

# 机器学习开题 -- 文本摘要自动提取器

高崇铭

January 8, 2018

## 1 研究问题

现有的推荐论文，主要突破和创新都在推荐方法上，试图用最新的科技，融合能收集到的跨域数据与知识，将数据结构表征和后续推荐结合起来，达到互利的效果。然而大多数方法对数据都有较强的假设，而现实中很多原始数据需要经过处理或转换才能完成进一步的挖掘工作。例如，从每天的人类出行、动物迁徙，以及飓风、洋流演化都能采集到大量时空属性的轨迹数据，而轨迹数据天生具有非结构性：其长度不固定，采样率不固定，并且含有很多噪声与不确定性，这使得后续的模式分析和地点、路径推荐等任务变得艰难。为此，在本文中，我将对轨迹数据基本处理，包括轨迹的表征、压缩、以及检索查询等关键步骤做出探讨，并提出对应的高效算法。此外，针对现有的地点推荐系统中面临的问题：数据稀疏、信息偏颇、迭代效率低下，以及不能充分利用多源进行很好的相互促进，我将融合用户的关系网络数据、访问地点的地理层级结构，考虑用户在特定时间、地点的签到信息（check-in），设计出一个全面的推荐系统。通过该推荐系统，用户的隐形偏好能被更好的挖掘，用户不喜欢的数据能够从海量的未访问样本中区分出来。为了克服传统的矩阵分解技术的解释性差、迭代慢的问题，我将采用分步分解的思想。最终，用户与地点的交互关系又将促进用户网络的演化，使得社团挖掘能够取得更好的结果。

1. 选题依据以及研究意义 近年来，随着移动设备飞快普及与硬件存储、计算能力的飞快提升，每天都有海量的数据以惊人的速度产生。高效地对这些数据进行表征和挖掘，将在现有的经济、环境、生物、医学、交通等领域中产生巨大的效益。具体地，在人们的日常生活中，由数据挖掘带来的效益表现为各式推荐服务。例如基于时间与空间的地点、路径推荐系统，都旨在抓住用户偏好以及领域特性，给用户推荐出合理、准确的结果。然而，涉及到时空的原始轨迹数据通常有着长度不固定、采样率不固定等天生的缺陷，虽然各种研究方法分别基于自己的假设提出了很多独立的轨迹特征提取以及预处理手段，但在富含多源语义的各种应用场景下，这些方法都只考虑了轨迹数据的单反面或少数方面信息，没有将多源信息有机地结合起来。如果能提出一种综合多源语义信息的轨迹表征方式，后续的挖掘任务，例如轨迹压缩、轨迹检索以及相似轨迹推荐都能被极大程度地促进。

另外，由于社交网络与地点推荐在某些软件下的高度结合，为用户推荐感兴趣的地点也是最近的热点问题。与传统商品、音乐、电影推荐不一样，在地点推荐中，人们的选择不再单纯地由地点的吸引程度来决定，还将由社交网络中用户与朋友的关系和朋友的偏好来决定，此外，用户的日常活动范围也将限制用户的选择。在这样的情境下，如何结合用户的社交网络、地点的地理信息，来给用户推荐准确地进行推荐感兴趣的地点，

是地点推荐系统中最重要的研究问题。另外，挖掘用户社交网络以及地点层级结构中的模式本身就是很重要的数据挖掘任务（社团挖掘、地理区域划分），如果能从用户与地点的交互中，以某种方式以促进社团挖掘以及地理区域划分过程，将更利于用户的地点推荐，形成一个互相促进的正反馈。这使得推荐系统与社团挖掘的性能提升的同时，也令信息的利用效率大大增高。

2. 国内外研究现状：二十一世纪以来，随着带有GPS、GSM及RFID等移动设备的普及，大量具有时间与空间属性的数据被收集与存储下来，研究如何从这些海量数据中提取出固定模式，催生了大量的轨迹挖掘任务。而后，由于社交网络的兴起与多领域数据的整合，轨迹与基于地点的签到数据 (check-in data) 被关联了大量的语义信息，如社交网络，如时间关联，人们将这样语义数据称为LBSNs (Location-Based Social Networks)，而轨迹数据在LBSNs中的语义分析与挖掘成为了关注的热点。这其中，POI (Point-of-interest) 推荐成为了一个非常热门且重要的任务。在现实中，进行POI推荐之前，轨迹需要良好的表征。然而，目前的POI推荐算法多将注意力集中到了推荐模块，而忽视了数据表征模块。事实上，数据表征的结果直接影响到推荐的有效性和准确性。于是，我们先对轨迹数据在LBSNs的表征做出总结：2.1 轨迹语义表征与挖掘

