基于位置的移动推荐系统效用评价研究

孟祥武 梁 弼 杜雨露 张玉洁
(智能通信软件与多媒体北京市重点实验室(北京邮电大学) 北京 100876)
(北京邮电大学计算机学院 北京 100876)

摘 要 近年来,基于位置的移动推荐系统已经成为个性化推荐服务研究领域的热门课题之一.如何在有限数据集和用户量的情况下,采用恰当的评价方法和指标来有效评估推荐系统性能,已成为移动推荐系统研究的关键任务.本文首先概括分析了基于位置的移动推荐系统效用评价在国内外的研究进展,并与传统推荐系统效用评价进行比较;然后重点从数据集、评价方法、评价指标三方面来对基于位置的移动推荐系统进行详细分析、比较和总结,并发现一些特殊的评价方法和评价指标;同时提出一种基于位置的移动推荐系统四层评价体系,它合理地将模型、数据集、评价方法、评价指标等有机地结合起来,并恰当呈现出这些评价要素之间的相互关系;最后对基于位置的移动推荐系统效用评价的有待深入研究的问题及发展趋势进行展望,并得出一些相关结论.

关键词 地理位置;移动推荐系统;数据集;评价方法;评价指标;评价体系
中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2019.02695

《计算机学报》

A Survey of Evaluation for Location-Based Mobile Recommender Systems

MENG Xiang-Wu LIANG Bi DU Yu-Lu ZHANG Yu-Jie
(*Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia*
(*Beijing University of Posts and Telecommunications*), *Beijing* 100876)
(*School of Computer Science*, *Beijing University of Posts and Telecommunications*, *Beijing* 100876)

Abstract With the rapid development of technologies such as mobile terminals, wireless communications and Internet of Things, location-based mobile recommender systems have been widely used and become increasingly popular in recent years, and this kind of recommender system has become a hot issue in the research area of personalized recommendation services. Its goal is to recommend appropriate items to users at any time and any place, so that users can enjoy satisfactory mobile services anywhere, anytime. Moreover, the recommended performance determines the service quality of the system, so location-based mobile recommender systems have a key task that is how to use appropriate evaluation methods and metrics to effectively evaluate their performance in the case of limited datasets and users. In this paper, we review the progress of research on the evaluation of location-based mobile recommender systems in recent years. Specifically speaking, we first summarize the current research status of evaluation for location-based mobile recommender systems both at home and abroad, and compare them with traditional recommender systems from three aspects of evaluation methods, evaluation metrics and datasets. Secondly, we provide an overview of datasets, not only to clarify the importance of datasets in evaluating location-based mobile recommender systems, but also to point out the relationship between datasets and recommendation models, evaluation methods and evaluation metrics. At the same time, the

收稿日期:2018-08-06;在线出版日期:2019-03-15.**孟祥武**,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为推荐系统、人工智能.E-mail: mengxw@bupt.edu.cn. **梁 弼**,博士研究生,副教授,主要研究方向为推荐系统、Web 服务、智能信息处理. **杜雨露**,博士研究生,主要研究方向为推荐系统、智能信息处理. **张玉洁**,副教授,主要研究方向为智能信息处理、推荐系统.

current usage circumstances of four common datasets, namely Foursquare, Yelp, Gowalla and DoubanEvent are compared and analyzed. Thirdly, we elaborate on the three evaluation methods (i. e. the subjective evaluation method, the objective evaluation method and the subjective-objective evaluation method), which are often used in location-based mobile recommender systems from the aspects of definition, process, application, advantages and disadvantages. In addition, we draw a conclusion that the subjective-objective evaluation method can make full use of the advantages of subjective evaluation and objective evaluation, so that the result of evaluation for the location-based mobile recommender system is more accurate and reasonable, but the excessive costs make it impossible to be commonly used in practical applications. Then, we summarize the evaluation metrics, and find out that some special evaluation metrics including the real-time, response time, adaptability and interactivity are more suitable for evaluating the location-based mobile recommender system. According to the efficacy of evaluation metrics, we divide the commonly-used evaluation metrics into four types (i. e., recommendation-centered, user-centered, system-centered and location-centered ones). It is concluded that the main metrics currently used in the location-based mobile recommender system are precision, recall, satisfaction and the real-time. In particular, according to the evaluation process of the location-based mobile recommender system, we propose a four-layer evaluation architecture for it, which consists of the scene layer, the model layer, the evaluation layer and the display layer. The functions of each layer are relatively independent but work together to complete the evaluation of the location-based mobile recommender system. The practice has proved that the four-layer evaluation architecture is correct, reasonable, feasible and universal, and those developers or researchers can refer to it while evaluating location-based mobile recommender systems. Finally, we discuss its development trends in the future and the problems to be further studied on evaluation for location-based mobile recommender systems, and draw some related conclusions.

Keywords geographical location; mobile recommender systems; datasets; evaluation methods; evaluation metrics; evaluation architecture

《计算机学报》

1 引 言

近年来,随着移动终端、传感器、无线通信、物联网以及人工智能等技术的快速发展,移动 APP (Application)得到广泛应用[1-5],基于位置的移动推荐系统也越来越受欢迎,如旅游推荐系统[6]、影视推荐系统[7]、餐馆推荐系统[8]以及社交媒体推荐系统[9]等.该类推荐系统目标是在任何时间和任何地点向用户推荐合适的项目,让用户随时随地都能享受到满意的移动服务,如景点推荐服务、电影推荐服务、活动推荐服务等[10].其形式化定义可描述为 F: $U\times I\times L\rightarrow R$,其中 F 表示基于位置的移动推荐系统模型函数,$U=\{u_1,u_2,u_3,\cdots,u_{|U|}\}$表示用户集合,$I=\{i_1,i_2,i_3,\cdots,i_{|I|}\}$表示项目集合,$L=\{l_1,l_2,l_3,\cdots,l_{|L|}\}$表示用户的位置集合,$R=\{r_1,r_2,r_3,\cdots,r_{|R|}\}$表示被推荐项目的有序集合.依据软件工程的思想,不论是哪类软件在被投于使用之前都需要进行评测和审核,这是软件生命周期中极其重要的环节,基于位置的移动推荐系统自然也不例外.而且,效用评价已成为设计、实现和优化基于位置的移动推荐系统过程中不可缺的环节,其基本评价过程如图 1 所示[11].

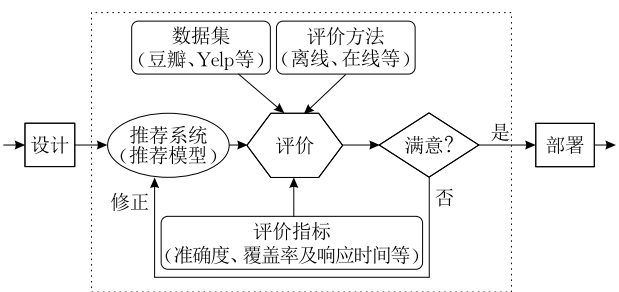

图 1 推荐系统效用评价过程

由图 1 可知，评价基于位置的移动推荐系统主要涉及到评价方法的选择、评价指标的确定以及数据集的筛选等任务. 并且，效用评价是基于位置的移动推荐系统层次化框架中非常重要的一层[12]，该层的基本功能是当把推荐结果展现给用户时，通过结合用户显式或隐式的反馈信息，利用准确度、满意度、实时性等评价指标来有效评测其性能优劣并发现其存在的问题，并根据具体需求对它进行扩展、完善和改进. 同时，在不同上下文环境中基于位置的移动推荐系统性能有着较大差别，甚至在相同环境下使用不同的推荐算法或模型，获得的推荐结果都不相同. 因此，如何恰当并有效地评价基于位置的移动推荐系统性能，已成为当前移动推荐服务研究中一项重要任务[13].

目前，面对众多不同类型的移动推荐系统，国内外不同研究人员或机构采用不一样的效用评价方案. 文献[14]使用问卷调查方式来评价基于上下文感知的移动推荐系统原型(ReRex)，虽然得到了用户真实评价结果，但参与人员和时间有限，评价结果具有一定的局限性. 文献[15-16]中 Yin 等人采用离线的方式来评测所构建的推荐模型和 POI(Point Of Interest)推荐系统，虽然方便快捷，但没有用户真正参与，不能获得真实用户评测结果，自然不能准确反映用户的真实情况. 文献[17]构建了位置和时间感知的社会协同检索模型(LTSCR)来实现连续 POI 推荐任务，并采用离线的方式在 Gowalla 和 Brightkite 数据集上对 LTSCR 进行评测，实验结果表明不论是准确率和召回率指标还是 MRR(Mean Reciprocal Rank)指标，该模型都优于现有的模型(如 POP、POP＋LR、PMF＋LR 和 FPMC＋LR)，但该模型未涉及到其它影响因素(如文本内容信息)，其推荐准确率较低. 文献[18]使用提出的概率实体推荐方法开发了一个位置时间感知的移动推荐系统(easylife)，并通过主客观结合方式来评估其性能，实验表明该系统能有效地为用户探索和发现当前所在位置周围的实体类型(如餐厅、酒店等)及实体(如我爱寿司、我喜欢住如家酒店等)，但该文献只考虑了用户-地点-时间因素，未考虑用户的社会关系等其它因素，故而其推荐的满意度不是很高. 文献[19]提出一个通用的推荐系统框架，它足以包含到目前为止的每一个推荐系统，同时保持了它们的具体特性，并使用一种通用方法，以新的评价指标 $\rho(\rho(S)=\frac{a}{|ev_r(S)|}$，$\rho$ 表示推荐系统性能指标，a 为常数，S 表示会话，$|ev_r(S)|$ 为推荐项目的个数)来评价这个通用框架范围内任何推荐系统(包括基于位置的移动推荐系统)性能. 这种方案虽然能评价目前所有的推荐系统，但评价指标过于单一(只有 ρ 指标)，所以有待进一步改进和完善. 由此可见，当前评价基于位置的移动推荐系统的方案已经很多，而且选择恰当的方法、指标和数据集来对它进行正确评价是必要的，这是基于位置的移动推荐系统进一步发展和推广非常重要的一步. 因为只有性能良好的推荐系统才能被用户所接受，才能被广泛使用，其生命周期才会更长.

而且，通过查阅文献，发现目前已有学者对推荐系统效用评价方面进行了综述. 譬如朱郁筱在文献[20]中对传统推荐系统现有的评价指标(如准确度、覆盖率、多样性和新颖性等)进行多方面多角度综述，并对它们各自的优缺点及适用环境进行深入剖析，而且指出目前评价覆盖率的指标还不够成熟，并给出该评价指标应符合的几条标准以供参考，同时比较了用户间的多样性和用户内的多样性. 但未对其它非传统推荐系统的评价指标进行综述，也未对与评价指标相关的数据集进行总结，因此该综述所论述内容不够全面. 刘建国在文献[21]中从准确度和准确度之外的其它指标两方面来对传统推荐系统评价指标进行综述，总结了不同准确性度量指标(如预测准确度、分类准确度等)及它们各自的优缺点，介绍了多样性、覆盖率、流行性和意外性等其它指标，指出这些指标存在的缺陷及未来改进方向. 但未涉及到对数据集和评价方法的综述，所以其综述内容也不够完善. 文献[22]专门对推荐系统中偶然性评价指标的近来研究情况进行综述，但未对其它评价指标、评价方法及数据集进行综述. 文献[23]从基于误差的度量(如 MAE、RMSE 等)和基于排名的度量(如准确率、召回率)来对评价推荐系统性能常用指标进行综述，但未涉及到其它方面的评价内容. 文献[24]主要对时间感知的推荐系统中现有评价协议(评价方法和评价指标)进行全面综述，并提出该类推荐系统评价条件的形式化描述框架，但未对相关的数据集进行综述，因此该文献所综述内容不够完整.

由上述可知，目前关于推荐系统评价方面的综述内容都还不够全面，而且没有学者专门对基于位置的移动推荐系统效用评价进行综述. 因此，本文将从数据集、评价方法、评价指标这三方面来对基于位置的移动推荐系统效用评价进行分析、比较和总结，

《计算机学报》

希望能为开发者或研究人员选择恰当的评价策略来评估基于位置的移动推荐系统提供参考，并希望能进一步促进移动推荐系统的发展. 其结构安排如下：第 2 节对传统推荐系统和基于位置的移动推荐系统的效用评价进行总结和比较；第 3 节主要概述基于位置的移动推荐系统中常用的数据集；第 4 节详细总结、分析和比较基于位置的移动推荐系统中主流的评价方法；第 5 节综述基于位置的移动推荐系统中通用的和特殊的评价指标，并对它们分类总结；第 6 节提出了基于位置的移动推荐系统四层评价体系原型结构；第 7 节分析目前研究中存在的问题并展望未来研究方向；最后对本文进行总结.

2 效用评价概述

近来，效用评价在国内外得到广泛的关注，而且传统推荐系统效用评价研究比较普遍，研究成果也较多. 但基于位置的移动推荐系统效用评价研究相对晚一些，目前还处于初级阶段，其评价指标、评价方法及数据集都还不完善，且缺少统一的评价体系. 以下对传统推荐系统和基于位置的移动推荐系统在效用评价方面进行概括性总结和比较.

2.1 传统推荐系统效用评价

传统推荐系统目前已经很常见，譬如在线电影推荐系统[25]、在线音乐推荐系统[26]、在线电视节目推荐系统[27]、在线图书推荐系统[28]、在线新闻推荐系统[29]、在线旅游推荐系统[30]、在线软件推荐系统[31]、在线科学引文推荐系统[32-33]以及各类社交网络等. 目前，对这些传统推荐系统的效用评价研究较多，经典文献[19]描述了协同过滤推荐系统中一些基本的评价方法及指标. 文献[34]专门综述了在推荐系统中多样性指标的定义、作用及相关算法等内容，并提出多样性指标未来的研究方向. 文献[35]总结了推荐系统目前主要的评测方法有离线评测、问卷调查和在线评测. 文献[11]使用两个不同大小的 MovieLens 数据集和一个 Jester 数据集来分别测试所提出的快速评价算法. 从这些已有的研究来看，当前传统推荐系统效用评价已取得较多成果，比如成熟的评价指标有准确度、覆盖率、多样性和新颖性等，常用的评价方法有离线评价和用户调查，经常使用的数据集有 MovieLens 数据集、Yahoo!Music 数据集、Netflix 数据集、Epinions 数据集、BookCrossing 数据集和电子商务交易数据集等. 总的来讲，目前传统推荐系统的评价指标及数据集已经非常丰富，评测方法比较成熟，并已经得到广泛应用. 而且，随着研究的不断进行，一些学者或研究机构也提出一些新的评价方法或指标，如文献[36]采用一种集成的评价方法来对四种不同的 CF(Collaborative Filtering)方法(IBCF、UBCF、SVD 和 CBCF)进行评测，即在对不同的准确度指标进行归一化处理后来评估不同的 CF 算法，通过在 MovieLens 数据集上进行实验，结果表明该方法可行且效果良好. 但目前还缺乏评价传统推荐系统的评价体系.

2.2 基于位置的移动推荐系统效用评价

由于平板电脑、智能手机等移动设备广泛地使用，位置感知系统变得越来越普遍，进而使得基于位置的移动推荐系统越来越流行，如组团旅游推荐系统[37]、组群活动推荐系统[38]等. 基于位置的移动推荐系统不仅具有移动互联网位置服务的社会化、本地化和移动性等信息服务特征，而且能根据不同用户的个性化需求进行信息过滤和主动推荐，因此在国内外已经赢得广泛关注，许多知名公司、企业、研究机构及研究人员对此领域已经展开深入研究并获得了许多研究成果和产品，如目前国内外流行的 Facebook、Fourquare、Yelp、Places、Gowalla、Twinkle、大众点评、豆瓣、陌陌、遇见等，以及 Kaminskas 实现的基于位置的音乐推荐系统[39]、Yang 等人构建的面向移动购物环境的位置感知推荐系统[40]、Yin 等人实现的基于位置＋内容感知的推荐系统(LCARS)和 POI 推荐系统等[15]，Colomo-Palaciosa 等人实现的考虑了社会化和上下文感知的移动旅游推荐系统[6]、Fang 等人开发的一种基于移动电话的室内购物移动推荐系统[41]等.

