

基于跨电商行为的交叉推荐算法

张亮^{1,2}, 柏林森¹, 周涛^{2,3}

(1. 北京百分点信息科技有限公司 北京 海淀区 100080; 2. 电子科技大学互联网科学中心 成都 610054;
3. 杭州师范大学阿里巴巴商学院 杭州 310036)

【摘要】利用百分点科技推荐引擎提供的原始数据,分析了用户跨电商的行为,提出了一种可在多个电商之间进行交叉推荐的算法。结果证明,该算法不仅在精确性上较完全冷启动的随机推荐有巨大的提高,而且所推荐的商品可以保持相当的多样性与新颖性。分析显示有约5%~10%的点击、收藏和购买行为发生在有交叉行为的用户身上,这些用户的活跃性明显强于非交叉用户。这些结果暗示交叉用户可能是网上购物的重度用户。该文展现了全新的研究思路,研讨了全新的分析对象,其思路 and 结果对于电子商务研究有重要价值。

关键词 冷启动问题; 交叉推荐; 电子商务; 跨电商行为; 推荐系统

中图分类号 TP301

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.01.031

Crossing Recommendation Based on Multi-B2C Behavior

ZHANG Liang^{1,2}, BAI Lin-sen¹, and ZHOU Tao^{2,3}

(1. Beijing Baifendian Information Technology Co., Ltd. Haidian Beijing 100080;

2. Web Sciences Center, University of Electronic Science and Technology of China Chengdu 610054;

3. Alibaba Business School, Hangzhou Normal University Hangzhou 310036)

Abstract Personalized recommendation has now been widely used in E-commerce, but there are still some problems to be solved such as cold-start problem, data sparsity, diversity-accuracy dilemma and so on. Existing literatures have focused on single data set, lacking a systematic understanding about the accessing behavior involving multiple web sites. Thanks to the real data, provided by Baifendian Information Technology recommendation engine, we analyze users' behavior on multi-B2Cs (business-to-customers) and propose a crossing recommendation algorithm which is able to recommend items of a B2C site to users according to the records of users in other B2C web sites. This algorithm largely improves accuracy compared with purely random recommendation under completely cold-start environment and can still keep high diversity and novelty.

Key words cold-start problem; crossing recommendation; E-commerce; multi-B2C behaviors; recommender systems

随着互联网上信息量的激增,人们对于信息的利用率反而降低了,这种现象称之为信息过载。作为目前解决信息过载问题最有前途的方案,个性化推荐技术在理论与实践上都有着巨大的价值^[1]。已经有许多种类的推荐算法被成功地部署并应用。但是,绝大多数的研究都在单一的数据集上进行,例如:MovieLens作为一个基于Web的研究型推荐系统,收集用户对电影的评分,并提供相应的电影推荐列表^[2]。Amazon的推荐系统为用户推荐可能感兴趣的商品,这些推荐据说带来了超过35%的销售额^[3]。其他的推荐系统还有Ringo^[4]的音乐推荐系统,

Jester^[5]的笑话推荐系统,Phoaks^[6]的网页信息推荐系统等等。

推荐系统虽然得到了广泛应用,但也存在诸多问题,如冷启动问题^[7]、数据稀疏性问题^[8]、如何兼顾算法多样性与精确性问题^[9]等。其中,如何对新商品或新用户进行推荐的冷启动问题最令人头痛:对于一个电子商务网站来说,一个新用户到来时没有携带任何历史信息,也就无法根据他的历史行为做出个性化推荐。与此同时,一个用户作为新用户进入某个网站之前可能已经产生了一些其他的行为。以图1为例,用户②和用户④同时在两个电商网

收稿日期: 2011-12-15; 修回日期: 2012-03-27

基金项目: 国家自然科学基金面上项目(60973069); 中央高校基本科研业务费(ZYGX2010Z001)

作者简介: 张亮(1988-),男,主要从事数据挖掘方面的研究。

站留下了历史行为信息。定义这种在一个以上电商有行为信息的用户为交叉用户。对于用户①这样的非交叉用户, 虽然只有一个电商网站的行为记录, 仍有可能通过其他用户建立他与其他电商商品的联系。目前还没有类似利用跨电商行为进行个性化推荐的相关研究。本文利用为独立电商提供SaaS (Software-as-a-Service)模式服务的百分点科技推荐引擎提供的原始数据, 分析了用户跨电商的行为, 并提出了一种可应用于在多个电商之间进行交叉推荐的算法。



