

基于电商评价数据的农产品用户画像分析

——以安化黑茶为例

李勇 谭小玲 陈晓婷 管慧¹

(长沙学院 经济与管理学院, 湖南 长沙 410000)

【摘要】: 通过构建安化黑茶产品的用户画像, 分析其用户群体的消费行为特征, 可为商家提供多维度的产品用户信息, 便于商家了解用户, 进而改进产品与服务, 为安化黑茶电商营销策略提供决策依据。以京东、天猫、淘宝三大电商平台为基础, 使用 python 的 scrapy 第三方库采集淘宝、京东电商平台安华黑茶的评价数据, 并对数据进行清洗、去空值、去停用词等预处理, 运用 jieba 第三方库对评论数据进行分词、词频分析, 完成对安化黑茶用户特征的完整刻画。结合客户评论数据分析, 可知安化黑茶近 3/4 用户为男性; 山东省为安化黑茶网购大省; 网购渠道依次为淘宝、京东、天猫; 关注安化黑茶的人更喜欢教育、汽车服饰、综艺等; 消费者对安化黑茶诉求如“减肥”、“抗衰老”、“降血脂”等比例升高。

【关键词】: 用户画像 安化黑茶 电商营销 数据分析 文本挖掘

【中图分类号】 F724.6; F323.7 **【文献标识码】** A

1 引言

近年来, 电子商务发展迅速。湖南安华黑茶借助电商平台的便捷化优势, 拓展了线上营销渠道, 其线上电商平台的商品评论区也逐渐成为消费者发表商品诉求的交流平台。因此可以通过获取用户的评论数据, 来挖掘用户的需求, 以此帮助湖南安化黑茶提升品牌价值、增强市场竞争力。而大数据环境下的用户画像分析是近年来深度挖掘用户需求的重要方法。国内外学者不断关注用户画像方面的研究, 其中用户画像精准营销是当前研究的热点。

用户画像这一概念最早由交互设计之父 Alan Cooper 提出, 其认为用户画像是真实用户的虚拟代表是建立在真实数据之上的目标用户模型。Teixeira C、Pinto JS、Martins JA 等国外研究者认为用户画像是一个独立的描述用户需求、偏好和兴趣的模型, 是从海量数据中提炼的个人数据集合。中国科学技术信息研究所信息资源中心主任曾建勋认为, 用户画像是指获取用户的专业背景、文化程度、知识获取习惯、兴趣偏好、特长任务等与用户需求趋向相关的信息, 以此为基础进行模型化表示, 为用户制定特定标签。虽然研究者对用户画像的概念表述存在一定差异, 但研究者一致认同用户画像是面向用户展开的基于真实数据进行的研究。Zoratti S 等提出了如何利用数据挖掘事物间的关联性并应用到精准营销中; Kennedy DS 从移动互联网的角度提出精准营销的具体实施方法; Zhen Y 等提出了精准营销的决策框架, 通过识别不同类别客户之间的潜在特征, 提出适当的精准营销策略。用户画像技术还逐步涉及个性化搜索、城市计算、预测用户年龄、兴趣爱好、性别、地理位置等。用户画像有两个层次的定义。第一种是从产品设计、运营过程中从用户群体中抽象出来的典型用户, 能够表达用户群体的主要需求和期望。第二种是基于特定使用情境下真实的用户行为数据所形成的描述用户属性及其行为的标签集合。第二种层次的用户画像更具有真

¹**作者简介:** 李勇 (1979-), 男, 湖南长沙人, 博士后, 副教授, 研究方向: 乡村振兴、舆情传播、复杂网络。
基金项目: 湖南省自然科学基金面上项目 (2019JJ40328)。

实性以及较强的时效性。

近几年，关于用户画像精准营销的决策框架的研究越来越多，但涉及具体案例及具体应用的研究较少，因此，本文构建用户画像是针对第二层次，根据客户的评价数据，抓住其需求特征，进行用户画像分析与可视化的构建，为企业确定目标客户，制定精准的营销方式、迎合不同消费者的需求提供依据。研究内容包括两部分，第一部分是获取用户的评论数据来搜集用户属性信息，将属性信息标签化以便对用户进行画像。第二部分是基于用户画像挖掘消费者对于安化黑茶的需求，实现对企业精准营销方式的进一步完善。