对这些新兴的基于位置的移动推荐系统效用评价，通过查阅最新文献，发现这些文献所使用的评价方法主要有客观评价(如离线评价)、主观评价(如在线评价)和主客观结合评价(如离线和在线相结合评价)，常用的数据集有 Foursquare 数据集、Yelp 数据集、Gowalla 数据集、DoubanEvent 数据集和 Twitter 数据集等，经常使用的评价指标有准确度、覆盖率、实时性和满意度等. 譬如 Oppokhonov 等人通过使用准确率、召回率评价指标，在 Foursquare 和 Gowalla 数据集上对所提出的基于当前位置的 POI 推荐模型(CLB)进行离线评估，实验结果表明该推荐模型的性能优于基于 CF 的 POI 推荐模型，能显著改善 POI 推荐系统性能[42]. 而且，一些学者已提出一些新的评价指标和评价方法，如 Olmo 等人采用所提出的新评价方法和新评价指标来对通用

《计算机学报》

推荐系统模板来进行评价[19]. 但目前仍然缺乏评价基于位置的移动推荐系统的一套评价体系.

2.3　二者效用评价的异同

通过上述分析，我们可以得出基于位置的移动推荐系统与传统推荐系统都缺少统一的评价体系，而且在效用评价策略方面有些是类似的，如某些评价方法和部分评价指标. 但由于基于位置的移动推荐系统更为复杂，所以具体实施评价的细节不一样. 同时，它们在评价时考虑的侧重点也不一样，如基于位置的移动推荐系统不仅侧重于准确度还侧重于实时性，而传统推荐系统则一般侧重于准确度. 并且，二者最为明显区别是所使用的数据集不一样，基于位置的移动推荐系统的数据集里必须含有相关的地理位置信息，而传统推荐系统就未必含有这些位置信息. 它们总体的对比情况如表 1 所示.

表 1　效用评价比较

推荐系统	评价方法	评价指标	数据集
传统推荐系统	离线评价，问卷调查，在线评价.	准确度，覆盖率，多样性，新颖性，满意度等.	MovieLens 数据集，Netflix 数据集，Yahoo!Music 数据集，Epinions 数据集，BookCrossing 数据集等.
基于位置的移动推荐系统	客观评价，主观评价，主客观结合评价.	准确度，覆盖率，实时性，交互性，自适应性等.	Foursquare 数据集，Yelp 数据集，Gowalla 数据集，Twitter 数据集，DoubanEvent 数据集，MobileServices 数据集，GeoLife GPS Trajectories 数据集等.

后续内容将从数据集、评价方法和评价指标三方面来综述和具体分析基于位置的移动推荐系统效用评价近年的研究状况，并对所提出的基于位置的移动推荐系统四层评价体系进行论述.

3　数据集

据理论分析与实验证明，数据集与推荐模型、评价方法及评价指标是相互关联的，不同类型的模型所需数据集、评价方法和评价指标可以不一样[43-47]. 譬如，POI 模型的数据集中必须含有地理位置信息，时空主题模型的数据集中需要时间和空间等信息. 文献[16]中 Yin 等人在 Yelp 和 Foursquare 两个不同数据集上采用离线的方式来评测所设计的 POI 推荐模型(ST-LDA)，并通过准确率指标与其它推荐模型(如 Geo-SAGE)进行比较，其实验结果表明该模型在这两个数据集上所得到的推荐准确率都比其它模型高. 文献[43]中 Liu 等人在含有文本信息和地理坐标等信息的 Twitter US 数据集上对所设计的宏观时空主题模型(MSTTM)进行离线评测，通过计算准确率指标证明该模型在进行 POI 推荐时比其它类似模型(如 USTTM)都高一些.

而且，实验进一步证明不同数据集对模型的适合性不一样，其内容直接影响着模型的评测结果，即一般情况下，同一模型在不同数据集上对同一指标的评价结果是不一样的. 如文献[44]在 Twitter NYC 和 Gowalla NYC 数据集上对同一模型 USTTM 进行离线评测，其实验结果表明在 Gowalla NYC 上得到的准确率高于 Twitter NYC，这说明 Gowalla NYC 数据集比 Twitter NYC 数据集更适合 USTTM 模型. 文献[45]在 Yelp 和 TripAdvisor 数据集上对同一模型 QPSO 进行离线评测，实验结果证明在 Yelp 上得到的准确率高于 TripAdvisor，这说明 Yelp 数据集比 TripAdvisor 数据集更适合 QPSO 模型.

并且，不同模型在不一样的数据集上进行评价，其所得的同一指标数值互有高低. 譬如文献[46]采用 NDCG 指标对所设计的两种模型 GeoMF++和 GeoMF 分别在两个基于位置的社交网络数据集 Gowalla 和 Jiepang 上进行评测，所得到的 NDCG 指标值总是 GeoMF++高于 GeoMF，而且从 NDCG 指标值可得出 Gowalla 数据集更适合这两种模型. 同样，文献[47]采用 Accuracy 指标对所提出的两种模型 EW^4 和 W^4 在含有位置信息的两个数据集 Twitter 和 GeoText 上进行评价，所得到的 Accuracy 指标值都是 EW^4 高于 W^4，这说明 EW^4 优于 W^4. 由此可得，不同推荐模型在不同数据集上所得到同一评价指标数值是不一样的，数据集的适合性在一定程度上决定了指标值的高低.

但以上所有模型都是在研发人员自己设置的评测环境下进行的，其评价结果具有一定的片面性. 由此可得，数据集、推荐模型、评价方法及评价指标彼此之间互相影响，通过优化数据集的内容以及增加数据集的数量可以在一定程度上提高模型或系统的推荐准确率. 而且，数据集在评价基于位置的移动推荐系统中起着非常重要的作用，对其进行综述是非常必要的，这对研究人员或开发者在选择数据集时有一定的参考价值和指导意义.

据查阅，随着移动通信网和移动互联网的发展已经催生了各种移动数据，它涵盖了移动终端、移动网络和应用服务产生的所有数据，数据中除了保存用户信息、项目信息、用户与项目间关联等信息外，

《计算机学报》

最重要的还存储了相关的地理位置信息，这是评价基于位置的移动推荐系统所需数据集中必不可少的内容. 而且，与传统网络数据相比，基于位置的移动数据具有许多显著的特性（如时效性强、位置可靠、类型多样以及数量庞大等），这些特性有利于基于位置的移动推荐服务的研究与发展. 但根据查阅，当前在基于位置的移动推荐系统中还没有公开可用的标准数据集，为了评估该类型的移动推荐系统，国内外许多研究者或机构已经收集了志愿者的真实数据或取得授权的用户数据[48]以及模拟生成的数据[49]. 真实数据集和用户数据集主要是由电子商务网站和社交网站等移动网络站点来获取的用户与项目交互数据，而模拟数据集一般根据一些预定的规则来生成或基于部分真实数据进行扩展生成.

当前，用于评价基于位置的移动推荐系统的数据集有 Foursquare 数据集、Yelp 数据集、Gowalla 数据集、Twitter 数据集、DoubanEvent 数据集、Jiepang 数据集、Reality 数据集、GeoLife 数据集、GPS 轨迹数据集、MobileServices 模拟数据集、中国移动应用商城数据集、MIT 数据集、洛桑数据集、Nodobo 数据集以及 WorldPop 数据集等[50-52]，这些数据集都是公开可用但非标准的. 下面主要介绍几种国内外典型且常用的含有位置信息的数据集.

(1) Foursquare 数据集. Foursquare 是当前流行的一个在线 LBSN（Location-Based Social Network），它为开发人员提供一个公开可用的 LBSN 数据. 截至 2010 年 11 月，Foursquare 拥有 2000 万多个用户和约 3500 万个签到记录. 该网站本身没提供一个公共 API（Application Programming Interface）来访问用户的签到数据，但它为用户提供一种替代方式，即用户可以将他们的 Twitter 帐户与 Foursquare 连接起来，并将签到消息作为 tweets 共享到 Twitter[53]. 文献[54]将搜集到的 Foursquare 数据划分为四类：第一类是用户基本信息，即用户 ID、姓名和住址等；第二类是场地基本信息，即场地 ID、名称、地址、GPS 坐标及其类别等；第三类是用户历史信息，由用户在系统中留下的所有点评 tips 来表示，且每个点评都与场地 ID、评论和时间戳联系在一起；第四类是用户社交关系信息，由用户 ID 组成. 目前，该数据集已成为国内外使用最为频繁的 LBSN 数据集[55-56].

(2) Yelp 数据集. 该数据集包含了来自 4 个国家 10 个城市的 366 000 名用户和 61 000 个 POI，包括英国的爱丁堡、德国的卡尔斯鲁厄、加拿大的蒙特利尔和滑铁卢、美国的匹兹堡、夏洛特、菲尼克斯、拉斯维加斯、香槟和麦迪逊. 这里约有 157 万的签到记录，并且每个签到记录存储了用户的 ID、POI 的 ID、POI 的位置、POI 的内容和签到日期. 其社交网络中的每条记录都保存了用户 ID、好友 ID 信息，即社交关系信息，但此数据集不包含确切的签到时间，只提供粗略的签到日期[16,57]. 文献[58]使用该数据集对所提出的空间主题模型（ST）进行评测，实验所得到的平均精准度指标比地理主题模型（GTM）至少提高 300%.

(3) Gowalla 数据集. Gowalla 是一个类似于 Foursquare 的 LBSN，用户通过该平台可以与朋友、家人分享所见所闻，包括发现新的地方、活动、旅行线路等信息. 目前，Gowalla 淡化签到功能，并向旅游和说故事的理念方向发展. Gowalla 数据集也是当前常用来评价基于位置的移动推荐模型的 LBSN 数据集，它包含 196 591 名移动用户在 2009 年 2 月至 2010 年 10 月期间，对 1280 969 个 POI 的 6 442 890 条签到记录，以及 950 327 条用户社交关系[59]. 文献[60]使用该数据集对所提出的具有多中心高斯模型的融合矩阵分解框架（FMFMGM）进行离线评价，实验结果表明 FMFMGM 的性能优于多中心高斯模型（MGM）至少 50%.

(4) DoubanEvent 数据集. DoubanEvent 是目前中国最大的一个在线 EBSN（Event-Based Social Network），通过该网站用户可以发布和参加各种社会活动. 关于豆瓣活动，它是由用户指定某个活动在何时何地发生，其他用户可以通过在线签到方式来表达加入该活动的意图. 此数据集截至目前已经包括 100 000 个用户、300 000 个活动和 3 500 000 个签到，而且大多数签到记录位于中国四大城市：北京、上海、广州和深圳[61]. 它采集数据时记录的信息主要有：①用户信息，包括用户 ID、用户名和用户所在城市；②活动信息，包括活动 ID、活动名称、活动经纬度、活动总结与活动分类；③用户反馈信息，包括用户 ID 和活动 ID；④用户社交关系，包括多个用户 ID[62]. 文献[15]使用该数据集来对所提出的 LCA-LDA 模型进行离线评价，所得的评价指标 Recall 高于 LDA、IKNN 及 USG 等其它基准模型.

综合上述内容，表 2 列举了目前常用数据集 Foursquare、Yelp、Gowalla 和 DoubanEvent 的使用情况. 并且，从表 2 可以看出，这四种数据集都适用于离线评价基于位置的移动推荐模型或系统，且所适用的评价指标为 Recall、Precision 及 NDCG 等.

表 2　常用数据集使用情况

数据集	文献	用户数	项目数	签到数	POI 数	社交关系数	推荐模型	评价方法	评价指标
Foursquare	文献[42]	1034	—	204428	16365	—	CLB	离线评价	Recall，Precision
	文献[55]	11326	—	1385223	182968	47164	GSBPR	离线评价	Precision，Recall，NDCG，MAP
	文献[56]	11326	—	1385223	182968	47164	iGSLR	离线评价	Precision，Recall
Yelp	文献[16]	366000	—	1569264	61000	—	ST-LDA	离线评价	Precision
	文献[55]	70817	—	335022	15579	303032	GSBPR	离线评价	Precision，Recall，NDCG，MAP
	文献[57]	30367	—	146456	25728	—	DCPR	离线评价	Recall，Precision，$F1$-score
Gowalla	文献[43]	5363	—	327198	4500	—	USTTM	离线评价	Accuracy，Perplexity
	文献[59]	196591	—	6442890	1280969	950327	CoRe	离线评价	Precision，Recall，Running time
	文献[60]	53944	—	4128714	367149	306958	FMFMGM	离线评价	Precision，Recall
DoubanEvent	文献[15]	100000	300000	3500000	—	—	LCA-LDA	离线评价	Recall，$F1$-score
	文献[61]	100000	300000	3500000	—	—	MLTRS	离线评价	MAP，Accuracy
	文献[62]	100553	19708	1596235	5202	1163403	GEM	离线评价，在线评价	Accuracy，Processing Time

总的来讲，当前用于评价基于位置的移动推荐系统所需的数据集还没有特定的，只要数据集中包含有用户以前的历史位置信息即可. 特别地，目前 Foursquare、Yelp、Gowalla 和 DoubanEvent 数据集是比较稳定可靠的，它们已成为国内外研究者用来评价基于位置的移动推荐系统或模型的主流数据集. 而且，这些数据集与推荐模型、评价方法及评价指标相互联系且互相影响，数据集的具体内容和数量将直接影响推荐系统的评价结果.

4　评价方法

通过查阅大量最新文献，我们发现传统推荐系统效用评价方法一般划分为离线评价、在线评价和问卷调查，而基于位置的移动推荐系统效用评价方法目前主要有四种，即在线评价、问卷调查、离线评价和主客观结合评价. 其中，在线评价需要用户亲自参与系统的评测，而且人力、物力和时间成本需求较高. 问卷调查和在线评价一样也需要用户参与，但比在线评价风险性低一些. 由于在线评价和问卷调查主要由用户当时的主观臆断来决定，只是具体实施细节不一样，因此本文将二者归于主观评价方法. 离线评价无需真实用户参与，没有用户的主观判断在里面，并能以较低的成本对大量推荐算法或模型进行有效评估，它由算法或模型本身及硬件环境来客观决定，因此本文将其称为客观评价方法. 主客观结合评价方法则是主观评价和客观评价的有机结合，它能充分利用主观评价和客观评价二者的优点来评测推荐系统，进而获得更好更可靠的评测结果，而且该方法是目前较为新颖的一种方法.

4.1　主观评价方法

主观评价是目前基于位置的移动推荐系统中常用的一种评测方法. 尽管它会受到用户性格、年龄、心情、个性、健康甚至敏感性的影响，但它能直接体现用户使用推荐系统的真实感受，评测结果具有一定的参考价值，所以已成为现在产业界主流的一种评测方法. 当前基于位置的移动推荐系统常用的主观评价方法有两种，即在线评价和问卷调查.

(1) 在线评价

在线评价实际上是设计在线用户实验，根据用户在线实时反馈结果来评测推荐系统的性能，其基本评测过程如图 2 所示.

