

面向社交用户的商业大数据挖掘 研究简述

作者简介：赵鑫，中国人民大学信息学院讲师。师从北京大学李晓明教授，博士期间专注于研究面向文本内容的社交用户话题兴趣建模（北京大学优秀博士论文），曾获得过两个科研资助：2011 年谷歌博士奖研金和 2012 年微软亚洲博士奖研金。近五年内在国内外著名学术期刊与会议上以第一作者或者通讯作者身份发表论文近 40 篇（含已录用论文），其中包括信息检索领域顶级学术期刊 ACM TOIS 和学术会议 SIGIR、数据挖掘领域顶级学术期刊 IEEE TKDE 和学术会议 SIGKDD、自然语言处理顶级会议 ACL 和 EMNLP。所发表的学术论文取得了一定的关注度，据 Google Scholar 统计，已发表论文共计被引用 1100 余次，其中以第一作者发表的《Comparing Twitter and Traditional Media Using Topic Models》被引用 539 次。担任多个重要的国际会议 (WWW, IJCAI, CIKM, ACL, EMNLP 等) 和期刊 (ACM TOIS/TKDD/TIST, IEEE TKDE 等) 评审，曾担任 CCL 2015/2016、AIRS 2016 出版主席。目前为中国中文信息学会青年工作委员会以及社交媒体处理专业委员会委员。

研究背景简介

最近几年，电子商务网站快速发展，典型的商业巨头包括国内的京东、淘宝，国外的亚马逊 (Amazon)、易趣 (eBay) 等。这些电子商务网站克服了传统消费中地理位置和时间的限制，在有效的物流机制支持下，使得商务交易可以在任何地方、任何时间发生，极大地满足了用户的购物需求。

从一个方面来看，电子商务平台网站本身累积了大量的用户数据，如用户购买记录、搜索记录、评论记录等。围绕着这些电商平台的数据，很多科研和工程人员努力改善电子商务网站的服务，使其更好地满足用户的需求。常用技术主要用于建立基于电子商务平台的搜索系统以及推荐系统。尽管这些方面已经受到了国内外学者的高度重视，大部分研究都是针对某个电子商务网站量身打造，所开发的产品推荐系统会受到该电子商务网站本身所提供的信息量和信息来源的制约。因此，能否重新梳理和拓宽研究视角至关重要。

从另一个方面来看，随着互联网技术的快速发展，社交媒体服务在用户的真实生活中发挥着越来越重要的作用，得到了广泛使用。同一用户可能同时拥有多个社交媒体网站的账号，分别对应着不同的网络社区身份。以这些社区身份为基础，用户可以同时参与到多个社交媒体平台，享受其中提供的应用服务。同时，随着社交媒体服务平台向移动客户端的迁移，一个智能移动通讯工具（如手机、平板电脑等）往往会绑定多种应用服务，使得这种拥有多社区身份的发展趋势进一步得到加强。因此，能否同时围绕用户的“真实身份”与“在线社交身份”打造电子商务服务，是非常重要的思维创新。同时利用**电子商务平台上的数据**以及**社交媒体平台上的用户数据**，能够解决一些之前电子商务平台网站很难解决的技术挑战，如冷启动推荐问题等。

本综述将充分结合两个方面的数据，简述作者在这一方向已完成的一些工作与进展。

重要任务以及相应方法概览

本文主要考虑三个重要任务：用户画像构建、用户意图检测和用户需求推荐。

用户画像构建

用户画像旨在为社交用户构建起一个可量化的特征表示，包括简单的属性特征（如年龄、性别）以及复杂的模式特征（如网络隐含表示）。下面分两个方面进行讨论。

1. 简单特征抽取与表示

首先考虑简单的社交信息特征抽取。很多社交网站要求用户在注册时填写一些身份信息，例如，在新浪微博中，用户可以填写年龄、性别、省份、教育和职业信息。这些信息可以直接作为用户画像的信息输入。Zhao 等人^[1,2]利用微博用户注册的属性信息进行用户画像构建，包括性别、年龄、职业、兴趣等，并且将得到的用户画像应用到产品推荐领域。在社交网站上，用户个人属性信息存在缺失和虚假现象，如用户隐去了年龄或者填写了错误的年龄。解决这些问题的基础是对数据质量的评估，进行有效清洗和补充。对于社交用户的属性信息进行清洗和补充是一个非常重要的研究问题，已经受到了研究学者的高度关注，这里略去具体的研究工作。

