



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
iCity & Big Data—Article

非独立同分布推荐系统：推荐范式转换的综述和框架

Longbing Cao

Advanced Analytics Institute, University of Technology Sydney, Sydney, NSW 2007, Australia

ARTICLE INFO

Article history:

Received 23 December 2015

Revised 4 May 2016

Accepted 12 June 2016

Available online 30 June 2016

关键词

独立同分布

非独立同分布

异构性

关系耦合

耦合学习

关系学习

独立同分布学习

非独立同分布学习

推荐系统

推荐

非独立同分布推荐

摘要

虽然推荐系统在我们的生活、学习、工作和娱乐中扮演着越来越重要的角色，但是很多时候我们收到的推荐都是不相关的、重复的，或者包含不感兴趣的产品和服务。这些差的推荐系统产生的原因来源于一个本征假设：传统的理论和推荐系统认为用户和物品是独立同分布的 (IID)。另一个明显的现象是，虽然投入了很多的精力模拟用户或者物品的特殊属性，但用户和物品的总体属性及它们之间的非独立同分布性 (non-IID) 被忽略了。本文先讨论了推荐系统的非独立同分布性，紧接着介绍了非独立同分布性原理，目的是从耦合和异构性的角度来深入阐述传统的推荐系统的固有本质。这种非独立同分布推荐系统引起了传统推荐系统范式的转化——从独立同分布向非独立同分布进行转化，希望能够形成高效的、相关性高的、个人定制和可操作的推荐系统。这种系统创造了令人兴奋的能够解决包含冷启动、以稀疏数据为基础、跨域、基于群组信息和欺诈攻击等各种复杂情况的新的研究方向和解决方案。

© 2016 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

推荐[1,2]是大数据方法的一种重要应用[3,4]。它在核心商业和新经济领域扮演着越来越重要的角色，尤其是涉及社交媒体、手机服务、在线商业、学习和生活等方面。近年来，在很多领域中推荐研究成为研究重点，这些领域包括推荐系统、信息检索、社交媒体、社交网络、机器学习、数据挖掘和数据工程。

高质量的推荐服务应该在合适的时间给合适的人群推荐最相关的产品。在这个方向人们投入了很多精力，

尤其是在推荐和信息检索领域，通过考虑特殊的因素，比如社会关系、朋友关系、用户对所购买产品的评价、对相似特征进行分类以及通过另一个领域推荐产品等方式来提高推荐水平。

然而在大多数情况下，我们通过新闻门户网站、在线购物网站和手机应用等渠道得到一些并不相关，甚至对品牌有所损害的推荐。比如一个著名的搜索引擎报道在希腊发生的民众示威游行时，在新闻旁边放了一个建议去希腊海滩旅游的广告链接。另一个推荐系统对一个对kiwis感兴趣的用户推荐了各种不同的水果，系统假

* Corresponding author.

E-mail address: longbing.cao@gmail.com

设用户喜欢猕猴桃(kiwi fruit)。在线书店网站经常罗列一些重复的或者用户已经购买过的书或者根本不相关的书。

推荐领域所面临的至关重要的问题包括：为什么会推荐不相关，或者重复的产品和服务？更重要的是，如何开展下一代的推荐？为了回答这些问题，需要研究如下这些基本问题：

- 现有的推荐理论和系统还缺乏哪些基本的技术从而导致较差的推荐？
- 如何建立高效的、相关的、私人订制的以及可操作的推荐系统？
- 如何提高推荐系统的质量，从而把个人或者群体用户感兴趣或喜欢的产品推荐给他们？
- 哪种新的推荐方法可形成一个能够抓住推荐的固有特征以及复杂性的统一的理论框架？
- 推荐系统如何转型，才能促成下一代推荐系统的研究？
- 下一代推荐系统形成的基础是什么？
- 对于下一代的推荐理论和系统，新的研究方向是什么？
- 哪种新的推荐理论体系能够解决传统推荐系统存在的冷启动、稀疏性、跨域、群体用户推荐和欺诈攻击等典型挑战？

为了回答上面的问题，有很多研究方向需要进一步探索，本文中所感兴趣的方向是进一步了解推荐用户和物品，以及一个用户对于一个物品给予的评分与用户和物品的特性之间的关系。这涉及进一步理解推荐系统的固有属性和复杂性，也就是说评分、用户特性和物品特性的异构性和耦合性(也叫做非独立同分布[5,6])，以及这三个方面之间的异构性和耦合性。

在现有的推荐系统研究中，在宏观方面投入了很多精力，比如用户对于物品的评价、用户的社会关系以及人们对物品的评价。这些工作大体上分为如下几个方面：①基于现有评分预估未来评分；②把用户对于物品的评价结合到模型中；③把用户友好性结合到模型中；④模拟群组喜好；⑤跨领域进行用户喜好转化的。用户对于物品的评论的喜好也输入到模型中[7,8]。近来对于物品和用户组的耦合关系也被考虑进推荐系统模拟中[9-11]，把微观的信息作为驱动力给动态评分估计带来进一步的提升。

然而最新的推荐系统研究[2]基于用户、产品和评分是独立同分布这样的假设，产生了独立同分布模型和方

法[5]。很少有研究考虑到特定用户和产品的非常微观的非独立同分布的信息。从这个角度来说，评分系统的最基础的动力被简化和忽视了，本文作者认为正是这点导致了现有推荐系统和服务的质量很低。比如传统的矩阵解析方法是推荐模型中广泛应用的模型。然而，如果房子和汽车微观的特性没有考虑进去的话，并且如果房子和汽车被视为独立同分布的，这种方法会对房子和汽车产生类似的结果。这会在通用的宏观模型和与推荐的用户和产品相联系的微观信息之间产生巨大的差距。

本文的研究重点是讨论这类信息在获取推荐的本质方面的驱动作用，进而提高推荐质量。通过对文献中关于非独立同分布推荐理论和系统简单论述的扩充[12]，提出了一套了解推荐本质的体系框架并进行了深入的解释。本文讨论现有分析存在的问题，进而检验下一代推荐理论和方法是否有必要阐述推荐体系的非独立同分布和原则。提出了非独立同分布的一个大致的学习框架，这个框架能够捕捉宏观的评分系统以及用户、物品和他(它)们的非独立同分布本质的微观的特殊信息。

非独立同分布包括推荐系统的耦合关系以及异构性。耦合关系包括用户内、用户间、物品内、物品间及用户和用品间主客观的交互关系(包括显式和隐式的关系)。异构性从用户传播到物品[包括他(它)们的性质]。非独立同分布包括：①推荐系统中所涉及的非独立同分布用户和物品范畴的显式属性；②用户间和物品间的异构性；③在用户和物品间的分级耦合关系[6]；④用户和物品间潜在的关系。