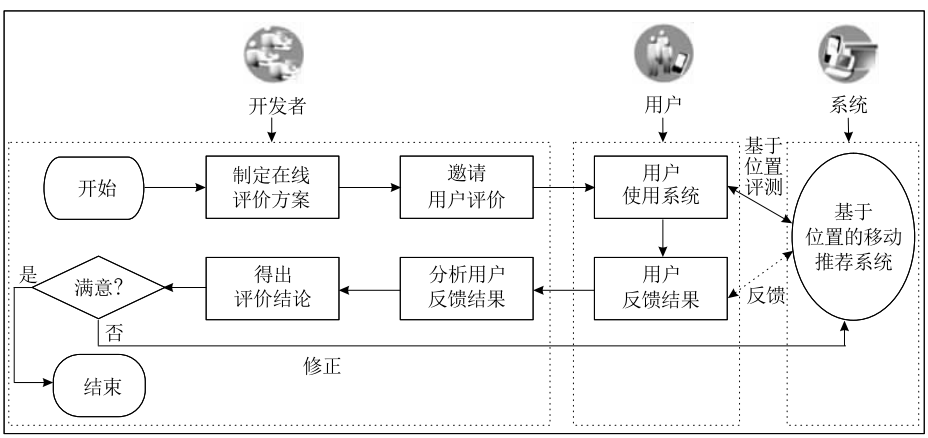

图 2　在线评价基本过程

目前常用的在线评价方法主要是 A/B 测试，其基本核心思想是：① 多个方案并行测试；② 每个方案只有一个变量不同；③ 以某种规则优胜劣汰[20]. 从第 2 点可以看出 A/B 测试只适合单变量的应用范围，所以当 A/B 测试应用于基于位置的移动推荐系统的评价中就只对应唯一变量——推荐算法或模型. 同时，A/B 测试名字中虽然只包含 A、B 两个方案，但并不是只用于比较两个不同方案优劣，实际上可以设计多个方案来进行测试，如 A、B、C、D 测试.

Takeuchi 等人[63]最早设计了在移动推荐中考虑位置的推荐系统——CityVoyager，它用来为用户推荐在东京购物的地方. 该系统设计在当时是创新的，但只邀请了两个用户来评估此推荐系统，所以评估结果不具有普遍性和可靠性. Bao 等人[54]实现了一个基于位置和偏好感知的移动推荐系统，该推荐系统不仅方便人们出行他们所在的城市，也适合出行新的城市. 他们通过邀请纽约和洛杉矶两个城市的 200 名用户来对该推荐系统进行在线评测，结果证明该推荐系统效率快，用户满意度高. Kaminskas 等人[39]采用用户主观在线评测方式来评估音乐与 POIs 的真实相关性，进而实现基于位置的音乐推荐. 由于用户在一次评测期间不能进行大量的判断，所以他们限制了数据集的大小，只包括来自欧洲 17 个大城市旅游景点的 25 个著名的 POIs，进而保证用户每一次评估是正确合理的. Wu 等人[64]邀请零售店主、手机用户以及软件开发专家采用在线的方式对所实现的移动短消息广告推荐系统(LARSMA)进行其有用性、易用性和隐私等方面的评测，并收集他们的反馈信息然后对 LARSMA 进行改进和完善.

Hagen 等人[65]对 421 名真实游客是否使用基于位置的旅游推荐系统来寻找旅游目的地进行现场测试. 其中 142 名年龄为 50 岁左右的游客使用一个旅游推荐系统(DTG Explorer)，137 名年龄为 48 岁左右的游客使用另一个旅游推荐系统(DTG Planner)，而 142 名 54 岁左右的游客则通过传统信息手段来寻找目的地，但携带了 GPS 记录器来跟踪他们的行踪. 通过对这三组人群进行现场测试，实验结果表明，与 GPS 记录器跟踪的人群相比，两个移动推荐系统通过提供一种非常类似于传统导游的服务而获得很高的接受度，而且游客发现的景点数多了四倍，在景点观赏时间也延长两倍，测试结果证明游客对两个移动推荐系统的满意度很高. Arana-Llanes 等人[66]使用不同技术组合形成一种以用户为中心的基于位置的移动推荐系统评价方法，并使用该方法对所实现的上下文移动推荐系统(CARS)的可信度、满意度进行测试. Noguera 等人[67]招募 27 名测试人员(其中男性 19 名，女性 8 名，年龄在 24 岁到 48 岁之间，平均年龄为 30 岁)来对所实现的三维上下文感知的移动推荐系统进行可用性在线评测，并取得好的评测结果.

由此可见，在线评价方法因具有及时响应、用户交互性强、能直接获得用户对基于位置的移动推荐系统的满意度、惊喜度和新颖性等优点而被一些研究人员采用. 但由于从设计实验到实施实验的整个过程的成本较高，且周期较长，所以大部分研究者较少使用在线评价方法来评测基于位置的移动推荐系统，但产业界使用相对多些.

(2) 问卷调查

问卷调查方式也需要真实用户参与到基于位置的移动推荐系统的评测，在推荐系统上完成相应的任务，通过他们的行为记录，并与他们进行恰当沟通，可以获得关于移动推荐系统有关推荐质量方面的宝贵信息. 当然，参与评测的用户不能随便选择，应尽量保证评测用户的分布与真实情况相接近(如男女各半)，年龄段、活跃度、敏感度与现实要尽量相当. 其实施的基本过程如图 3 所示.

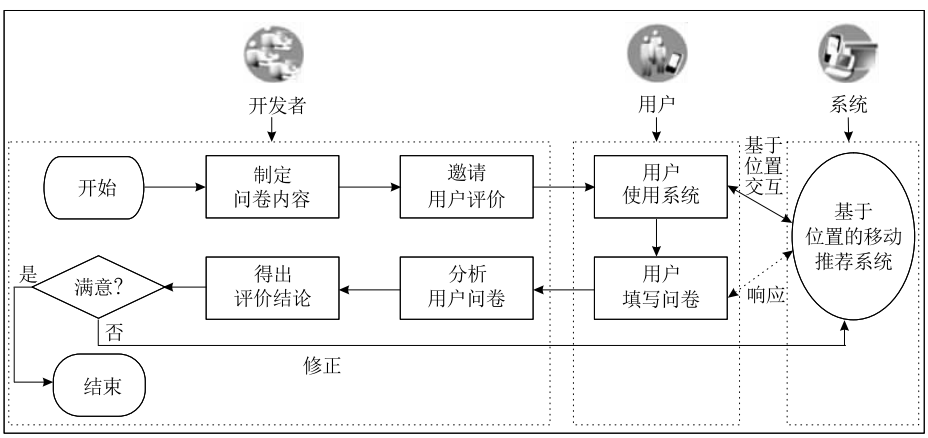

图 3 问卷调查基本过程

Braunhofer 等人[68]通过对 30 名真实用户且年龄在 18～34 岁的学生、同行、同事和运动员进行问卷调查，通过回答 10 个调查问卷来对所实现的移动 POI 推荐系统(STS)进行可用性评估和分析，评估结果表明该系统的总体可用性处于“好”和“优”之间，满足了当前用户的需求. Chen[69]使用问卷调查和深入访谈的方式对所实现的基于位置的智能手机图书推荐系统(IMLBRS)进行评测，并对不同类型参与者设置不一样的问题，评测结果令人满意. Baltrunas 等人[14]利用一些上下文感知的知识来评测移动推荐系统 ReRex 的用户满意度，并用 16 个可行的问题来进行问卷调查. 实验结果表明，虽然问卷调查的评测人员数量有限，但与没有上下文感知因素的移动推荐系统相比，所开发的此移动推荐系统获得更高的用户满意度. Horowitz 等人[70]从用户感知的有用性和易用性等多方面出发，采用问卷调查方式来对所开发的上下文感知移动推荐系统(EventAware)进行评估，结果表明 99.3% 用户对本系统表示满意. Meng 等人[71]同样采用问卷调查的方式对所提出的基于 POI 类别的旅程推荐方法进行满意度测试，测试结果表明用户满意度平均达到 80%.

由此可见，用户调查评价方式可以更多地获得用户主观感受指标(如满意度、易用性以及有用性等)，而且相对在线评价风险性偏低. 但其缺点是召集测试用户成本较高，并且测试过程中用户的行为也有可能与真实环境不一致，所以会出现测试过程中收集到的用户行为在真实环境中无法再现的可能性. 虽然问卷调查存在这些缺点，但它是一种比较简单且实用的主观评价方式，所以在基于位置的移动推荐系统评价中经常被采用.

综上所述，主观评价的两种方法即在线评价和用户调查各有其优缺点，但它们都能有效反映出真实用户对推荐系统经实际体验后的感受，如满意度、稳定性、有用性和易用性等，因此二者目前已被产业界、研究机构用来评价基于位置的移动推荐系统. 但它的样本较少、成本太高而且不稳定，并存在一定的主观性、偏见性和危险性，所以如何更经济、更可控、更合理、更可靠地来评价基于位置的移动推荐是主观评价未来研究关键问题.

4.2　客观评价方法

目前基于位置的移动推荐系统主流客观评价方法是离线评价. 所谓的离线评价是指根据待评价的基于位置的移动推荐系统在测试数据集上的表现，然后依据计算出来的评价指标来衡量推荐系统的质量，其基本过程如图 4 所示.

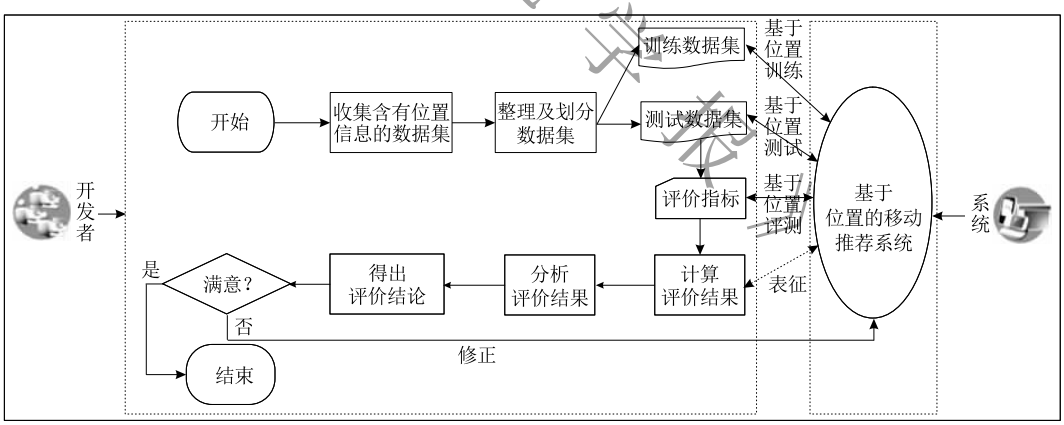

图 4　客观离线评价基本过程

离线评价方法其实源于机器学习和信息检索，即把数据集划分为训练集和测试集，在训练集上构建推荐模型，然后在测试集上测试推荐模型的性能，并且常采用的离线测试方法是 k-折交叉验证[21]. 总的来看，离线评价主要有三个环节：数据的收集、数据集的划分以及评价指标的选择. 其中，收集的用户数据应该与用户在实际移动推荐系统上的数据要足够相似，即以后用于离线实验的数据集要与未来部署后的系统所产生的数据尽量接近. 目前数据划分方式常用的是随机划分，而离线评价指标主要有准确率、召回率、实时性和多样性等.

(1) 基于地点的评价

基于地点的客观评价使用的数据集存储了用户以前一些与时序无关的地点位置数据，尤其是用户感兴趣的 POI 位置信息[72-73]. Yin 等人在文献[15]中首先设置两类待查询城市的地点集合，一类是对用户来说是陌生的城市，另一个类是用户所居住城市. 然后将用户历史活动数据划分为一个训练集和一个测试集，并且针对这两类待查询城市集合采用了两种不同的划分策略. 对于第一类城市集合，文献

[15]选择非本地城市用户访问过的所有空间项目作为测试集,使用其它城市用户的历史活动作为训练集;对于第二类城市集合则随机选择 20%本地用户访问的空间项目作为测试集,使用其余用户个人历史活动作为训练集.接着,根据所设计的数据划分方案来对所提出的推荐模型(LCA-LDA)和推荐算法(TA)进行具体训练和测试,并采用召回率指标来评测其推荐优劣且获得较理想结果. Yin 等人还在文献[16]中采用类似的评测方法来对所提出的时空主题模型(ST-LDA)和推荐算法(AP)进行训练和测试,并采用准确率指标来评测其推荐优劣且得到较高的准确率.同样,Yin 等人在文献[74]仍采用类似的评测方法来对所提出的推荐模型(JIM)进行训练和测试,也采用准确率指标来评估其推荐优劣且得到更准确的推荐结果.

除了 Yin 等人研究团队外,Chen 等人[75]为了评估所提出的在 POI 推荐中能自动调整权重算法的性能,通过在五个真实的数据集上进行离线评测,并计算出准确率、召回率和多样性三个指标,实验数据表明这种方法可以很好地平衡位置相关性和多样性两个因素,并能为用户提供更好的 POI 推荐. Gao 等人[76]通过在 Foursquare 数据集上对所提出的融合地理-社会信息的个性化 POI 推荐模型(GeoEISo)进行离线测试,并通过计算准确率和召回率指标进行客观衡量,实验结果表明 GeoEISo 比目前其它 POI 推荐模型具有更好的性能. Si 等人[77]在 Foursquare 和 Gowalla 数据集上采用离线的方式对所提出的自适应 POI 推荐方法(CTF-ARA)进行评测,实验结果表明,相比于其它 POI 推荐方法,CTF-ARA 提高了 POI 推荐的准确率和召回率.同样,Ying 等人[78]采用准确率和召回率评价指标在 Foursquare 数据集上对所实现的时间感知 POI 推荐系统进行离线评测,评测结果表明该系统具有好的性能. Sang 等人[79]在超过 100 K 实际签到记录和 20 K POIs 离线评测中提出概率推荐方法,并通过评价指标准确度和 MAE 来权衡,实验结果证明了该方法的有效性. Yuan 等人[47]采用基于地点的离线评价方式来评价为移动用推荐所建立非参数贝叶斯模型 EW^4,并在 Twitter 和 GeoText 两个真实数据集上以预测精度(ACC)和平均误差距离(DIS)作为量化指标来与目前主流的模型(如 KL,TR,ST,HG,W^4,$EW^4_{\backslash T}$,$EW^4_{\backslash w}$)进行对比实验,实验结果表明 EW^4 能更好地发现移动用户感兴趣的时空主题.

由此可见,基于地点的客观离线评价方法由于其简便、可靠等优点已经广泛应用于基于位置的移动推荐系统中,尤其是在评价目前流行的 POI 移动推荐系统中,而且数据集中的用户历史地点信息对评估所构建的基于位置的移动推荐模型起到非常重要的作用,它直接影响着推荐系统的最终推荐效果.但由于所使用的测试数据集有限,其评测结果具有一定的局限性,并且无用户亲自参与评测,因而不能很好地体现真实用户实际感受.

(2)基于轨迹的评价

基于轨迹的客观评价所使用的数据集保存了用户一些与时序有关的 GPS 位置数据,这些数据记录了用户的历史轨迹信息[80]. Lee 等人[81]使用 TrafficPulse 和 GeoLife 中的历史轨迹数据,来离线对所提出的一种基于简化了的移动用户 GPS 轨迹数据的拼车推荐方法进行测试,实验结果表明该拼车方法能有效处理高容量且详细的移动轨迹数据,它既能减少数据存储空间又能降低计算成本,且效率较高. Zheng 等人[82]将用户的历史位置和活动建模作为用户-位置-活动评级张量,使用 119 个用户大约 2.5 年的真实 GPS 数据集来客观离线评估所提出的三种基于协同过滤的算法(CLAF,PCLAF 和 RPCLAF),并计算出相应的均方根误差(RMSE)和 ROC 曲线下的面积(AUC)度量指标,测试数据表明三种算法优于目前主流算法(如 AF),并且 RPCLAF 算法优于其它两个算法,能实现更准确的个性化移动推荐. Yoon 等人[83]使用 125 名用户从 2007 年 5 月到 2009 年 8 月的 17745 GPS 轨迹数据来离线评估所提出的旅程推荐方法,评估结果表明该推荐方法比其它方法更为有效. Shafique 等人[84]根据用户在兴趣区域(ROI)的停留时间、频率和转折点等历史轨迹,提出一个贪婪近似方法为每个 ROI 找到最流行的路径推荐给移动用户,并在 Flickr 数据集上进行离线测试以证明该方法的有效性. Rodríguez-Hernández 等人[85]使用 DataGen-CARS 生成的数据,采用离线的方式对所构建的基于轨迹和用户偏好的博物馆艺术品推荐系统进行测试,并计算出平均绝对误差(MAE)评价指标值,其值较小且推荐准确度较高,这很好地满足了游客的需求.