图1 交叉推荐应用场景示意图

1 实证分析

1.1 基本数据

B2C (business to consumer) 是一种以网络零售为主的电子商务模式。为了证实交叉推荐算法的有效性, 本文利用百分点科技的电商消费行为数据平台, 选取了3个比较典型的垂直型B2C网站数据进行实证分析。果皮网(www.guopi.com)专注于护肤美容类商品的网络销售, 麦包包(www.mbaobao.com)是箱包在线直销网站, 名鞋库(www.s.cn)主要进行鞋类商品的在线销售, 同时对其销售品牌旗下的服装和箱包类商品也有所涉及。我们可以确定并记录同一用户在多电商上的行为信息, 这些行为包括用户对商品的点击、收藏和购买活动。以上行为都可以在一定程度上反映出当前用户对于特定商品的偏爱程度。在一般的情况下, 会加权处理不同类型的行为, 以区别不同行为产生影响的大小。但是本文的目的在于初步研究这种跨电商行为的可用价值, 而非提高某些已知算法的精度。故不再考虑权重分配等细节问题而对这3种不同行为赋予同样的权重。

本文将2011年4月份上述三个网站所包含的用户行为信息, 进行了相应的数据预处理, 并进一步划分为果皮网-麦包包(下表中简称G-M)和麦包包-名鞋库(下表中简称M-S)两个独立数据集。为了统一

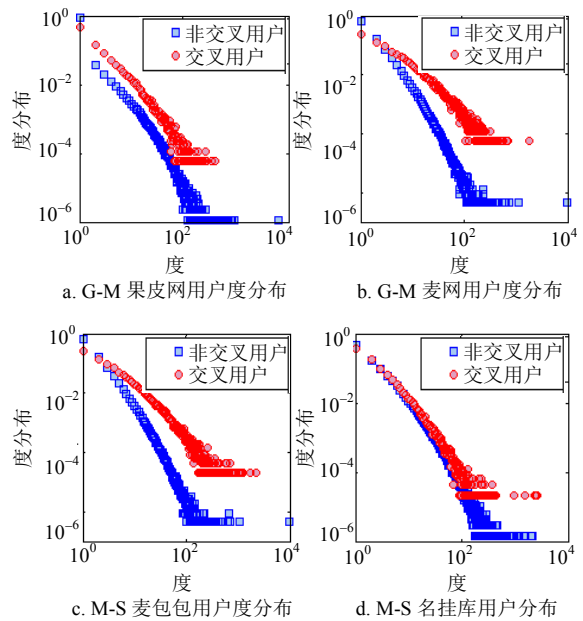
描述不同数据集上的商品信息, 简称数据集第一个电商网站为A电商, 第二个电商网站简称为B电商, 下同。表1中, 独立电商的行为数指仅在单独一个电商有过行为记录的非交叉用户的行为数, 跨电商的行为指在一个以上电商有行为信息的用户为交叉用户的行为数。虽然交叉行为数量并不多, 但是作为两个不同品类的B2C电商, 交叉行为还是有着5%~10%的占比, 这个结果还是很令人吃惊的。

表1 数据集的基本统计信息

数据集	用户数	商品数		行为数			
		A电商	B电商	总行为数	独立A电商	独立B电商	跨电商
G-M	3 128 796	11 940	11 440	6 550 339	1 592 065	4 619 098	339 176 (5.18%)
M-S	2 973 801	11 606	7 140	8 558 177	4 612 698	3 068 686	876 793 (10.25%)

1.2 度分布情况

文献[10-12]利用二部分图建立用户-商品关联关系, 将用户集 U 与商品集 I 之间的关系表示为一个二部分图 $G(X, E)$, 其中顶点集 $X = U \cup I$, 如果用户 U_m 通过点击、收藏和购买行为表现出对于特定商品 I_n 的偏爱, 则为 U_m 与 I_n 建立一条连边。网络中节点的度是指该节点连接的其他节点的数目。图2为在G-M, M-S两个数据集上分别统计果皮网、麦包包和名鞋库三个电商网站用户行为数据中用户与商品的度, 得到的节点度在双对数坐标下的分布图。从图中可以明显的看出用户与商品的长尾特性。交叉用户与非交叉在用户活性、价值、异质性上存在着明显的差异。交叉用户中行为数较多的那一部分用户所占比例更高, 相比非交叉用户具有更强的活跃度。



