

个性化推荐的十大挑战

周 涛

百分点信息科技有限公司

特邀专栏作家

关键词：个性化推荐 挑战

“个性化推荐”对于很多读者来说并不陌生，但其中不乏认识上的误区。有的人觉得个性化推荐就是细分市场和精准营销，实际上细分市场和精准营销往往是把潜在的用户分成很多群体，这固然比基于全体的统计有了长足进步，但是距离“给每一个用户量身定做的信息服务”还有很大差距。所以，个性化推荐其实是细分市场的极致！

信息服务实际上经历了两次理念上的变革，第一次是从总体到群体，第二次是从群体到个体。第二次变革正在进行中，其核心技术就是本文要讨论的个性化推荐技术。

有些读者觉得个性化推荐等同于协同过滤，这可能是因为协同过滤应用比较广泛并且比较容易为大众理解。协同过滤其实只是个性化推荐技术中的一个成员。它与很多更先进技术相比，就好像流行歌曲和经典音乐，前者广受欢迎，而且一般人也可以拿个麦克风吼两声，但是说到艺术高度，流行歌曲就差一些了。当然，流行歌曲的经济价值可能更大，这也是事实。总的来说，协同过滤只是个性化推荐技术中的一款轻武器，远远不等于个性化推荐技术本身。

有些读者可能不了解个性化推荐，我可以先推荐一

些阅读材料。中文综述可以看我们在2009年《自然科学进展》上的文章^[1]。通过这篇文章可以较快地得到很多信息，了解个性化推荐研究的概貌。有了这个基础，如果想要了解突出应用的算法和技术，我推荐项亮和陈义合著的《推荐系统实践》^[2]。百分点信息科技有限公司出版过一本名为《个性化：商业的未来》的小册子^[3]，应用场景和商业模式介绍得比较细致，而在技术上涉及很少，附录里面介绍了一些主流算法和可能的缺陷，或许能够对读者稍有启发。国外的专著建议关注最近出版的两本文献^[4,5]，其中文献^[4]是很多文章的汇总，因为写这些文章的都是达人，所以质量上佳。爱多麦克逊思（Adomavicius）和塔兹林（Tu-

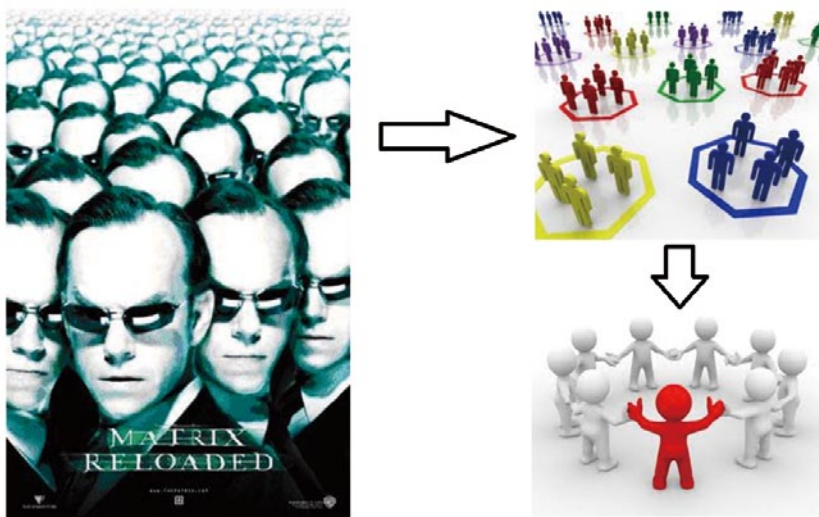


图1 信息服务的两次变革：从总体到群体，从群体到个体

zhilin)的大型综述特别有影响力,不仅系统回顾了推荐系统研究的全貌,还提出了一些有趣的开放性问题^[6]——尽管我个人不是很喜欢他们对于推荐系统的分类方法。我们在2012年也发表了一篇大综述,应该是目前最全面的综述,所强调的不仅仅是算法,还有很多现象和思路^[7]——大家有兴趣不妨看看。

有些读者认为个性化推荐技术的研究已经进入了成熟阶段,没有什么特别激动人心的问题和成果。恰恰相反,现在个性化推荐技术面临很大的挑战,这也是本文力图让大家了解的。接下来进入正题!我将列出十个挑战(仅代表个人观点),其中有些是很多年前就认识到但是没有得到解决的长期问题,有些事实上是不可能彻底解决的,只能提出改良方案,还有一些是最近的研究提出来的焦点问题。特别要提醒读者注意的是,这十个挑战并不是孤立的,一个方向上的突破极有可能使若干重大挑战都取得进展。

挑战一:数据稀疏性问题

现在待处理的推荐系统规模越来越大,用户和商品(也包括其他物品,譬如音乐、网页、文献……)数目动辄百千万计,两个用户之间选择的重叠非常少。如果用用户和商品之间已有的选择关系占所有可能存在的选择关系的比例来衡量系统的稀疏性,那么我们平时研究最多的MovieLens数据集的稀疏度是4.5%,Netflix是1.2%。这些其实已经都是非常密的数据了,Bibsonomy是0.35%,Delicious是0.046%。想想淘宝上号称有近10亿商品,平均而言一个用户能浏览1000件吗,估计不能,所以稀疏度应该在百万分之一或以下的

量级。数据非常稀疏,使得绝大部分基于关联分析的算法(譬如协同过滤)效果都不好。在这种情况下,通过珍贵的选择数据让用户和用户、商品和商品之间产生关联的重要性,往往要比用户之间对商品打分的相关性还重要^[8]。举例来说,一个用户看了一部惊悚片,就暴露了用户的兴趣,并且使其和很多其他看过同样片子的用户关联起来了——至于他给这个片子评价高还是低,反而不那么重要了。事实上,我们最近的分析显示,稀疏数据情况下给同一个商品分别打负分(低评价)和打正分的两个用户是要看作正相关的而非负相关的,就是说负分扮演了“正面的角色”^[9]——我们需要重新严肃地审查负分的作用,有的时候负分作用甚至大于正分^[10]。

对于这个问题从本质上说是无法完全克服的,但是有很多办法可以在一定程度上缓解它的影响。首先可以通过扩散的算法,从原来的一阶关联(两个用户有多少相似打分或者共同购买的商品)到二阶甚至更高阶的关联^[11-13],甚至通过迭代寻优的方法,考虑全局信息导致的关联^[14,15]。这些方法共同的缺点是都建立在相似性本身可以传播的假设上,并且计算量往往比较大;其次在分辨率非常高的情况下,例如考虑单品,数据可能非常稀疏。但是如果把这些商品信息粗粒化,譬如只考虑品类,数据就会立刻变得稠密。如果能够计算品类之间的相似性,就可以帮助进行基