由此可见,基于轨迹的客观离线评价方法同样以其简单、经济等优点已广泛应用于基于位置的移

《计算机学报》

动推荐系统中，且主要用在面向大量 GPS 轨迹的移动推荐系统中，用户历史位置的轨迹信息在评估所构建的基于位置的移动推荐模型占非常重要的地位，其质量优良直接影响着移动推荐模型的性能好坏. 但该方法仍然无用户真正参与系统评测，因而不能直接获得用户真实评测结果.

综上所述，客观离线评价是目前基于位置的移动推荐推荐系统中一种重要且常用的评测方法. 虽然它不能直接体现用户的真实感受结果，但它采用定量的方式来计算推荐系统的评价指标，通过指标大小来衡量推荐系统优劣. 所以，客观评价实质上是对主观评价的一种预测，其稳定性、可靠性及可预测性使得它成为当前基于位置的移动推荐系统中主流的评价方法. 很明显，相对于主观评价来说，客观评价最大的优点是不需要真实用户的参与，因此该方法更简便且更经济，一旦选定所需要的数据集，只需要将待评测的基于位置的移动推荐模型或系统放于此数据集上运行，便能快捷地评测各种不同推荐算法的性能，帮助设计者过滤掉不恰当的算法，从而确保最终用于推荐的算法是最佳的. 所以该方法已普遍应用于基于位置的移动推荐系统的评测，但它仍存在一些问题，这些问题都是后续需要研究的内容，如很难获得商家实际所关注的各种商业指标，具体有如下几点[35]：

(1) 数据集的稀疏性限制其使用范围. 客观离线评价只能评价用户“已经评分”过的项目，它不能使用一个不包含用户任何历史记录的数据集，来评价基于位置的移动推荐系统对该用户的推荐结果.

(2) 评价结果的客观性不能准确地反映出用户真实感受. 由于用户自我感觉的主观性，所以不管离线评测的结果如何，都不能得出用户是否真正喜欢该移动推荐系统推荐的结果.

(3) 客观离线评价所使用的测试数据集数量及内容有限，不能包括所有相关数据集的所有数据内容，因此所得到的结果具有一定的片面性.

(4) 客观离线评价只能测评推荐算法的评分预测的准确度或 Top-N 推荐准确率等客观指标，难以准确测评出满意度、惊喜度和交互性等主观性指标.

(5) 客观离线评价难以找到离线评价指标和在线实时反馈之间的关联关系. 由于离线评价和在线评价方法的本质不同，所以很难获得两者之间的关联关系.

(6) 客观离线评价很难模拟在线环境中上下文条件，如上下文位置、上下文时间等，从而导致评价结果具有一定的局限性.

4.3　主客观结合评价方法

由于主观评价和客观评价各有优缺点，所以相对二者来讲，能较好地评测基于位置的移动推荐系统的方法是将两者有机地结合起来(即主客观结合评价法)，这样便能充分利用二者的优点并避免它们的缺点，其基本评价过程如图 5 所示.

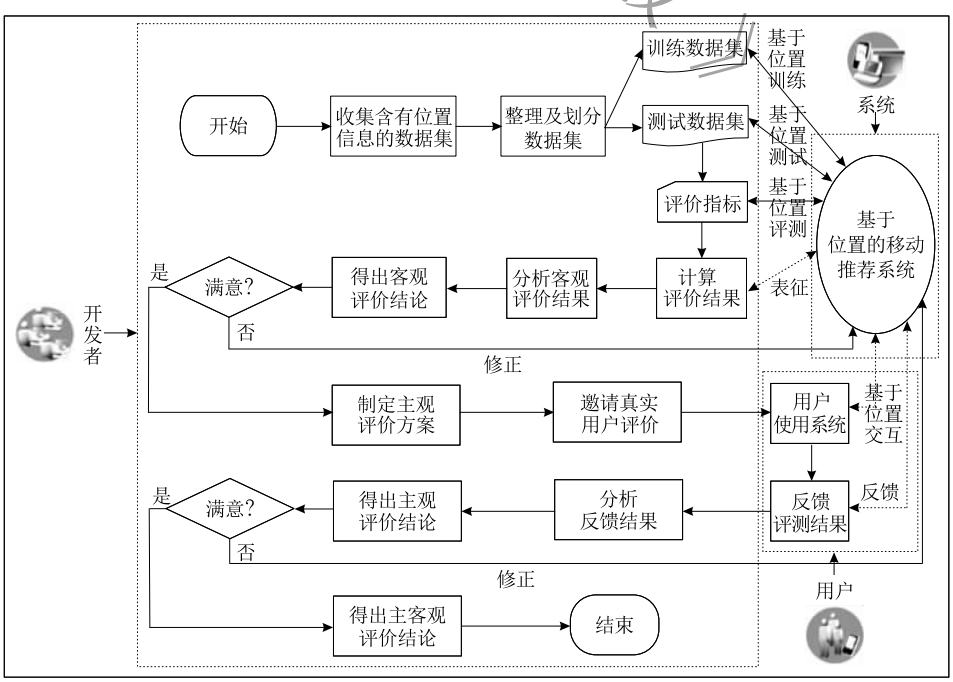

图 5　主客观结合评价基本过程

由图 5 可见，主客观结合评价方法的过程由主观评价过程和客观评价过程组合而成，即先对基于位置的移动推荐系统进行客观评价，当客观评价结果满意后再进行主观评价. 其具体过程如下：

首先按照客观评价方法的过程收集当前推荐系统所需的数据，并将收集到的数据划分为训练集和测试集，使用训练集数据对所设计的基于位置的移动推荐系统进行训练，训练结束后再在测试集上对它进行客观测试，通过计算并分析其相应的评价指标得出该系统的客观评价结论，若结论满足要求则进入主观评价阶段，若不满足要求则需要修正推荐系统模型，再重复上述过程直至满足客观评价的要求.

当推荐系统通过客观评价后便进入主观评价阶段，其评价过程与主观评价方法类似，即开发人员先制定好主观评价方案，然后邀请真实用户来亲自使用已经通过客观评价的基于位置的移动推荐系统，并将体验结果反馈给开发人员，开发人员通过分析用户反馈的结果得出该系统主观评价结论，如结论满意则一次完整的主客观评价结束，若不满意则还需要修正系统模型并重复上述过程直至满意为止. 只有客观评价结果和主观评价结果都满足要求才能得到主客观结合评价的最终结论，至此主客观评价才结束.

文献[86]较早地使用主客观结合评价方式来评测所构建的移动 Web 个性化搜索系统，并取得了好的测试结果. 文献[40]首先通过志愿者的主观评价，来判断所开发的面向移动购物环境的位置感知推荐系统所提供的服务是否满足移动用户的需求，然后在客观测试时要求移动用户对学习得到的每项移动用户偏好进行满意度评价，通过最终的满意度来判断该推荐系统是否能很好地满足移动用户需求，这有效地增强了满意度指标的可靠性. Sang 等人在文献[87]采用主客观结合评价方式对所开发的一个基于本地的移动推荐应用程序进行评估，客观地计算了该推荐系统的 Top-N 准确度和平均绝对误差(MAE)，并邀请 12 名参与者来亲身体验该系统并做出主观评价. Amoretti 在文献[88]首先通过模拟数据和真实数据来分别对所实现的智能移动推荐系统(UTravel)进行准确度评估，主要涉及准确率、召回率和 $F1$ 评价指标的计算，然后采用问卷调查的方式邀请 20 名用户在大约 3 周内对安装在他们手机上的 UTravel 进行可用性主观评价，并获得了满意的评测结果.

综上可得，主客观结合评价的总体思路是，先通过客观离线评价分析比较不同推荐模型在某些指标上的性能，快速过滤掉评测结果不好的推荐模型，并留下较优的候选模型，然后采用在线评价进一步评测和调优候选模型获得更好的推荐结果. 由此可见，主客观结合评价方式能恰当利用主观评价评价和客观评价的优势，使其对基于位置的移动推荐系统评测更准确合理. 它是目前较理想的评价方法，也是比较新颖的一种方法，但其过高的成本使它在具体应用中使用较少，而且目前对于基于位置的移动推荐还没有一种量化的方式来计算主客观综合评价结果，也没有主客观量化的特有评价指标，这些都是后续需要研究的内容.

4.4 评价方法比较

由上述可见，基于位置的移动推荐系统与传统推荐系统所用评价方法相似，但在实施评价的某些细节方面却有所不同，主要区别有以下两点：第一，在使用客观离线评价方法时是否需要考虑数据集中含有位置信息，由于离线评价基于位置的移动推荐系统要求所用的数据集必须含有地理位置数据，因此需要考虑该问题，而离线评价传统推荐系统时则不需要考虑该问题；第二，在使用主观在线评价方法时评价人员是否需要改变当前所处位置，基于位置的移动推荐系统由于需要实时获取用户当前位置来进行推荐，因此为了有效评测其推荐质量，需要评价人员在实施评测的过程中不停地改变位置，但传统推荐系统不一定根据位置进行推荐，所以评价人员可以在某个固定位置实施在线评价.

并且，根据以上所论我们可以得出基于位置的移动推荐系统三种评价方法常用的评价指标、优点及缺点，具体如表 3 所示.

表 3 评价方法比较

评价方法	评价指标	主要优点	主要缺点
主观评价	满意度，惊喜度，信任度，交互性，新颖性，可解释性等.	交互强，响应快，结果真实，更接近实际等.	样本量少，耗时多，成本高，不稳定，用户参与，主观性强，用户偏见高等.
客观评价	准确率，召回率，$F1$，多样性，鲁棒性，实时性，响应时间等.	无真实用户参与，简单快捷，易实施，成本低等.	依赖于测试数据集，计算易过时，算法复杂，难获取真实反馈等.
主客观结合评价	目前主流的客观评价指标及主观评价指标.	评价结果更准确，更符合实际情况等.	实施复杂，耗时最多，成本最高等.

《计算机学报》

为了应对大数据时代的挑战，有些学者提出一些新的评价方法来评测推荐算法，这些方法仍然适合评价基于位置的移动推荐系统. 如朱奕健等人在文献[52]探讨了相似公式背后的数学规律，引入一种新的数据结构，并设计线性时间算法来评测推荐系统. 实验证明该算法能缩短推荐系统评估时间，并有助于提高推荐系统设计和开发的效率. 但现有的这些评价方法要么需要邀请评测人员亲自参与评价（如在线评价），要么需要研发人员手工选择所需的数据集、模型参数、评价指标等（如离线评价），虽然比较简单但缺乏自动评价能力，因此研发一种自动或智能评价方法是必要的，这样不仅能快速得到准确可靠的评测结果，还能有效节约评测时间、人力、物力和财力.

5　评价指标

面对当前大量的推荐算法或模型，如何客观有效地评价推荐系统的优劣是颇具挑战的问题，评价指标恰是衡量推荐系统效用的标准. 而且，不同类型的推荐系统在选择评价指标时考虑的着重点不一样，新类型的推荐系统需要新的评价指标，新的评价指标将促进推荐系统发展. 通常情况下，研发人员需要从多方面、多角度对推荐系统的效用进行综合评价，并采用多种评价指标来进行衡量，在得到这些评价指标结果后，如果结果不理想，应根据具体评价指标分析推荐中哪些环节需要改进，以此来优化推荐系统性能，从而实现更好的推荐效果. 一直以来，我们推荐系统研究团队都是按此方法来对所实现的推荐系统或模型进行评价和优化.

我们团队成员 Liu 等人[89]使用 MAE 和 P@R 两个指标，在 MIT 数据集上采用离线方式对所提出的基于移动用户位置的网络服务推荐方法进行准确度衡量；Wang 等人[90-91]在 Yelp 学术数据集上，采用 MAE、RMSE、Precision、Recall、Efficiency 及 Speedup 等指标，对所提出的新位置用户偏好抽取方案及挖掘算法 MUPNL 进行性能评估；Chen 等人[92-93]在 Twitter 数据集上，除了恰当使用 Precision、Recall、$F1$ 和 MAP 等传统指标，对所提出的基于位置的新闻推荐算法 LP-ESA 和 LP-DSA 进行离线性能评价外，还正确使用实时性、点击率等指标来评价移动新闻推荐系统；Hu 等人[94]在 Movielens 数据集上，采用 Coverage、Diversity 及 NDCG 指标对所提出的偏好融合组推荐方法进行离线客观评价；Lv 等人[29]使用 NewsREEL 数据集来对所设计的混合新闻推荐模型 Fere 进行效用评价，不但采用经典的 Precision、Recall 等指标来衡量其性能，而且针对新闻数据的稀疏性提出了稀疏度评价指标 Sav 用来精确地表征数据的稀疏率，还根据标准差提出了 St 指标用来测量 P@K 和 R@K 的稳定性等等. 不仅我们团队在评价指标方面取得了一定的研究成果，其他团队同样得到了很多研究成果，如 Shani 等人[95]根据图书的作者，通过度量推荐项目与已评分项目之间的距离来评价图书推荐系统的惊喜度. 而且，离线评测所得到的评价指标值在一定程度会受当时所选择的数据集数量及内容影响[45].

以下本文从通用的推荐系统评价指标、特殊的基于位置的移动推荐推荐系统评价指标和评价指标分类总结这三方面来对评价指标进行综述.

5.1　推荐系统评价指标

在众多的评价指标中，准确度是其最重要且最常用的指标，它可以很容易地通过离线评测方式计算出来，现已广泛应用于各类推荐系统中（包括基于位置的移动推荐系统）. 由于准确度具体指标较多，一般将其分为四种类型[85]：预测评分准确度、预测评分关联性、分类准确度和排序准确度. 其中，预测评分准确度含平均绝对误差、平均平方误差和均方根误差等指标，预测评分关联性包含 Pearson 关联、Spearman 关联和 Kendall's Tau 等指标，分类准确度包含准确率、召回率和 $F1$ 等指标，排序准确度包含平均排序分等指标，具体内容如表 4 所示[20].

另一种常用评价指标是多样性，它用来衡量推荐系统为用户推荐项目的差异程度，其目的是满足不同用户的不同需求，提高用户获得满意项目的概率[86]. 还有一种常用的评价指标就是满意度，它是目前流行的一种主观评价标准.

除了上述三种常用的通用评价指标外，目前还存在覆盖率、信任度、惊喜度、新颖性、鲁棒性、流行性、普适性、扩展性、稳定性、保密性、可靠性、可解释性、可访问性以及用户偏好的隐式反馈和部分知识（如 Mean Percentage Ranking、Percentile Ranking、AP Correlation、Expected Discounted Rank Correlation）等评价指标且同程度地应用于各类推荐系统评价[96-100].