除了用户画像以外，还可以考虑对于物品（例如产品）建立受众画像。例如一款产品的受众特征可以刻画为“单身未婚女性、年龄在 18 至 24 区间、大学文化程度”。如果能够获得受众用户的群体属性特征，将对于产品推荐具有重要意义。Zhao 等人^[1,2]主要考虑利用以下两种社交数据进行产品受众特征的学习。

a) 利用电商平台的评论信息：第一种资源是在线产品评论。用户有时候会在评论内容中显式地提及与受众属性信息相关的信息。例如，在一条评论数据中“这款手机不错，给儿子买一个”暗示当前的产品适合该评论作者的儿子，也就是她的儿子是该产品的一个潜在受众，同时也可以推断知道[年轻][男士]是一个受众的两个特征。

b) 利用微博平台的关注信息：在微博中，用户可以自由地表达自己对于某款产品或者品牌的情感。如果情感取向为“正”（褒），就可以把当前用户当作一个潜在的产品受众。通过搜集这样的正向情感用户，然后聚合他们的个人属性信息，用来推断该产品的受众特征。主要采用如下两种用户行为来捕捉用户对于某一产品的正向情感，包括关注关系(following)和提及关系(mentioning)：如果一个用户对某一产品感兴趣或者已经使用过该产品，她很有可能通过发表状态文本的形式来表达自己的情感取向。给定一个产品，使用产品名来检索得到所有包含该产品名字的微博，然后进一步使用基于机器学习的方法来判定每条微博中的用户情感取向{正向，负向}。

2. （相对）复杂特征表示

除了简单的特征抽取外，还可以对于已知的社交信息进行初步或者较为复杂的加工学习。例如，Xiao 等人^[3]利用微博上用户的权威度（例如 PageRank 值）以及文本相关度（与知乎提问的语义相似性）来改进知乎上最佳答案的预测。基本的假设如下：给定一个

问题，如果一个知乎用户在新浪微博中的权威度越高，那么他的答案就会越有可能成为最佳答案；如果一个知乎用户所发表的新浪微博内容与问题越相关，那么他的答案就会越有可能成为最佳答案。该方法对原始的用户文本信息以及网络结构信息进行初步地加工学习。在[4]中，Zhao 等人进一步拓展了 KDD 2014 的工作，对于文本信息以及网络结构信息使用了分布式表示学习的方法来抽取特征表示。

特别地，最近在网络表示学习里面所提出了网络嵌入式表示(“Network Embedding”)的方法。该方法对于具有网络结构的数据刻画能力非常好，值得用户画像建模时所考虑。网络嵌入式表示在很多任务中都比传统的方法取得了很大的提高，其中一个重要的提升原因就是引入了分布式表示的思想，以及有效(效果和速度上)的训练方法。对比第一种简单抽取的方法，不难发现复杂特征表示的表示能力更强，但是缺点是可解释性通常较差。因此，在实际应用中，可以考虑将两种类型的特征表示混合使用。

用户意图检测

在本文，用户意图特定指的是用户进行相关**商业消费**的潜在意图。传统的电子商务网站捕捉用户意图的方法主要是利用用户的搜索日志以及用户的浏览/消费记录。随着社交媒体平台的快速发展，可以从社交内容中进行用户意图的捕捉与挖掘。在社交网站中，用户可以实时地发表状态文本，如朋友圈内的状态、微博平台中的短文本等等，这些文本统称为状态文本。状态文本直接表达了用户的观点和需求，对于检测实时的用户商业意图具有很大的价值。整体来说，用户意图可以分为两大类，即显式意图与隐式意图。