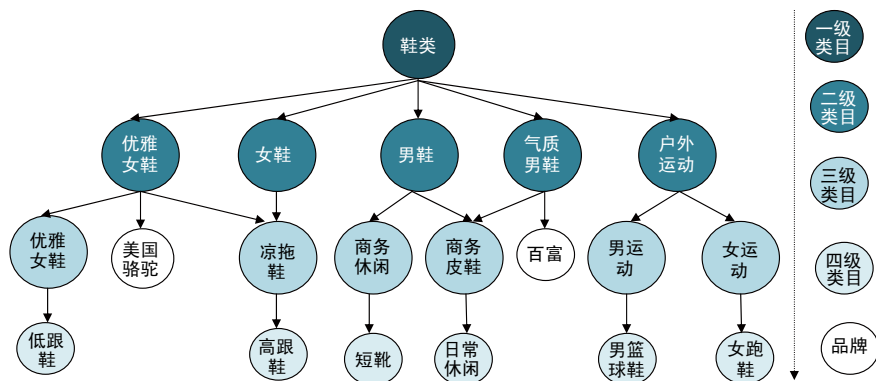


图2 品类树的示意图

于品类的推荐（图2是品类树的示意图）。在语义树方面有过一些尝试^[16]，但是很不成熟，要应用到商品推荐上还需要在理论和技术上的进步。另外，还可以通过添加一些缺省的打分或选择，提高相似性的分辨率，从而提高算法的精确度^[17]。这种添加既可能是随机的，也可能来自于特定的预测算法^[18]。

随机的缺省打分或随机选择为什么会起到正面的作用呢？仅仅是因为提高了数据的密度吗？我认为经过仔细思考会否决这个结论。对于局部热传导的算法^[19]，添加随机连接能够整体把度最小的一些节点的度提高，从而降低小度节点之间度差异的比例（原来度为1的节点和度为3的节点度值相差2倍，现在都加上2，度为3的节点和度为5的节点度值相差还不到1倍），这在某种程度上可以克服局部热传导过度倾向于推荐最小度节点的缺陷。与之类似，随机链接可以克服协同过滤或局部能量扩散算法^[20]过度倾向于推荐最大度节点的缺陷。总之，如果缩小度的比例差异能够在某种程度上克服算法的缺陷，那么使用随机缺省打分就能起到提高精确度的作用。

挑战二：冷启动问题

新用户因为罕有可以利用的行为信息，很难给出精确的推荐。反过来，新商品由于被选择次数很少，也难以找到合适的办法推荐给用户——这就是所谓的冷启动问题。如果我们能够获得商品充分的文本信息并据此计算商品之间的相似性，就可以很好解决冷启动的问题^[21]，比如我们一般不担心图书或者论文推荐会遇到冷启动的问题。不幸的是，大部分商品与图书和文章不同，本身就是丰富的内容。在这种情况下，通过人工或者自动搜索爬取的方法，可以获得商品相应的描述，也会有一定的效果。与之相似，通过注册以及询问得知一些用户的属性信息，譬如年龄、居住城市、受教育程度、性别和职业等等，能够得到用户之间属性的相似度，从而提高冷启动时推荐的精确度^[22,23]。

最近标签系统（tagging systems）的广泛应用提供了冷启动问题潜在的解决方案^[24]。因为标签既可以看作是商品内容的萃取，同时也反映了用户的个性化喜好——比如对《桃姐》这部电影，有的人打上标签“伦理”，有的人打上标签“刘德华”，两个人看的电影一样，但是兴趣点可能不尽相同。当然，利用标签也只能是提高有少量行为的用户的推荐准确性，对于纯粹的冷启动用户，是没有帮助的，因为这些人还没有打过任何标签。系统也可以给商品打上标签，但是这里面没有个性化的因素，效果会打折扣。从这个意义上讲，利用标签进行推荐、激励用户打标签以及引导用户选择合适的标签，是非常重要的^[25]。

要缓解冷启动的问题，一种有效的办法是尽可能快地了解用户的特点和需求，所以如何设计问卷调查本身以及如何利用其中的信息也是一门大学问^[26]。与之相对应，对于一个新商品，怎样让用户，特别是有影响力的用户，给出高质量的评价，对于解决冷启动问题也有重大价值^[27]。如何在保证一定推荐精度的情况下，让新用户和新商品的特性尽快暴露，是一个很有意义的研究难题^[28]。

最近一个有趣的研究显示，新用户更容易选择特别流行的商品^[29]——这无论如何是一个好消息，说明使用热销榜也能获得不错的结果。冷启动问题还可以通过多维数据的交叉推荐和社会推荐的方法予以部分解决，其精确度和多样性又远胜于热销榜，这一点在后面会进一步介绍。

挑战三： 大数据处理与增量计算问题

尽管数据很稀疏，大部分数据都包含数以百万计的用户和商品，与此同时，新商品也不断加入系统，新用户不停进入系统，用户和商品之间还不停产生新的连接。数据量不仅大，而且数据本身还动态变化，如何快速高效处理这些数据成为迫在眉睫的问题。在这个大前提下，算法时间和空间的复杂性，尤其是前者，受到了空前重视。一般而

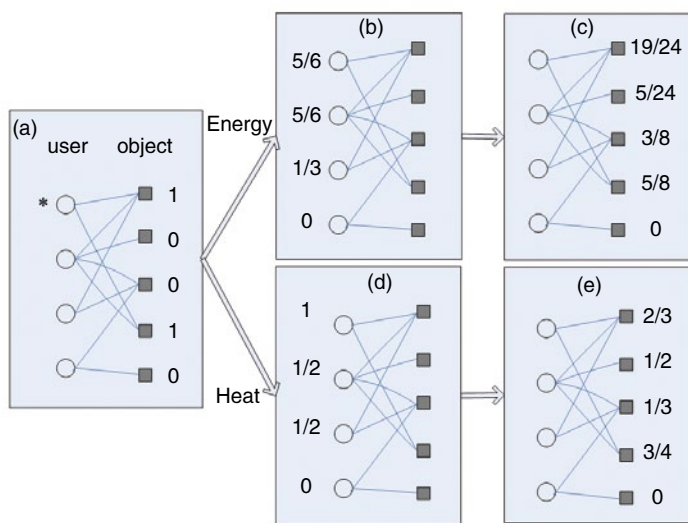


图3 兼顾精确性和多样性的混合扩散推荐算法示意图

言,一个高效的算法,要么自身复杂性很低,要么能够很好地实行并行化,要么两者兼具。

提高算法的效率,有很多途径。大致上可以分为两类:精确算法和近似算法。需要注意的是,精确算法中“精确”一词,并不是指算法的推荐精确度有多大,而是相对于近似算法而言,强调这个算法并不是以牺牲算法中某些步骤的精确性而提高效率的。比如说计算 n 的阶乘,可以有不同的高精度算法,凡是得出最后精确值的就是精确算法,而如果利用斯特林公式进行计算,就属于近似算法了。一般而言,近似算法的效率会明显高于精确算法。

通过巧妙的方法,可以设计出效率很高的精确算法。例如文献[30]设计了一种可以用于潜层狄利克雷分配(latent dirichlet allocation, LDA)算法的新的采样方法,比传统吉布斯采样算法快8倍。文献[31]设计了一种预测算法,只考虑一个用户与其他用户打分的差异以及一个商品与其他商品得分的差异,这个算法远远快于协同过滤算法,却能够得到比标准的基于用户的协同过滤算法更精确的预测效果,其预测精度有时候甚至可以和SVD分解的方法媲美。提高精确算法的另外一条途径就是并行化——很多算法的并行化,一点都不简单。谷歌中国成功将LDA算法并行化并应用于Orkut的推荐