由不同设备采集到的原始轨迹数据只是一个带着时间戳的序列，其中包含的信息需要进一步筛选和提取。不同的研究工作根据后续的挖掘任务，提出了不同的轨迹特征提取方法以及语义表达模型，此处将常见的轨迹表达模型分类为三种：

(1) 基于关键点提取的轨迹表征 将轨迹数据用有限且紧凑的关键点来表达是一种最常见的方式。而关键点有两种定义方式：(a) 全局关键点。这种类型的关键点又被成为Points-of-Interest (POIs)，通常来说，这些点是提前选定的，比如地图上的学校、餐馆、加油站等标志性点或区域，也可以简单地将地图均匀地划分为小方格得到[1, 2, 3, 4, 5, 6, 7, 8]。然而也有很多研究工作致力于从原始轨迹中提取出这样的全局关键点（区域），比如一些工作[9, 10]采用基于划分的K-means的思想来对轨迹集中的GPS点进行聚类得到全局关键点，同样地，另外一些工作则使用基于密度方法来聚类，例如DBSCAN, OPTICS, KDE, 以及重力模型[11, 12, 13, 14, 15, 16]。(b) 基于单条轨迹的关键点。这种方式通常是基于后续工作的需求，对轨迹的某种特征进行提取。例如文献[17, 18]提取每一条轨迹的stay points来代表该轨迹，stay points即轨迹进行停留了一段时间的点，能结合停留时间与地点场景，刻画轨迹产生者的行为，如购物、吃饭、睡觉等。而文献[19]则提取出重要的驻点以及拐点来作为一条轨迹的关键点。

(2) 基于关键线段提取的轨迹表征 类似于关键点表征，关键线段表征也是一种直观的轨迹表示方法。轨迹的关键线段同样可分为两种：(a) 全局关键线段。在城市中，车辆的交通轨迹应当被道路所约束，因此将轨迹表示为一系列道路段的拼接是一种直观而合理的方式，这也催生了一些路网匹配的方法[20, 21, 22]，同样，也有从轨迹集合中提取出线段的工作，例如文献[23]用分段聚类的方法，用最小描述长度 (MDL) 作为评价因子将轨迹中的代表性线段聚类找出。(b) 基于单条轨迹的关键线段。这一类的方法也较为直观，例如文献[24, 15]将每条轨迹划分为行走段和非行走段，而文献[25, 26, 27] 则用压

缩的思想，将原轨迹段集合表示为最少的特征段，并在这个过程中保持最少的误差。

(3) 轨迹高层结构提取与表达 根据后续的挖掘任务，很多工作直接从轨迹数据中提取出高层的数据结构。例如为了轨迹检索，轨迹被投影到了树的结构上[28, 29, 30]，文献[31]则提出了用户与位置的二分图与树形层次图结构来表征原始轨迹中用户与地点的关系。文献[5, 4, 32]则用用户与地点和时间的三阶张量来对轨迹进行信息抽取。

现有的轨迹表征方法通常针对某种特定轨迹挖掘任务而设计的，其只考虑了某些特定领域的信息，没有将各种语义信息综合进表征中。且在预处理过程中，每条轨迹都是分开对待的，随着轨迹数据的飞速增长，相似的轨迹与地点将被冗余地处理多次，这种分开处理的策略将占用越来越多的存储空间和处理时间，因此，我们需要提出一种基于全局的轨迹表征。为了使这种全局表征方式更加完备，轨迹的相似性度量以及各种语义都应以一种方式融入到这种表征中来。

### 3.1 POI推荐的特性

#### (1) 地理影响

如同1970年的Tobler第一地理定律指出：“任何东西都有关联的东西，但距离近的东西关联的更多”[33]。对于LBSNs，Tobler第一定律表明用户会更为偏好距离近的POI而非距离远的POI，且喜欢已经喜欢的POI附近的POI的概率更大。事实上，地理影响对用户的偏好影响是最大的。