并且，以上这些指标都有其适用的场景和推荐算法. 譬如，文献[29]中使用 Recall 指标在 NewsREEL 新闻数据集上对所提出的 Fere 新闻推荐算法

《计算机学报》

表 4 准确度指标

类型	指标	公式	注释	说明	推荐质量
预测评分准确度	平均绝对误差（MAE）	$MAE=\frac{1}{\lvert E^p\rvert}\sum_{(u,a)\in E^p}\lvert r_{ua}-r'_{ua}\rvert$	r_{ua} 为用户 u 对项目 α 的真实评分.	该类指标是计算预测评分和真实评分的差异程度.	值越小，预测准确度越高，推荐质量越好.
	平均平方误差（MSE）	$MSE=\frac{1}{\lvert E^p\rvert}\sum_{(u,a)\in E^p}(r_{ua}-r'_{ua})^2$	r'_{ua} 为用户 u 对项目 α 的预测评分.		
	均方根误差（$RMSE$）	$RMSE=\sqrt{\frac{1}{\lvert E^p\rvert}\sum_{(u,a)\in E^p}(r_{ua}-r'_{ua})^2}$	E^p 为测试集.		
	标准平均绝对误差（$NMAE$）	$NMAE=\frac{MAE}{r_{\max}-r_{\min}}$	$r_{\max}$ 为用户 u 对项目 α 的最高评分. $r_{\min}$ 为用户 u 对项目 α 的最低评分.		
预测评分相关性	Pearson 相关（PCC）	$PCC=\frac{\sum_a(r'_a-\bar{r}')(r_a-\bar{r}')}{\sqrt{\sum_a(r'_a-\bar{r}')^2}\sqrt{\sum_a(r_a-\bar{r}')^2}}$	r_a 为项目 α 的真实评分. r'_a 为项目 α 的预测评分.	该类指标用来衡量预测评分和真实评分之间的相关性.	值越大，预测评分相关性越大，系统预测结果越好，推荐质量越佳.
	Kendall's Tau（τ）	$\tau=\frac{C-D}{C+D}$	C 为正序对的数目. D 为逆序对的数目.		
分类准确度	准确率（$P(L)$）	$P(L)=\frac{1}{M}\sum_u P_u(L)$	L 为系统推荐的项目数. M 为测试用户的数量. N 为系统中项目的总数. B 为用户喜好的项目的总数.	该类指标用来衡量推荐系统能够正确预测用户喜好或不喜好某个项目的能力.	值越大，准确度越高，推荐质量越好.
	召回率（$R(L)$）	$R(L)=\frac{1}{M}\sum_u R_u(L)$			
	准确率提高率（$e_p(L)$）	$e_p(L)=P(L)\frac{MN}{B}$			
	召回率提高率 $e_R(L)$）	$e_R(L)=R(L)\frac{N}{L}$			
	F1 指标（$F_1(L)$）	$F_1(L)=\frac{2P(L)R(L)}{P(L)+R(L)}$			
排序准确度	平均排序分（RS）	$RS_{ua}(L)=\frac{l_{ua}}{L_u}$	L_u 为用户 u 待排序项目数量. l_{ua} 为待预测项目 α 在用户 u 推荐列表中的排名.	该类指标用来度量算法得到的有序推荐列表和用户对项目排序的统一程度.	排序分值越小，系统越趋向于把用户喜好的项目排在前，其推荐质量越佳.
	排序偏差准确率（RBP）	$RBP(p,L)=(1-p)\sum_{i=1}^{L}r_i p^{i-1}$	r_i 为排在第 i 位的项目是否是用户喜好的. p 为用户依次浏览推荐列表下一个项目的概率.		值越大，推荐质量越好.

计算机学报

进行评价，所得到的 Recall@5 平均值高于 PCF-sum 推荐算法 0.831. 文献[57]同样使用 Recall 指标对所设计的 DCPR 推荐算法在位置数据集 Yelp 上进行评测，所得到的 Recall@20 高于 FM 模型约 0.208. 这表明 Recall 指标既适用于评价传统的新闻推荐系统，又适用于评价基于位置的移动系统. 文献[101]使用 NDCG 指标在 Last.fm 数据集上对所提出的动态音乐推荐算法进行评价，所得的 NDCG 值平均比 CKF 算法高 0.021. 文献[46]也使用 NDCG 指标对所设计的 GeoMF++模型在含有位置的数据集 Jiepang 上进行评测，所得到的 NDCG 值总体较 PD 模型高 0.019. 这说明 NDCG 指标同样适合评估传统的音乐推荐系统，也适合评估基于位置的移动系统. 由此可得，Recall、NDCG 等指标是既适用于传统推荐系统评价也适用于基于位置的移动推荐系统评价，因此它们具有通用性和普适性.

5.2 基于位置的移动推荐系统评价指标

据查新，目前基于位置的移动推荐系统本身暂无特有的评价指标，只是现有评价指标在该类推荐系统具体应用中的使用的频率不一样，如准确率、召回率、NDCG 及满意度等指标经常使用，而响应时间、自适应性和交互性等指标使用得相对少一些，但这些指标早已提出并有定性或定量的描述. 下面主要总结在评价基于位置的移动推荐系统时需要侧重考虑的几种特殊评价指标.

(1) 实时性

实时性是基于位置的移动推荐系统一大特征，一般来说包括两方面内容，第一方面是实时地根据用户所在位置为用户推荐喜好的项目，另一方面是实时地将新添加项目推荐给可能喜好的用户. 现实

中许多项目都具有实时性，如新闻、用户所在位置、用户的临时偏好等. 目前流行的移动旅游 APP 可以根据游客当前所在位置实时地为游客推荐可能喜欢的景点、路线、餐饮及酒店等，这样既为游客提供了便利服务，也带动了本地区经济发展[102]. 并且，在线计算时间在一定程度上决定了基于位置的移动推荐系统实时性能，所以对于实时推荐可以考虑基于并行计算的框架，如通过 Mahout 来提升移动推荐系统的实时性[99]. 由此可见，实时性是评价基于位置的移动推荐系统性能一个重要的考核指标.

(2) 响应时间

由于移动网络的特点，需要及时为用户提供满足其个性化需求的移动服务，所以响应时间是衡量基于位置的移动推荐系统重要评价指标，一般情况要求响应时间适中. 若系统响应时间过长，用户就会感到不安和沮丧，进而影响用户对系统的访问率. 若响应时间过短，有时会造成用户加快操作节奏，从而导致错误结果. 目前已有研究人员把响应时间应用于评价基于位置的移动推荐系统中，譬如文献[103]邀请 9 位用户在线评价基于 CoMeR 模型的 Cont-AwareMovie 智能手机电影推荐系统，通过增加电影数据量(从 40 部增加到 200 部)来评测该系统响应时间，最终得到响应时间大约为 0.3 s，该响应时间是普通用户可以接受的，这有效解决了以前随着电影数据库的增大而系统响应时间变长(即电影数据量与响应时间成正比关系)的问题. 文献[7]对所开发的移动电影推荐系统 RecomMetz 在非冷启动和冷启动场景下进行评测，所得到响应时间在两种情况下都比较稳定，其值在 595 ms～650 ms 范围内. 并且，两种情况的平均响应时间分别为 615 ms 和 610 ms，这较好地满足了用户需求，没有对用户带来负面影响. 同样，文献[104]使用响应时间、访问时间等来评测所设计的推荐算法.

(3) 自适应性

自适应是系统在处理和分析数据的过程中，根据被处理数据的数据特征自动调整处理方法、处理顺序、处理参数、边界条件或约束条件，使其与所处理数据的统计分布特征、结构特征相适应，以取得最佳的处理效果的过程. 由于基于位置的移动推荐服务位置不固定，随时可能发生变化，为了能及时根据当前位置为用户推荐所偏好的项目，系统需要具有较好的自适应能力来随时调整用户偏好，这样才能满足移动用户实时需求，因此自适应是基于位置的移动推荐系统另一个重要评价指标. 文献[105]中实现了基于代理的个性化旅游活动推荐系统 Tourist@，它采用基于代理的模块设计方式灵活地构建不同用户活动，并在用户的移动设备中实现位置的实时感知，进而实现在不同位置对用户旅游活动的自适应推荐，并通过分析用户行为(如查询或评测等)来及时更新用户偏好. 文献[104]构建了基于用户位置和偏好的混合推荐算法，该算法通过参数设置能适合于不同的环境，具有较好的自适应能力.

(4) 交互性

对于屏幕尺寸小、输入/输出能力差、终端处理能力弱、无线网络不稳定的移动手机来说，基于位置的移动推荐系统的交互性(如语音、动作及视频等)也是需要考虑的一个主观性评价指标，所有交互性良好的推荐系统能提高用户使用该系统的效率，反之会降低系统的使用率. 在文献[7]中实现了一个多平台的本地移动用户交互界面，并充分利用移动设备的传感能力，实时地向用户推荐电影节目，这有效增强了用户体验，进而提高了用户观看电影的概率. 文献[106]在问卷中对所实现的 iTravel 系统的用户界面交互性效果进行调查，调查结果显示，用户对界面友好性问题的平均分为 3.865(总分为 5 分)，其较低的分数证明该系统的交互性有待进一步提升. 文献[68]对所实现的 STS 系统的用户交互接口进行评测，其简单和直观特性有效满足了用户需求.

(5) 可用性

可用性是基于位置的移动推荐系统的另一个特殊指标，它是指系统对用户来说有效、易学、高效、少错和令人满意的程度，即用户能否用该系统完成他所需完成的任务，以及完成的效率如何，用户的主观感受怎样. 所以，实际上它是从用户角度看到的系统质量，是系统竞争力的核心. 文献[68]邀请 30 位用户对所实现的 POI 系统推荐系统 STS，采用问卷调查方式来进行可用性评价，得到的 SUS(System Usability Scale)平均分为 77.92，远高于 68 的基准分，而且 99%置信区间的 SUS 得分误差为 2.84，所以在 99%置信度下，STS 系统的真实 SUS 得分在 75.08 和 80.76 之间，明显高于基准分. 由此可得该 STS 系统具有较好的可用性. 文献[106]对所构建的基于位置的移动旅游推荐系统(iTravel)，邀请 95 名自愿者到鲁康旅游并亲自对 iTravel 系统进行为期两天的可用性现场评测，旅游结束后请自愿者填写一份含有 16 个问题的调查问卷表，从调查结果来看自愿者对该系统的可用性评分平均为 4.085(总分为 5 分)，较好地满足了用户需求. 由此可见，可用

《计算机学报》

性适用于评价基于位置的移动推荐系统.

5.3 评价指标分类与总结

由上述可见,目前用于评价基于位置的移动推荐系统的指标比较丰富,它不仅覆盖了评价传统推荐系统的通用指标(如准确率、覆盖率及满意度等),还有与自身相关的特殊指标(如实时性、自适应性及响应时间等),并且这些指标中某些指标之间是存在一定的关系(如用户的信任度与系统的鲁棒性、稳定性和实用性是正相关的关系等)[95].根据评价指标的功效本文将上述两类指标进一步划分为四种类型,即以推荐为中心、以用户为中心、以系统为中心和以位置为中心.其中,以推荐为中心的指标主要用来客观表征推荐算法/模型的推荐性能,以用户为中心的指标主要体现用户使用推荐系统的主观感受,以系统为中心的指标主要用来表征推荐系统本身性能,以位置为中心的指标主要体现推荐系统定位效果,其具体划分如表 5 所示[107-109].而且,这四种类型指标一起合作能全面地对基于位置的移动推荐系统进行综合性评估.

表 5 评价指标分类

类型	指标	定量/定性
以推荐为中心	准确度	定量
	覆盖率	定量
	多样性	定量
以用户为中心	信任度	定性
	新颖性	定量
	惊喜度	定量
	满意度	定性
	风险性	定性
	交互性	定性
	可用性	定量/定性
以系统为中心	鲁棒性	定量
	可靠性	定量
	可扩展性	定量/定性
	稳定性	定量
	保密性	定量/定性
	响应时间	定量
	实用性	定量/定性
	易用性	定性
以位置为中心	实时性	定量
	自适应性	定量/定性

但实际上,要如实判断一个基于位置的移动推荐系统是否真正良好,最有意义的指标可能是用户是否能从该推荐系统中获得利益,以及用户是否愿意再次使用该推荐服务,并且该指标已经被提出并称为推荐系统的用户接受[11].而且 Armentano 等人在文献[110]中已设计一个技术接受模型(TAM)来具体处理推荐系统的接受问题.除此之外,对于不同推荐任务选择恰当的评价指标来评测也是非常重要的.文献[52]将推荐系统的任务划分为三类:预测任务、推荐任务和效用最大化任务.并为每个任务分配了一组评价指标,实验证明不同任务选择恰当地评价指标的确是重要的,因为它直接影响到推荐系统的推荐效果.

总的来讲,目前基于位置的移动推荐系统主要指标是准确率、召回率、满意度以及实时性等,但一个真正优秀的基于位置的移动推荐系统不仅仅取决于这些硬性指标,而更多地取决于系统真正能帮助用户完成任务的多少.所以,我们若要对一个基于位置的移动推荐系统进行评价,首先需要清晰地定义该系统自身所要完成的任务,然后针对特定的任务选择适当的指标来对推荐算法进行有效评价.譬如要对基于位置的旅游推荐系统的景点推荐任务进行评价,一般选择准确率和满意度指标来对它所采用的景点推荐算法进行评价,若推荐的准确率高并满足了用户需求,则能有效帮助用户免去查询当前想参观的旅游景点任务,这样便为旅游者节约了时间和精力,从而实现了推荐服务的真正目的.

6 评价体系

依据上述所论,可以得出基于位置的移动推荐系统的总体评价思路是:研发人员根据通过应用场景所构建的推荐模型当前需求,选择恰当的评价方法及评价指标,在含有位置的数据集上对其进行评价,并通过多种可视化方式把所得到的评价结果展示出来.依据该评价思路,本文总结并提出基于位置的移动推荐系统的四层评价体系原型如图 6 所示.

由图 6 可知,该基于位置的移动推荐系统四层评价体系由场景层、模型层、评价层和显示层组成,各层的含义及功能如下.

第一层为场景层,该层处于最底层,用来描述当前应用场景的具体需求内容,如考虑了上下文影响的 Context-aware 活动推荐[10]、考虑了社会化影响的 Social-aware 旅游推荐[6],考虑了位置信息感知的 Location-Aware 新闻推荐[92],考虑了实时情境的 Time-aware POI 推荐[17]及自适应的 POI 推荐[111]等.

第二层为模型层,它是指待评价的基于位置的移动推荐模型.此层根据下层不同的应用场景采用合理的建模方式来构建推荐模型(如基于主题的模型、基于内容的模型、基于规则的模型、基于知识的模型以及基于深度学习模型[112]等),并在一个数据

计算机学报

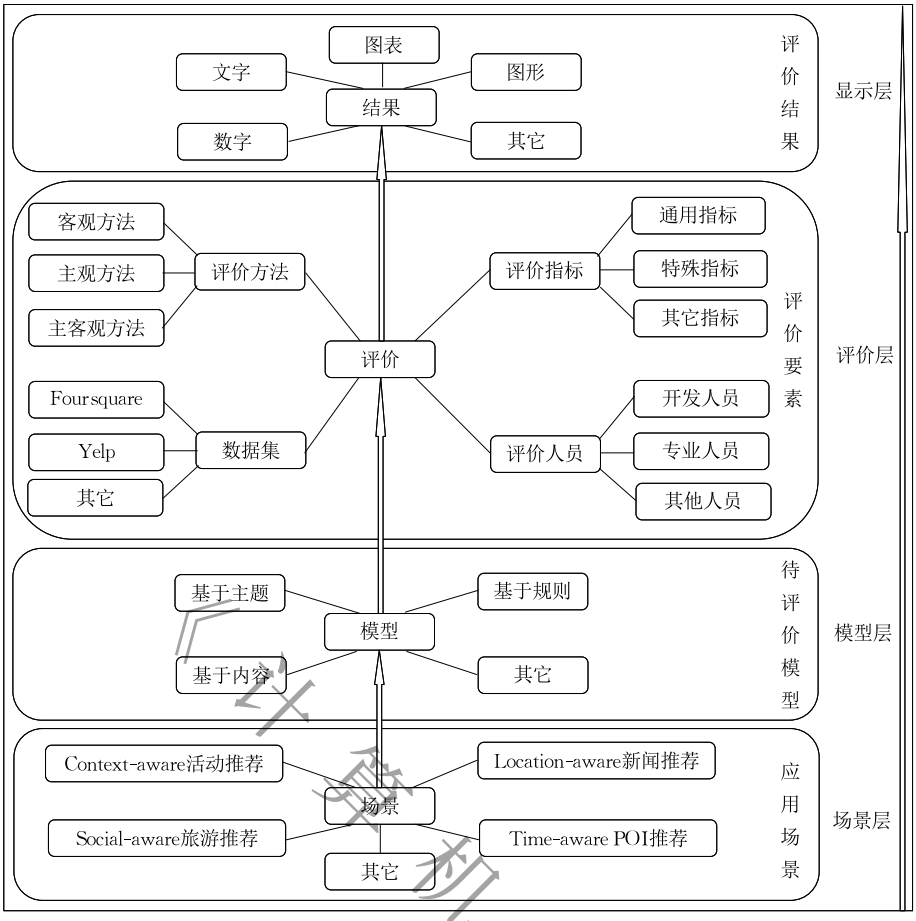

图 6　基于位置的移动推荐系统评价体系原型图

集上完成模型训练，然后等待上层（评价层）来进行评测.