中,取得了很好的效果^[32]。最近文献[33]提出了一种随机梯度下降法,可以并行分解百万行列的矩阵,该方法可以应用在包括推荐在内的若干场景下。

近似算法往往基于增量计算,也就是说当产生新用户、新商品以及新的连接关系时,算法的结果不需要在整个数据集上重新进行计算得到,只需要考虑所增加节点和连边局部的信息,对原有的结果进行微扰,快速得到新结果^[34]。一般而言,随着加入的信息量的增多,这种算法的误差会积累变大,最终每过一段时间还需要利用全局数据重新进行计算。理想的算法是能够保证其误差不会累积,也就是说其结果与利用全部数据重新计算的结果之间的

差异不会单调上升。我们不妨把这种算法叫做自适应算法,它是增量算法的一个加强版本^[35],其设计要求和难度更高。增量算法在业界已经有了应用,比如百分点推荐引擎中的若干算法都采用了增量技术,使得用户每次新浏览、收藏或者购买商品后其推荐列表立刻更新。但是自适应算法目前还只是在比较特殊的算法上才能实现,更谈不上在工业界应用了。

挑战四: 多样性与精确性的两难困境

如果要给用户推荐他喜欢的商品,最“保险”的方式就是给他特别流行的商品,因为这些商品有受到欢迎的可能性更大(否则也不会那么流行),往坏了说,也很难特别被讨厌(不要举凤姐的例子)。但是,这样的推荐产生的用户体验并不一定好,因为用户很可能已经知道这些热销流行的产品,所以得到的信息量很少,并且用户不会认同这是一种“个性化的”推荐。麦克尼(Mcnee)等人已经警告大家,盲目崇拜精确性指标可能会伤害推荐系统——因为这样可能会导致用户得到一些信息量为零的“精准推荐”,并且视野变得越来越狭窄^[36]。事实上,让用

户视野变得狭窄也是协同过滤算法存在的一个比较主要的缺陷。已经有一些实证研究显示,多样性、新颖性、偶然性这些从未获得过如精确性一般重要地位的因素,对于用户体验都十分重要——譬如用户希望音乐推荐更多样更偶然^[37]。与此同时,应用个性化推荐技术的商家也希望推荐中有更多的品类出现,从而激发用户新的购物需求。多样性和新颖性的要求在大多数情况下具有一致性,一些商家更喜欢引导用户关注一些销量一般的长尾商品(这些商品的利润往往更多),这种新颖性的要求往往和多样性的要求一致。还有一些特别的需求非常强调多样性和新颖性,比如类似于“唯品会”这样的限时抢购模式或者最近非常流行的团购模式,广受欢迎的热销商品很快就抢购/团购一空,推荐引擎能够发挥作用的只能是推荐那些不太被主流消费者关注的小众产品。对于新浪微博这类的社会网络,相当一部分新用户很快就不活跃了,在很大程度上是因为得不到其他人关注。与此类似,在世纪佳缘和百合网这类网站中,一个用户如果很长时间没有机会得到任何异性的青睐,也会失去动力。在这种情况下,我们要考虑的不仅仅是向某些用户推荐,而是如何把一些至少目前还不是特别受欢迎的用户推荐出去——这时候,在多样性和新颖性上表现出色的算法意义更大。最近乌干德尔

(Ugander)等人的工作显示,一个用户要向其他若干用户推广某种互联网活动,在同等推广力度下(用推荐的朋友数目衡量),其所选择的推荐对象的结构多样性越大,效果往往越好^[38]。

保证推荐的多样性很有价值,但是,推荐多样的商品和新颖的商品与推荐的精确性之间存在矛盾。前者存在很大风险——推荐一个没什么人看过或者打分较低的东西,很可能被用户憎恶,从而效果更差。很多时候,这是一个两难的问题,只能通过牺牲多样性来提高精

确性,或者牺牲精确性来提高多样性。一种可行之策是直接对算法得到的推荐列表进行处理,从而提升其多样性^[39~41]。文献[41]就是在推荐算法得到的前N个商品中进行一次组合优化,找出L个商品($L < N$),使得这L个商品两两之间平均相似度最小。目前,百分点推荐引擎所使用的方法与之类似。这种方法固然在应用上是有效的,但是没有任何理论的基础和优美性可言,只能算一种野蛮而实用的招数。更好的办法是在设计算法的同时,就考虑到推荐的多样性和精确性,这可以通过精巧混合能量扩散和热传导算法^[19],或者利用有偏的能量扩散^[42]和/或有偏的热传导来实现^[43]。图3是能量扩散与热传导混合算法的示意图。这个算法认为目标用户选择过的商品具有一定的“推荐能力”,在能量扩散过程中它们被赋予初始能量1,而在热传导的过程中它们被认为是初始热源,具有温度1。能量扩散是一个守恒的过程,每一个时间步节点上的能量都均匀分配给所有邻居(图3上半部分);与之相对,热传导过程中每一个节点下一个时间步的温度等于其所有邻居温度的平均值(图3下半部分)。前者倾向于推荐热门商品,后者倾向于挖掘冷门商品,两相结合,精确性和多样性都能明显提高^[19]。尽管上面提到的这些算法效果很好,似乎也

		商 品						
		1	2	3	4	5	6	7
用 户	a	+	-		+	+		+
	b	-	+	+	-	-		-
	c	+	-	+		-	-	-
	d	-	+	+	-			
	e	-		-	-	-		-
	f	+	-	+	+	+		+
	g		-	+	+	-	-	+
	h	+	-	+	+	+		?
	i	+	-	+		-	-	-
	j	-	+	+	-			-
	k	-		-	-	-		-
	l	+	-	+	+	+		-
	m		-	+	+	-	-	-

真实评分

攻击对象

虚假评分

图4 对推荐系统实施攻击的示意图

能够比较直观地进行理解,但是我们还没有办法就相关结果提供清晰而深刻的见解。多样性和精确性之间错综复杂的关系和隐匿其后的竞争,到目前为止还是一个棘手的难题。

挑战五: 推荐系统的脆弱性问题

受推荐系统在电子商务领域重大的经济利益的驱动,一些心怀不轨的用户通过提供一些虚假恶意的信息,故意增加或者压制某些商品被推荐的可能性^[44]。图4展示了一个攻击的实例。假设我们现在要决定是否向用户h推荐商品7,如果系统中只有那些合法用户(a~g),通过上表我们发现用户a和f与用户h的品味比较相似,由于用户a和f都喜欢商品7,那么系统应该把商品7推荐给用户h。如果受到恶意攻击,系统会发现大多数由攻击者生成的虚假用户(i~m)的品味都与用户h相似,并且他们对商品7都给了负面的评价,那么在这种情况下,系统就不会把商品7推荐给用户h。这样一来,就达到了那些攻击者降低对商品7推荐可能性的目的。上面的例子仅仅是众多攻击方案中比较简单的一例,布克(Burke)等人2011年的研究报告中就分析了4大种类8种不同的攻击策略^[45]。此外,攻击者还通过将攻击对象和热销商品或特定用户群喜欢的商品,绑定而提高攻击效果,甚至通过持续探测猜测系统的计算相似性的算法,有针对性地开展攻击。