这种非独立同分布的推荐理论观点开启了我们的思路转换，并提供了下一代基础研究和高质量推荐研究的新方向。实际上，在大数据学习[5,6]中非独立同分布是数据学科和大数据分析的理论基础和实践挑战[3,13-15]，但在相关领域并没有引起足够的重视，比如计算机、信息学和统计学，原因是现有分析和学习理论体系主要基于独立同分布的假设。希望本文有关非独立同分布的推荐原理和体系能够对基础研究有所启发，并期望在其他的分析学、学习和信息处理领域得到好的应用。

本文内容如下：第二部分讨论推荐系统的固有本质，第三部分展现非独立同分布系统理论的原则，第四部分综述主要问题，尤其是与现有推荐系统和理论相关的独立同分布的假设，第五部分阐述推荐领域中对于研究对象以及推荐结果产生的思维方式的转化，第六部分介绍非独立同分布推荐理论框架，第七部分分享一些主

要的非独立同分布推荐体系的案例，第八部分描述了非独立同分布系统的前景，第九部分为结论。

2. 推荐的本质

如图1所示，本文把推荐相关的各种途径获得的信息结合起来，将推荐系统分为4个维度，每一维度包含4种不同的信息：

(1) 表(A)包含评分信息，由用户对物品的评分构成，其中表达了用户评分偏好，表(A)反映了推荐系统的主观信息和成果；

(2) 表(B)包含用户信息，该表反映了驱动评分偏好的用户特征、属性和关系；

(3) 表(C)包含吸引和影响用户喜好和评分的物品的属性、特性和关系；

(4) 表(D)包含隐式的用户和物品的相互关系。这和表(A)~(C)的来源不同，表(A)~(C)的信息来源是显式的，而表(D)的信息来源是隐式的，从用户和物品的属性方面反映了他(它)们之间的相互关系，表(D)既包含连接用户和物品的主观信息，也包含相应的客观信息。

如上的系统观(图1)并没有考虑：①环境(E)(如图中外面的边框)，需要在推荐系统研究中考虑环境(E)；②表中(A)~(D)的内容与环境(E)之间的交互性。为讨论推荐系统的本质，在推荐系统中关注这个4个表格是至关重要的。图1中的4个表格阐述了如下几个方面的重要信息。

(1) 特定推荐任务中所涉及的用户和物品。用户和物品的信息表现了他(它)们的特性、属性和相互关系。由于他(它)们各自的特性、特征值以及对象在本质上都是不同的，因此用户和物品是特殊和异构的。用户和物品是因人(物)而异，有他(它)们自己独特的性质和相互

关系。

(2) 他(它)们彼此之间如何相互作用，影响相互的评分行为和偏好。用户彼此耦合，并且因为一个或多个原因或多或少地相互影响，物品之间也是同样的道理。除了这种用户群体内或者群体(全体)间的影响，用户和物品之间的特殊相关性在很大程度上也会相应地受个人用户和物品的影响。对于全局和局部特性的相互作用，偏好之间相互影响的平衡关系的建模是至关重要的。

(3) 揭示用户和物品之间的相互影响。一个用户喜欢一种产品(通过评分展现)是因为用户的特性和相应的物品之间存在着特殊的耦合关系，虽然这种耦合是复杂的和隐式的。

(4) 利用客观和主观信息通过评分促成最终的决策。通过连接表(A)中主观评价和表(B)、(C)中客观的用户和物品的信息，以及表(D)中用户和物品之间主客观的关联关系，我们能够获得反映推荐系统中的连接、驱动因子和动态性的完整的信息，而这些信息通常都能从评分反映出来。

这4个表格或维度的视角与传统的推荐系统本质上是不一样的，传统的推荐系统只包含表(A)~(C)。除了评分之外，也涉及用户/物品的部分或者特殊信息。虽然对评分如何产生和产生的原因进行深入的了解很重要，但是表(D)中所包含的潜在的用户和物品的相互关系在相应的领域还缺乏深入的研究。表(D)中的相互关系不像其他表格中的那么明显，但是它们包含用户属性和物品特性之间隐式的相互关系(我们在这里命名为耦合[6])，详见6.5部分的讨论。

表格(D)中复杂的隐式关系(如CP代表物品价格和用户城市间复杂的关系)是由潜在的人口、行为、社会、经济或文化等方面驱动的，这些都包含在用户和物品的

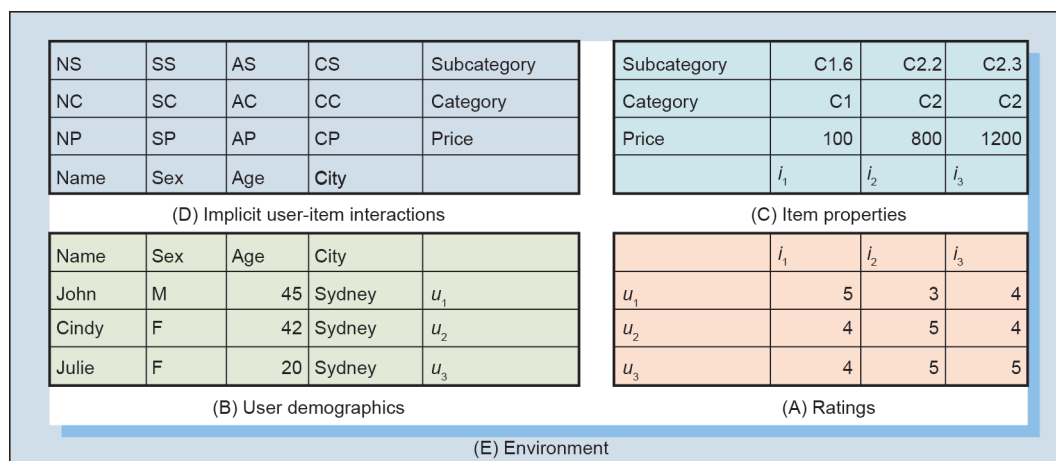


图1. 推荐系统的系统观。

特性及他(它)们之间的耦合关系中。用户之间和物品之间的异构性导致个性化的评分行为。因此在这4个表格中存在着很强的非独立同分布, 这形成了表(A)中评分和偏好的潜在驱动力。

实际上, 图1中所提出的4个维度不仅仅呈现了如何建立全面理解推荐问题普适的数据结构, 同时也表述了推荐系统驱动因子的显式和隐式的非独立同分布性。

3. 推荐系统的非独立同分布性

根据上文推荐系统的4个表格或维度的视角, 这一部分着重讨论推荐问题中本质上所存在的非独立同分布特性。实际上, 任何被推荐的物品和用户都是非独立同分布的, 也就是说, 在用户之间、物品之间以及用户和物品之间存在不同层次的耦合关系, 异构性存在于用户间和物品之间。下面讨论推荐系统所包含的这两方面的内容。