2 客户评价数据的采集与处理

构建用户画像，应从用户留在网络上的“数据足迹”的采集、整理与归类开始。用户发表评论的时间、内容属性、追加评论、评论图片是用户数据采集的主要来源，采集数据之后需要对用户数据进行有效过滤，去除无效、杂乱数据，过滤之后对海量的用户评论数据进行清洗、分词、词频分析等处理，发现其中隐含的有价值的信息，再通过文本处理工具与情感分析工具构建用户画像，主要从用户特征、购买渠道、地域分布、关联产品这几个方面进行可视化呈现，具体如图 1 所示。

2.1 客户评价数据采集

据公开资料显示，中国网民已经达 6.88 亿，其中 90.1%的通过手机上网，手机网民达 6.2 亿，超过 70%网购消费发生在移动端，电商网络逐渐成为主流。而在众多电商平台中，天猫、淘宝电商平台年度活跃用户达 5.76 亿，天猫商家店铺数量达 23 万，淘宝商家店铺数量达 1000 万；京东电商平台年度活跃用户达 4.18 亿，已注册京东商城的商家超过 20 万。

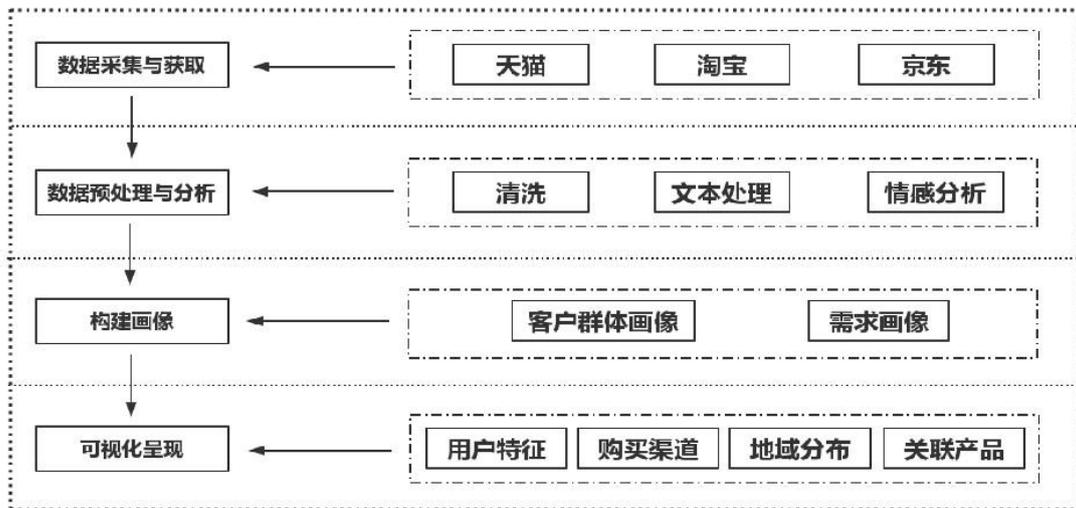


图 1 安化黑茶用户画像构建流程图

据此，本文以淘宝、京东为数据采集对象，使用 python 的 scrapy 第三方库爬取对淘宝、天猫、京东有关安化黑茶的客户评论数据。由于淘宝、京东平台上的安化黑茶商家众多且经营商品复杂，不利于后期数据分析工作。因此本文根据店铺的好评率、收藏人数、经营年数将天猫、淘宝店铺分成“金色皇冠”商家、“蓝色皇冠”商家、“蓝色钻石”商家并从这三类商家中分别选取三个最有影响力的商家，从京东旗舰店、自营店、第三方卖家三类商家中分别选取三个最有影响力的商家，选取这六个店铺的客户评价作为数据样本，具体店铺选取情况如表 1 所示。