1. 显式意图检测

在显式意图里面，我们进一步定义两种不同的用户意图。

- 个体意图检测

首先，介绍个体意图检测，即特定用户自身所表达的消费意图。例如，一名新浪微博用户发表了一条微博“我想要换个新手机，求推荐”。该用户直接表达了消费意图，这种通过社交网站所表达的消费意图还没有被大型电子商务网站所重视。目前，一些小型企业(特别是创业公司)开始利用这些具有消费意图的微博进行产品的定向推广。Zhao 等人^[1]在 KDD 2014 年的论文中首次提出使用微博数据进行用户的消费意图检测，并且将用户意图检测任务刻画为一个二分类问题，即有商业意图和无商业意图；进一步，为了解决这个二分类问题，使用微博的文本特征以及微博用户的人口统计学属性信息。

尽管具有商业意图的社交文本比例相对较低，但是由于社交网站中的文本数量巨大，因此即使比例很小，最后的绝对数字仍然很大，值得电商平台思考去进一步挖掘与利用。Wang 等人^[5]对于上述问题进行了一个泛化，不再是简单考虑二分类问题，而是利用 Twitter 中的状态文本建立了一个消费意图体系。该分类体系主要是基于国外团购网站

Groupon¹的分类体系进行修改得到的。在该文里面，作者对于数千条微博进行了人工标注以及分类，最后得到如下的类别体系（图 1）：

意向类别	数量 (百分比)	例子
食物	245 (11.50%)	hungry...i need a salad.....four more days to the BEYONCE CONCERT...
出行	187 (8.78%)	I need a vacation really bad. I need a trip to Disneyland!
事业与教育	159 (7.46%)	this makes me want to be a lawyer RT @someuser new favorite line from an ...
商品和服务	251 (11.78%)	mhmmm, i wannna a new phoneeee. ... i have to go to the hospital. ...
事件和活动	321 (15.07%)	on my way to go swimming with the twoon @someuser; i love her so muchhhh!
琐事	436 (20.47%)	I'm so happy that I get to take a shower with myself. :D
没有意向	531 (24.92%)	So sad that Ronaldo will be leaving these shores...http://URL

图 1 Twitter 中消费意图体系与比例

在微博中，获得有标注的用户意图数据非常困难。因此在[5]中，作者设计了一个基于图正则化的半监督标注算法，可以有效利用意图关键词以及微博之间的语义关系来缓解有标注数据的稀疏性。

- 群体意图检测

上述主要介绍了基于个体的消费意图检测。对比个体意图检测，群体意图检测主要关心一个群体中的用户所表达出的整体意图模式。以下图（图 2）为例，在“大黄鸭”事件之后，淘宝搜索引擎很快就已经生成了一些相关的定制查询，这些查询是人们集中所关心的一些购买产品。这一个例子说明了群体消费意图很有可能是由于特定事件或者话题所引起的。再举一个例子，如“北京雾霾”这一事件带来的群体性消费意图，可能是口罩、空气净化器、绿植等除霾产品的热销。

¹ www.groupon.com



图2 “大黄鸭” 热点话题之后所激发的购物热潮

针对热点事件/话题对于群体性消费趋势的影响，Wang 等人^[6]给出了量化的统计与验证。具体方法为：对于新浪排行榜某一时间段内的上榜话题，人工检测是否在淘宝中存在着对应产品，如果存在的话，就说明该话题催生了群体性消费意图。统计中考虑了五个类别（商业、人物、体育、国内以及电影），如图 3 所示，最后得到的结论为**国内**类别内部的话题更有可能催生更多数量的群体购买意图（绝对数量），**电影**类别内部的话题所催生的购买意图比例最高（相对比例）。那么给定一个热门话题，如何提前预知哪些产品会成为相关热销产品呢？Wang 等人^[6]继续提出了一个新颖的思路，首先将热点话题作为查询去检索相关微博，然后识别检索得到的微博所包含的产品名字（例如，“又雾霾了，赶快买**口罩**”），最后利用产品间的关联性加强相关度的判断，取得了不错的效果。这种解决方法巧妙地利用了群体智慧以及社交平台的及时性。