(2) 隐式反馈与稀疏性 在传统的推荐系统中，用户通常将自己对商品（书、电影、音乐等）的偏好表达为一个评分矩阵，其中每一个元素都是一个固定范围内的数值（如[1,5]），越大则表示某用户对该商品的喜好程度越高。而在POI推荐系统中，没有这样的评分矩阵数据，用户对于地点的交互数据只是一个访问频率，这个数值是离散的整数。这也就造成一个问题，大多数用户频繁访问的地点都是极少的，这些地方可能访问量高达上千次。而对于其他地方，用户可能仅仅访问过一到两次，这并不能体现用户对该地点的喜爱程度。举一个例子，在Netflix电影推荐数据集中，数据的空缺值占99%，但在Gowalla签到数据中，有值的地方仅占 $2.08 \times 10^{-4}$ 。这种访问频率的极大反差表现了POI推荐中的一大挑战。

(3) 社交属性 通过LBSNs中的数据可以观察得到假设：用户会从其朋友处得到并接受喜好推荐的建议，传统推荐系统将社交关系与评分洗好结合起来以增强推荐系统的性能。一些工作[34, 35]也表明了社交关系数据的融入确实能增强推荐系统的性能。然而在POI推荐系统中，有工作[36]表明96%的用户分享了不到10%的公共地点，这表明大量的用户并没有分享POI给自己的朋友，因此，社交网络在POI推荐中的影响并不如传统推荐那样明显。

### 3.2 结合社交网络的地点推荐系统

#### (1) 基于纯地点的推荐系统

传统的推荐系统的思想很朴实，给用户推荐商品，并不用局限于LBSNs下地点推荐系统。然而，由于用户在社交网络中的签到数据的地理信息可以挖掘出用户某些偏好，2011年左右，文献[37, 38]开始将传统推荐系统算法应用于推测POI中。此时的推荐仍然是考虑

最经典的协同过滤思想，将POIs视为商品，那基于用户的推荐系统[39]和基于商品的推荐系统[40, 41]分别考虑相似用户和相似商品进行推荐。之后，随着基于矩阵分解的推荐方法开始流行[42, 43, 44]，文献[38]提出了加约束项的矩阵分解的POI推荐方法，文献[45]基于概率矩阵分解（PMF）与概率因子模型（PFM）的POI推荐系统。这些方法仅仅基于LBSNs中用户与地点的交互数据，没有考虑其他方面的语义。因此，后续工作开始加入其他方面的语义信息来增强推荐性能。

## （2）考虑地理影响的POI推荐系统

在地点推荐系统中，自然要考虑地理信息对用户决策的影响。1970年的托比第一地理定律指出：“任何东西都有关联的东西，但距离近的东西关联的更多”[33]。这给了地点推荐系统两个指示：（a）用户访问地点应该遵循就近原则。（b）用户对自己喜欢的地点周边的地点更加感兴趣。同时，大量的文献[46, 47, 37, 48, 49]也在研究中发现了空间位置聚集的现象，并把这些规律应用于地点推荐系统中，对已有模型进行了改进。其中文献[48, 37]将用户访问两个地点之间的概率建模为两个地点之间距离的幂律分布。而[45]则将这一概率建模为多维混合高斯分布，且用实验表明其性能好于纯地点推荐系统中的PMF与PFM方法。而[49]则引入了核函数来建模这一概率，且用实验表明其性能好于用幂律分布来进行建模的POI推荐系统。而文献[47]进一步提出了基于贝叶斯概率的非负矩阵分解算法（BNMF）来对地理影响进行刻画，且用实验表明其性能更好于PMF与RMF算法。

## （3）考虑用户网络的POI推荐系统

事实上，基于用户社交关系的推荐系统在LBSN概念前久被广泛使用了，分为基于存储的[50, 51, 52]与基于模型[34, 35]的两种方法。之后，文献[45, 36]将这些方法推广到了地点推荐系统中，并取得了良好的效果。具体地，文献[36]提出了基于朋友的POI协同过滤推荐算法（FCF），其考虑了朋友的偏好，而非LBSNs中其他非朋友用户的爱好。注意FCF更加注重推荐的准确性而非可行性，这意味着FCF可能不能产生很多推荐，但能保证推荐的质量。而文献[45]提出了加入社交约束的概率矩阵分解方法（PMFSR），做出了朋友之间的隐向量应该尽量相似的假设，并将这化为一个约束带入到普通的矩阵分解方法中。

## （4）考虑时间影响的POI推荐系统

在传统的推荐系统中存在考虑时间影响的工作，比如有基于矩阵分解[53]的方法和随机游走[54]的方法。在这些传统方法中，时间影响是以衰减因子的形式加入到推荐系统中的。相反地，在POI推荐中，时间影响则是以推荐不同时间阶段的POI作为出发点。其中文献[48]考虑了用户倾向于在不同时间阶段访问不同类型的POI，于是提出了时间敏感的POI推荐系统，并在实验中表明了加入时间的POI推荐系统系统性能好于不加的。而文献[55]则进一步加入不同时间阶段的用户偏好约束：用户在一天中的不同阶段考虑的POI类型是不同的，而在相邻时间阶段的偏好则是相似的，于是他们对用户在不同时间阶段的偏好做了约束，并取得了不错的效果。