将以上三家天猫、淘宝店铺的安化黑茶客户评价作为关键词，采集客户评价数据并将数据采集结果保存至 Excel 表格中。本文共采集了 10 个字段，主要涉及电商消费者发布的评价内容及时间等，共采集到 13571 条数据，数据采集结果如图 2 所示。将采集的三家京东店铺的安华黑茶客户评价数据保存至 Excel 表中，共采集到 12037 条数据，局部数据采集结果如图 2 所示。

会员	评价星级	评价内容	时间	点赞数	页面网址	页面标题	采集时间
D***k	star5	茶收到了，泡了一杯。茶汤清淡，茶味很好。不过黑茶还是熬煮好	2019-03-11 16:32	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:21.9194723
波***魔	star5	去年秋季买来二袋，马上热水泡了这茶，一喝感觉非常好，口感好	2019-01-20 20:19	2	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.1477841
张***3	star5	给妈妈买的黑茶，一直喝的这个品牌，这次购物很满意，快递小哥	2019-01-27 20:51	2	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.2984946
J***f	star5	掌柜的服务态度真好，发货很快。商品质量也相当不错。太喜欢了	2019-03-22 19:09	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.3636008
J***a	star5	不错，物流很快，前天下单的，昨天就到了，而且包装还是很严实	2019-04-08 09:41	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.4644514
使***8	star5	茶叶口感纯正，很好喝，物美价廉，反复冲泡多次仍然入口醇香，	2019-04-20 20:39	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.5853564
w***2	star5	量很足，还没有喝，据说很好，可以减脂，刮油。第一次喝黑茶。	2019-04-27 14:41	1	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.8001244
小***李	star5	老爸特别喜欢喝茶，通过朋友知道了还有黑茶，我们没喝过就买来	2019-05-07 16:24	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:22.9434309
J***9	star5	味道很香醇，外形很完整，没有碎，物流很快。一直不知道安化黑	2018-12-07 15:45	1	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:23.1165110
格***箱	star5	茶叶收到，口感我们认为很好。以前都是喝熟普洱，现在想试试喝	2018-10-29 09:14	4	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:23.1953384
J***z	star5	发货和快递很快，拆开包装就能闻到一股茶香，味道也很不错！	2019-05-15 12:21	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:38.4615655
J***d	star5	黑茶不错，很匀，无碎末末，味道醇厚，后味不错，满满的清香，	2019-04-07 20:43	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:38.6376328
e***使	star5	包装很好 看着不错 还没来得及泡呢 喝过再来追评。	2019-05-15 14:50	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:38.8236404
格***箱	star5	前几天我刚买了茶叶，口感很好。这次是给老父亲买的，还没有喝	2018-11-02 19:05	1	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:38.9737905
C***冬	star5	茶不错 比以前喝的味道浓 很好 为什么Q5标不一样呢 这	2018-07-13 12:54	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.0704151
J***j	star5	不错，味道很香，很浓，物流速度快，服务态度好，喝完再订几包	2018-12-06 09:16	2	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.1816209
琴***6	star5	到货就煮茶，然后品茶？，茶汤纯正，味道醇厚，很满意的一次购	2018-12-23 18:51	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.3046083
J***b	star5	宝贝收到了，样子还是很好看，还没有泡泡水，不好说，好评吧。	2018-12-01 12:21	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.4327974
J***4	star5	不错不错不错，以前买过，这次再买，味道还不错，一样的，茶汤	2018-04-14 21:52	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.5721847
J***x	star5	茶叶已收到，送货很快，茶叶还不错我是泡着喝的，煮着喝效果点	2018-11-25 08:26	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:39.7325224
WJXTY	star5	质量挺好，挺耐泡的，味道纯正，清新，香郁，物美价廉，喝完继	2019-05-08 07:01	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:55.1173063
J***0	star5	快递很给力，东西还没尝，不过看包装挺好的，物美价廉	2018-11-25 16:06	4	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:55.2223104
7512837376	star5	茶叶收到了，包装很好。色泽鲜绿。茶叶大小均匀。无碎渣。口感	2019-04-04 13:15	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:55.3940730
J***b	star5	宝贝刚刚收到，打开看了一下，质量好，味道不错！值得购买！	2018-08-15 17:28	0	https://item.jd.com/1713	安化黑茶2013年老茶天尖茶叶	湖南转 2019-05-25 20:19:55.4954871

