 485 条, 占比 17.3%, 症状类交互类型共计 244 条, 占比 8.7%, 反馈类交互行为类型 177 条, 占比 6.3%, 提醒类交互行为类型共计 244 条, 占比 8.7%, 离题类交互类型和情感类交互类型也均较少, 仅有 216 条和 157 条, 占比分别为 7.7% 和 5.6%。这说明大多数用户在健康社区中想通过用户之间的交流

互动来得到其他用户的建议和帮助,寻求正确的治疗方法,治疗疾病是鼻炎贴吧中用户交互行为的主要目的,其次是关于疾病治疗的咨询和求助,进一步反应了百度鼻炎吧中用户信息需求主要集中于有关鼻炎的治疗和求助。而症状、提醒等交互行为主要是为了配合治疗所产生的中间环节,在百度贴吧这个社区中占据较少的成分。值得一提的是,离题的交互行为超过了情感、反馈等交互内容,为此有学者提出离题行为不一定不利于健康社区的发展,有些离题行为涉及的话题可能比较轻松,会促进用户之间的交流互动,如 Huang K Y 基于在线癌症社区的讨论区版块发现离题行为有助于促进用户更好地融入社区中<sup>[23]</sup>。在线社区为用户情绪的表达和宣泄提供了场所,用户之间的情感交互行为有利于用户在社区中寻求归属感,用户之间的相互鼓励和依赖有利于鼻炎贴吧中用户的留存。

表5 交互行为类型统计

交互类型	次数(次)	占比(%)
预防	56	2.0
求助	485	17.3
症状	244	8.7
治疗	1226	43.7
反馈	177	6.3
提醒	244	8.7
情感	157	5.6
离题	216	7.7

## 5 研究结论

本文将百度贴吧鼻炎社区的用户作为研究对象,从用户自身角度和用户交互角度两个维度构建用户活跃度的分类指标,使用 K-Means 算法识别活跃用户,进而使用内容分析法细化活跃用户交互行为类型,具体结果如下。

### 5.1 在线健康社区活跃用户的识别

本文利用百度鼻炎吧的用户数据进行用户活跃度分类从而识别在线健康社区的活跃用户。在选取用户活跃度分类指标时,本文在以往研究的基础上提出了从用户自身角度和用户交互角度两个维度构建用户活跃度的分类指标,进一步将用户活跃度的分类维度进行完善。在识别活跃用户时,本文根据 R 语言中的 wssplot 函数绘制的组内平方和随聚类数目的变化趋势图确定最佳分类数目,根据结果本文确定了用户分类的最佳聚类数为 3 类,在此基础上识别出活跃用户。

在线健康社区的发展不仅依靠网络技术的支持,更需要用户的积极参与以及建立用户之间的长期联系<sup>[24]</sup>,在线社区中成员关系网主要是通过几个中心成员支撑起来的<sup>[25]</sup>。活跃用户作为在线社区中的意见领袖,是在线社区中主要的内容创造者和信息提供者,活跃用户的存在不仅有利于在线社区新知识的创造和分享,并且能在一定程度上引起其他用户的积极回应,提高非活跃用

户的互动积极性,有助于在线健康社区中用户的留存。因此在线健康社区在今后运营过程中,应充分发挥活跃用户的积极作用,带动社区中其他用户进行健康信息的创造和分享,疏导患者用户的不良情绪;对于积极回复、提供有效健康信息的用户进行奖赏,提升用户对社区的归属感,增强社区中用户黏性,提高用户网络互动程度。

## 5.2 在线健康社区活跃用户交互行为探索

本文通过内容分析法将百度鼻炎吧这一在线健康社区中用户交互行为细分为预防、求助、症状、治疗、反馈、提醒、情感、离题 8 种交互行为,统计结果发现治疗类交互行为是百度鼻炎吧中最主要的交互行为类型,求助类交互行为次之,由此可以发现百度鼻炎吧中用户交流互动的主要目的是希望通过在线健康社区获取有效的治疗方法,用户信息需求主要集中于鼻炎的治疗和求助。以往在线健康社区中用户行为研究中比较集中于研究高血压、糖尿病等慢性病的健康社区,这些健康社区中用户的交互内容以疾病日常管理、疾病确证和治疗为主<sup>[26][27]</sup>。高血压、糖尿病等作为一种慢性病,患者在平常生活的自我调理就显得很重要,因此这两种病的患者用户更关注疾病日常管理和预防<sup>[28][29]</sup>,而鼻炎贴吧中,由于该疾病的特征,在本研究中的内容交互的侧重点主要为治疗,而对疾病日常管理和预防内容的交互相对较少,仅占比为 2%。