现有方法大多数是利用社交网络、地理信息来对地点进行推荐，然而在推荐的过程中却没有考虑到其他信息的数据是否全面，是否与用户地点交互数据一致的问题。于是一个问题浮现：能否考虑一个地点推荐系统，不仅将社交网络以及地理信息用于推荐，反



之还能用推荐信息来补齐社交网络以及地理信息（社团挖掘问题、区域划分问题），这样能使得信息最大效益的利用，而不同的任务也能互相促进。另一方面，无论是社交网络，还是用户地点交互数据都是非常稀疏的，对于check-in数据，其交互元素大多是用户访问特定地点的次数，而大量的空缺值中，如何将用户不感兴趣的负样本从潜在的未访问的元素中区分开来，是一个困扰推荐系统的问题。

#### 4. 存在挑战与拟解决的关键问题：

##### （1）现有轨迹表征方法未考虑LBSNs中的多源语义

现有的基于LBSNs进行轨迹挖掘或者POI推荐的工作都将轨迹表征问题进行简化，其仅仅考虑单方面的语义信息，这使得轨迹的表征单调，并且没有良好的解决稀疏、长度不固定和噪声多的问题。这使得不同轨迹挖掘或POI推荐的工作的数据集不能共享，也使得其挖掘或推荐的性能依赖于表征的好坏。如何融入LBSNs中丰富的语义信息是一个急需解决的问题。

##### （2）未充分挖掘POI在地理分布中的层级结构

几乎所有的POI推荐工作都将POI视为地图中的点集，而在事实上，POI的分布是具有层次性的。比如一个POI属于一条特定街道，而这条街道属于某个区，进而属于某个城市。在推荐中，城市、区和街道对POI也是有影响的，这种影响应该用一个层次结构刻画并融入到POI的推荐工作中。

（3）LBSNs中的社交网络数据稀疏，提升POI推荐效果不明显 为了解决POI推荐中用户与地点交互的稀疏问题，很多工作引入LBSNs中的社交网络。但是，社交网络的信息本身也是稀疏和缺失的，这就造成了上文提到的社交网络不能很好的促进POI推荐这一现象。要使得社交网络能够促进POI推荐算法的正确性，需要首先考虑解决社交网络数据本身的稀疏和缺失性问题。

基于以上三个问题，我将借鉴目前已有算法，分别提出相应的解决方法，并将在后续的毕设中实现。我将用综述中提到的现有算法作为对比算法，在多个数据集上综合测试，并在测试中不断改进自己的算法。

#### 技术路线：

##### （1）提出一个考虑多源语义的轨迹表征模型

由于现有轨迹表征都是根据具体的挖掘工作而提出的，故不能广泛使用于大多数挖掘模型，且仅考虑了很少的语义信息。本文拟提出一种结合所有语义信息的轨迹表征方法，使得轨迹的长度不一致、采样率不一致和噪声多等问题被克服。由于各种全局统计语义信息的存在，我们的表征模型应该是基于整个数据集，而不是对单条轨迹提出。我将借鉴数据挖掘中的同步模型[56, 57, 58, 59, 60]以及文本挖掘任务中词汇嵌入的技术[61, 62, 63, 64]来对轨迹进行规整和数值表征。

##### （2）充分考虑空间POI分布的层次性来加强POI表征

针对现有轨迹挖掘和POI推荐算法没有良好的利用POI的分布的层次性，我将在轨迹表征方法中，考虑POI的层次性。基于一个区域应该综合考虑并结合其包含的所有POI的语义信息这样的假设，我将把区域和POI的关系表征为树结构，根节点为整个数据集（一个

城市)，叶节点为不能细分的POI。对于每一个节点的语义表征向量，应该和其父节点与其所有子节点的语义表征向量相差不大。加入这个约束后，不同区域的POI表征将被区分开来，同时，对于用户来到新的区域，给其推荐POI的问题也能很好的解决。