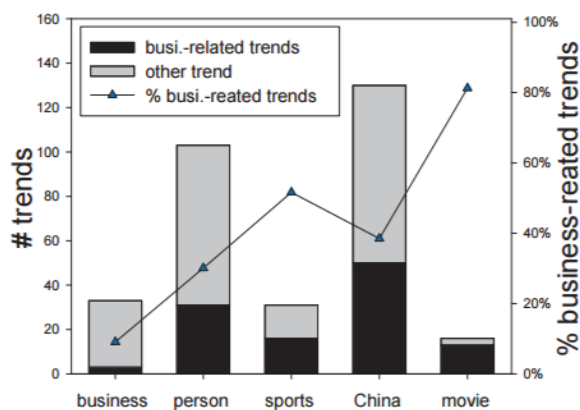


图3 新浪话题榜单中的话题所对应的商业购买意向比例

2. 隐式意图检测

对于隐式意图来说，是指用户在社交文本中没有显式提及任何商品名称，也没有直接提及任何购买行为，但是具有一定的购买暗示性与潜在性。如，“刚生下来的小 baby 总

喜欢尿床，太让人崩溃了”这条微博说明该用户为一个新生儿的父母，尽管没有流露出任何购买意图，但是可以推断得知，他们可能具有购买婴儿纸尿裤产品的倾向。目前来说，捕捉这种隐式的话题与产品之间的关联非常具有挑战性，需要深层次的推理机制和算法，同时需要特定领域的知识图谱或者先验知识的支持。对于研究学者来说，更大的挑战是，很难进行精准的量化评测。例如，在上面的例子中，我们无法得知该新生儿的家长在真实生活中是否购买了纸尿裤，从而无法断定这条微博是否一定具有商业意图。

用户需求推荐

用户画像和意图检测解决了“用户是谁”和“用户想要什么”，用户需求推荐进一步将用户需求与提供的服务进行关联。例如，根据用户的购买需求，推荐所需要的产品。下面我们对于一些典型的推荐算法进行讨论。

1. 基于人口统计学的推荐方法

在[1,2]中，Zhao 等人提出使用人口统计学特征（从新浪微博获取）来进行产品推荐。主要想法是将用户和产品表示在相同的属性信息维度上：首先，从公开的社交账号信息中提取用户的个人属性信息；然后，从社交媒体网站中学习产品受众聚合（即上述所介绍的产品受众画像）的属性信息。经过上述两个步骤，可以将用户和产品表示在相同的属性维度，从而可以很好地进行相似度学习。基于这些特征，可以使用基于机器学习的排序算法 (Learning to Rank) 实现精准的产品推荐。学习排序算法是在学习阶段给定一些查询 (query) 和检索得到的相关文档，并且这些文档相对于一个查询的相关程度也已经给定，试图要构建一个排序函数 (或者说模型) 最小化训练数据的损失。在检索 (测试) 阶段，给定一个查询，系统会返回一个候选文档的排序列表，该列表是按照相关度分数递减进行排序得到的。通过简单的类比，可以将学习排序的方法应用到产品推荐任务上。

2. 基于浅层分布式表示学习的推荐方法

最近几年，浅层神经网络模型得到了广泛应用，特别是词嵌入表示模型。词嵌入表示模型 (embedding models) 最早应用在自然处理领域中，利用背景信息构建词汇的分布式表示。嵌入式表示模型往往简单而又比较有效，因此很快被人们应用到推荐系统中。其核心思想就是同时构建用户和物品（或者其他背景信息）的嵌入式表示，使得多种实体的嵌入式表示存在于同一个隐含空间，进而可以计算两个实体之间的相似性。

在[4]中，Zhao 等人使用 doc2vec 模型来同时学习用户和物品的序列特征表示，然后将其用在传统基于特征的推荐框架中，引入的嵌入式特征可以在一定程度上改进推荐效果。在[7]中，嵌入式表示模型被用来进行地理位置 (POI) 推荐，其基本框架就是刻画一个地理位置的条件生成概率，考虑了包括用户、轨迹、邻近的地点、类别、时间、区域等因素。整体来说，这类模型的目标函数都是试图构建数据点的条件概率 $\Pr(\text{item}|\text{context})$ 或者背景信息的条件概率 $\Pr(\text{context}|\text{item})$ ，其中 item 和 context 分别表示推荐物品以及其对应的背景信息，因此一个关键问题就是如何构建背景信息 (context)。在嵌入式表示模型中，多个实体的嵌入式特征可以进行多种组合（包括 max pooling、average pooling 等），对于搭建分布式表示模型非常灵活。上述工作都利用了分布式表示模型具有较强的相关性