3.1. 异构性

在图1的4个表格中从不同的维度展现了异构性, 用户和物品都不是同分布的, 下面列举了用户和物品的异构性, 以及用户和物品间的异构性的不同场景。

(1) 用户的异构性。每个用户具有自己独特的属性、特征、喜好、行为以及评分的倾向。对所有用户以同分布来理解每个用户独特的特性、独特的需求和倾向就未免太简单了。

(2) 物品的异构性。每个物品和另一个物品在类型、属性、类别、应用领域等方面是不同的。物品特殊的属性对于不同的用户和用户评价的吸引力是不同的。

(3) 用户或物品属性的异构性。每个用户和物品的属性是不同的, 每个用户的属性独特表达了用户的人口

特征、特性、喜好、行为和倾向等。同样, 每个物品的属性表达了物品的类别、类型、特性、领域等。每一项用户或物品属性不是同分布的, 他(它)有自己独特的分布, 因此需要单独对待。

(4) 用户和物品之间的异构性。用户对物品的态度是不同的, 因此不能假设它们遵从同样的分布。所以假设它们采用相似的关系矩阵或者用同样的模型来获取用户和物品间的特殊属性有些过于简单。

现存的很多方法, 如基于分解矩阵的方法, 在处理如上关于推荐系统中需要考虑的异构性方面的问题可能不会产生有意义的结果, 甚至有可能产生令人误解的推荐。当模型中忽略了个人的特征, 也可能不能提供个性化的推荐。

3.2. 耦合

在推荐系统中考虑到异构性是推荐系统的一个进步, 虽然它没有抓住推荐系统的所有特征和复杂性。另一个问题是抓住显式和隐式的耦合关系——通常是有层次的。因此耦合指的是两个或者更多方面之间的关系或者相关性(可以是输入之间的, 也可以是输入、输出之间的)[6]。

像图1显示的, 4个表格内含不同的耦合关系。推荐系统的耦合问题代表用户之间、物品之间, 以及用户和物品之间, 无论何种原因或在任何方面存在的清晰和隐式的关系。进一步的解释详见图2。

(1) 用户和用户之间的耦合。这些指在图2中在用户内部和用户之间的耦合关系, 进一步包含: ①用户内部属性间的耦合, 展现了用户属性价值的关系, 比如用户属性、群体、领域、行为和社会关系之间的耦合; ②用户之间属性的耦合, 表现了用户属性之间的关系, 比如用户的年龄和他们的位置; ③用户和用户群之间的用户

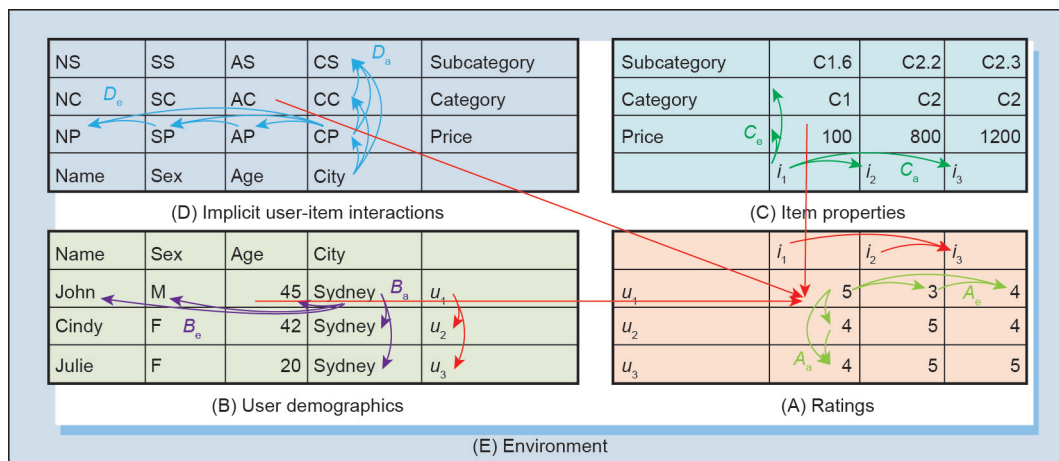


图2. 推荐系统的非独立同分布。

耦合。

(2) 物品和物品之间的耦合，这些与用户和用户之间的耦合类似，表(C)中物品内部和物品之间的耦合包含：①物品属性内部的耦合；②物品之间属性的耦合；③物品和物品群之间的耦合。

(3) 用户和物品之间的耦合。这些指的是用户-物品组或者群内部和之间的耦合，包含如下方面：①表(A)中用户对于物品评分和评论体现的用户和物品间的耦合；②隐式的用户和物品间的耦合，像表(D)中所展现的用户属性对用户相关物品属性的影响，以及用户属性和用户相关物品属性之间的关系。

除了上述所讨论的耦合，耦合经常通过一定的层次表现出来，比如表2中，耦合存在于属性价值、属性(对于用户和物品)、目标(用户和群体)和目标群(用户群或者物品分类)。尤其是，用户与物品的耦合会在不同的级别间出现，比如说从评分表格之中的耦合[表(A)]到不同用户组评分之间的耦合关系，从用户属性和物品属性之间的耦合到用户属性矩阵和物品属性矩阵之间的耦合等。

4. 现有推荐系统存在的问题

基于上述关于推荐系统本质的讨论，本章讨论现存推荐系统研究存在的问题，以及传统推荐理论和体系经常使用的有关独立同分布性的假设。

4.1. 推荐系统的相关研究

现存的推荐系统的算法和矩阵大致可以分为如下四类：协同过滤推荐系统(CF)、基于内容的推荐系统(CBF)、组合的推荐系统和问题导向型推荐系统。

协同过滤矩阵[16]通过一个用户自身或者其他用户的评价结果来预估评价。当其考虑到其他用户的行为后，协同过滤矩阵引入与有关用户或者邻居特征相似的用户组的行为和喜好[17]。然而，这种矩阵是基于物品的[18,19]。

基于内容的推荐系统利用了用户对物品的评价、物品的内容、用户阅读到的有关物品的相关资料。这种额外的信息会被引入评价系统中。

通过将协同过滤矩阵和基于内容推荐系统通过平行或串行的各种组合形成了一些组合推荐方法。组合可以将协同过滤矩阵推荐系统和内容推荐系统分开，接着将它们的结果结合起来或者把基于内容推荐的矩阵应用于具有类似观点的用户，随后应用协同过滤矩阵进行对评分的推荐。不少研究人员通过在模型中引入特殊的因子

(比如协同过滤矩阵)，提出了不同的方法，比如基于社交网络的模型[20,21]。

耦合矩阵分解框架[22]的提出是为了用广义线性连接函数对每个关系矩阵进行因子化，并且无论何时，一个实体模型与不止一个关系相关，而且能够将模型中不同的因子连接在一起。这个方法通过矩阵分解相关关系和矩阵来捕捉因子之间存在的简单且潜在的关系，但是它不能获取微观的数据特征和复杂性。