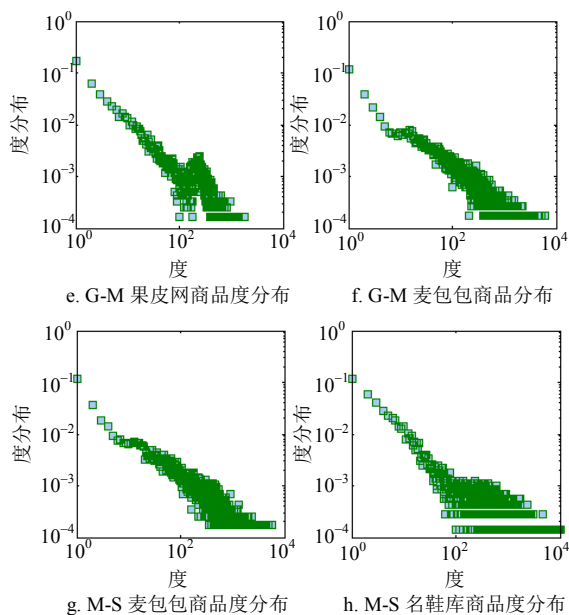


图2 用户与商品的度分布图

2 算法描述

交叉推荐算法关注解决两个问题：1) 对于某用户，系统已经获得其在目标网站以外的其他网站的行为信息，如何在用户首次登陆目标网站后进行推荐；2) 在用户浏览时推荐其他合作垂直电商中该用户可能感兴趣的物品。这两个问题本质上可以划归为同一类问题，即利用经销不同品类的垂直电商的历史信息做跨电商交叉推荐。在一个独立B2C网站发现内部物品之间的关联相对来说是很容易的，但却很难挖掘出跨电商物品之间的关联。交叉推荐算法将通过建立推荐映射表的方式挖掘跨站物品之间的关联强度。

2.1 基于二部分图资源分配的推荐算法

最近，一些研究者将物质扩散与热传导理论引入到个性化推荐算法的研究工作中，提出了众多基于网络结构的推荐算法^[10-15]。本节将简要介绍基于二部分图资源分配的推荐算法，这些方法把用户和物品看成抽象的节点，所有算法利用的信息都藏在用户与物品的选择关系中。该算法假定在一个由 m 个用户和 n 个物品构成的推荐系统中，使用一个具有 $(m+n)$ 个节点的二部分图表示，如果用户 i 选择过物品 j ，则在 i 和 j 之间建立一条连边，记录 $a_{ij}=1(i=1,2,\dots,m;j=1,2,\dots,n)$ ，否则 $a_{ij}=0$ 。对于任意目标用户 i ，推荐算法的目的是把所有 i 没有选择过的物品按照 i 喜欢的程度进行排序，并且把排名靠前的那些物品推荐给 i 。算法的思想在于，假设 i 选择过的物品，都具有某种向 i 推荐其他物品的能力，这

个抽象的能力可以看作位于相关物品上的某种可分配的资源^[16]。对于一个有 m 个用户和 n 个物品的一般推荐系统。如果用 ω_{ij} 表示物品 j 愿意分配给物品 i 的资源比例，可以得到 ω_{ij} 的一般表达式^[10]：

$$\omega_{ij} = \frac{1}{k_j} \sum_{l=1}^m \frac{a_{il}a_{jl}}{k_l}$$

用户 l 的度。对于给定的目标用户，使用他选择过的物品信息构建向量 f ，其中，选择过的物品设定初始资源为1，否则设定为0。这个资源分配构型，携带了不同用户的个性化信息。通过上述公式，最终的资源分配向量可以表示为 $f' = Wf$ 。把目标用户没有看过的所有物品按照向量 f' 中对应元素的大小排序，数值越大表示用户越喜欢，于是系统可以将排序靠前的物品推荐给用户。实验结果显示，基于二部分图资源分配的推荐算法在精确性和多样性都好于经典的协同过滤算法^[10]。