建模能力。很多推荐任务，本质可以转换为相关度排序问题，因此嵌入式表示模型是一种适合的候选方法。一般来说，浅层的嵌入式表示模型的训练非常高效，因此在大规模数据集上有效性和复杂度都能达到不错的效果。

此外，Zhao 等人^[8]还借鉴了网络嵌入式表示学习(network embedding)中的方法进行物品推荐。具体思路就是将用户与物品同时刻画为图上的节点，进一步每个用户的采纳行为（例如，一个用户买了一个产品），都可以刻画为图上的一条边（用户与产品分别对应着图上的节点），这样推荐问题转变为图上节点相似度的计算问题，可以直接应用已有的网络嵌入式表示模型。

3. 基于序列建模的推荐方法

在很多推荐场景下，用户行为对应着一个生成的序列。实际上，上面提到的基于分布式表示学习的想法，就是一种简单的序列建模方法，只不过在这一方法中，背景信息的序列性被忽略了。基于用户行为的数据特点，可以考虑应用或者借鉴多种序列建模方法。在[9]中，Yang 等人同时结合 RNN 及其变种 GRU 模型来分别刻画用户运动轨迹的长短期行为模式，通过实验验证，在“next location”推荐任务中取得了不错的效果。如图 4 所示，给定一个用户的轨迹序列，在预测下一个地点时，邻近的短期访问背景和较远的长期访问背景都被考虑进模型中。

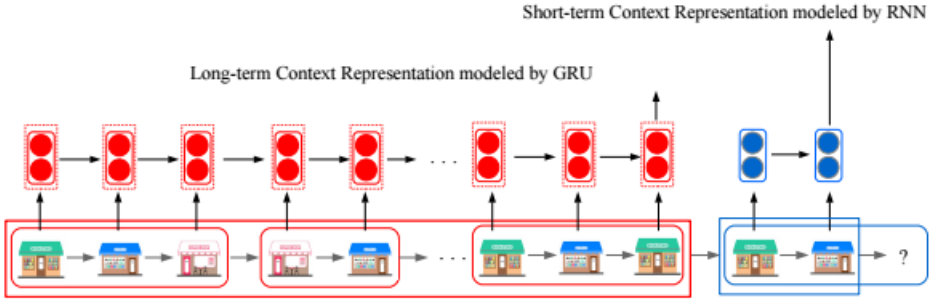


图 4 长短期访问背景的序列建模 [23].

在[10]中，Zhao 等人提出一个基于隐马尔可夫模型的主题模型来刻画用户轨迹行为，从用户轨迹行为中抽取隐含转移特征模式，来进行朋友关系的预测。其基本的想法就是，每个用户在一次出行的时候，应该有一个特定的意图，每个特定的意图都会对应着一个隐马尔科夫模型。每次地点的生成都会对应着一个意图和一个状态，从而可以挖掘特定意图下的用户隐含行为模式。

4. 面向购买受众的推荐方法

传统的推荐算法是针对特定用户进行产品推荐，一个常用的假设就是该用户所购买的产品都是自己本人使用。随着电子商务的快速发展，代购行为已经非常普遍了。Zhao 等人^[11,12]首次利用京东的在线评论数据进行代购行为的定量统计。基于一个 1.39 亿条购物记录（包括时间与评论信息）的京东评论数据集，发现大概 11%（一个比例的下界估计）的购物行为都是代购行为，这一统计数字说明代购行为非常普遍。进一步又发现 49%的产

品收到了至少十条带有受众 mention 的用户评论，25%的用户发表过至少十条以上的受众 mention。传统的推荐算法只能为一个用户构建单一的特征表示，无法融入受众信息。为了刻画受众信息，Zhao 等人^[11,12]继续提出抽取受众短语（mention）并且将其组织成类别，从而分别建立类别与特定短语的购买倾向。该工作一共考虑了六种类别：**小孩**、**年轻女士**、**成年女士**、**年轻男士**、**成年男士**和**公司**（图 5）。上述工作说明推荐算法不应只为一个用户设置单一的偏好表示，Wang 等人^[13]进一步将此想法进行泛化，使用了基于主题模型的多偏好表示的评分预测方法。