通过对在线健康社区用户交互行为的细分,并结合已有研究中不同健康社区用户的交互行为关注点的不同,在线健康社区在今后的管理与运营中应该结合用户交互行为的特点,关注不同疾病用户的健康信息需求,使用户享受优质的信息服务。例如在线健康社区可以针对用户信息需求设置预防类、治疗类、症状类等不同需求的健康信息版块,方便用户在纷杂的信息中快速找到所需的健康信息;针对在线健康社区中用户大量的求助和治疗类需求,可以邀请权威医学专家在线根据用户的差异化需求,为用户提供专业化、个性化的健康知识服务;也可以通过人工智能化分析手段,根据用户的健康信息搜索行为实现智能化、精准化的信息推送,如当用户搜寻相关疾病预防时,平台可以基于用户输入的文本信息通过自然语言处理技术,为用户智能推送相关疾病的症状、治疗,相关医院、医生等信息。

## 6 研究展望

本文选取了国内在线健康社区中的百度鼻炎吧作为研究对象,百度贴吧用户基数大、主题相对单一,在今后研究中可以考虑不同类型在线健康社区以及国内外在线健康社区中用户行为的对比研究;并且随着用户需求的不断变化,用户信息需求以及交互行为呈现动态化发展,因此今后可以进一步研究在线健康社区中用户行为的动态化,发现用户行为的变化规律。

### 【参考文献】

- [1] 中国互联网络信息中心. 第 44 次中国互联网络发展状况统计报告 [R/OL]. (2019-08-30)[2020-03-26]. [http://www.cac.gov.cn/2019-08/30/c\\_1124939590.htm](http://www.cac.gov.cn/2019-08/30/c_1124939590.htm)
- [2] FOX S. The Social Life of Health Information [R/OL]. (2011-05-12)[2018-01-01]. <http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011>.