第三层为评价层，它是基于位置的移动推荐系统评价体系的核心层，负责完成对基于位置的移动推荐模型或系统的评价. 本层包括的评价要素有评价人员、数据集、评价方法、评价指标以及其它软硬件等. 其中，评价人员涉及到研发模型或系统的开发人员、专业人员和其他人员（如普通用户、邀请的评测人员等）；数据集包括含有位置信息的 Yelp 数据集、Foursquare 数据集和 DoubanEvent 数据集等，而且，某个模型的评测数据集一般与训练数据集是同一类型；评价方法有主观评价方法、客观评价方法和主客观结合评价方法. 对于评价指标，它是底层模型或系统的评价量化标准. 目前主要用于评价基于位置的移动推荐系统的评价指标有三类，即通用评价指标、特殊评价指标和其它评价指标. 通用评价指标（如准确率、召回率等）面向任何推荐系统，特殊评价指标（如实时性、响应时间等）主要面向基于位置的移动推荐系统，其它评价指标针对某具体的基于位置的移动推荐系统需要而自定义的相关指标. 并且，由于不同应用场景需要的推荐模型不同，需要的评价方法和侧重的评价指标也可能不同，自然需要的评价人员也不一样.

第四层为显示层，该层处于基于位置的移动推荐系统评价体系最上层. 它通过多种可视化方式将其下层（评价层）的评价结果显示出来. 目前评价结果的显示方式多种多样，譬如数字、文字描述、图表展示等，这使得评价结果显示更加生动形象，从而有利于研发人员理解和接受模型的评测结果. 研发人员通过分析这些不同形式的评价结果，进而对基于位置的移动推荐系统性能的优劣作出判决.

根据上述评价体系各层功能，可将其基本应用过程概括为，首先研发人员需要分析底层的应用场景需求，然后根据分析结果建立恰当的推荐模型或系统，并将其作为模型层的内容等待评价；接着研发人员邀请评测人员（包括研发人员本人）来对该模型进行评价，并根据模型特征选择合适的评价方法、评价指标和数据集，而且同一模型上下层所使用的数

据集应该是相同类型;然后评测人员按照研发人员要求对模型实施具体评测,并将评测结果反馈给研发人员;最后研发人员将评测结果通过多种可视化方式进行展示并进行分析,通过分析这些评测结果来判定所设计的推荐模型或系统的优劣.

按照评价体系的应用过程,文献[10]根据活动推荐的应用场景,通过考虑活动上下文影响因素,开发了一套基于 Android 平台的上下文感知的活动推荐系统.为验证该系统推荐活动的优劣,文献[10]采用主观评价方法来对它进行评价.首先,研发人员邀请自愿者安装该活动推荐系统,并让他们使用该系统所提供的推荐服务.当自愿者为期两周的亲身体验结束后,再向其中 16 名自愿者发放调查问卷.问卷主要从 4 个方面(推荐算法、影响上下文感知的方法、用户界面和系统总体效果)的 10 个问题对该系统进行调查,并将调查结果以图的形式展示.从展示结果可得出,该系统能为用户推荐有价值的活动,因此用户对该活动推荐系统的满意度较高.

同样,文献[111]根据自适应 POI 推荐的应用场景需求,结合用户活动和空间特征构建了自适应 POI 推荐模型 APRA-SA,并将训练好的 APRA-SA 作为待评价模型;然后研发人员在评价层中根据该模型的特征及需要,采用客观离线评价方法在 Foursquare 和 Gowalla 两个数据集上通过 Precision 和 Recall 指标来对它进行评测;评测完成后将所得到的评价指标结果以柱状图的形式展示.展示结果表明,所提出的 APRA-SA 算法在 Gowalla 上和 Foursquare 数据集上的 Precision 和 Recall 指标都优于其它基准算法,从而证明该推荐模型在自适应 POI 推荐方面目前是最佳的.

类似地,文献[6-7,17]以及文献[92]等都采用此四层评价体系来对所设计的基于位置的移动推荐模型或系统进行评价,并获得较理想的评测结果.由此证明该评价体系是正确可行的,且具有一定的通用性.

而且,该评价体系各层之间互相影响,即场景层影响模型层,接着模型层影响评价层,最后评价层影响显示层.当基于位置的移动推荐系统具体场景不一样时,其构建的模型自然不一样,所适合的评价要素可能不同,最终得到的评测结果也不同.

譬如文献[43]根据用户签到数据,构建了 USTTM 和 MSTTM 两个时空主题模型,并通过 POI 推荐来评价其优劣.首先使用 Accuracy 指标在 Gowalla NYC 和 Twitter NYC 两个数据集上对所提出的 USTTM 进行离线评测,所得到 POI 推荐的 Accuracy@10 平均值分别约为 0.509 和 0.448,前者高于后者 0.061 左右,所以 Gowalla NYC 更适合评测 USTTM.接着同样使用 Accuracy 指标在 Twitter US 数据集上对 MSTTM 和 USTTM 进行离线评测,所得到 POI 推荐的 Accuracy@10 平均值分别约为 0.702 和 0.346,明显前者高于后者,所以 Twitter US 数据集更适合评测 MSTTM.由此可得,推荐模型不一样,所适合的评价要素(如数据集等)也可能不一样,最后得到的评价结果也不相同.

综上可得,该四层评价体系是正确、合理、可行且通用的.它通过四层架构有机地将应用场景、待评价模型、数据集、评价方法、评价指标等结合起来,各层功能相对独立但又一起协作完成对基于位置的移动推荐系统评价.该评价体系使推荐系统评价过程更为清晰,为开发者或研究人员评价基于位置的移动推荐系统提供了参考.

7 问题及研究方向

近来,尽管国内外研究者或机构对基于位置的移动推荐系统效用评价研究较为普遍,对所使用的数据集、评价方法与评价指标已进行了探究和广泛应用,并取得一些实验性成果,但仍面临着许多问题和难点亟待解决,概括来讲主要有以下这些问题需要进一步研究.

(1)有关基于位置的评价指标方面的问题

①基于位置的专用评价指标

虽然目前应用于基于位置的移动推荐系统评价指标较多,但根据其本身特点和用户实际需求,它缺少特定的专用评价指标,如有关用户定位方面的评测指标、位置与用户关联的评测指标、位置与时间关联的评测指标、用户间位置关联的评测指标以及用户与组团间位置关联的评价指标等.可见,目前评价指标虽然很多但仍不够完善,而且缺乏专用的位置评价指标来有效评估基于位置的推荐模型或系统.因此,针对该系统特征提出新的专用评价指标是一个有潜力的研究方向,而且所提出专用指标需要结合具体移动推荐系统的领域特征,譬如基于位置的新闻推荐系统与基于位置的旅游推荐系统的专用评价指标应该有所不一样,因为两个系统的关注点不同.

②基于位置的动态评价指标

当前应用于基于位置的移动推荐系统的评价指

《计算机学报》

标大多不能实时地对它进行评测，因为这些指标属于静态评价指标. 然而，基于位置的移动推荐系统本身在不同时刻、不同地点的性能不一样，并且为用户所推荐项目的满意度也不同，所以采用恰当的动态指标来对该系统进行实时评价是非常必要的. 该动态评价指标是一类随时间变化的指标函数，可以形式化表示为 $m=f(t,l)$（m 为指标，t 为时间，l 为位置），它能对系统在不同位置向用户所推荐的项目进行实时评测，这样更符合基于位置的移动推荐特征，并能确保其性能，从而能实时地满足用户需求. 因此，对基于位置的移动推荐系统相关的动态评价指标研究是未来一个研究方向.

③ 基于位置的评价指标间关系问题

当前常用于评价基于位置的移动推荐系统的大量指标中某些指标之间存在一定的关系，当一个指标发生变化时另外一个或几个指标也会随之发生变化，譬如准确度与新颖性、多样性成反相关的关系，用户的信任度与系统的鲁棒性、稳定性成正相关的关系，以及在某些时候系统向用户推荐的项目准确度越高但用户的满意度未必就越强等. 通过研究不同指标间的这些关系可以减少该类系统中评价指标的计算量及复杂性，进而提高其工作效率. 所以，研究基于位置的评价指标间所存在的关系是未来一个值得研究的问题. 而且，根据多个评价指标间不同的关系，本文将其进一步细化为如下几个研究点：

(i) 基于位置的评价指标重要性排序问题

为了使基于位置的移动推荐系统所推荐的项目更能满足用户需求，在具体实验时都会采用多种评价指标来对它进行评测，并通过调整参数或者模型使其评测指标结果更理想. 但对具体某个移动推荐系统，它的具体目标不同，对指标的要求及侧重点也不同，所以需要根据自身特征对不同评价指标进行重要性排序，并且排列的顺序不一样其用户满意程度也不相同. 譬如基于位置的移动新闻推荐系统的评价指标按其重要性排序一般为准确率、新颖性、实时性等. 由此可见，基于位置的移动推荐系统指标重要性排序是未来一个值得研究的指标定性分析问题.

(ii) 基于位置的评价指标权重分配问题

当评价基于位置的移动推荐系统时可以采用多种指标来对其进行评价，譬如采用主客观结合方式来评价该类系统时，可以使用准确率和新颖度等指标来一起评测系统. 由于不同指标都有各自的评测结果，系统最终评测结果则需要根据其具体需要来综合这些已得到的指标结果，这时就存在不同指标的评价结果权重分配的问题. 一般来说，对该系统越重要的评价指标其权重分配就越多，否则就越少. 而且，权重分配越恰当将有利于提高推荐系统的用户满意度，否则将影响推荐系统的使用率. 因此，评价指标权重分配问题是基于位置的移动推荐系统未来指标定量研究的一个问题.

(iii) 基于位置的多评价指标组合问题

当使用单个指标来评价基于位置的移动推荐系统时总存在一定的不足性，为了增强评价结果的全面性，人们常常采用多评价指标组合的方式来对其进行评测. 并且，针对不同的基于位置的移动推荐系统它们各自的特殊性，所组合的评价指标不一样，同时侧重点也不同. 为了使所组合的评价指标能较好地表征基于位置的移动推荐系统性能，我们采用合理的融合策略使多种指标恰当地组合为一种评价指标是必要的，这样不仅可以提高系统的开发效率，还能提升系统性能. 为了实现该目标，除了选择评价系统所需的一系列评价指标外，还需重点考虑这些指标的融合策略. 因此，多评价指标组合问题将是未来基于位置的移动推荐系统需研究的一个问题.

(2) 有关基于位置的评价方法方面的问题

① 基于位置的专门评价方法

目前评价基于位置的移动推荐系统方法仍然是常用的客观离线评价、在线主观评价和问卷调查，没有专门的方法来评价基于位置的移动推荐系统，也没特定的方法来评测基于位置的移动推荐模型或推荐算法，同样也没有专门的方法来对位置特征进行有效评测. 因此，对基于位置的专门评价方法研究是一个值得研究的问题，只有恰当的位置评价方法才能有效地评估出基于位置的移动推荐系统性能，才能保证其质量.

② 基于位置的评价算法或评价模型

要准确合理地评价基于位置的移动推荐系统除了使用现有的评价方法外，还可以使用特定的评价算法或模型，使其具有一定的智能性. 该问题侧重于评价算法和模型的研究，而不是直接使用现有的评价方法和评价指标来进行，其难度远高于评价指标的简单应用. 它需要对基于位置的移动推荐系统的评价策略进行特定建模，所得到的评价结果更为准确可靠，这样更能满足用户用户需求. 因此，对基于位置的移动推荐系统评价算法或评价模型研究是未来研究的一个关键问题.

③ 基于位置的评价系统开发

现今已有各种各样的评价系统，譬如移动电话服务评价系统、导游服务评价系统以及网易云客户评价系统等.要想获得用户使用基于位置的移动推荐系统的真实体验，开发一套与其对应的评价系统是十分必要的，这样不仅可以方便获得用户使用该推荐系统的真实感受，而且可以大大降低评价的成本，提高评价的效率，并能取代传统评价方法.因此对该问题的研究也是基于位置的移动推荐系统效用评价未来的一个研究方向.

(3)有关基于位置的数据集方面的问题

①位置的实时准确获取及评价

基于位置的移动推荐系统最重要的是要实时准确地获取用户当前位置.目前实时计量一般以时间单位秒(s)为标准，实时位置数据的采集间隔根据具体需要来指定(如 2 s)，若位置在指定的时间范围内不能及时获取甚至获取错误，则不能为用户提供实时正确的个性化推荐服务.并且，对于获取到的用户当前位置是否正确，目前一般采用问卷调查的方式进行评测，其评测结果具有一定的主观性，如何采用客观有效的评价方法来对其评测是需要深入研究的问题.由此可见，位置的实时准确获取及评价对于基于位置的移动推荐系统来说是必须首先研究的关键问题，该问题直接影响着推荐的准确度.

②基于位置的数据集标准化

对于基于位置的移动推荐系统来说，位置是影响预测结果准确性的一个至关重要因素，尤其在目前流行的 POI 推荐中特别突出，因此该类推荐系统的数据集中必须含有位置数据.目前基于位置的数据集已经比较多，并且大多数是公开可用的，但当前还没有一个统一的标准来有效获取、存储和管理这些数据，所获取数据存在不完整、不一致、冗余、冲突、错误及稀疏等问题，这些问题都可以归结为数据标准化问题.有效解决这些问题的办法是制定出基于位置的数据集标准化方案，为以后采集数据提供一种规范，也为评价基于位置的移动推荐系统提供有力保障.

③用户位置数据的隐私保护

基于位置的移动推荐系统为了给用户在当前位置提供及时准确的推荐，最基本就是要挖掘出移动用户在不同时间的位置、行为及偏好等信息，甚至实时地获取用户活动轨迹.而位置、活动轨迹等都属于用户个人的特殊隐私，它本身包含用户许多敏感信息或由此推出的个人私密信息(如家庭住址、工作地点、休闲场所等)，这些数据信息对用户本人来说是非常保密的.所以如何有效保护用户位置隐私是基于位置的移动推荐系统发展过程中必须解决的问题，这也是评价该推荐系统时需要考虑的一个评测内容，从而确保用户个人安全.

④位置数据的适合性研究

由于不同位置数据所适合的模型不一样，同一模型使用不一样的位置数据集进行评测其结果也有所不同，所以对于不同的基于位置的移动推荐系统，它们所适应的位置数据是不一样的，而且越适合待评价系统的位置数据将有利于提高系统评测的准确度，因此准确度可以作为评价位置适合性的一个评价标准，准确度越高说明该位置数据越适合此类系统，否则就不适合.那么，面对目前大量的位置数据，如何从中选择恰当适合的位置数据来对同一类基于位置的移动推荐系统进行评价是一个值得研究的问题，通过研究找出与之适合的位置数据来更准确地评价基于位置的移动推荐系统或模型，进而提高该类系统推荐的准确度.