图 2 电商平台客户评价数据存储界面

2.2 客户评价数据预处理

通过本次研究的客户对于安化黑茶的评价，对应的客户对于安化黑茶的喜好程度与改善建议，从而对安化黑茶的客户做用户画像，因此只有客户评价内容不为空才对本次研究有意义，所以将数据中空字段评论过滤掉。同时，用户的评论列表里可能存在部分过短的评论信息，经过调查研究，发现 8 个字以上的评论信息为有效的评论信息，故以此清洗掉少于 8 个字的用户评论，最终得到天猫、淘宝有效评论数据 12065 条，京东有效评论数据 10087 条数据。

通过数据预处理，集成数据样本，对数据进行描述性统计分析和分类数据统计，再进行可视化。

根据预处理的数据，分析用户语义画像，分析移动端用户的网络业务类型，再根据流量业务跳转规律和用户上网偏好分析，用特定的数据结构将用户画像数据存储在数据库中，然后选择相应的数据可视化工具，将可视化数据转化成 JSON 格式并存储。

2.3 客户评价数据分析

通过对客户评价数据进行分词、词频的分析，可以发现天猫、淘宝消费者中，除了“茶叶”、“服务”、“满意”等基础性的关键词之外，还出现了“口感”、“很快”、“物流”、“下次”等关键词。对这些关键词进行统计后，利用 Tagxedo 可视化工具进行处理，得到高频关键词（频次 50 以上）可视化如图 3(a) 所示。在京东消费者中，“快递”、“包装”、“质量”、“正宗”等这类关键词的比例在升高，得到高频关键词可视化如图 3(b) 所示。



图 3(a) 天猫、淘宝客户

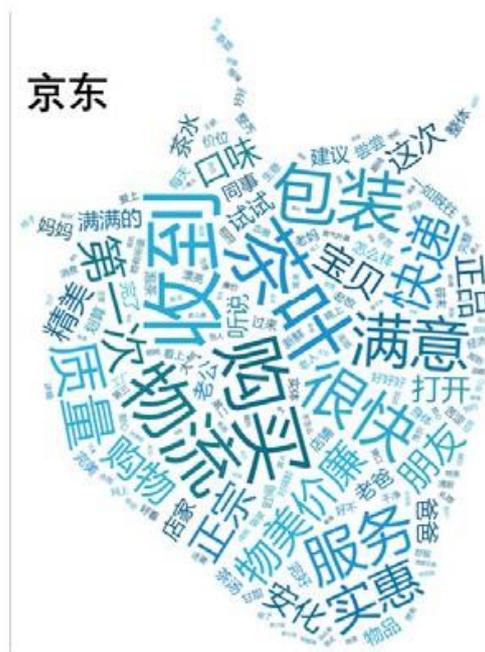


图 3(b) 京东客户评价数据词云图

3 基于电商评价数据的安化黑茶用户画像

用户画像主要为揭示消费者的消费行为，如能通过从数据存储、集成、处理到数据分析以及可视化，并分析可视化呈现的用户画像模型。有利于商家开展对某一类群体精确的营销推广服务。具体可通过对消费者群体用户画像结果进行细分，深度挖掘活跃消费者、普通消费者与潜在消费者，并分析具体消费者的消费行为与消费趋向，以及在营销推广过程中存在哪些问题。