一些建模工具和评价矩阵已经被提出来用于衡量推荐质量，比如协同过滤矩阵通常会用皮尔逊相关系数，聚类被用来对相似的物品或用户分组。传统的数据挖掘和机器学习方法比如 k -均值、 k -模式、隐式的 C 聚类、矩阵分解模型、自适应共振理论、概率聚类(最大期望算法)、贝叶斯置信网、马尔可夫链和Rocchio分类通常被用在推荐系统中。

近年来特殊的推荐问题比如冷启动、跨域、基于群组信息和欺诈攻击引起越来越多的关注[23]。群组推荐从一个域中用户对物品的评价，来给其他类别的对物品的评价提供信息[24,25]。一些研究工作[26]专注于特殊的数据、形成推荐的结构和方法。

最近由于我们在非独立同分布学习[5]和推荐系统的研究宣传[12]，一项新兴起的推荐系统的研究方向[6]专注于对用户内部、用户间、物品内部、物品间、用户和物品间的耦合关系建模，并把它们与评分系统结合起来。基于此做了一些基本的工作，包括基于物品耦合关系的推荐[11]、基于物品耦合关系的协同过滤推荐系统、具有物品耦合关系的基于用户群的推荐[27]。本文是基于这些先前的研究和文献[12]中的一些主要结论来构成关于非独立同分布推荐学习的一个全面的、系统框架的讨论。

4.2. 传统推荐理论独立同分布假设

大部分现有推荐理论和系统仅仅或者主要涉及表(A)中的评分信息，从评分的角度专注于学习用户对于物品的喜好。典型的推荐算法(包括协同过滤推荐系统和矩阵分解模型)通常忽略了相关用户喜好的潜在原因[18,28,29]，这部分潜在原因大部分可从表(B)和(C)中有关用户和物品的信息中提取出来。他们通常把用户和物品看做独立同分布的，不考虑用户是谁以及用户能够评价什么样的产品。本文作者所了解到的信息是目前还没有研究能够完整地引进像图2中所示的评分、用户和物品的内部和之间及用户和物品之间的耦合关系。

现对传统的协同过滤推荐系统和矩阵分解模型假设

的独立同分布做分析[30,31]。

基本协同过滤模型建立了一个过滤过程，其中涉及对象之间的协同作用，采用了基于用户的或是基于物品的协同过滤模型。基于用户的协同过滤模型假设如果用户 a 和用户 b 对一件物品评分相同，那么用户 b 对另一个物品的评分也会和用户 a 对此物品的评分相似。同理，基于物品的协同过滤模型假设，对于物品 x 感兴趣的用户也会对物品 y 感兴趣。虽然对于相应的问题都已经提出了不同的协同过滤模型，然而原始基于记忆的协同过滤模型的基本假设可以由方程(1)来表示[28]。方程(1)是基于用户 i 对于物品 j 的评价来推测用户 a 的评价 $r_{a,j}$ ，平均评分 \bar{r}_i 衡量了用户 i 和用户 a 相似性的权重。

$$r_{a,j} = \bar{r}_a + k \sum_{i=1}^n w_{a,i} (r_{i,j} - \bar{r}_i) \quad (1)$$

其中， $w_{a,i}$ 假设 i 和 a 之间只有较弱的相关性。基本的协同过滤模型会有如下的假设：①用户 i 对所有物品进行独立的评价，将这些物品视为独立而且同分布的，忽略对于两个物品 j_1 和 j_2 的评分 r_{i,j_1} 和 r_{i,j_2} 之间的联系；②两个用户独立评价物品，用户视为独立同分布的，忽略两个用户和两个用户评分 $r_{i_1,j}$ 和 $r_{i_2,j}$ 间的关系；③用户关于物品的评价不会相互影响，也就是忽视了用户和物品之间的联系；④这个算法并没有涉及用户属性和物品属性，确切地说，它只使用了评分信息。

让我们进一步分析矩阵分解方法。矩阵分解算法假设评分 R 是两个矩阵 P 和 Q 的近似因式分解，分别代表用户和物品的潜变量矩阵。因此用户 i 对物品 j 的评分 $\hat{r}_{i,j}$ 的估计可以通过如下方法实现：

$$\hat{r}_{i,j} = \sum_{k=1}^k p_{ik} q_{kj} \quad (2)$$

其中，向量 p_{ik} 和 q_{kj} 抓住了用户 i 对于物品 j 潜在的变量，同时评分会受类似用户 k 的影响。

基本矩阵分解模型做了如下方面的假设：①评分估计既不取决于用户也不取决于物品特性；②它假设评分是由用户和物品潜在因子驱动的，因此忽略了用户之间、物品内部以及用户和物品之间的耦合(影响)关系和异构性。

上述对于协同过滤推荐系统和矩阵分解模型原理的分析表明，这两种方法假设用户和物品是独立而且同分布的，虽然提出了这两种算法的变型，但在评分估计中还是缺乏对于用户和物品属性这些驱动因子的考虑。如果像图1中非独立同分布信息没有被考虑进去，那么可以推断其作出的推荐也不会表达个人的喜好。

5. 推荐系统研究的思维转化

以下讨论现有推荐研究的特色以及提出推荐系统发展的四个时代。

5.1. 推荐系统的特色

虽然推荐系统的发展经历了一些重要的时期，在这些时期关注重点放在了推荐系统研究上。根据第四部分对于已有推荐系统的讨论，当前推荐系统的典型特征可归纳为：

- (1) 假设用户和物品是非独立分布的；
- (2) 把焦点放在可以观测到的因素和范畴上；
- (3) 结合潜在因素但是忽略了隐式的用户/物品变量，反之亦然；
- (4) 忽略或者简化用户和物品的显式和隐式变量之间的关系；
- (5) 缺乏对主观因素深入的挖掘，缺乏主客观因素的结合；
- (6) 缺乏对核心动因及用户和物品内部和之间隐式关系的挖掘。

5.2. 推荐系统分类

对于如何将推荐系统分类存在很多不同的观点。对于推荐系统具有代表性的问卷调查大致展示了如下从不同出发点和不同兴趣点出发的推荐系统研究的现状分类。

(1) 文献[32]显示了混合网络系统的分类方法，基于知识来源一共分成四类：协同过滤推荐系统、基于内容的、人口统计的和基于知识来源的。在上述四类中通过结合七种不同的方法结合(权重的、混合的、转移的、融合各种特征的、串联的、特征扩充和元级[33])，最后产生了53个可能的混合方法。