2.2 跨电商交叉推荐算法

跨电商交叉推荐的关键点在于，利用交叉用户的跨电商行为，把对其他电商中每一个物品的访问，等价映射为对目标电商中的若干物品的访问并赋予不同的关联权重。以图3为例，假定全局存在A和B两个B2C网站，分别有A1、A2、A3、A4、B1、B2、…、B6共10件物品和5个不同用户。下面就整个推荐算法流程做详细说明。

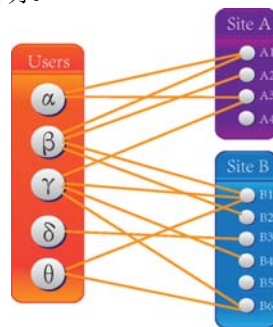


图3 交叉推荐二部分图示意图

2.2.1 建立物品映射表

算法为全局的每一个物品建立一张物品映射表，在推荐系统中，这个过程一般采用离线计算的方式，每积累一定量新数据再进行更新，亦可采用实时增量算法更新，不在本论文讨论范围之内。对应每一个物品。物品映射表的规模为 $S \times (N-1)$ 项。其中 S 为一个物品映射为其他电商物品的最大数目， N 为推荐平台所包含的电商数。遍历所有物品，其中每处理一个物品时，都将此物品初始资源设为1，而

将其他商品初始资源设为0。此时, 由当前商品构成的初始资源向量 $f^{(0)}$ 含有且仅含有一项数值为1。以图3所示的二部分图为基础, 利用2.1节的算法, 可进行一步资源传递, 记为 $f^{(1)} = Wf^{(0)}$ 。如果 $f^{(1)}$ 中对应于除当前商品所属电商外其他每个电商的非零资源数均不小于 S , 则资源传递过程停止, 否则重复进行资源传递过程。由于本文只考虑两个电商交叉推荐, 如果当前商品属于A电商, 其他每个电商就是指B电商。这个计算过程在每一个垂直电商找到满足数量的商品获得由初始商品分配的资源。多次传递过程中, 每一次传递与第一次传递过程类似。如 t 次传递后的资源向量为: $f^{(t)} = Wf^{(t-1)} = W^t f^{(0)}$ 。可能存在一些商品需要相当多的传递次数才能满足条件或经过无限次传递也无法满足条件, 考虑到算法复杂度与系统的硬件资源, 需要设定阈值 M , 当传递次数大于该阈值, 传递过程会被强制停止。参数的选取问题在第5节说明。

结束资源传递后, 对于资源向量中除当前商品所在电商外的其余 $N-1$ 个电商, 分别选出它们商品中资源值排名前 S 的商品(非零值不足 S 的只考虑非零值所对应商品), 将它们的资源值归一化, 保证每个电商被选出商品的资源和都为1。这 $S \times (N-1)$ 个商品及其对应的资源值(权重)就构成了当前商品的映射表。

2.2.2 生成推荐列表

在上一步获得的商品映射表基础之上, 交叉推荐系统将在线实时响应用户的要求。由于在线计算的原因, 这部分计算需要较小的时间复杂度。算法将获得用户的历史访问信息(这里我们假设用户在目标电商网站未曾购买或访问过任何商品, 即纯粹的冷启动用户), 并将每个用户访问过的商品在目标电商网站的商品映射表权重相加, 取权重排序的前 K 个商品推荐给目标用户。

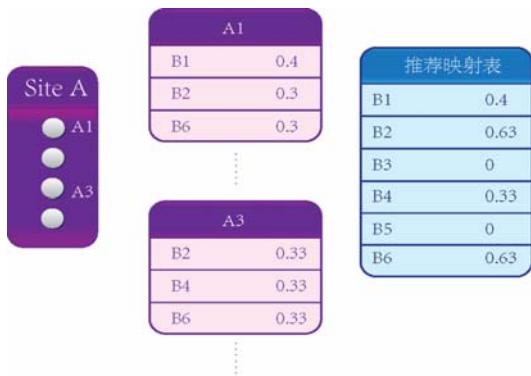


图4 推荐映射表示意图

例如, 对于图3中用户 α , 根据他在A中偏爱的商品A1和A3, 查询A1与A3的推荐映射表在目标网站B的权值, 可以得到如图4所示的权重列表。假设 α 第一次访问B网站, 将B所对应商品的权值相加可以得到针对 α 的推荐映射表, 根据推荐列表长度的实际情况, 系统将依次推荐商品B2、B6、B1、B4。