- [3] 周军杰, 左美云. 线上线下互动、群体分化与知识共享的关系研究——基于虚拟社区的实证分析 [J]. 中国管理科学, 2012, 20(6): 185-192.
- [4] 周军杰. 用户在线参与的行为类型——基于在线健康社区的质性分析 [J]. 管理案例研究与评论, 2016, 9(2): 173-184.
- [5] SANDARS J, HOMER M, PELL G, et al. Web 2.0 and social software: the medical student way of e-learning [J]. Medical Teacher, 2008, 30(3): 308-312.
- [6] 霍英男. 不同信任环境下社交网络用户影响力、活跃度及自我表露关系研究 [D]. 北京邮电大学, 2015.
- [7] 刘卉. 基于网络社区用户活跃度的研究 [D]. 上海: 上海师范大学, 2012.
- [8] YAGUANG Shi, YI Yuan. Research on Information Communication Model of Social Networking [J]. Library Tribune, 2009, 11 (6): 220-223.
- [9] YAN X, YAN L. Gender Classification of Weblog Authors [C]// AAAI spring symposium: computational approaches to analyzing weblogs, Boston, 2006: 228-230.
- [10] 黄炜, 李总苛, 李岳峰. 微信用户活跃度影响因素分析 [J]. 湖北工业大学学报, 2015, 30(6): 29-32+63.
- [11] 刘璇, 汪林威, 李嘉, 等. 在线健康社区中用户回帖行为影响机理研究 [J]. 管理科学, 2017, 30(1): 62-72.
- [12] 林鑫, 周知. 用户认知对标签使用行为的影响分析——基于电影社会化标注数据的实证分析 [J]. 情报理论与实践, 2015, 38(10): 85-88.
- [13] 曹磊. 新浪微博与腾讯微博个人页比较研究 [D]. 武汉: 华中科技大学, 2012.
- [14] 邓岩. 基于 SNS 的用户人气影响因素分析 [J]. 中国新通信, 2013, 15(22): 34-35.
- [15] 张鑫. 在线健康社区用户参与行为的类型及偏好研究 [J]. 情报资料工作, 2019, 40(5): 84-91.
- [16] HUANG K, INDUSHOBHAC, RAN W. Not just for support: Companionship activities in health care virtual support communities [J]. Communications of the Association for Information Systems, 2014, 29(34): 561-594.
- [17] 赵栋祥. 国内在线健康社区研究现状综述 [J]. 图书情报工作, 2018, 62(9): 134-142.
- [18] 张昱, 谢怀军. 手机银行用户活跃度与粘度的影响因素与业务发展研究 [J]. 财经界 (学术版), 2012(2): 168-170.
- [19] 周静, 李季. 从人口统计特征和生活方式探讨 SNS 社交网站用户的使用行为及其活跃度 [J]. 科技与管理, 2011, 13(2): 72-77.
- [20] URBANOSKI K, VAN Mierlo T, CUNNINGHAM J. Investigating patterns of participation in an online support group for problem drinking: a social network analysis [J]. International Journal of Behavioral Medicine, 2017, 24(5): 703-712.
- [21] 李纲, 王馨平, 巴志超. 微信群中会话网络结构及用户交互行为分析 [J]. 情报理论与实践, 2018, 41(10): 124-130+14.
- [22] NAMBISAN P. Information seeking and social support in online health communities: impact on patients' perceived empathy [J]. Journal of the American Medical Informatics Association, 2011, 18(3): 298-304.
- [23] HUANG K Y, CHENGALUR-Smith I S, RAN W. Not just for support: companionship activities in healthcare virtual support communities [J]. Communications of the Association for Information Systems, 2014, 34(1): 29.
- [24] IRIBERRI A, LEROY G. A life-cycle perspective on online community success [J]. ACM Computing Surveys (CSUR), 2009, 41(2): 1-29.
- [25] 刘瑛, 孙阳. 弱势群体网络虚拟社区的社会支持研究——以乙肝论坛“肝胆相照”为例 [J]. 新闻与传播研究, 2011, 18(2): 76-88+111-112.
- [26] 唐晓波, 李津. 在线健康社区信息需求主题分析 [J]. 数字图书馆论坛, 2019(2): 12-17.
- [27] 李重阳, 翟姗姗, 郑路. 网络健康社区信息需求特征测度——基于时间和主题视角的实证分析 [J]. 数

字图书馆论坛, 2016(9):34-42.

[28] 邓胜利, 刘瑾. 基于文本挖掘的问答社区健康信息行为研究——以“百度知道”为例 [J]. 信息资源管理学报, 2016, 6(3):25-33.

[29] 金碧漪, 许鑫. 社会化问答社区中糖尿病健康信息的需求分析 [J]. 中华医学图书情报杂志, 2014, 23(12):37-42.

# Analysis of Active User Identification and Interactive Behavior in Online Health Community

DONG Wei<sup>1</sup> LI Jianhong<sup>2</sup> TAO Jinhui<sup>1</sup>

(1. School of Education, Tianjin University, Tianjin 300350, China;

2. College of Management and Economics, Tianjin University, Tianjin 300072, China)

---

**Abstract:** [ **Purpose/significance** ] As a platform for users to share medical and health related information and solve problems, online health community plays an increasingly significant role. The research on the characteristics of user interaction behavior in online health community has a certain reference value for further discovering the rules of user interaction in online community and the construction of related platform.

[ **Method/process** ] Using data mining related tools to obtain information about user interaction behavior in online health community, taking Baidu rhinitis bar as an example, the active users are identified and analyzed based on the relevant indicators of users themselves and interaction. On this basis, the content analysis method is used to code the content of active users' comments and further classify the types of user interaction behavior. [ **Result/conclusion** ] It is found that user interaction can be classified into eight types of prevention, consultation and help seeking, symptom, treatment, feedback, reminder, emotion and digression. Treatment, consultation and help seeking interaction are the main types of interaction in Baidu rhinitis bar. Therefore, based on the above research, online health community should combine the characteristics of different types of user interaction behaviors in the future management and operation, pay attention to the different health information needs of users with different diseases, and provide users with accurate and high-quality information services in appropriate ways.

**Keywords:** Online health community; Active users; Interaction behavior type

---

( 本文责编: 周 霞 )