在对用户个体或群体做出的精确刻画和描述的基础上，进行相似需求、兴趣偏好、使用习惯和活跃程度的用户间关联的建立，形成基于客户评价数据的可视化用户画像。

3.1 主要用户特征画像

购买安化黑茶的消费者中，女性占比 33%，男性占比 67%。据分析，虽然从性别这一方面看，男性消费者人群是女性消费者的一倍多，但是女性消费者日益增多，且女性消费者评价中较多出现“健康养生”、“满意”、“给妈妈”等字段，说明了安化黑茶逐渐成为女性养生的产品，女性将是安化黑茶一个庞大的潜在消费群，男女比例如图 4 所示。

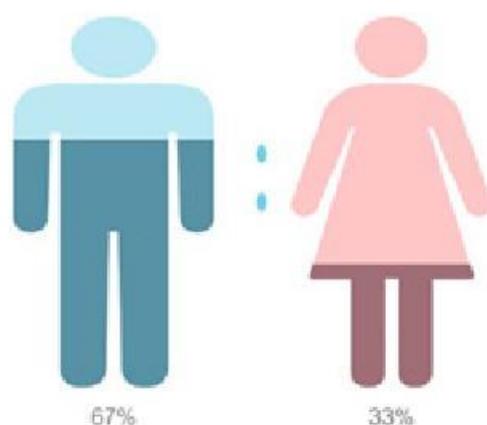


图 4 安化黑茶性别占比图



图 5 安化黑茶用户地域分布图

客户评价数据分析得到山东成为网购安化黑茶的第一大省，其次广东、河南、江苏以及河北对于安化黑茶的购买力度日益上升，东北、贵州、湖北等则增长迅速，市场空间潜力大。相较于广东、福建、安徽、四川、北京和上海这些爱喝茶的省份，新疆、西藏、甘肃等省份对于安化黑茶的关注度较低，用户地域分布情况如图 5 所示。

3.2 主要购买渠道画像

根据安化黑茶客户所提及的购买渠道关键词，用户在淘宝购买的比例为 45.7%；紧接着是京东电商平台购买比例为 29.5%；天猫的购买比例也较高，达到了 15.2%；占比最低的是其他方式，购买比例为 9.6%，具体如图 6 所示。

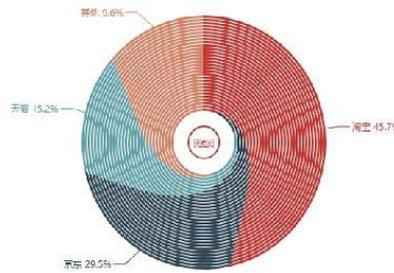


图 6 安化黑茶购买渠道

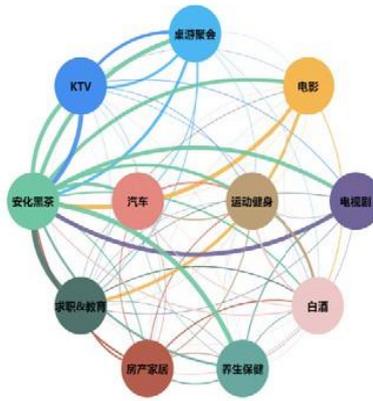


图 7 安化黑茶关联产品

3.3 关联产品特征画像

根据不同电商平台安化黑茶客户的评价数据，利用知识图谱技术对已形成的画像进行分析并建立用户间的关联和交互。如图 7 所示，关注安化黑茶的人更喜欢教育、汽车服饰、综艺和王者荣耀，且安化黑茶人群较年轻，白领居多。可帮助安化黑茶企业寻找各类消费者群体的相似需求及潜在需求，揭示深层次的消费倾向，以便开展根据不同人群的特点制定的营销推广服务。

电商营销中，顾客的数据库对于商家来说是一个极其有利的优势，依附现有的电商平台技术，客户的消费记录数据、用户轨迹、搜索关键词等用户数据都可以集成形成数据库，助于了解用户意向，满足顾客一系列的需求的同时也应该保护客户的隐私，通过顾客属性筛选并建立个人的标签。在电子商务营销中，客户的数据库对于商家来说是非常有利的优势。根据现有的电子商务平台技术，客户的消费记录数据，用户轨迹，搜索关键字和其他用户数据可以集成到数据库中，并通过客户属性过滤和建立个人标签，了解用户的意图并满足一系列客户需求，但应该保护客户的隐私。