Category	Example product adopter mentions
Children	小孩(kids) 她小孩(her kid) 新生儿(new-born baby) 婴儿(baby) 小侄女(little niece) 小外甥(little nephew)
Young female	妹妹(younger sister) 表妹(cousin) 孕妇(pregnant woman) 女朋友(girlfriend) 媳妇(daughter-in-law) 姐姐(elder sister)
Old female	妈妈(mother) 老妈(mother) 阿姨(aunt) 姑姑(aunt) 岳母(mother-in-law) 丈母娘(mother-in-law)
Young male	弟弟(younger brother) 哥哥(elder brother) 表弟(cousin) 哥们(buddy) 兄弟(brothers) 男朋友(boyfriend)
Old male	爸爸(father) 老爸(father) 叔叔(uncle) 爷爷(grandfather) 岳父(father-in-law) 公公(father-in-law) 舅舅(uncle)
Colleagues	公司(company) 办公室(office) 员工(employee) 分公司(branch) 厂里(factory) 工人(worker)

图 5 受众短语分类示意图

5. 面向进化物品的推荐方法

传统的推荐算法有一个相对较强的假设，即物品本身是固定的。而在实际应用系统中，同一个物品经常会发生进化和改变，例如，iPhone 手机会迭代更新多次、同一款 APP 应用连续推出多个版本。一种处理方法就是将物品的多个版本当做完全不同的物品处理，但是这种方法会忽略掉同一物品间版本与物品间的关联。Yao 等人^[14]系统地对于手机应用 APP 的多个版本进行了定量化的研究，发现（1）同一个 APP 的多个版本收到的用户平均打分可能存在很大的差异；（2）同一个用户对于同一个 APP 的多个版本的评分可能存在很大的差异。这些发现都说明不能以一种静态的观点去设计推荐算法。Yao 等人^[14]提出了一种新颖的矩阵分解算法用来进行评分预测，可以刻画同一个 APP 的多个版本，引入了版本间的时态关联以及相似 APP 之间的关联。能够在一定程度上对于一个 APP 的新版本进行冷启动的评分预测，取得了不错的效果。这一工作启示我们，在很多情况下，需要以一种动态的观点来设计推荐系统，特别是现实应用系统中，物品是变化更新的。

研究展望

基于上面的讨论，可以看到目前围绕社交大数据的工作还有很多可探索的空间，在未来会有更多更广泛的尝试。下面对于未来三个可能的研究方向进行简要介绍。

“互联网+”：社交账号趋于统一化

“互联网+”代表着一种新的经济形态，它指的是依托互联网信息技术实现互联网与传统产业的联合，以优化生产要素、更新业务体系、重构商业模式等途径来完成经济转型和升级。依托于快速发展的社交媒体平台，很多传统服务业已经开始发生改变（例如网上购物等）。同时，个人网络身份正在逐步趋于统一化，例如，微信可以提供各种生活消费等，还有很多服务平台支持第三方登录。在可预见的未来，用户的网络身份一定会逐步“现实化”与“统一化”，多种账号会打通联系，与用户真实身份的联系将会越来越显著。因此，电子商务平台所面临的服务场景不再局限于单一用户信息领域和单一物品信息领域。实现信息的跨网站联合利用，对于跨平台的信息融合与聚合尤为重要。

深度模型的有效利用

目前已有工作中的商业数据挖掘算法大部分还是基于传统非深度学习的模型与方法。在一些初步的商业信息挖掘工作中，开始有学者尝试使用深度模型，并且取得了一定的效果；但是相比语音图像以及自然语言领域等，这些效果上的提升以及应用的广度还远远不够。如何在商业挖掘这个特定领域发挥深度模型的潜在能量，是一个很有意义的研究思路。同时，在设计复杂模型的同时，能否兼顾考虑可解释性对于应用系统具有重要的意义。