### (3) 提出充分利用社交网络的POI推荐算法

针对现有POI推荐算法信息利用率低，没有良好的利用社交网络信息，以及社交网络的信息本身也是稀疏和缺失等问题，我将提出新的推荐思路：不仅用社交网络的信息来做POI推荐，反过来也将POI推荐的信息用于社交网络的链路预测，从而解决社交网络信息不全这问题，使得POI推荐系统有更好的性能和解释性。

## References

- [1] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007, pp. 330--339.
- [2] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 195--203.
- [3] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 1027--1036.
- [4] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, 2014, pp. 25--34.
- [5] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 3, pp. 712--725, 2015.
- [6] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. Xu, "Destination prediction by sub-trajectory synthesis and privacy protection against such prediction," in Data Engineering (ICDE), 2013 IEEE 29th International Conference on, 2013, pp. 254--265.
- [7] B. Zheng, N. J. Yuan, K. Zheng, X. Xie, S. Sadiq, and X. Zhou, "Approximate keyword search in semantic trajectory database," in Data Engineering (ICDE), 2015 IEEE 31st International Conference on, 2015, pp. 975--986.

- [8] E. Cho, S. A. Myers, and J. Leskovec, Friendship and mobility: user movement in location-based social networks," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, 2011, pp. 1082--1090.
- [9] D. Ashbrook and T. Starner, Learning significant locations and predicting user movement with gps," in Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on, 2002, pp. 101--108.
- [10] -----, Using gps to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous computing*, vol. 7, no. 5, pp. 275--286, 2003.
- [11] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, Mining periodic behaviors for moving objects," in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 1099--1108.
- [12] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, Recommending friends and locations based on individual location history," *ACM Transactions on the Web (TWEB)*, vol. 5, no. 1, p. 5, 2011.
- [13] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen, Discovery of convoys in trajectory databases," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 1068--1080, 2008.
- [14] Z. Chen, H. T. Shen, and X. Zhou, Discovering popular routes from trajectories," in Proceedings of the 27th International Conference on Data Engineering, 2011, pp. 900--911.
- [15] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, Understanding mobility based on gps data," in Proceedings of the 10th international conference on Ubiquitous computing. ACM, 2008, pp. 312--321.
- [16] Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui, Regularity and conformity: Location prediction using heterogeneous mobility data," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1275--1284.
- [17] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, Mining user similarity based on location history," in Proceedings of the 16th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, 2008, p. 34.
- [18] Y. Zheng and X. Xie, Learning travel recommendations from user-generated gps traces," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 1, p. 2, 2011.

- [19] N. Adrienko and G. Adrienko, Spatial generalization and aggregation of massive movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205--219, 2011.
- [20] J. S. Greenfeld, Matching gps observations to locations on a digital map," in 81th annual meeting of the transportation research board, vol. 1, no. 3, 2002, pp. 164--173.
- [21] W. Chen, M. Yu, Z. Li, and Y. Chen, Integrated vehicle navigation system for urban applications," 2003.
- [22] P. Newson and J. Krumm, Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2009, pp. 336--343.
- [23] J.-G. Lee, J. Han, and K.-Y. Whang, Trajectory clustering: a partition-and-group framework," in *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 2007, pp. 593--604.
- [24] Y. Zheng, L. Liu, L. Wang, and X. Xie, Learning transportation mode from raw gps data for geographic applications on the web," in *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 247--256.
- [25] W.-C. Lee and J. Krumm, Trajectory preprocessing," in *Computing with Spatial Trajectories*, 2011, pp. 3--33.
- [26] D. H. Douglas and T. K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112--122, 1973.
- [27] R. Bellman, On the approximation of curves by line segments using dynamic programming," *Communications of the ACM*, vol. 4, no. 6, p. 284, 1961.
- [28] A. Guttman, R-trees: a dynamic index structure for spatial searching. ACM, 1984, vol. 14, no. 2.
- [29] L. Wang, Y. Zheng, X. Xie, and W.-Y. Ma, A flexible spatio-temporal indexing scheme for large-scale gps track retrieval," in *9th International Conference on Mobile Data Management*, 2008, pp. 1--8.
- [30] D. Pfoser, C. S. Jensen, Y. Theodoridis et al., Novel approaches to the indexing of moving object trajectories." in *VLDB*, 2000, pp. 395--406.