从上面的介绍可以看出,一个推荐算法能否在一定程度上保持对恶意攻击的鲁棒性,是一个需要认真思考的问题。以最简单的关联规则挖掘算法为例,Apriori算法的鲁棒性就远胜于k近邻算法^[46]。有一些技术已经被设计出来提高推荐系统面对恶意攻击的鲁棒性,譬如通过分析对比真实用户和疑似恶意用户之间打分行为模式的差异,提前对恶意行为进行判断,从而阻止其进入系统或赋予疑似恶意用户比较低的影响力^[47~49]。总的来说,这方面的研究相对较少,缺乏系统性的分析,反而是攻击策略层出不穷,有一种“道高一尺,魔高一丈”的感觉。

挑战六: 用户行为模式的挖掘和利用

深入挖掘用户的行为模式有望提高推荐的效果或在更复杂的场景下进行推荐。例如,新用户和老用户具有很不一样的选择模式:一般而言,新用户倾向于选择热门的商品,而老用户对于小众商品关注更多^[29],新用户所选择的商品相似度更高,老用户所选择的商品多样性较高^[50]。上面曾经介绍过的能量扩散和热传导的混合算法^[19]可以通过一个单参数调节推荐结果的多样性和热门程度。在这种情况下,可以考虑给不同用户赋予不同的参数(从算法结果的个性化到算法本身的个性化),甚至允许用户自己移动一个滑钮调节这个参数——当用户想看热门的时候,算法提供热门推荐;当用户想找点很酷的产品时,算法也可以提供冷门推荐。

用户行为的时空统计特性也可以用于提高推荐或者设计针对特定场景的应用(关于人类行为时空特性的详细分析请参考文献[51])。最简单的例子是,在推荐前考虑用户从事相关活动随时间变化的活跃性。例如,在进行手机个性化阅读推荐的时候,如果以往的数据显示某个用户只在7点到8点之间有1个小时左右的手机阅读行为(可能是上班途中的地铁或者公交车上),那么9点钟发送一个电子书阅读的短信广告就是很不明智的选择。从数据中还可以分析出影响用户选择的长期和短期的兴趣,通过将这两种效应分离出来,可以显著提高推荐的精确度^[52~54]。事实上,简单假设用户兴趣随时间呈指数递减,也能够得到改进的推荐效果^[55,56]。随着移动互联网的飞速发展以及GPS及其他手机定位技术的发展和普及,基于位置的服务成为一个受到学术界和业界广泛关注的问题。基于位置信息的推荐可能会成为个性化推荐的一个研究热点和重要的应用场景,而这个问题的解决需要对用户的移动模式有深入的理解,包括预测用户的移动轨迹和判断用户在当前位置是否有可能进行餐饮购物活动等^[57,58],同时还要有定义的方法去定义用户之间以及地点之间的相似性^[59,60]。事实上,

评价指标	名称	符号	偏好	是否依赖于推荐列表长度	备 注
预测评分 准确度	平均绝对误差	MAE	小	否	适用于比较关注精确的预测评分的系统
	平均平方误差	MSE			
	均方根误差	RMSE			
	标准平均绝对误差	NMAE			
预测评 分关联 准确度	Pearson关联	PCC	大	否	适用于不关注精确预测评分的系统，其中NDMP适用于弱排序
	Spearman关联	ρ	大		
	Kendall's Tau	τ	大		
	基于距离的标准指标	NDMP	小		
分类 准确度	准确率	$P(L)$	大	是	除AUC外，其他不适用于没有明确二分喜好的系统
	召回率	$R(L)$		是	
	准确率提高率	$e_p(L)$		是	
	召回率提高率	$e_R(L)$		是	
	F1指标	$F_1(L)$		是	
	ROC曲线面积	AUC		否	
排序准确度	平均排序分	RS	小	否	适用于对推荐排序要求严格的系统
基于排序 加权的指标	半衰期效用指标	HL(L)	大	是	考虑了具体的推荐排序值，更合理些
	折扣累计利润	DCG(b,L)			
	排序偏差准确率	RBP(p,L)			
覆盖率	预测覆盖率	COV _p	大	否	这些指标单独使用没有意义，应与准确度指标一起考虑。欲计算种类覆盖率指标需要先对商品种类分类
	推荐覆盖率	COV _R (L)		是	
	种类覆盖率	COV _C		是	
多样性	Inter-user diversity	H(L)	大	是	这些指标单独使用没有意义，应与准确度指标一起考虑。欲计算种类覆盖率指标需要先对商品种类分类
	Intra-user diversity	I(L)	小		
新颖性	推荐商品平均度	N(L)	小	是	
	系统的自信息量	U(L)	大		
	推荐的新颖率	UE	大		
	考虑排序的推荐新颖率	UER	大		

图5 推荐系统评价指标

即便把位置信息简单地作为一个单独属性加以考虑，也可以明显提高广告推荐^[61]和朋友推荐^[62]的精确度。特别要提醒各位读者，知道了用户的位置信息，并不意味着可以随时向用户推荐近处的餐饮购物等场所，因为频繁而不精确的推荐会让用户有一种被窥探和骚扰的感觉。从这个意义上讲，把握推荐的时间和地点非常重要！一般而言，在用户经常出没的地点，比如工作地点、学校、住家等，进行推荐，效果往往是比较差的。因为用户对于这些地点比系统还熟悉，而且很难想象用户在上下班的路上会有特别地情致购物或者进餐。实际上可以预测的时空信息往往是商业价值比较低的，而用户在吃饭时间去了一个不常去的地方，往往有更大的可能是和朋友聚会就餐。这就要求系统更加智能，能够对用户当前行为所蕴含的信息量进行估计（要同时考虑时间和空间），并且在信息量充分大的时候进行推荐。

另外，不同用户打分的模式也很不一样^[63,64]，用户针对不同商品的行为模式也不一样^[65,66]（想象你在网上下载一首歌和团购房子时的区别），这些模式都可以挖掘刻画并用于提高推荐的效果。总而言之，推荐引擎要做的是针对合适的对象在合适的时间和合适的地方推荐合适的内容（4S标准）。通过分析不同用户在选择、评价、时间、空间等行为模式上的不同，我们最终的目的是猜测目标用户当前的意图，并且针对不同的意图匹配或组合不同的算法结果，将其推荐

给用户。这不仅需要更高级的数据分析能力，还需要有经验丰富了解业务逻辑的工作人员配合完成。这种称为“情境计算”的思路，有可能较大程度提高推荐及其他信息服务（比如搜索）的质量，百分点的推荐引擎就是在这种思路下架构的^[67]。

挑战七：推荐系统效果评估

推荐系统的概念提出已经有几十年了，但是怎么评价推荐系统，仍然是一个很大的问题。常见的评估指标可以分为准确度、多样性、新颖性和覆盖率四大类。每一类下辖很多不同的指标，例如准确度指标又分为预测评分准确度、预测评分关联、分类准确度、排序准确度四类。以分类准确度为例，又包括准确率、召回率、准确率提高率、召回率提高率、F1指标和AUC值。图5总结了文献中曾经出现过的几乎所有的推荐系统指标^[68]。之所以对推荐