(4)其它问题

①基于位置的移动推荐系统评价体系完备性问题

本文仅仅提出了基于位置的移动推荐系统评价体系原型，该四层评价体系原型恰当地将评价基于位置的移动推荐系统所涉及的数据集、评价方法、评价指标以及评价人员等要素有机结合起来，为研究人员提供了一个总体评价思路和参考模板，并弥补了移动推荐系统效用评价研究领域的一个缺口，但它未涉及到评价要素的具体内容(譬如评价指标的选择、数据集内容的筛选等)，其完备性较弱.因此，进一步扩充和完善该基于位置的移动推荐系统四层评价体系是必要的，使其更加完整可行，以便为研究人员提供更为全面的参考，进而提高他们对该类推荐系统的评价效率.

②基于位置的参数敏感性问题

评价基于位置的移动推荐系统时涉及到许多参数，不仅包括推荐模型本身的各种参数还包括相关评价指标中的参数等，对这些参数设置不一样所获得的系统评测结果也不一样.因此，参数调节的好坏在一定程度上直接影响着基于位置的移动推荐系统性能，尤其是基于位置的推荐模型本身的参数所起的作用更大，对其敏感性分析在评价该类推荐系统显的尤为重要，只有这些参数敏感性越低，才能使基于位置的移动推荐系统越稳定，其性能才会越好.因此如何降低参数敏感性使其评测结

《计算机学报》

果更加稳定，这是以后该类移动推荐系统研究的一个方向.

③ 基于位置的上下文因素问题

离线评价基于位置的移动推荐系统重点是评价所设计的推荐模型，推荐模型的性能在一定程度上代表推荐系统的性能，因此对推荐模型的研究是当前推荐系统研究的一个热点. 并且，伴随着上下文因素（譬如位置、时间以及地理区域等）对基于位置的推荐模型影响越来越大，其恰当的设计能有效提升推荐系统性能，因此在评测基于位置的移动推荐系统时，对模型中的这些上下文因素进行评测是必要的，且设计不一样的上下文因素所得到的系统评测结果也是不一样的，所以进一步研究不同上下文因素对基于位置的移动推荐系统性能的影响是未来的一个研究问题.

④ 基于位置的评测结果泛化能力问题

为了保证基于位置的移动推荐系统具有良好的性能，其评测结果应该是稳定可靠的，当它获得了满意的评测结果后才能投之于使用，才能为用户及时推荐喜好的项目. 同时，不论采用主观在线评价还是客观离线评价来评测该类移动推荐系统时，都能得到与之对应的评测结果，这些评测结果在一定程度上代表该系统在当时的性能. 但采用相同的评价方法及指标在不同环境条件下所得到的评价结果并不一定完全相同（尤其在主观评价中最为突出），从而使得评测结果泛化力较弱，不能真正有效地表征该推荐系统性能. 所以，如何使基于位置的移动推荐系统评测结果具有更好的泛化能力是后续有待研究问题之一，其泛化能力强弱将直接影响着推荐系统推荐结果的稳定性以及用户的体验效果.

⑤ 基于位置的移动群组推荐系统评价问题

随着组团活动的越发流行（譬如组团旅游、组团购物等），基于位置的移动群组推荐系统倍受用户青睐和关注[113]. 而群组推荐与个性化推荐方法不能完全等同，因为个性化推荐面向的是具体某个用户，而群组推荐则面向的是一组成员. 由于不同用户对同一项目的兴趣偏好不一样，当对一组用户推荐时需要进行成员兴趣融合或推荐融合，而个性化推荐则不需要这些融合. 于此类似，当对基于位置的移动群组推荐系统进行评价时所采用的评价方法和指标也不完全一样，它侧重考虑与成员位置相关的内容，因此如何采用有效的方式来对它评测是未来研究的一个热点问题，通过研究该问题使此类系统具有更佳的性能，为群组用户推荐更好的服务.

8 总　结

伴随着移动推荐服务的发展，基于位置的移动推荐系统已成为当今最热门且发展速度最快的移动应用程序之一，为了提高它的性能，对其效用评价研究越发重要. 本文从数据集、评价方法和评价指标三个方面出发，分别对近几年来国内外研究者或机构关于基于位置的移动推荐效用评价研究情况进行总结、比较和分析，并提出评价体系原型. 通过对比分析得到如下这几点结论.

（1）专门用来评价基于位置的移动推荐系统的指标、方法等欠缺. 目前用于评价基于位置的移动推荐系统的指标及方法大多来源于通用的评价指标及方法，并且已有的这些静态评价指标中某些指标间存在一定的关系. 但针对基于位置的移动推荐系统的专用评价指标、评价方法却很少，同时对指标间所存在的关系、多指标组合策略等研究也很少. 根据基于位置的移动推荐系统的特殊性，提出并实现其专用的评价指标、评价方法、评价算法、评价模型及评价系统等都是必要的，为实现自动化或智能化评价该类推荐系统奠定了基础.

（2）用于评价基于位置的移动推荐系统所需的数据集不规范，并存在用户隐私泄露问题. 当前虽然已有各种类型的数据集，但根据基于位置的移动推荐系统的特征，该类数据集必须含有用户的个人地理位置数据，当对它进行评测时同样需要这些位置数据，但并不是所有数据集都含有这些数据，同时现有的数据集暂无具体的规范标准. 为了根据用户当前所处的位置实时地向用户准确推荐其所喜欢好项目，需要深度挖掘出用户以前及当前所处的位置，这在一定程度上会暴露用户个人的位置隐私，对用户的工作、生活甚至生命有可能造成不良影响，因此需要采用有效的方式来保护用户个人位置隐私，确保用户个人的安全.

（3）评价结果受多种因素影响，应努力提高评价结果的泛化能力. 基于位置的移动推荐系统评测结果受多种因素影响，包括数据量、数据特征（如数据的稀疏性、多样性）、模型因素（如时间、位置）、评测的环境因素（如软、硬件配置）以及相关参数设置（如模型的超参数）等. 对这些因素的设置不同将得到不一样的评测结果，其中模型因素及评测的环境因素对其评测结果影响较大. 一般来说，良好的推荐模型设计和评测环境设置将使评测结果更好，而且

计算机学报

用户的满意度也会更高.所以对评测基于位置的移动推荐系统所涉及到的各种影响因素进行深入研究是必要的,这不仅有利于提升评测结果的稳定性、可靠性以及重现性,而且能提高评测结果的泛化能力,进而提高系统的使用率及用户的满意度.

(4)数据集、模型、评价方法、评价指标和评价结果相互联系,而且这种联系可以通过四层评价体系恰当地呈现出来.数据集是训练模型必不可少的成分,同样也是评测模型的重要组成部分,不同类型的推荐模型所适应的数据集不一样.不仅如此,数据集与评价方法和评价指标也是相关的.一般来讲,离线评价推荐模型必须要数据集,而且数据集内容不一样,所适合的评价指标可能不一样,其评价结果也不相同.可见,数据集、模型、评价方法、评价指标和评价结果是互相联系且彼此影响的,它们可以通过本文所提出的四层评价体系有机地结合在一起,并且该体系结构能恰当地体现出它们之间的相互关系,这为评价基于位置的移动推荐系统提供了统一参考模板.

综上所述,对基于位置的移动推荐系统效用评价进行深入研究是必要的,同样也是重要的,因为它不仅对其自身有用,还对其它系统效用评价也有价值.最后,希望本综述对相关领域的科研人员和工程技术人员提供一定的参考和帮助.

参 考 文 献

[1] Liu X Z, Ai W, Li H R, et al. Deriving user preferences of mobile apps from their management activities. ACM Transactions on Information Systems, 2017, 35(4): 1-32

[2] He J N, Liu H Y. Mining exploratory behavior to improve mobile app recommendations. ACM Transactions on Information Systems, 2017, 35(4): 1-37

[3] Wang H Y, Li Y C, Guo Y, et al. Understanding the purpose of permission use in mobile apps. ACM Transactions on Information Systems, 2017, 35(4): 1-40

[4] Yao Y, Zhao W X, Wang Y J, et al. Version-aware rating prediction for mobile app recommendation. ACM Transactions on Information Systems, 2017, 35(4): 1-33

[5] Cao D, He X G, Nie L Q, et al. Cross platform app recommendation by jointly modeling ratings and texts. ACM Transactions on Information Systems, 2017, 35(4): 1-27

[6] Colomo-Palacios R, García-Peñalvo F J, Stantchev V, Misra S. Towards a social and context-aware mobile recommendation system for tourism. Pervasive and Mobile Computing, 2017, 38(2): 505-515

[7] Colombo-Mendoza L O, Valencia-García R, Rodríguez-González A. RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes. Expert Systems with Applications, 2015, 9(42): 1202-1222

[8] Zeng J, Li F, Liu H Y, et al. A restaurant recommender system based on user preference and location in mobile environment//Proceedings of the 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). Kumamoto, Japan, 2016: 55-60

[9] Liu Q, Ma H P, Chen E H, Xiong H. A survey of context-aware mobile recommendations. International Journal of Information Technology & Decision Making, 2013, 12(1): 139-172

[10] Herzog D, Wörndl W. Mobile and context-aware event recommender systems. Lecture Notes in Business Information Processing, 2017, 292: 142-163

[11] Zhang F, Gong T, Lee V E, et al. Fast algorithms to evaluate collaborative filtering recommender systems. Knowledge-Based Systems, 2016, 96: 96-103

[12] Liu Shu-Dong, Meng Xiang-Wu. Recommender systems in location-based social networks. Chinese Journal of Computers, 2015, 38(2): 322-336(in Chinese)
(刘树栋,孟祥武.基于位置的社会化网络推荐系统.计算机学报,2015,38(2):322-336)

[13] Az-Zahra H M, Pinandito A, Tolle H. Usability evaluation of mobile application in culinary recommendation system//Proceedings of the 2015 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob). Bandung, Indonesia, 2015: 89-94

[14] Baltrunas L, Ludwig B, Peer S, Ricci F. Context relevance assessment and exploitation in mobile recommender systems. Personal & Ubiquitous Computing, 2012, 16(5): 507-526

[15] Yin H Z, Cui B, Sun Y Z, et al. LCARS: A spatial item recommender system. ACM Transactions on Information Systems, 2014, 32(3): 1-37

[16] Yin H Z, Zhou X F, Cui B, et al. Adapting to user interest drift for POI recommendation. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(10): 2566-2581

[17] Zhang W, Wang J Y. Location and time aware social collaborative retrieval for new successive point-of-interest recommendation//Proceedings of the ACM International on Conference on Information & Knowledge Management. Melbourne, Australia, 2015: 1221-1230

[18] Zhuang J F, Mei T, Hoi S C H, et al. When recommendation meets mobile: Contextual and personalized recommendation on the go//Proceedings of the Ubicomp: Ubiquitous Computing, International Conference. Beijing, China, 2011: 153-162

[19] Olmo F H, Gaudioso E. Evaluation of recommender systems: A new approach. Expert Systems with Applications, 2008, 35(3): 790-804

计算机学报

[20] Zhu Yu-Xiao, Lv Lin-Yuan. Evaluation metrics for recommender systems. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175(in Chinese)
(朱郁筱，吕琳媛. 推荐系统评价指标综述. 电子科技大学学报，2012，41(2)：163-175)

[21] Liu Jian-Guo, Zhou Tao, Guo Qiang, Wang Bing-Hong. Overview of the evaluated algorithms for the personal recommendation systems. Computer Systems and Complexity Science, 2009, 6(3): 1-10(in Chinese)
(刘建国，周涛，郭强，汪秉宏. 个性化推荐系统评价方法综述. 复杂系统与复杂性科学，2009，6(3)：1-10)

[22] Kotkov D, Wang S, Veijalainen J. A survey of serendipity in recommender systems. Knowledge-Based Systems, 2016, 111: 180-192

[23] Shinde S, Potey M M A. Survey on evaluation of recommender systems. International Journal of Engineering and Computer Science, 2015, 4(2): 10351-10355

[24] Campos P G, Díez F, Cantador I. Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. User Modeling and User-Adapted Interaction, 2014, 24(1): 67-119

[25] Carrer-Neto W, Hernández-Alcaraz M L, Valencia-García R, García-Sánchez F. Social knowledge-based recommender system. Application to the movies domain. Expert Systems with Applications, 2012, 39(12): 10990-11000

[26] Tan S, Bu J, Chen C H, He X. Using rich social media information for music recommendation via hypergraph model. ACM Transactions on Multimedia Computing, Communications and Applications, 2011, 7S(1): 1-22

[27] Barragáns-Martínez A B, Costa-Montenegro E, Burguillo J C, et al. A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. Information Sciences, 2010, 180(22): 4290-4311

[28] Núñez-Valdéz E R, Cueva-Lovelle J M, Sanjuán-Martínez O, et al. Implicit feedback techniques on recommender systems applied to electronic books. Computers in Human Behavior, 2012, 28 (4): 1186-1193

[29] Lv P T, Meng X W, Zhang Y J. FeRe: Exploiting influence of multi-dimensional features resided in news domain for recommendation. Information Processing & Management, 2017, 53(5): 1215-1241

[30] Subramaniyaswamy V, Vijayakumar V, Logesh R, Indragandhi V. Intelligent travel recommendation system by mining attributes from community contributed photos. Procedia Computer Science, 2015, 50: 447-455

[31] Bobadilla J, Ortega F, Hernando A, Gutiérrez A. Recommender systems survey. Knowledge-Based Systems, 2013, 46: 109-132

[32] Tang X, Wan X, Zhang X. Cross-language context-aware citation recommendation in scientific articles//Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval. Gold Coast, Australia, 2014: 817-826

[33] Sassi I B, Mellouli S, Yahia S B. Context-aware recommender systems in mobile environment: On the road of future research. Information Systems, 2017, 72: 27-61

[34] Kunaver M, Požrl T. Diversity in recommender systems—A survey. Knowledge-Based Systems, 2017, 123: 154-162

[35] Liu Pan, Chen Min-Gang. Evaluation of personalized recommender systems. Journal of Nanchang University (Natural Science), 2016, 40(2): 143-150(in Chinese)
(刘攀，陈敏刚. 个性化推荐系统评估. 南昌大学学报，2016，40(2)：143-150)

[36] Boddu R S K. An integrated assessment approach to different collaborative filtering algorithms//Proceedings of the 2016 IEEE International Conference on Big Data. Washington, USA, 2016: 3954-3956

[37] Garcia I, Sebastia L, Onaindia E. On the design of individual and group recommender systems for tourism. Expert Systems with Applications, 2011, 38(6): 7683-7692

[38] Gottapua R D, Monangi L V S. Point-of-interest recommender system for social groups. Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical, Procedia Computer Science, 2017, 114: 159-164

[39] Kaminskas M, Ricci F, Schedl M. Location-aware music recommendation using auto-tagging and hybrid matching//Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong, China, 2013: 17-24

[40] Yang W S, Cheng H C, Dia J B. A location-aware recommender system for mobile shopping environments. Expert Systems with Applications, 2008, 34(1): 437-445

[41] Fang B, Liao S Y, Xu K Q, et al. A novel mobile recommender system for indoor shopping. Expert Systems with Applications, 2012, 39(15): 11992-12000

[42] Oppokhonov S, Park S, Ampomah I K E. Current location-based next POI recommendation//Proceedings of the International Conference on Web Intelligence. Leipzig, Germany, 2017: 831-836

[43] Liu Y, Ester M, Qian Y, et al. Microscopic and macroscopic spatio-temporal topic models for check-in data. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(9): 1957-1970

[44] Liu Y, Ester M, Hu B, Cheung D W. Spatio-temporal topic models for check-in data//Proceedings of the IEEE International Conference on Data Mining. Atlantic, USA, 2015: 889-894