3 评价指标

必须有相应的指标对算法进行评测, 才可以反映出推荐的质量。但是, 针对特定目的, 清晰的鉴别算法的优劣是一个具有挑战性的课题, 在如何选择恰当的评价指标对算法的表现进行评估方面还没有达成共识^[17-18]。目前现有的评价指标覆盖了包括准确度、覆盖度、多样性、新颖性等众多方面。本文通过离线实验的方法, 将从准确度, 多样性, 新颖性等多个方面对算法进行评测。下面各节将分别介绍这些评价指标。

3.1 准确度

准确度一般可以分为预测准确度和排序准确度两类。分别应用于预测评分的系统和top-K的推荐系统中, 后者在电子商务领域应用更加普遍。在离线实验中, 将使用一个用户与项目交互的数据集, 选择一些用户作为测试用户并隐藏这些用户的一些行为信息。推荐系统将给出 K 个他们可能感兴趣的物品, 此时可以通过准确率(Precision)和召回率(Recall)来度量推荐的准确度^[17-18]。对于某个测试集中的用户 u , 其准确率定义为推荐的 K 个商品中, 与用户 u 相关联的商品数量(用户实际选择但被隐藏的商品数)与推荐列表长度的比值。召回率定义为推荐的 K 个商品中, 与用户 u 相关联的商品数量与用户关联的所有对象数的比值。对于整个系统而言, 系统的准确率和召回率定义为测试集中所有用户的准确度和召回率的平均值。即:

$$\text{Precision} = \frac{1}{U} \sum_u \frac{|R_u \cap T_u|}{|R_u|} \quad \text{Recall} = \frac{1}{U} \sum_u \frac{|R_u \cap T_u|}{|T_u|}$$

式中, 对于用户 u , R_u 是推荐列表中商品的集合; T_u 是测试集中 u 选择过的商品的集合; U 是测试用户数。

3.2 新颖性

除了准确度以外, 新颖性也是影响用户体验的重要指标, 标志向用户推荐非热门商品的能力。如系统推荐了非常流行的商品给用户, 虽然准确度可能很高, 但是用户可能已经购买过或者通过其他渠

道了解到这些商品,因此不会认为这样的推荐系统是有价值的。文献[7]提出利用推荐商品的平均度对推荐的新颖性进行度量。推荐列表中商品的平均度越小,对于用户来说,其新颖程度就越高,定义为:

$$\langle d \rangle = \frac{1}{mK} \sum_{i=1}^n \sum_{O_r \in O_r^i} d(O_r)$$

式中, $K = |O_r^i|$ 为用户 i 的推荐列表长度; m 为用户数; $d(O_r)$ 是收藏对象 O_r 的用户人数。

3.3 多样性

个性化推荐系统的设计宗旨就是针对不同用户的偏好和需求给出不同的推荐,而准确度高的系统不一定能照顾到不同用户的不同需求^[19]。文献[7]提出使用平均海明距离度量推荐列表的多样性,用户 i 与用户 j 的推荐列表的海明距离被定义为:

$$H_{ij} = 1 - \frac{Q_{ij}}{L}$$

式中, K 为推荐列表长度; Q_{ij} 为系统推荐给用户 i 和用户 j 推荐列表中相同商品的数量。平均海明距离定义为 H_{ij} 的平均值 $\langle H \rangle$, 它在数值上介于 0 与 1 之间, 分别表示对所有用户推荐相同的商品和推荐完全不一样的商品。

4 结果及分析

4.1 数据集划分方式

在已有的果品网-麦包包数据集G-M以及麦包包-名鞋库数据集M-S上,对用户进行训练集和测试集的划分。其中,训练集将用于产生商品映射表和推荐列表,测试集将用于进行算法指标的评估。由于交叉推荐的测试集仅能从交叉用户中选取,采取以下的策略进行数据集划分:将单独在电商A或者电商B记录行为信息的非交叉用户全部划归在训练集中;对于交叉用户随机划分为50%~50%两部分,将其中一部分划归训练集,另一部分作为测试集。这样划分的结果保证训练集中一部分为交叉用户,另一部分为非交叉用户,测试集中用户类型全部为交叉用户。在以上划分策略的基础之上,可以得到以下的两组训练集-测试集,分别标示为G-M, M-S。在以上的每一对训练集-测试集分别可以进行根据A中行为推荐B中商品和根据B中行为推荐A中商品两种形式的算法评估。