(2) 文献[34]中提出了一种推荐系统分类方法，这个方法考虑了简明性、降低维度、扩散、社会过滤、元方法和性能评价。

(3) 文献[35]中提出的协同过滤推荐系统考虑到推荐系统从算法到对推荐系统用户体验的相关问题，以及有关质量、存在潜在的危险和用户控制等开放的问题。

(4) 参考文献[36]提供了推荐系统分类方法，一共包含四个级别：基于内存的(评分)、基于内容的(用户和物品的特性，对应传统网络)、基于社会学(关系和可靠性，对应社交网络)以及基于情境的(用户和物品的位置，对于物流网)，这种分类方法既考虑隐式的数据，也考虑

显式的数据, 以及用户和物品的数据。

(5) 从2001年到2010年相关的文献分类和进展如文献[37]中所讨论的, 把推荐系统分为8个应用领域(书籍、文献、图片、电影、音乐、购物、电视和其他)和8种数据挖掘技术(关联规则、聚类、决策树、 k 最近邻分类算法、链分析、神经网络、回归和其他启发式方法)。

(6) 除了文献[32]中所引用的分类方法, 文献[2]中共28章分为四部分讲述了分类方法: 推荐方法、推荐系统评价、人机交互和更高级的题目。最近的手册中没有提供有关推荐系统有价值的分类。

这篇文献集中讨论了下述七大类推荐技术:

(1) 基于记忆推荐方法, 主要关注评分估计, 通过传统的矩阵解析方法和价值分解隐式对从用户到物品评价或对物品隐式评价[36]。

(2) 协同过滤推荐系统, 主要考虑用户与用户之间的关系, 以及图1中表(B)中用户信息用户(或物品)相邻的关系, 相当于基于相似用户或者物品的推荐。

(3) 用户分析推荐或者基于模型的推荐, 主要考虑用户人口信息, 目的是获得相似的用户, 关于相似的人口信息, 就是所说的个性化订制的推荐, 尤其要关注于表(B)中相关用户的信息。

(4) 基于内容的推荐, 尤其是涉及物品关键词、描述、表(C)中物品信息语义索引。

(5) 基于群组的推荐, 涉及表(B)中的社会和朋友的关系, 目的是对相关的用户组推荐物品, 或者给一组用户建议物品类别。

(6) 基于知识的推荐, 主要涉及: ①领域知识来衡量某些物品特征如何符合用户的需求和喜好, 以及一个物品如何符合一个用户的喜好, 比如通过学习表(D)中在相关用户的属性和物品属性所存在的耦合关系来达成基于案例的推荐; ②将相关的用户需求和表(D)中物品的属性联系起来构成应用规则, 达成所需的推荐。

(7) 混合推荐, 将上述方法结合起来, 比如将协同过滤推荐系统和基于内容的推荐结合起来。

上述分类方法综合了信息驱动的观点(大部分方法都是基于信息的)及基于功能和目标的方法。这些方法没有解决最关键的挑战(如以稀疏数据为基础和欺诈攻击), 它们缺失了一些重要方面的研究[比如表(D)中提到的可视化和关系挖掘]。

5.3. 推荐系统研究分类法

推荐方法的分类学如图3所示, 一共包含7个层次:

应用、来源、目标、挑战、技术、交付物(成果)以及推荐评价。

(1) 应用。这是指领域问题以及推荐的应用——推荐产品、服务、渠道等。推荐的典型应用包含: 手机应用和服务、社交媒体、网络应用、在线商业和服务(包括购物、新闻、娱乐、服务、食物和饮料服务)、工作流程和政策建议、健康和医疗服务推荐、旅游服务、市场营销和客户服务、商业和工业服务、工业优化、物流和运输服务、数字生活(包括虚拟现实和动画)以及生活服务。

(2) 这个指的是推荐系统所涉及的数据来源, 包括核心数据和辅助数据, 可以是主观的, 也可以是客观的, 显式的或者隐式的。辅助数据可以包含反馈数据、场境(上下文)数据、外部数据、领域知识、系统数据以及从网络获得的信息。

(3) 目标。该项对推荐系统的目的进行归类。商业和科技目标可能都和推荐系统有所联系。从商业观点出发, 推荐系统可以应用于改进营销和销售额、客户关系和使用体验、服务目标、经济财务目标、人机交互, 以及网站界面设计, 从而激励出新的商业机遇(如新用户、创新产品、新服务)。从科技角度出发的推荐系统可能集中于提升评价预测、成本效益、最优化、创新性、多样性、可预测性、稳健性、信用、风险管理, 以及建议的可实践性。

(4) 挑战。该项和诸多方面有关, 包括推荐系统来源的特征和复杂性(创新性、多样性、领域交叉、团体和社群针对性、动态和在线性质)、用户行为和满意度(冷启动、大众偏好、欺诈攻击的影响、个性化满意度、人类智能)、环境(相互关联、限制、社会文化议题)、基础设施(可扩展性、效率)、性能(质量、准确度、误差率、可用性、实用性、枝节问题、可实施性), 诸如此类。

(5) 技术。依据推荐系统的引擎、基础设施、算法、成果以及性能优化, 跨领域方法和技术已经被纳入推荐系统的研究当中。典型技术包括: CF、基于内容的推荐、数据挖掘和机器学习、数学和统计、相似学习、动态在线学习、经济金融模型、社会科学方法、情境感知技术、可视化, 以及多方法的结合。

(6) 成果。推荐系统的输出取决于推荐系统的目标和技术, 而这部分目标和技术又来自于对数据和挑战的认知。从推荐系统所获取的推荐成果可能包含相似的用户或产品、新用户、产品服务以及新应用和新策略, 提问或是对于问题的解答, 推荐有关的团体或社群, 以及