[45] Logesh R, Subramaniyaswamy V, Vijayakumar V, et al. A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city. Future Generation Computer Systems, 2018, 83: 653-673

《计算机学报》

[46] Lian D, Zheng K, Ge Y, et al. GeoMF++: Scalable location recommendation via joint geographical modeling and matrix factorization. ACM Transactions on Information Systems, 2018, 36(3): 1-29

[47] Yuan Q, Cong G, Zhao K Q, et al. Who, Where, When, and What: A nonparametric bayesian approach to context-aware recommendation and search for twitter users. ACM Transactions on Information Systems, 2015, 33(1): 1-33

[48] Pálovics R, Benczúr A A, Kocsis L, et al. Exploiting temporal influence in online recommendation//Proceedings of the 8th ACM Conference on Recommender Systems. Foster City, USA, 2014: 273-280

[49] Diaz-Aviles E, Drumond L, Schmidt-Thieme L, et al. Real-time top-*n* recommendation in social stream//Proceedings of the 6th ACM Conference on Recommender Systems. Dublin, Ireland, 2012: 59-66

[50] Yuan Q, Cong G, Ma Z Y, et al. Who, Where, When and What: Discover spatio-temporal topics for twitter users//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013: 605-613

[51] Chen C, Lukasiewicz T, Meng X W, Xu Z H. Location-aware news recommendation using deep localized semantic analysis//Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA 2017). Suzhou, China, 2017: 507-524

[52] Zhu Yi-Jian, Zhang Zheng-Qing, Huang Yi-Qing, et al. Random forest based telco out-calling recommendation system. Computer Science, 2016, 43(11): 557-563(in Chinese)
(朱奕健，张正卿，黄一清等. 基于随机森林模型的电信运营商外呼推荐系统. 计算机科学，2016，43(11)：557-563)

[53] Bauer S, Noulas A, Clark S, Mascolo C. Talking places: Modelling and analysing linguistic content in foursquare//Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing. Amsterdam, Netherlands, 2012: 348-357

[54] Bao J, Zheng Y, Mokbe M F. Location-based and preference-aware recommendation using sparse geo-social networking data//Proceedings of the 20th ACM International Conference on Advances in Geographical Information Systems. Redondo Beach, USA, 2012: 199-208

[55] Gao R, Li J, Du B, et al. Exploiting geo-social correlations to improve pairwise ranking for point-of-interest recommendation. China Communications, 2018, 15(7): 180-201

[56] Zhang J D, Chow C Y. iGSLR: Personalized geo-social location recommendation — A kernel density estimation approach//Proceedings of the ACM Sigspatial International Conference on Advances in Geographic Information Systems. Orlando, USA, 2013: 1-10

[57] Wang F J, Qu Y Z, Zheng L, et al. Deep and broad learning on content-aware POI recommendation//Proceedings of the 2017 IEEE 3rd International Conference on Collaboration and Internet Computing. San Jose, USA, 2017: 369-378

[58] Hu B, Ester M. Spatial topic modeling in online social media for location recommendation//Proceedings of the 7th ACM Conference on Recommender Systems (ACM RecSys'13). Hong Kong, China, 2013: 25-32

[59] Zhang J D, Chow C Y. CoRe: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations. Information Sciences, 2015, 293: 163-181

[60] Cheng C, Yang H, King I, Lyu M R. Fused matrix factorization with geographical and social influence in location-based social networks//Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Canada, 2012: 17-23

[61] Wang S F, Gong M G, Li H L, et al. Memetic algorithm based location and topic aware recommender system. Knowledge-Based Systems, 2017, 131: 125-134

[62] Yin H, Zou L, Nguyen Q V H, et al. Joint event-partner recommendation in event-based social networks//Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE). Paris, France, 2018: 929-940

[63] Takeuchi Y, Sugimoto M. CityVoyager: An outdoor recommendation system based on user location history//Proceedings of the International Conference on Ubiquitous Intelligence and Computing. Springer, Berlin, Heidelberg, 2006: 625-636

[64] Wu C H, Kao S C, Wu C C, Huang S. Location-aware service applied to mobile short message advertising: Design, development, and evaluation. Information Processing and Management, 2015, 51: 625-642

[65] Hagen K T, Gretzel U. Effectiveness of mobile recommender systems for tourist destinations: A user evaluation//Proceedings of the IEEE Workshop on Intelligent Data Acquisition & Advanced Computing Systems: Technology & Application. Dortmund, Germany, 2007: 663-668

[66] Arana-Llanes J Y, Rendón-Miranda J C, González-Serna J G, Alejandres-Sánchez H O. Design and user-centered evaluation of recommender systems for mobile devices-Methodology for user-centered evaluation of context-aware recommender systems//Proceedings of the 2014 International Conference on Computational Science and Computational Intelligence. Las Vegas, USA, 2014: 277-280

[67] Noguera J M, Barranco M J, Segura R J, Martínez L. A mobile 3D-GIS hybrid recommender system for tourism. Information Sciences, 2012, 215(18): 37-52

[68] Braunhofer M, Elahi M, Ricci F. Usability assessment of a context-aware and personality-based mobile recommender system//Proceedings of the International Conference on Electronic Commerce & Web Technologies. Munich, Germany, 2014: 77-88

《计算机学报》

[69] Chen C M. An intelligent mobile location-aware book recommendation system that enhances problem-based learning in libraries. Interactive Learning Environments, 2013, 21(5): 469-495

[70] Horowitz D, Contreras D, Salamó M. EventAware: A mobile recommender system for events. Pattern Recognition Letters, 2018, 105(1): 121-134

[71] Meng X W, Lin X Y, Wang X D. Intention oriented itinerary recommendation by bridging physical trajectories and online social network. KSII Transactions on Internet & Information Systems, 2012, 6(12): 3197-3218

[72] Zhao G S, Qian X M, Kang C. Service rating prediction by exploring social mobile users' geographical locations. IEEE Transactions on Big Data, 2017, 3(1): 67-78

[73] Divya K S, Tech M, Kumar D P. Service rating prediction by exploring social mobile client's geographical places. International Journal of Advanced Engineering Research and Science, 2017, 3: 66-69

[74] Yin H Z, Zhou X F, Shao Y X, et al. Joint modeling of user check-in behaviors for point-of-interest recommendation//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia, 2015: 1631-1640

[75] Chen B L, Yu S B, Tang J, et al. Using function approximation for personalized point-of-interest recommendation. Expert Systems with Applications, 2017, 79: 225-235

[76] Gao R, Li J, Li X F, et al. A personalized point-of-interest recommendation model via fusion of geo-social information. Neurocomputing, 2017, 273: 1-12

[77] Si Y L, Zhang F Z, Liu W Y. CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features. Knowledge-Based Systems, 2017, 128: 59-70

[78] Ying Y K, Chen L, Chen G C. A temporal-aware POI recommendation system using context-aware tensor decomposition and weighted HITS. Neurocomputing, 2017, 242: 195-205

[79] Sang J T, Mei T, Sun J T, et al. Probabilistic sequential POIs recommendation via check-in data//Proceedings of the International Conference on Advances in Geographic Information Systems. Redondo Beach, California, 2012: 402-405

[80] Yang C, Sun M S, Zhao W X, et al. A neural network approach to jointly modeling social networks and mobile trajectories. ACM Transactions on Information Systems, 2017, 35(4): 1-36

[81] Lee D W, Liang S H L. Crowd-sourced carpool recommendation based on simple and efficient trajectory grouping//Proceedings of the 4th ACM SIGSPATIAL International Workshop on Computational Transportation Science. Chicago, Illinois, 2011: 12-17

[82] Zheng V W, Zheng Y, Xie X, Yang Q. Towards mobile intelligence: Learning from GPS history data for collaborative recommendation. Artificial Intelligence, 2012, 184(2): 17-37

[83] Yoon H, Zheng Y, Xie X, Woo W. Social itinerary recommendation from user-generated digital trails. Personal & Ubiquitous Computing, 2012, 16(5): 469-484

[84] Shafique S, Ali M E. Recommending most popular travel path within a region of interest from historical trajectory data//Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. Burlingame, California, 2016: 2-11

[85] Rodríguez-Hernández M D C, Ilarri S, Hermoso R, Trillo-Lado R. Towards trajectory-based recommendations in museums: Evaluation of strategies using mixed synthetic and real data. Procedia Computer Science, 2017, 113: 234-239

[86] Goenka K, Arpinar I B, Nural M. Mobile Web search personalization using ontological user profile//Proceedings of the 48th Annual Southeast Regional Conference. Oxford, Mississippi, 2010: 1-6

[87] Sang J T, Mei T, Xu C S. Activity sensor: Check-in usage mining for local recommendation. ACM Transactions on Intelligent Systems and Technology, 2015, 6(3): 1-24

[88] Amoretti M, Belli L, Zanichelli F. UTravel: Smart mobility with a novel user profiling and recommendation approach. Pervasive and Mobile Computing, 2017, 38: 474-489

[89] Liu Shu-Dong, Meng Xiang-Wu. Approach to network services recommendation based on mobile users' location. Journal of Software, 2014, 25(11): 2556-2574(in Chinese)
(刘树栋，孟祥武. 一种基于移动用户位置的网络服务推荐方法. 软件学报，2014，25(11)：2556-2574)

[90] Wang F, Meng X W, Zhang Y J. An adaptive user preferences elicitation scheme for location recommendation. Chinese Journal of Electronics, 2016, 25(5): 943-949

[91] Wang F, Meng X W, Zhang Y J, Zhang C H. Mining user preferences of new locations on location-based social networks: A multidimensional cloud model approach. Wireless Networks, 2018, 24(1): 113-125

[92] Chen C, Meng X W, Xu Z H, Lukasiewicz T. Location-aware personalized news recommendation with deep semantic analysis. IEEE Access, 2017, 5(99): 1624-1638

[93] Meng Xiang-Wu, Chen Cheng, Zhang Yu-Jie. A survey of mobile new recommend techniques and Applications. Chinese Journal of Computers, 2016, 39(4): 685-703(in Chinese)
(孟祥武，陈诚，张玉洁. 移动新闻推荐技术及其应用研究综述. 计算机学报，2016，39(4)：685-703)

[94] Hu Chuan, Meng Xiang-Wu, Zhang Yu-Jie, Du Yu-Lu. An enhanced group recommendation method based on preference aggregation. Journal of Software, 2018, 29(10): 3164-3183 (in Chinese)

《计算机学报》

(胡川，孟祥武，张玉洁，杜雨露. 一种改进的偏好融合组推荐方法. 软件学报，2018，29(10)：3164-3183)

[95] Shani G，Gunawardana A. Evaluating recommendation systems //Proceedings of the Recommender Systems Handbook. Boston，MA：Springer，2011：257-297

[96] Villegas N M，Sánchez C，Díaz-Cely J，Tamura G. Characterizing context-aware recommender systems：A systematic literature review. Knowledge-Based Systems，2018，140(1)：173-200

[97] Xu K，Zhang W D，Yan Z. A privacy-preserving mobile application recommender system based on trust evaluation. Journal of Computational Science，2018，26(5)：87-107

[98] Bobadilla J，Gutiérrez A，Ortega F，Zhu B. Reliability quality measures for recommender systems. Information Sciences，2018，442：145-157

[99] Parra D，Sahebi S. Recommender systems：Sources of knowledge and evaluation metrics//Velásquez J D，et al.，Eds. Advanced Techniques in Web Intelligence-2. Berlin，Germany：Springer，2013：149-175

[100] Ge M，Persia F. Evaluation in multimedia recommender systems：A practical guide//Proceedings of the 2018 IEEE 12th International Conference on Semantic Computing (ICSC). Laguna Hills，USA，2018：294-297

[101] Zheng E，Kondo G Y，Zilora S，Yu Q. Tag-aware dynamic music recommendation. Expert Systems with Applications，2018，106：244-251

[102] Gavalas D，Konstantopoulos C，Mastakas K，Pantziou G. Mobile recommender systems in tourism. Journal of Network and Computer Applications，2014，39(4)：319-333

[103] Yu Z W，Zhou X S，Zhang D Q，et al. Supporting context-aware media recommendations for smart phones. IEEE Pervasive Computing，2006，5(3)：68-75

[104] Celdrán A H，Pérez M G，Clemente F J G，Pérez G M. Design of a recommender system based on users' behavior and collaborative location and tracking. Journal of Computational Science，2016，12(12)：83-94

[105] Batet M，Moreno A，Sánchez D. Turist@：Agent-based personalised recommendation of tourist activities. Expert Systems with Applications，2012，39(8)：7319-7329

[106] Yang W S，Wang S Y. iTravel：A recommender system in mobile peer-to-peer environment. The Journal of Systems and Software，2013，6(86)：12-20

[107] Castells P，Vargas S，Wang J. Novelty and diversity metrics for recommender systems：choice，discovery and relevance//Proceedings of the International Workshop on Diversity in Document Retrieval at the ECIR 2011：the 33rd European Conference on Information Retrieval. Dublin，Ireland，2011：29-37

[108] Ge M，Delgado-Battenfeld C，Jannach D. Beyond accuracy：Evaluating recommender systems by coverage and serendipity //Proceedings of the 2010 ACM Conference on Recommender Systems. Barcelona，Spain，2010：257-260

[109] Avazpour I，Pitakrat T，Grunske L，Grundy J. Dimensions and metrics for evaluating recommendation systems// Robillard M，Maalej W，Walker R，Zimmermann T，eds. Recommendation Systems in Software Engineering. Berlin，Germany：Springer，2014：245-273

[110] Armentano M G，Abalde R，Schiaffino S，Amandi A. User acceptance of recommender systems：Influence of the preference elicitation algorithm//Proceedings of the 2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP'14). Corfu，Greece，2014：72-76

[111] Si Y L，Zhang F Z，Liu W Y. An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features. Knowledge-Based Systems，2019，163：267-282

[112] Huang Li-Wei，Jiang Bi-Tao，Lv Shou-Ye，et al. Survey on deep learning based recommender systems. Chinese Journal of Computers，2018，41(7)：1619-1647(in Chinese)
(黄立威，江碧涛，吕守业等. 基于深度学习的推荐系统研究综述. 计算机学报，2018，41(07)：1619-1647)

[113] Lu Z Y，Li H，Mamoulis N，Cheung D W L. HBGG：A hierarchical bayesian geographical model for group recommendation//Proceedings of the 2017 SIAM International Conference on Data Mining. Houston，USA，2017：372-380

计算机学报

MENG Xiang-Wu，Ph. D.，professor，Ph. D. supervisor. His research interests include recommendation systems and artificial intelligence.

LIANG Bi，Ph. D. candidate，associate professor. His research interests include recommendation systems，Web services and intelligent information processing.

DU Yu-Lu，Ph. D. candidate. His current research interests include recommendation systems and intelligent information processing.

ZHANG Yu-Jie，associate professor. Her research interests include intelligent information processing and recommendation system.

Background

In recent years, mobile applications have been widely used by people with the rapid development of mobile terminals, wireless communications, Internet of Things and artificial intelligence. And location-based mobile recommendation systems are becoming increasingly popular. The goal of this kind of recommendation system is to recommend right items to users at any time and any place, so that users can enjoy a large number of mobile services anywhere, anytime. How to effectively evaluate the performance of systems by using appropriate evaluation methods and evaluation metrics in the limited datasets and the number of users has become a research hotspot of location-based mobile recommendation systems.

This paper gives a comprehensive overview of the current study on evaluation for location-based mobile recommender systems, including comparison of the evaluation between traditional recommendation systems and location-based mobile recommender systems, datasets, evaluation methods and evaluation metrics. Furthermore, some special evaluation methods and evaluation metrics are found, and a four-layer evaluation architecture for location-based mobile recommender systems is proposed. At the end of the paper, the future research issues are presented, and some conclusions are drawn. It is hoped that the review will provide some references for researchers and engineers in this field.

《计算机学报》