4.2 仿真推荐结果

与交叉推荐算法(Crank)相比较,本文同时计算了两种对冷启动不敏感的推荐算法:全局度排序

(Grank)和随机推荐(Rrank)算法。Grank根据商品的流行程度,将商品按照度降序排列。该算法具有很低的算法复杂度,但是完全非个性化。Rrank从目标电商网站内部全部商品中随机产生指定数量的商品,推荐给目标用户。

在选定的两组训练集-测试集上,利用训练集生成推荐列表的过程中设定推荐映射表规模 $S = 20$, 最大资源分配次数阈值 $M = 5$ 。在分配次数阈值参数的选取上,已经通过实验验证在数据集G-M上,82.90%的商品可以在 $M = 5$ 的阈值 F 获得20个商品相应的映射表,而获得20个商品的非零资源所需平均资源分配次数为1.67次。在数据集M-S上,93.21%以上的商品可以在 $M = 5$ 的阈值 F 内获得20个商品相应的映射表,平均所需资源分配次数为1.48次。表2给出了针对M-S数据集最大资源分配次数对于映射表中商品数目比例以及这些能够找到20个映射的商品所需要的最大传播次数。可以看出,随着 M 变大所能带来的变化量衰减非常快,超过5的更大的 M 值已经无法提高映射表商品占比,也无法提高推荐精确度。

表2 阈值 M 对于映射表商品占比和最大传播次数的影响

最大资源分配次数 M	获得映射表商品占比/(%)	满足商品最大传播次数
1	49.11	1
2	92.62	2
3	93.20	3
4	93.21	4
5	93.21	4
6	93.21	4
7	93.21	4
8	93.21	4
9	93.21	4

4.2.1 准确率和召回率

在准确率和召回率两个指标上,分别计算了两组训练集-测试集G-M和M-S电商A与电商B相互推荐的4种情况,如图5和图6所示,推荐列表 K 长度取值5至40。Crank在准确度方面较Rrank有巨大提高。

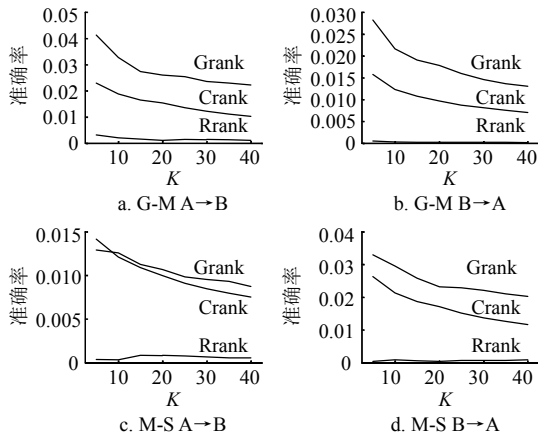


图5 准确率比较

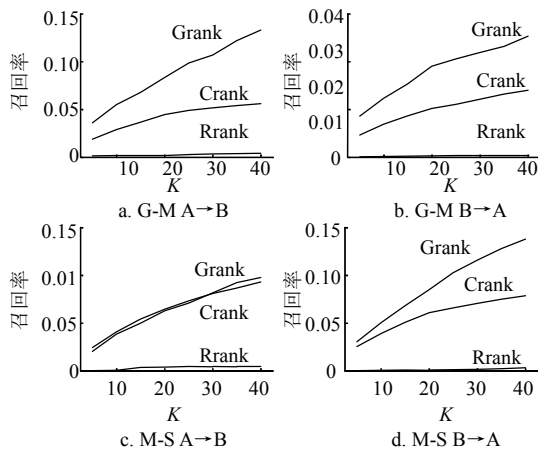


图6 召回率比较

4.2.2 推荐商品平均度

与算法的精确性评估类似, 在新颖性方面同样采用训练集-测试集G-M和M-S, 在电商A与电商B相互推荐的4种情况下计算平均度指标, 如图7所示。推荐列表K长度取值5~40。一般而言, 被推荐的产品平均度越小的系统具有更好的新颖性。Crank在新颖性方面较Grank有较大优势。