4.4 建立客户反馈机制

客户反馈机制不仅可以方便商家在市场验证后通过市场收集客户的意见，还可以根据客户的意见调整产品的营销，使定位更加准确，可以概括产品不受市场欢迎的原因，是未来准确定位产品的基础。

5 结语

本文以大数据背景下安化黑茶的用户画像为研究对象，通过 python 对天猫、淘宝、京东三大电商平台进行数据采集，对三大电商平台的不同等级安化黑茶店铺的客户评价进行文本挖掘与分析，再通过构建用户画像对不同的消费者进行消费行为可视化，在此基础上进行用户画像分析。从多维度对用户的潜在需求进行精确画像，不断提升营销与推广的精准性，不断为发展安化黑茶业务和改善公司形象、为大数据时代安化黑茶企业的发展降本增收。

参考文献:

- [1] 张小可, 沈文明, 杜翠凤. 贝叶斯网络在用户画像构建中的研究[J]. 移动通信, 2016(22):22-26.
- [2] Teixeira C, Pinto J S, Martins J A. User Profiles in organizational Environment [c]. funchal, Madeira, Portugal: the Fourth International Conference on Web Information Systems and Technologis, 2008: 329-332.
- [3] 曾建勋. 精准服务需要用户画像[J]. 数字图书馆论坛, 2017(12):1.
- [4] Zoratti S, Gallagher L. Precision Marketing: Maximizing Revenue Through Relevance[M]. Kogan Page, 2012.
- [5] Kennedy D S. The Ultimate Marketing Plan: Target Your Audience! Get Out Your Message! Build Your Brand! [M]. Adams Media, 2014.
- [6] Zhen Y, Si Y, Zhang D, et al. A decision-making framework for precision marketing[J]. Expert Systems with Applications, 2015 (7): 3357-3367.
- [7] Luo F, Wang J Z, Promislow E. Exploring local community structures in large net-works[J]. Web Intelligence and Agent Systems: An International Journal, 2008 (4): 387-400.
- [8] Yuan N J, Zhang F, Lian D, et al. We know how you live: exploring the spectrum of urban lifestyles[A]. Proceedings of the first ACM conference on Online social networks[C]. ACM, 2013: 3-14.
- [9] Rosenthal S, McKeown K. Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations[A]. Proceedings of the 49th Annual Meeting of the Association for Computations

Linguistics: Human Language Technologies[C]. Association for Computational Linguistics, 2011: 763-772.

[10] Chaabane A, Acs G, Kaafar M A. You are what you like! Information leakage through users' interests[A]. Proceedings of the 19th Annual Network & Distributed System Security Symposium (NDSS) [C]. Citeseer, 2012.

[11] Liu W, Ruths D. What's in a name? using first names as features for gender inference in twitter[A]. AAAI spring Symposium: Analyzing Microtext[C]. 2013.

[12] Burger J D, Henderson J, Kim G, et al. Discriminating gender on twitter[A]. Proceedings of the Conference on Empirical Methods in Natural Language Processing[C]. Association for Computational Linguistics, 2011: 1301-1309.

[13] Chen J, Liu Y, Zou M. Home location profiling for users in social media[J]. Information & Management, 2016 (1): 135-143.

[14] Jurgens D. That's what friends are for: Inferring location in online social media platforms based on social relationships[A]. Proceedings of the 7th International Conference on Weblogs and Social Media [C]. JCWSM, 2013: 273-282.

[15] Li R, Wang S, Deng H, et al. Towards social user profiling: Unified and discriminative influence model for inferring home locations[A]. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. ACM, 2012: 1023-1031.

[16] 张海涛, 崔阳, 王丹, 等. 基于概念格的在线健康社区用户画像研究[J]. 情报学报, 2018, 37(09): 912-922.