社交数据的深入挖掘与隐私保护

对于电子商务公司或者应用来说，如何充分利用用户的多种社交数据来构建更为深入的数据分析具有重要意义。典型的任务包括用户画像、用户收入水平预测、用户征信水平分析等。从一个方面来说，研究者希望用户在社交网站平台上留下尽可能多的痕迹，供数据挖掘算法使用；从另外一个方面来说，人们也尝试从中挖掘哪些痕迹容易导致用户的个人隐私泄露。在已有的工作中，前者是一个常见的研究题目，而后者受到的关注较少，值得继续深入研究。

致谢相关合作者

上述工作均为我与其他研究人员合作完成，或者在其支持下完成，在此对他们表示谢意。

参考文献

[1] Wayne Xin Zhao, Yanwei Guo, Yulan He, Han Jiang, Yuexin Wu, Xiaoming Li: We know what you want to buy: a demographic-based system for product recommendation on microblogs. KDD 2014: 1935-1944

[2] Wayne Xin Zhao, Sui Li, Yulan He, Liwei Wang, Ji-Rong Wen, Xiaoming Li: Exploring demographic information in social media for product recommendation. Knowl. Inf. Syst. 49(1): 61-89 (2016)

[3] Yang Xiao, Wayne Xin Zhao, Kun Wang, Zhen Xiao: Knowledge Sharing via Social Login: Exploiting Microblogging Service for Warming up Social Question Answering Websites. COLING 2014: 656-666

[4] Wayne Xin Zhao, Sui Li, Yulan He, Edward Y. Chang, Ji-Rong Wen, Xiaoming Li: Connecting Social Media to E-Commerce: Cold-Start Product Recommendation Using Microblogging Information. IEEE Trans. Knowl. Data Eng. 28(5): 1147-1159 (2016)

- [5] Jinpeng Wang, Gao Cong, Wayne Xin Zhao, Xiaoming Li: Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets. AAAI 2015: 318-324
- [6] Jinpeng Wang, Wayne Xin Zhao, Haitian Wei, Hongfei Yan, Xiaoming Li: Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs. EMNLP 2013: 1337-1347
- [7] Ningnan Zhou, Wayne Xin Zhao, Xiao Zhang, Ji-Rong Wen, Shan Wang: A General Multi-Context Embedding Model for Mining Human Trajectory Data. IEEE Trans. Knowl. Data Eng. 28(8): 1945-1958 (2016)
- [8] Wayne Xin Zhao, Jin Huang, Ji-Rong Wen: Learning Distributed Representations for Recommender Systems with a Network Embedding Approach. AIRS 2016: 224-236
- [9] Cheng Yang, Maosong Sun, Wayne Xin Zhao, Zhiyuan Liu, Edward Y. Chang: A Neural Network Approach to Joint Modeling Social Networks and Mobile Trajectories. CoRR abs/1606.08154 (2016)
- [10] Wayne Xin Zhao, Ningnan Zhou, Wenhui Zhang, Ji-Rong Wen, Shan Wang, Edward Y. Chang: A Probabilistic Lifestyle-Based Trajectory Model for Social Strength Inference from Human Trajectory Data. ACM Trans. Inf. Syst. 35(1): 8 (2016)
- [11] Wayne Xin Zhao, Jinpeng Wang, Yulan He, Ji-Rong Wen, Edward Y. Chang, Xiaoming Li: Mining Product Adopter Information from Online Reviews for Improving Product Recommendation. TKDD 10(3): 29 (2016)
- [12] Jinpeng Wang, Wayne Xin Zhao, Yulan He, Xiaoming Li: Leveraging Product Adopter Information from Online Reviews for Product Recommendation. ICWSM 2015: 464-472
- [13] Keqiang Wang, Wayne Xin Zhao, Hongwei Peng, Xiaoling Wang: Bayesian Probabilistic Multi-Topic Matrix Factorization for Rating Prediction. IJCAI 2016: 3910-3916
- [14] Yuan Yao, Wanye Zhao Xin, Yaojing Wang, Hanghang Tong, Feng Xu and Jian Lu. Version-Aware Rating Prediction for Mobile App Recommendation. To appear in ACM TOIS.