系统的评价很困难，是因为这些指标之间并不是一致的。一般而言，多样性、新颖性和覆盖率之间一致性较好，但不绝对，而这三者往往都和准确度有冲突。如前所述，解决多样性和精确性之间的矛盾本身就是一个重大的挑战！更不幸的是，即便是同一类指标，其表现也不完全一致。举个例子说，一些基于SVD分解的算法，以降低均方根误差（参考图5）为目标，在预测评分精确性方面表现上佳，但是在推荐前L个商品的准确率和召回率（见图5）方面则表现得很不如人意，有些情况下甚至还不如直接按照流行度排序的非个性化算法^[69]。

图5介绍的都是基于数据本身的指标，可以认为是第一层次。实际上，在真实应用时，更为重要的是另外两个层次的评价。第二个层次是商业应用上的关键表现指标，比如受推荐影响的转化率、购买率、客单价以及购买品类数等等，第三个层次是用户的真实体验。绝大部分研究只针对第一个层次的评价指标，而业界真正感兴趣的是第二个层次的评价（比如到底是哪个指标或者哪些指标组合的结果能够提高用户购买的客单价），而第三个层次最难，没人能知道，只能通过第二层次的效果来估计。如何建立第一层面和第二层面指标之间的关系，就成为了关键问题，这一步打通了，理论和应用之间的屏障就通了一大半。

挑战八：用户界面与用户体验

这个问题更多的不是一个学术性质的问题，而是真实应用中的问题。10年前就有学者指出^[70,71]，推荐结果的可解释性对于用户体验有至关重要的影响——用户希望知道这个推荐是怎么来的。基于相



图6 Facebook页面上用户注意力集中的区域的分布，其中红色的区域是用户注意力最集中的区域，黄色次之，绿色再次之，白色最次。

似性的推荐（比如协同过滤）在这个问题上具有明显的优势，例如，亚马逊基于商品的协同过滤的推荐^[72]，在发送推荐的电子邮件时会告诉用户之所以向其推荐某书，是因为用户以前购买过某些书，新浪微博基于局部结构相似性的“关注对象推荐”^[73]在推荐的同时，会说明你的哪些朋友也关注过它们。而矩阵分解或者集成学习算法就很难向用户解释推荐结果的起源。一般而言，用户更喜欢来自朋友的推荐而不是系统的推荐，这一点在后面讲社会推荐的时候还会详细提到。另外，好的界面设计能够让用户觉得推荐的结果更加多样化^[74]，更加可信^[75]。

实际应用时，推荐列表往往包括很多项，这些推荐项最好能够区分成很多类别，不同类别往往来自于不同的推荐方法。例如，看过还看过（浏览过本商品的客户还浏览过的商品）、买过还买过（购买过本商品的客户还购买过的商品）、看过最终购买（浏览过本商品的客户最终购买的商品）、个性化热销榜（个性化流行品推荐）、猜你喜欢（个性化冷门商品推荐）等等。当然，每个推荐项呈现的

结果往往都来自复杂的算法，绝不仅仅像它们的名字听起来那么简单。另外，推荐栏呈现的位置对于推荐的结果影响也很大，因为同一个网页上不同位置对于用户注意力的吸引程度大不一样。图6给出了EyeTrackShop针对Facebook个人页面不同位置受关注程度的示意，可以看出，不同的位置受到的关注相差很大。

如何更好地呈现推荐，是一个很难建立理论模型和进行量化的问题。对于不同的被推荐品而言，用户界面设计的准则也可能大不相同。建立一个可以进行A/B测试的系统（随机将用户分为两部分，各自看到不同的推荐页面和推荐结果），可以积累重要的实验数据，指导进一步的改进工作。

挑战九：多维数据的交叉利用

目前网络科学研究一个广受关注的概念是具有相互作用的网络的结构和功能。网络与网络之间的相互作用大体可以分成三类：一类是依存关系^[76]，譬如电力网络和互联网，如果发生了大规模停电事故，当地的自主系统和路由器也会受到影响，导致网络局部中断；第二类是合作关系^[77]，比如人的一次出行，可以看作航空网络、铁路网络和公路网络的一次合作；第三类是交叠关系^[78]，主要针对社会网络，这也是我们最关注的。我们几乎每个人都参与了不止一个大型的社会网络中，比如你可能既有新浪微博的账号，又是人人网的注册用户，而且还使用手机，那么你已经同时三个巨大的社会网络中了。与此同时，你可能还经常在淘宝、京东、麦包包、1号店、库巴网等地方进行网购，因此也是若干张用户-商品图中的成员。

想象一下，如果能够把这些网络数据整合起来，特别是可以知道每个节点身份的对应关系（不需要知道你的真实身份，只需要知道不同网络中存在的若干节点是同一个人），那么这会带来多么巨大的社会经济价值。举个例子，你可能已经在新浪微博上关注了很多数据挖掘达人的微博，并且分享了很多算法学习的心得和问

题。当你第一次上当当网购书的时候，如果主页向你推荐数据挖掘的最新专著并附有折扣，你会心动吗？交叠社会关系中的数据挖掘，或称多维数据挖掘，是有望真正解决系统内部冷启动问题的终极法宝——只要用户在系统外部的其他系统有过活动。单纯从个性化商品推荐来讲，可以利用用户在其他电商的浏览购买历史来提高在目标电商推荐的精确度——当然，每一个电商既是付出者，也是获利者，总而言之，大家能够通过提高用户体验和点击深度实现共赢。与此同时，可以利用微博和其他社会网络的活动提高商品推荐的精度，反过来还可以利用商品浏览历史来提高微博关注对象推荐的精度。给一个经常购买专业羽毛球和浏览各种专业羽毛球设备的用户推荐关注羽毛球的专业选手和业余教练的成功率应该很高，而且不会陷入“总在一个圈子里面来回推荐”的问题中。

从机器学习的角度，杨强等人提出的“迁移学习”算法有望解决这种跨邻域的推荐^[79]，因为这种算法最基本的假设就是在一个领域所学习的知识在其他领域也具有一定的普适性。诺瑞（Nori）等人最近的分析显示^[80]，在某系统中特定的行为（比如说在Delicious上收藏标签）可用于预测另外系统中的特定行为（比如说在Twitter上的信息选择）。其核心思想与杨强等人的一致。这种跨网的学习事实上已经被证明可以提高链路预测的效果^[81,82]。尽管有了上面的有利证据，我们还是需要特别注意，迁移学习不仅在不同领域间的效果差别很大，而且还依赖于相关系统内部连接产生的机制，并不是都能产生良好的效果，原因在于有时在一个商品品类上表现出高相似性的用户在另外一些商品品类上可能会表现出完全不同的偏好^[83]。

我们分析了百分点科技服务客户的真实数据，发现有很多用户具有交叉购物的习惯，即在多个独立的B2C电商中有浏览和购买行为，如图7所示^[84,85]。即便只考虑两个点，例如利用麦包包的浏览购买数据为名鞋库的用户进行个性化推荐（这些用户为名鞋库上是没有任何历史记录的新