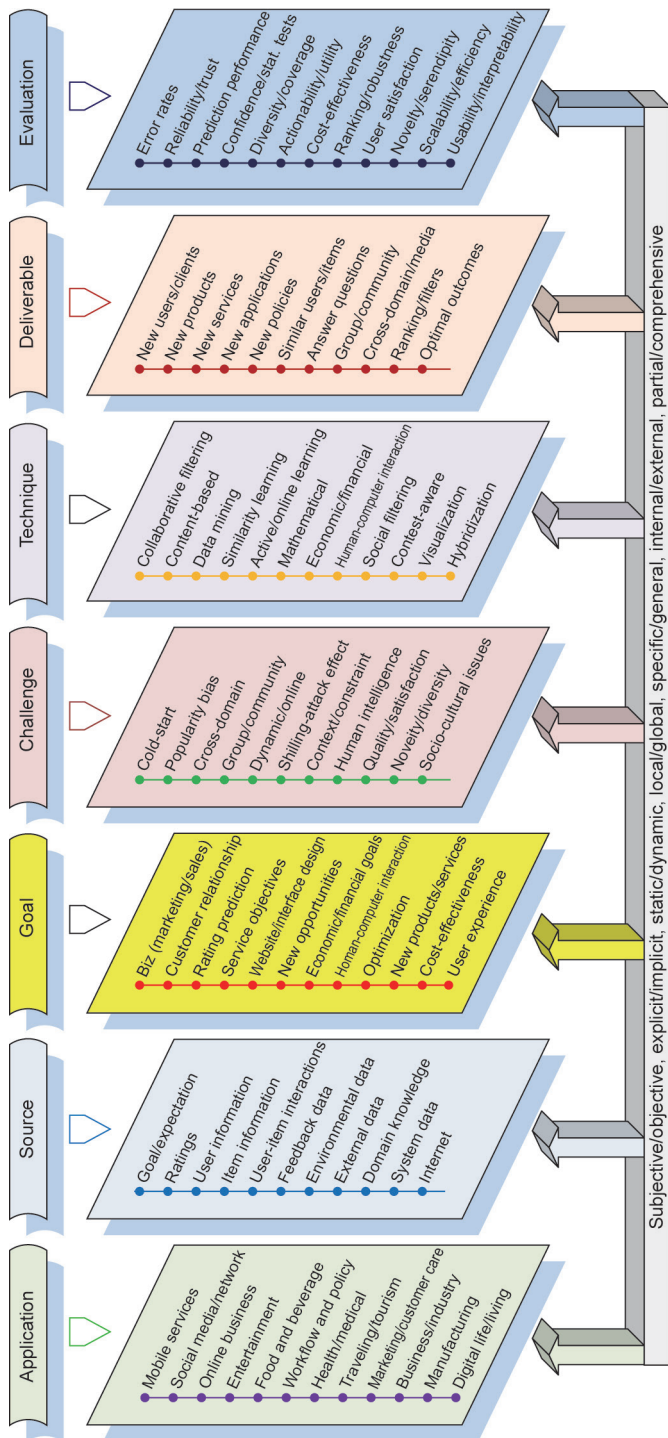


图3. 推荐研究的复层次模型。

跨领域跨媒体的机遇和体验，提供评价和过滤建议，从而获取最优结果。

(7) 评价。可以从商业和技术的角度来评估推荐系统的性能。商业指标可以包含用户满意度、新颖性和多样性、覆盖度、商业价值、可交互性和可解释性。技术指标可包含优化后的误差率、预测性能、可实现性、稳健性、对新状况的敏感性、可靠性、置信度和

统计测试性能、可实践性、效率，以及可扩展性。

一个有效的推荐系统必须在上述各层面间保持平衡。这种平衡是多方面的，包括：主观和客观、隐式和显式、局部和全局、静态和动态、内部和外部，以及7个层面的部分和整体。

5.4. 推荐系统的研发过程

该部分将针对推荐系统的研究归入四个主要阶段(图4)：

- (1) 第一代(1st G)：基于评分的推荐系统研究；
- (2) 第二代(2nd G)：基于用户/物品的推荐系统研究；
- (3) 第三代(3rd G)：跨用户/物品的推荐系统研究；
- (4) 第四代(4th G)：非独立同分布的推荐系统研究。

第一代以基于评分的推荐系统研究为主，相当于在评分表(表1)中通过直接通过模拟评分的过程(如MF法)或类比相似的评分行为表现或偏好(如经典CF法)来对评价动向进行建模和估算。也有一些研究关注基于记忆的方法和特定的评估特点，如对稀疏的评价数据的建模、冷启动评价，以及具有欺诈攻击效果的评价。在此阶段，图1表(A)部分中的评价信息主要取决于相关的模型。

第二代为基于用户/物品的推荐系统研究，相当于对评价动向建模，并做出基于用户或是物品的推荐，以及通过结合图1中表(B)、(C)中所显示的特定用户或物品信息建立基于内容的模型。典型例子包括用户间、不同类别或物品不同子类间(即跨领域或分级别推荐)的社会关系和过滤；依照评价行为或偏好对用户进行聚类；或对物品聚类以便推荐(即基于群体的推荐)。这类推荐系统将用户和物品信息纳入评价估值和用户/物品推荐中，进一步探索了通常遇到的挑战，如冷启动、稀疏的评价和欺诈攻击等，这类推荐模型也可以通过结合其他技术，诸如跨领域和基于群体的推荐。

第三代为跨用户/物品的推荐系统研究，通过纳入用户与物品间特定交互信息，如图1中表(B)、(C)所示，来建立评价模型并做出用户/物品推荐，这些交互信息包括用户对于产品的评价、用户的偏好以及特定产品类型或特点间的联系。一些现有的基于内容的模型属于这一类别，包括用户和物品信息，以及用户对于物品的评论、情绪和观点。

在现有文献中，和上述三代有关的研究都假设用户和物品/产品是独立同分布的，并没有考虑到用户与用户、产品与产品、用户与产品内部及其之间的“价值-客体”非独立同分布特征的存在[38,39]。一些方法，如基于MF的方法，日渐关注挖掘评价中的潜在变量。而当用户和产品信息结合起来时，其中的异质性和耦合关系[5]往往被忽略。

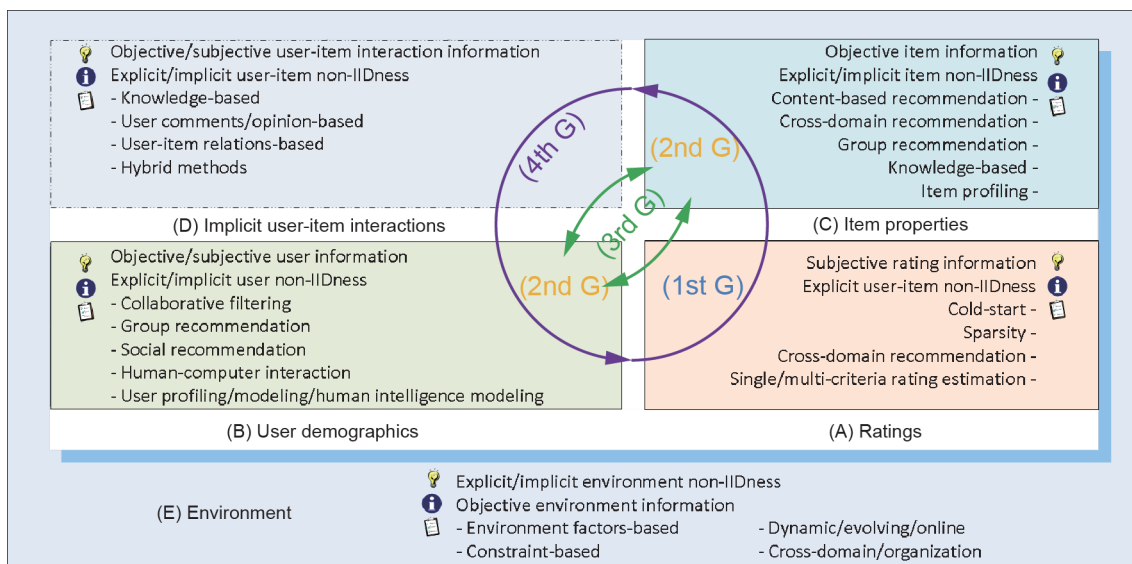


图4. 推荐系统研究的4个阶段。

第四代为基于非独立同分布的推荐系统研究，对用户内和用户间[表(B)]、产品内和产品间[表(C)]以及用户和产品间[图1中表(A)、(D)]的隐式/显式、主观/客观的非独立同分布性进行建模和协同增效。在这一阶段，我们假定用户和产品都是非独立同分布的，需要在不同的程度上从价值、属性、实体这些方面进行考虑，也要考虑用户属性和产品属性之间的交互作用。本文的讨论主要针对第四代推荐系统的研究展开(见图1中的系统概览)，这在已有文献中还没有被讨论过。