- [31] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, Mining interesting locations and travel sequences from gps trajectories," in Proceedings of the 18th International Conference on World Wide Web, 2009, pp. 791--800.
- [32] Y. Liu, C. Liu, B. Liu, M. Qu, and H. Xiong, Unified point-of-interest recommendation with temporal interval assessment," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1015-1024.
- [33] W. R. Tobler, A computer movie simulating urban growth in the detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234--240, 1970.
- [34] M. Jamali and M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks," in Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010, pp. 135--142.
- [35] H. Ma, H. Yang, M. R. Lyu, and I. King, Sorec: social recommendation using probabilistic matrix factorization," in Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008, pp. 931--940.
- [36] M. Ye, P. Yin, and W.-C. Lee, Location recommendation for location-based social networks," in Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, 2010, pp. 458--461.
- [37] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, Exploiting geographical influence for collaborative point-of-interest recommendation," in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011, pp. 325--334.
- [38] B. Berjani and T. Strufe, A recommendation system for spots in location-based online social networks," in Proceedings of the 4th Workshop on Social Network Systems. ACM, 2011, p. 4.
- [39] J. S. Breese, D. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering," in Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1998, pp. 43--52.
- [40] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th international conference on World Wide Web. ACM, 2001, pp. 285--295.
- [41] G. Linden, B. Smith, and J. York, Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76--80, 2003.

- [42] A. Mnih and R. R. Salakhutdinov, Probabilistic matrix factorization," in Advances in neural information processing systems, 2008, pp. 1257--1264.
- [43] D. D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization," in Advances in neural information processing systems, 2001, pp. 556--562.
- [44] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, 2009.
- [45] C. Cheng, H. Yang, I. King, and M. R. Lyu, Fused matrix factorization with geographical and social influence in location-based social networks." in Aaai, vol. 12, 2012, pp. 17--23.
- [46] H. Gao, J. Tang, and H. Liu, gscorr: modeling geo-social correlations for new check-ins on location-based social networks," in Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012, pp. 1582--1586.
- [47] B. Liu, Y. Fu, Z. Yao, and H. Xiong, Learning geographical preferences for point-of-interest recommendation," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013, pp. 1043--1051.
- [48] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, Time-aware point-of-interest recommendation," in Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013, pp. 363--372.
- [49] J.-D. Zhang, C.-Y. Chow, and Y. Li, igeorec: A personalized and efficient geographical location recommendation framework," IEEE Transactions on Services Computing, vol. 8, no. 5, pp. 701--714, 2015.
- [50] J. Golbeck, Generating predictive movie recommendations from trust in social networks," Trust Management, pp. 93--104, 2006.
- [51] P. Massa and P. Avesani, Trust-aware recommender systems," in Proceedings of the 2007 ACM conference on Recommender systems. ACM, 2007, pp. 17--24.
- [52] M. Jamali and M. Ester, Trustwalker: a random walk model for combining trust-based and item-based recommendation," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009, pp. 397--406.
- [53] Y. Koren, Collaborative filtering with temporal dynamics," Communications of the ACM, vol. 53, no. 4, pp. 89--97, 2010.

- [54] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. Sun, Temporal recommendation on graphs via long-and short-term preference fusion," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010, pp. 723--732.
- [55] H. Gao, J. Tang, X. Hu, and H. Liu, Exploring temporal effects for location recommendation on location-based social networks," in Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013, pp. 93--100.
- [56] J. Shao, X. He, C. Böhm, Q. Yang, and C. Plant, Synchronization-inspired partitioning and hierarchical clustering," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 4, pp. 893--905, 2013.
- [57] J. Shao, Z. Han, Q. Yang, and T. Zhou, Community detection based on distance dynamics," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1075--1084.
- [58] J. Shao, Q. Yang, H.-V. Dang, B. Schmidt, and S. Kramer, Scalable clustering by iterative partitioning and point attractor representation," ACM Transactions on Knowledge Discovery from Data, vol. 11, no. 1, p. 5, 2016.
- [59] J. Shao, F. Huang, Q. Yang, and G. Luo, Robust prototype-based learning on data streams," IEEE Transactions on Knowledge and Data Engineering, 2017.
- [60] J. Shao, C. Gao, W. Zeng, J. Song, and Q. Yang, Synchronization-inspired co-clustering and its application to gene expression data," in 2017 IEEE 11th International Conference on Data Mining (ICDM). IEEE.
- [61] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, A neural probabilistic language model," Journal of machine learning research, vol. 3, no. Feb, pp. 1137--1155, 2003.
- [62] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch," Journal of Machine Learning Research, vol. 12, no. Aug, pp. 2493--2537, 2011.
- [63] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality," in Advances in Neural Information Processing Systems 26, 2013, pp. 3111--3119.
- [64] J. Pennington, R. Socher, and C. Manning, Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532--1543.