图7 用户跨多个独立B2C电商网站浏览购物的示意图

用户，但是在麦包包上有浏览购买行为），就可以明显提高推荐的准确度（比完全冷启动的随机推荐高数十倍）^[84]，而如果利用3家或以上的外部电商的数据，其推荐的精确度可以明显高于热销榜（注意，热销榜一点个性化都没有），特别在团购类网站上表现非常好^[85]。拥有交叉用户使得不同系统之间的“迁移”更加容易（注意，“迁移学习”原始的方法^[79]不要求系统之间具有相同的用户和商品），萨赫比（Sahebi）和科恩（Cohen）最近测试了那些同时评价了书和电影的用户，也发现利用对书的评分信息可以在很大程度上预测出对电影的评分^[86]。虽然针对多维数据挖掘的研究刚刚起步，但我相信其必将成为学术研究和商业应用上的双料热点和双料难点。

挑战十：社会推荐

很早以前，研究人员就发现，用户更喜欢来自朋友的推荐而不是被系统“算出来的推荐”^[87]。社会影响力被认为比历史行为的相似性更加重要^[88,89]，例如通过社会关系的分析，可以大幅度提高从科研文献^[90]到网购商品^[91]，从个人博客^[92]到手机应用软件^[93]推荐的精确度。最近有证据显示，朋友推荐也是淘宝商品销售一个非常重要的驱动力量^[94]。来自朋友的社会推荐有两方面

的效果：一是增加销售（含下载、阅读……）^[95]，二是在销售后提高用户的评价^[96]。当然，社会推荐的效果也不是我们想象得那么简单：文献[95]在同一篇论文中指出针对不同类型的商品社会推荐的效果大不一样；文献[97]指出不同类型的社会关系对推荐的影响也不同；文献[98]指出朋友的负面评价影响力要大于正面评价等等。

在社会推荐方向存在的挑战主要可以分为三类：一是如何利用社会关系提高推荐的精确度^[89,99]；二是如何建立更好的机制以促进社会推荐^[100~102]；三是如何将社会信任关系引入到推荐系统中^[103~107]。社会推荐的效果可能来自于类似口碑传播的社会影响力，也可能是因为朋友之间本来就具有相似的兴趣或者兴趣相投更易成为朋友，对这些不同的潜在因素进行量化区别，也属学术研究的热点之一^[108]。■



周涛

百分点信息科技有限公司预研部，主要研究方向为推荐算法设计与分析。
zhutou@ustc.edu

参考文献

- [1] 刘建国, 周涛, 汪秉宏, 个性化推荐系统的研究进展, 自然科学进展19 (2009) 1 ~ 15
- [2] 项亮, 陈义, 推荐系统实践, 图灵出版社, 2012
- [3] 苏萌, 柏林森, 周涛, 个性化: 商业的未来, 机械工业出版社, 2012
- [4] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, Recommender Systems Handbook: A Complete Guide for Scientists and Practitioners, Springer, 2011
- [5] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, Recommender Systems: An Introduction. Cambridge University Press, 2011
- [6] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering 17 (2005) 734 ~ 749

- [7] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender Systems, Physics Reports (to be published).<http://dx.doi.org/10.1016/j.physrep.2012.02.006>
- [8] M.-S. Shang, L. Lü, W. Zeng, Y.-C. Zhang, T. Zhou, Relevance is more significant than correlation: Information filtering on sparse data, EPL 88 (2009) 68008
- [9] W. Zeng, Y.-X. Zhu, L. Lü, T. Zhou, Negative ratings play a positive role in information filtering, Physica A 390 (2011) 4486 ~ 4493
- [10] J. S. Kong, K. Teague, J. Kessler, Just Count the Love-Hate Squares: a Rating Network Based Method for Recommender Systems, in: Proceedings of the KDD Cup' 11. ACM Press, New York, 2011
- [11] Z. Huang, H. Chen, D. Zeng, Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering, ACM Transactions on Information Systems 22 (2004) 116 ~ 142
- [12] T. Zhou, R.-Q. Su, R.-R. Liu, L.-L. Jiang, B.-H. Wang, Y.-C. Zhang, Accurate and diverse recommendations via eliminating redundant correlations, New Journal of Physics 11 (2009) 123008
- [13] J.-G. Liu, T. Zhou, H.-A. Che, B.-H. Wang, Y.-C. Zhang, Effects of high-order correlations on personalized recommendations for bipartite networks, Physica A 389 (2010) 881 ~ 886.
- [14] J. Ren, T. Zhou, Y.-C. Zhang, Information filtering via self-consistent refinement, EPL 82 (2008) 58007
- [15] D. Sun, T. Zhou, J.-G. Liu, R.-R. Liu, C.-X. Jia, B.-H. Wang, Information filtering based on transferring similarity, Physical Review E 80 (2009) 017101
- [16] 田久乐, 赵蔚, 基于同义词词林的词语相似度计算方法, 吉林大学学报(信息科学版) 28 (2010) 602 ~ 608
- [17] J. S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998, 43 ~ 52
- [18] I. Esslimani, A. Brun, A. Boyer, Densifying a behavioral recommender system by social networks link prediction methods, Social Network Analysis and Mining 1 (2011) 159 ~ 172
- [19] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J.R. Wakeling, Y.-C. Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems, Proceedings of the National Academy of Sciences of the United States of America 107 (2010) 4511 ~ 4515
- [20] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Physical Review E 76 (2007) 046115
- [21] M. J. Pazzani, D. Billsus, Content-Based Recommendation Systems, Lect. Notes Comput. Sci. 4321 (2007) 325 ~ 341
- [22] A. I. Schein, A. Popescul, L. H. Ungar, D. M. Pennock, Methods and metrics for cold-start recommendations, in: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, New York, 2002, 253 ~ 260
- [23] X. N. Lam, T. Vu, T. D. Le, A. D. Duong, Addressing cold-start problem in recommendation systems, in: Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, 2008, 208 ~ 211
- [24] Z.-K. Zhang, C. Liu, Y.-C. Zhang, T. Zhou, Solving the cold-start problem in recommender systems with social tags, EPL 92 (2010) 28002
- [25] Z.-K. Zhang, T. Zhou, Y.-C. Zhang, Tag-Aware Recommender Systems: A State-of-the-Art Survey, Journal of Computer Science and Technology 26 (2011) 767~777
- [26] A. De Bruyn, J. C. Liechty, E. K. R. E. Huizingh, G. L. Lilien, Offering online recommendations with minimum customer input through conjoint-based decision aids, Marketing Science 27 (2008) 443~460
- [27] S. S. Anand, N. Griffiths, A Market-based Approach to Address the New Item Problem, in: Proceedings of the 5th ACM Conference on Recommender Systems, ACM Press, New York, 2011, pp. 205~212
- [28] J.-L. Zhou, Y. Fu, H. Lu, C.-J. Sun, From Popularity to Personality—A Heuristic Music Recommendation Method for Niche Market, Journal of Computer Science and Technology 26 (2011) 816~822
- [29] C.-J. Zhang, A. Zeng, Behavior patterns of online users and the effect on information filtering, Physica A 391 (2012) 1822~1830
- [30] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, M. Welling, Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2008, pp. 569~577
- [31] F. Cacheda, V. Carneiro, D. Fernández, V. Formoso, Comparison of Collaborative Filtering Algorithms Limitations of Current Techniques and Proposals for Scalable, High-Performance Recommender Systems, ACM Transactions on the Web 5 (2011) 2.
- [32] W.-Y. Chen, J.-C. Chu, J. Luan, H. Bai, Y. Wang, E. D. Chang, Collaborative Filtering for Orkut Community: Discovery of User Latent Behavior, in: Proceedings of