图4进一步绘制了图1内容的系统视图，涵盖了推荐系统研究的四个阶段。从一定的意义上而言，第四代实际上包含了前三代：①前三代理论和方法需要一个从独立同分布到非独立同分布的思维方式的转换；②非独立同分布推荐系统必须涵盖四个表——表(A)~(D)，以及非独立同分布假定下的环境(E)。

6. 非独立同分布推荐系统理论框架

这部分提出非独立同分布推荐系统的理论框架，并提出用户、物品及其显式的和隐式的关联关系的非独立同分布的公式。

6.1. 非独立同分布推荐系统框架

为了有效地获取上述推荐系统问题中非独立同分布特征，有必要引入一个新的非独立同分布推荐系统理论框架以建立非独立同分布推荐系统的理论体系。该非独立同分布系统框架的原则是抓取广泛存在于用户(物品)属性内部

及其之间、用户内部及其之间、物品内部及其之间，以及用户和物品之间的异质性和联系(即非独立同分布性)。

该非独立同分布推荐系统框架的目标为以下几点：

(1) 整合异质性和关联性，即将推荐系统的非独立同分布性整合进其算法和体系；

(2) 抓取显式的非独立同分布性，如图2表中(B)所示的用户之间的联系，以及隐式的非独立同分布性，如图2中表(D)所示的用户和物品之间的联系；

(3) 获取主观的非独立同分布性，如图2中表(A)所示的评价，和客观的非独立同分布性，如表(B)和表(C)中所示的用户和物品。

图5阐明了非独立同分布推荐系统的理论框架，其观点为推荐系统的非独立同分布性是内嵌在这四个表格中的。

(1) 用户的非独立同分布性。联系和差别共同存在于用户内部及其之间，具体存在于用户的属性、属性值、用户本身和他们的组群。这一点由图2中的表(B)所显示，可以用一个包含了用户属性和相应值的用户信息矩阵B来表述。

(2) 物品的独立同分布性。联系和差别共同存在于物品内部及其之间，具体存在于物品的属性、属性值、物品本身和物品类别中。这一点可由图2中的表(C)显示，可以用一个包含了物品属性和相应值的物品信息矩阵C来表述。

(3) 显式用户-物品非独立同分布性。存在于用户-物品联系的内部及其之间的显式非独立同分布性体现在用户对于物品的评价之中，如图2中的表(A)所阐释。然而由于基于评价的用户-物品非独立同分布性是主观的，

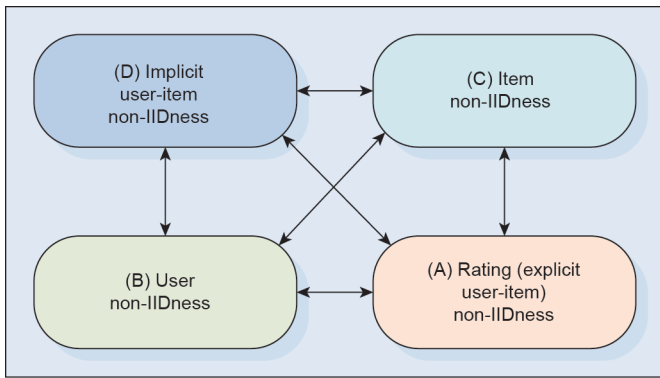


图5.非独立同分布推荐系统框架。

因此它也被称为主观用户-物品非独立同分布性。用户对物品的评价可表示为一个评价信息矩阵 \mathbf{A} 。

(4) 隐式用户-物品非独立同分布性。用户-物品联系的内部及其之间的隐式非独立同分布性存在于他(它)们的隐式属性之间的交叉, 如图2中的表(D)所示。可以用一个用户/物品隐式信息矩阵 \mathbf{D} 来表示。由于建立在属性交叉和个性化特征上, 隐式的用户-物品的非独立同分布性是客观的。因此, 隐式用户-物品非独立同分布性也称为客观的用户-物品非独立同分布性。

表(A)中的最终评价反映出了表(B)~(D)中所体现的信息、交互及其协同作用的复合效应。

6.2. 用户的非独立同分布性

如图2中的表(B)所示, 用户的非独立同分布性 \mathbb{RS}^B 是嵌入到用户信息表中的。它从用户本身、用户属性值、用户属性和用户组群等方面抓取了用户之间的交互、联系和影响。例如, 如表(B)所示, 用户 u_1 和用户 u_2 的年龄和性别不同, 但是仍然有所联系, 因为他们都住在悉尼而且年龄相仿。

相应地, 表(B)的用户信息包含:

(1) 用户内非独立同分布性 $B_a(\cdot)$, 它在用户属性值之中抓取了非独立同分布性的价值矩阵(类似于耦合行为分析中的观念内部耦合行为[38]; 感兴趣的读者可以参考文献[39]中所提到的属性值及其相似矩阵之间的内部属性相似性);

(2) 用户间非独立同分布性 $B_c(\cdot)$, 它表示用户属性的非独立同分布矩阵(类似于耦合行为分析中所提到的行为相互耦合的概念; 感兴趣的读者可以参考文献[39]中所提到的属性值及其相似矩阵之间属性类似性的概念);

(3) 用户整体非独立同分布性 $B(B_a(\cdot), B_c(\cdot))$, 它将用

户内非独立同分布性和用户间非独立同分布性整合在一起, 此处 $B(\cdot)$ 表示将 $B_a(\cdot)$ 和 $B_c(\cdot)$ 整合在一起的函数(感兴趣的读者可以参考文献[38]中的耦合行为矩阵和文献[39]中的耦合目标相似矩阵)。