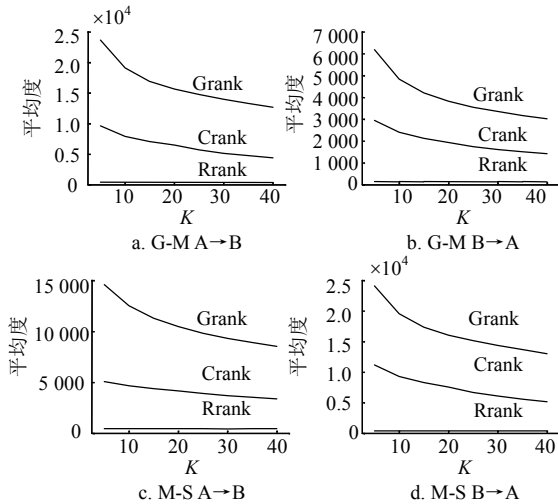


图7 推荐商品平均度比较

4.2.3 推荐商品平均海明距离

表3 平均海明距离比较

数据集	推荐方向	随机推荐		交叉推荐	
		K=10	K=20	K=10	K=20
G-M	A→B	0.999 124	0.998 251	0.922 264	0.918 131
	B→A	0.999 162	0.998 326	0.874 561	0.875 148
M-S	A→B	0.998 599	0.997 199	0.918 739	0.893 807
	B→A	0.999 138	0.998 277	0.894 709	0.884 461

Grank对于每个用户给出完全相同的推荐, 不是一种个性化的推荐方式。根据海明距离计算公式, 该方法平均海明距离数值上等于0; Crank在上述的4种推荐情况下, 平均海明距离约为0.9, 意味着对于任意两个用户, 推荐结果中约有90%的内容是彼此不相同的, 可以认为具有相当好的多样性。平均海明距离比较如表3所示。

5 结论与讨论

冷启动是信息推荐中最困难的问题。尽管对于没有任何已知信息的用户, 网站可以向用户给出按照流行程度排名的推荐, 或者干脆随机生成一些推荐, 但是前者完全无法体现个性化的思想, 后者精确性非常差。比上述两种方法好很多的算法尽管也有不少, 但是巧妇难为无米之炊, 没有用户信息, 也就没有用武之地了。本文另辟蹊径, 考虑用户在多个不同B2C网站访问的交叉数据, 从而对于某网站而言的完全没有信息的用户, 可能已经在其他网站有过很多访问历史, 可以形成有相当精确性和个性化的推荐。事实上, 仿真试验显示, 哪怕只利用一个站外电商数据进行交叉推荐, 其精确度可以数十上百倍超过随机推荐, 与此同时保持和随机推荐相似的多样性和新颖性。通过考虑不同行为的权重, 以及引入更多的站外电商数据, 本算法的精确度可以得到进一步的提高。在描述算法时, 已将算法描述为可以利用多个站外电商行为推荐目标电商的商品, 最近的研究也显示, 增加更多的站外电商数据, 可以大幅度提高推荐精确度, 特别是当目标电商为团购电商的时候, 效果特别好。事实上, 本文的精髓并不在于某种特定的算法(这个方法本身还是不错的), 而是在于提供了全新的可供分析的数据对象和全新的研究视角, 这必将开拓相关研究人员的视野!

个性化将成为未来商业的核心理念和技术^[20]。

个性化时代风雨满楼，呼之欲出！人们现在所能接受的个性化服务，往往只是通过一个网站内部的数据记录得到的。事实上，当你进入一家红酒的网络零售平台时，你在微博上的言论、你在人人网上的朋友圈子和他们的爱好、你在其他电子商务网站浏览的服装和购买的首饰，都蕴涵着你的品味和喜好，泄漏可能你自己都没有觉察的红酒方面的偏好。真正的个性化时代，需要在充分尊重个人隐私的前提下利用用户在不同平台和服务中留下的数据轨迹！本文迈出了这种技术的第一步，其所讨论的问题将成为个性化时代的重要支撑技术。