- the 18th International Conference on World Wide Web, ACM Press, New York, 2005, pp. 681-690.
- [33] R. Gemulla, P. J. Haas, E. Nijkamp, Y. Sismanis, Large-scale matrix factorization with distributed stochastic gradient descent, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2011, pp. 569-577.
- [34] B. Sarwar, J. Konstan, J. Riedl, Incremental singular value decomposition algorithms for highly scalable recommender systems, in: International Conference on Computer and Information Science, 2002, pp. 27-28.
- [35] C.-H. Jin, J.-G. Liu, Y.-C. Zhang, T. Zhou, Adaptive information filtering for dynamics recommender systems, arXiv:0911.4910.
- [36] S. M. Mcnee, J. Riedl, J. A. Konstan, Being accurate is not enough: how accuracy metrics have hurt recommender systems, in: Proceedings of the CHI'06 Conference on Human Factors in Computing Systems, ACM Press, New York, 2006, pp. 1097-1101.
- [37] Y. C. Zhang, D. O. Seaghdha, D. Quercia, T. Jambor, Auralist: introducing serendipity into music recommendation, in: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, ACM Press, New York, 2012, pp. 13-22.
- [38] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg, Structural diversity in social contagion, Proceedings of the National Academy of Sciences of the United States of America 109 (2012) 5962-5966.
- [39] B. Smyth, P. McClave, Similarity vs. diversity, in: D.W. Aha, I. Watson (Eds.), Case-Based Reasoning Research and Development, Springer, 2001, pp. 347-361.
- [40] C.-N. Ziegler, S.M. Mcnee, J.A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, in: Proceedings of the 14th International Conference on World Wide Web, ACM Press, New York, 2005, pp. 22-32.
- [41] N. Hurley, M. Zhang, Novelty and diversity in top-N recommendation—analysis and evaluation, ACM Transactions on Internet Technology 10 (2011) 14.
- [42] L. Lü, W. Liu, Information filtering via preferential diffusion, Physical Review E 83 (2011) 066119.
- [43] J.-G. Liu, T. Zhou, Q. Guo, Information filtering via biased heat conduction, Physical Review E 84 (2011) 037101.
- [44] B. Mobasher, R. Burke, R. Bhaumik, C. Williams, Towards trustworthy recommender systems: an analysis of attack models and algorithm robustness, ACM Transactions on Internet Technology 7 (2007) 23.
- [45] R. Burke, M. P. O'mahony, N. J. Hurley, Robust Collaborative Recommendation, in: F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), Recommender Systems Handbook, Part 5, Springer, 2011, pp. 805-835 (Chapter 25).
- [46] J. J. Sandvig, B. Mobasher, R. Burke, Robustness of collaborative recommendation based on association rule mining, in: Proceedings of the 2007 ACM Conference on Recommender Systems, ACM Press, 2007, pp. 105-112.
- [47] S. K. Lam, D. Frankowski, J. Riedl, Do You Trust Your Recommendations? An Exploration of Security and Privacy Issues in Recommender Systems, Lecture Notes in Computer Science 3995 (2006) 14-29.
- [48] P. Resnick, R. Sami, The influence limiter: provably manipulation-resistant recommender systems, in: Proceedings of the 2007 ACM Conference on Recommender Systems, ACM Press, 2007, pp. 25-32.
- [49] C. Shi, M. Kaminsky, P. B. Gibbons, F. Xiao, DSybil: Optimal Sybil-Resistance for Recommendation Systems, in: Proceedings of the 30th IEEE Symposium on Security and Privacy, IEEE Press, 2009, pp. 283-298.
- [50] M.-S. Shang, L. Lü, Y.-C. Zhang, T. Zhou, Empirical analysis of web-based user-object bipartite networks, EPL 90 (2010) 48006.
- [51] 刘怡君, 周涛, 社会动力学, 科学出版社, 2012.
- [52] S.-H. Min, I. Han, Detection of the customer time-variant pattern for improving recommender systems, Expert Systems with Applications 28 (2005) 189-199.
- [53] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, 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 Press, New York, 2010, pp. 723-732.
- [54] N. N. Liu, M. Zhao, E. Xiang, Q. Yang, Online evolutionary collaborative filtering, in: Proceedings of the 4th ACM Conference on Recommender Systems, ACM Press, New York, 2010, pp. 95-102.
- [55] J. Liu, G. Deng, Link prediction in a user-object network based on time-weighted resource allocation, Physica A 39 (2009) 3643-3650.
- [56] Y. Koren, Collaborative filtering with temporal dynamics, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2009, pp. 447-456.
- [57] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility, Science 327 (2010) 1018-1021.
- [58] E. Cho, S.A. Myers, 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, ACM Press, New York, 2011, pp. 1082-1090.
- [59] V. W. Zheng, Y. Zheng, X. Xie, Q. Yang, Collaborative location and activity recommendations with GPS history data, in: Proceedings of the 19th International Conference on World Wide Web, ACM Press, New York, 2010, pp. 1029-1038.
- [60] M. Clements, P. Serdyukov, A. P. De Vries, M. J. T. Reinders, Personalised travel recommendation based on location co-occurrence, arXiv:1106.5213.
- [61] T. H. Dao, S. R. Jeong, H. Ahn, A novel recommendation model of location-based advertising: context-aware collaborative filtering using GA approach, Expert Systems with Applications 39 (2012) 3731-3739.
- [62] S. Scellato, A. Noulas, C. Mascolo, Exploiting Place Features in Link Prediction on Location-based Social Networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2011, pp. 1046-1054.
- [63] Y. Koren, J. Sill, OrdRec: An ordinal model for predicting personalized item rating distributions, in: Proceedings of the 5th ACM Conference on Recommender Systems, ACM Press, New York, 2011, pp. 117-124.
- [64] Z. Yang, Z.-K. Zhang, T. Zhou, Uncovering Voting Patterns in Recommender Systems (unpublished).
- [65] J. Vig, S. Sen, J. Riedl, Navigation the tag genome, in: Proceedings of the 16th International Conference on Intelligent User Interfaces, ACM Press, New York, 2011, pp. 93-102.
- [66] L. Chen, P. Pu, Critiquing-based recommenders: survey and emerging trends, User Modeling and User-Adapted Interaction 22 (2012) 125-150.
- [67] T. Lau, 张韶峰, 周涛, 推荐引擎: 信息暗海的领航员, 中国计算机学会通讯 8(6) (2012) 22-25.
- [68] 朱郁筱, 吕琳媛, 推荐系统评价指标综述, 电子科技大学学报 41 (2012) 163-175.
- [69] P. Cremonesi, Y. Koren, R. Turrin, Performance of Recommender Algorithms on Top-N Recommendation Tasks, in: Proceedings of the 4th ACM Conference on Recommender Systems, ACM Press, New York, 2010, pp. 39-46.
- [70] R. Sinha, K. Swearingen, The role of transparency in recommender systems, in: Proceedings of the CHI'06 Conference on Human Factors in Computing Systems, ACM Press, New York, 2002, pp. 830-831.
- [71] A. D. J. Cooke, H. Suján, M. Suján, B. A. Weitz, Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations, Journal of Marketing Research 39 (2002) 488-497.
- [72] G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Computing 7 (2003) 76-80.
- [73] L. Lü, T. Zhou, Link prediction in complex networks: a survey, Physica A 390 (2011) 1150-1170.
- [74] R. Hu, P. Pu, Enhancing recommendation diversity with organization interfaces, in: Proceedings of the 16th international conference on Intelligent user interfaces, ACM Press, New York, 2011, pp. 347-350.
- [75] G. Lenzini, Y. van Houten, W. Huijsen, M. Melenhorst, Shall I trust a recommendation? Lecture Notes in Computer Science 5968 (2010) 121-128.
- [76] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature 464 (2010) 1025-1028.
- [77] C.-G. Gu, S.-R. Zou, X.-L. Xu, Y.-Q. Qu, Y.-M. Jiang, D.-R. He, H.-K. Liu, T. Zhou, Onset of cooperation between layered networks, Physical Review E 84 (2011) 026101.
- [78] M. Mognani, L. Rossi, The ML-model for multi-layer social networks, in: Proceedings of 2011 International Conference on Advances in Social Networks Analysis and Mining, IEEE Press, 2011, pp. 5-12.
- [79] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (2010) 1345-1359.
- [80] N. Nori, D. Bollegala, M. Ishizuka, Exploiting user interest on social media for aggregating diverse data and predicting interest, Artificial Intelligence 109(B3) (2011) 241-248.
- [81] M. A. Ahmad, Z. Borbora, J. Srivastava, N. Contractor, Link Prediction Across Multiple Social Networks, in: Proceedings of the 2010 IEEE International Conference on Data Mining Workshop, IEEE Press, 2010, pp. 911-918.
- [82] B. Cao, N. N. Liu, Q. Yang, Transfer learning for collective link prediction in multiple heterogeneous domains, in: Proceedings of the 27th International Conference on Machine Learning, 2010.
- [83] B. Xu, J. Bu, C. Chen, D. Cai, An exploration of improving collaborative recommender systems via user-item subgraphs, in: Proceedings of the 21st International Conference on World Wide Web, ACM Press, New York, 2012, pp. 21-30.
- [84] 张亮, 柏林森, 周涛, 基于跨境电商行为的交叉推荐算法, 电子科技大学学报 (已接收).
- [85] T. Zhou, L. Gong, C. Li, L.-S. Bai, L. Zhang, S. Huang, S. Guo, M.-S. Shang, Information filtering in interacting

- networks (unpublished).
- [86] S. Sahebi, W. W. Cohen, Community-Based Recommendations: a Solution to the Cold Start Problem (unpublished).
- [87] R. Sinha, K. Swearingen, Comparing recommendations made by online systems and friends, in: Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, 2001.
- [88] M.J. Salganik, P.S.Dodds,D.J.Watts, Experimental study of inequality and unpredictability in an artificial culturalmarket, Science 311 (2006) 854-856.
- [89] P. Bonhard,M.A. Sasse, "Knowingme knowing you"—using profiles and social networking to improve recommender systems, BT Technology Journal 24 (2006) 84-98.
- [90] S.-Y. Hwang, C.-P.Wei, Y.-F.Liao, Coauthorship networks and academic literature recommendation, Electronic Commerce Research and Applications 9 (2010) 323-334.
- [91] P. Symeonidis, E. Tiakas, Y.Manolopoulos, Product recommendation and rating prediction based onmulti-modal social networks, in: Proceedings of the 5th ACM Conference on Recommender Systems, ACM Press, New York, 2011, pp. 61-68.
- [92] A. Seth, J. Zhang, A social network based approach to personalized recommendation of participatory media content, in: Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, AAAI Press, 2008, pp. 109-117.
- [93] W. Pan, N. Aharonym, A. S. Pentland, Composite Social Network for Predicting Mobile Apps Installation, in: Proceedings of the 25th AAAI Conference on Artificial Intelligence, AAAI Press, 2011, pp. 821-827.
- [94] S. Guo, M. Wang, J. Leskovec, The role of social networks in online shopping: information passing, price of trust, and consumer choice, in: Proceedings of the 12th ACM Conference on Electronic Commerce, ACM Press, New York, 2011, pp. 157-166.
- [95] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, ACM Transactions on Web 1 (2007) 5.
- [96] J. Huang, X.-Q.Cheng, H.-W.Shen, T. Zhou, X. Jin, Exploring social influence via posterior effect of word-of-mouth recommendations, in: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, ACM Press, New York, 2012, pp. 573-582.
- [97] Q. Yuan, L. Chen, S. Zhao, Factorization vs. Regularization: Fusing Heterogeneous Social Relationships in Top-N Recommendation, in: Proceedings of the 5th ACM Conference on Recommender Systems, ACM Press, New York, 2011, pp. 245-252.
- [98] Z. Abbassi, C. Aperjis, B. A. Huberman, Friends versus the crowd: tradeoff and dynamics (unpublished).
- [99] H. Ma, D. Zhou, C. Liu, M. R. Lyu, I. King, Recommender Systems with Social Regularization, in: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, ACM Press, New York, 2011, pp. 287-296.
- [100] M. Medo, Y.-C.Zhang, T. Zhou, Adaptive model for recommendation of news, EPL 88 (2009) 38005.
- [101] T. Zhou, M. Medo, G. Cimini, Z.-K.Zhang, Y.-C. Zhang, Emergence of scale-free leadership strcuture in social recommender systems, PLoS ONE 6 (2011) e20648.
- [102] G. Cimini, D.-B. Chen, M. Medo, L. Lü, Y.-C. Zhang, T. Zhou, Enhancing topology adaptation in information-sharing social networks, Physical Review E 85 (2012) 046108.
- [103] J. O'Donovan, B. Smyth, Trust in recommender systems, Proceedings of the 10th international conference on Intelligent user interfaces, ACM Press, 2005, pp. 167-174.
- [104] P. Massa, P. Avesani, Trust-aware recommender systems, in: Proceedings of the 2007 ACM conference on Recommender systems, ACM Press, 2007, pp. 17-24.
- [105] H. Ma, I. King, M. R. Lyu, Learning to Recommend with Social Trust Ensemble, in: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, ACM Press, New York, 2009, pp. 203-210.
- [106] M. Jamali, 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 Press, New York, 2009, pp. 397-406.
- [107] M. Jamali, M. Ester, A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks, in: Proceedings of the 4th ACM Conference on Recommender Systems, ACM Press, New York, 2010, pp. 135-142.
- [108] J. He, W. W. Chu, A Social Network-Based Recommender System (SNRS), Data Mining for Social Network Data 12 (2010) 47-74.