感谢郭盛收集整理数据，感谢与尚明生、张子柯和张韶峰有益的讨论。

参 考 文 献

- [1] LÜ L, MEDO M, YEUNG C H, et al. Recommender Systems[J]. *Physics Reports*, 2012(519): 1-49.
- [2] DAHLEN B J, KONSTAN J A, HERLOCKER J L, et al. Jumpstarting movielens: user benefits of starting a collaborative filtering system with "dead data", TR 98-017[R]. Minnesota, USA: University of Minnesota.
- [3] LINDEN G, SMITH B, YORK J. Amazon.com recommendations item-to-item collaborative filtering[J]. *IEEE Internet Computing*, 2003, 7(1): 76-80.
- [4] SALTON G. Automatic text processing[M]. New York: Addison-Wesley, 1989.
- [5] GOLDBERG K, ROEDER T, GUPTA D. Eigentaste: a constant time collaborative filtering algorithm[J]. *Information Retrieval J*, 2001, 4(2): 133-151.
- [6] TERVEEN L, HILL W, AMENTO B, et al. PHOAKS: A system for sharing recommendations[J]. *Communications of the ACM*, 1997, 40(3): 59-62.
- [7] ZHANG Z K, LIU C, ZHANG Y C, et al. Solving the cold-start problem in recommender systems with social tags[J]. *EPL*, 2010, 92(2): 28002.
- [8] HUANG Z, CHEN H, ZENG D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filter[J]. *IEEE Trans Information Systems*, 2004, 22(1): 116-142.
- [9] ZHOU T, ZOLTÁN K, LIU J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. *PNAS*, 2010, 107(10): 4511-4515.
- [10] ZHOU T, REN J, MEDO M, et al. Bipartite network projection and personal recommendation[J]. *Phys Rev E*, 2007(76): 046115.
- [11] ZHOU T, JIANG L L, SU R Q, et al. Effect of initial configuration on network-based recommendation[J]. *EPL*, 2008, 81(5): 58004.
- [12] HUANG Z, ZENG D, CHEN H. Analyzing consumer-product graphs: empirical findings and applications in recommender systems[J]. *Management Science*, 2007, 53(7): 1146-1164.
- [13] ZHANG Y C, BLATTNER M, YU Y K. Heat conduction process on community networks as a recommendation model[J]. *Phys Rev Lett*, 2007(99): 154301.
- [14] ZHANG Y C, MEDO M, REN J. Recommendation model based on opinion diffusion[J]. *EPL*, 2007, 80(6): 68003.
- [15] LIU Z, ZHANG Q M, LÜ L, et al. Link prediction in complex networks: a local naïve Bayes model[J]. *EPL*, 2011, 96(4): 48007.
- [16] 刘建国, 周涛, 汪秉宏. 个性化推荐系统研究进展[J]. *自然科学进展*, 2009, 19(1): 1-15.
LIU Jian-guo, ZHOU Tao, WANG Bing-hong. Review on personalized recommendation[J]. *Prog Nat Sci*, 2009, 19(1): 1-15.
- [17] 刘建国, 周涛, 郭强, 等. 个性化推荐系统的评价方法综述[J]. *复杂系统与复杂性科学*, 2009, 6(3): 1-8.
LIU Jian-guo, ZHOU Tao, GUO Qiang, et al. Survey on metrics about personalized recommendation[J]. *Complex Systems and Complexity Sciences*, 2009, 6(3): 1-8.
- [18] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. *电子科技大学学报*, 2012, 41(2): 163-175.
ZHU Yu-xiao, LÜ Lin-yuan. Survey on the metrics about recommender systems[J]. *Journal of University of Electronic Science and Technology of China*, 2012, 41(2): 163-175.
- [19] ZIEGLER C N, MCNEE S M, KONSTAN J A, et al. Improving recommendation lists through topic diversification[C]//www'05 Proceedings of the 14th International Conference on World Wide Web. New York: ACM, 2005.
- [20] 苏萌, 柏林森, 周涛. 个性化: 商业的未来[M]. 北京: 机械工业出版社, 2012.
SU Meng, BO Lin-sen, ZHOU Tao. Personalization: the future of business[M]. Beijing: China Machine Press, 2012.

编辑 税红