上述用户非独立同分布性的几个方面在整合之后, 形成表(B)所示的非独立同分布用户空间 \mathbb{RS}^B :

$$\mathbb{RS}^B = B(B_a(\cdot), B_c(\cdot)) \quad (3)$$

6.3. 物品的非独立同分布性

物品的非独立同分布性 \mathbb{RS}^C 表示在物品信息——表(C)中。它体现在物品、物品属性、物品属性值和物品类别之间的联系和影响之中。例如, 在表(C)中, 物品 i_2 和 i_3 尽管都属于同一类别 C_2 , 却在价格和子类别中有所差异。

同样地, 物品信息, 即表(C), 嵌入了物品内非独立同分布性 $C_a(\cdot)$ 、物品间非独立同分布性 $C_c(\cdot)$ 以及物品整体非独立同分布性 $C(C_a(\cdot), C_c(\cdot))$, 此处 $C(\cdot)$ 表示将 $C_a(\cdot)$ 和 $C_c(\cdot)$ 整合在一起的函数。

物品非独立同分布性在上述几个方面的整合之后, 形成表(C)所示的用户非独立同分布空间 \mathbb{RS}^C :

$$\mathbb{RS}^C = C(C_a(\cdot), C_c(\cdot)) \quad (4)$$

6.4. 用户-物品显式非独立同分布性

表(A)用户-物品交互作用中的用户-物品非独立同分布性 \mathbb{RS}^A 通过评价反映了用户同物品间的显式而主观的相互作用的影响。表(A)中的用户-物品评价非独立同分布性 $A(\cdot)$ 可以进一步分解为用户-用户评价的非独立同分布性和物品-物品评价的非独立同分布性。例如, 在表(A)中, 用户 u_2 和 u_3 给予物品 i_1 和 i_2 相同评价, 却给予 i_3 不同评价。

相应地, 可以将 $A(\cdot)$ 按类别分为用户内部评价的非独立同分布性 $A_a(\cdot)$ 、物品之间评价的非独立同分布性 $A_c(\cdot)$, 以及整体评价的非独立同分布性 $A(A_a(\cdot), A_c(\cdot))$, 此处 $A(\cdot)$ 表示将 $A_a(\cdot)$ 和 $A_c(\cdot)$ 整合在一起的函数。完整的显式非独立同分布下的用户-物品相似性 \mathbb{RS}^A 可以表示为

$$\mathbb{RS}^A = A(A_a(\cdot), A_c(\cdot)) = A(A_{j_1, j_2}, A_{i_1, i_2}) \quad (5)$$

式中, 非独立同分布 A_{j_1, j_2} 代表对于物品 j_1 和 j_2 之间所获评价的非独立同分布; 非独立同分布 A_{i_1, i_2} 代表对于用户 i_1 和 i_2 之间评价的非独立同分布。

现有的研究大都忽视了评价的非独立同分布性。因此, 显式非独立同分布的用户-物品相似性 \mathbb{RS}^A 简化为

$$\mathbb{RS}^A = A(A_{ij}) \quad (6)$$

式中, A_{ij} 代表偏好评价矩阵; $A(\cdot)$ 为总体函数。

6.5. 用户-物品隐式的非独立同分布性

推荐系统中最有趣也是最复杂的非独立同分布存在于表(D)中, 即用户-物品的隐式、客观的非独立同分布 \mathbb{RS}^D 。它体现了用户属性和物品属性之间隐式的但是客观的相互作用。它的复杂性在于, 表(D)中可能有着层次型的非独立同分布, 它们反映着表(B)和表(C)之间的交互和影响。

在表(D)中, 具有下标的单元, 如“ i_1j_1 ”, 称为一个用户-物品耦合单元。每个单元的非独立同分布, 即 $D_{i_1j_1}$, 是由矩阵 C_a 和矩阵 B_a 的积(C_a 为表示某个特定物品属性 q_{j_1} 的矩阵, B_a 为表示某个特定用户属性 p_{i_1} 的矩阵)来表示的。表(D)阐释了单元内和单元间的非独立同分布性。

用户-物品耦合单元 $D_{i_1j_1}$ 的隐式非独立同分布 $\mathbb{RS}^{D_{i_1j_1}}$ 或许可以理解为一个矩阵。它由两部分构成: ①用户 i_1 的特定属性在所有具有属性值 j_1 的物品上的非独立同分布, 即 $D_a(D_{i_1j_1})$ ($1 \leq j_1 \leq J$); ②物品 j_1 的特定属性在所有具有属性值 i_1 的用户上的非独立同分布, 即 $D_e(D_{i_1j_1})$ ($1 \leq i_1 \leq I$)。

例如, 在图1中, SP表示某位用户的性别和某种物品的价格之间的隐式关联。此外, SP也可能被表(D)中的其他关联所影响, 如SC、NP和AP。例如, 这三位用户可能来自同一家庭, 其中, Cindy和John可能是夫妻因而会影响对方做出评价。Julie可能是他们的女儿, 也许对价格不敏感但是会更多地受到来自母亲的影响。可能John关注质量而Cindy对价格更敏感, Julie则倾向于做出平衡二者后的决策。此外, i_2 和 i_3 可能属于同一类别的物品。

这样, 用户-物品耦合单元的隐式非独立同分布性由以下表达式来度量:

$$\mathbb{RS}^{D_{i_1j_1}} = \mathbb{RS}^{D_{i_1j_1}}(D_a(D_{i_1j_1}), D_e(D_{i_1j_1})) \quad (7)$$

最后, 隐含在表(D)中的总体用户-物品隐式非独立同分布 \mathbb{RS}^D 为所有用户-物品耦合单元的非独立同分布的集合, 可表达为

$$\mathbb{RS}^D = \mathbb{RS}^D(\mathbb{RS}^{D_{i_1j_1}}) \quad (8)$$

此处, $(i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J)$ 。

表(D)中用户-物品整体隐式非独立同分布由两部分构成: ①单个用户 i_1 在所有物品属性上的非独立同分布 $D_a(\cdot)$, 代表着不同物品属性和同一用户属性之间的非独立同分布; ②单个物品 j_1 在所有用户属性上的非独立同分布 $D_e(\cdot)$, 代表着不同用户属性和某一特定物品属性之间的非独立同分布。用 \mathbb{RS}^D 来表示这两部分的耦合性。

$$\mathbb{RS}^D = \mathbb{RS}^D(D_a(\cdot), D_e(\cdot)) \quad (9)$$

\mathbb{RS}^D 无法像 \mathbb{RS}^A 、 \mathbb{RS}^B 和 \mathbb{RS}^C 一样用一个简单的矩阵来表达。从承载量和不同属性、层面、形式之间的交互来说, 它都比后者承载了更大量的信息。到目前为止, 还没有对非独立同分布 \mathbb{RS}^D 展开研究的工作。

用一个例子来说明表(D)中的非独立同分布, 它包含表(D)中的单元CP。CP由 CP_a 和 CP_e 构成, 那么 CP_a 表示某位用户所在城市中所有物品价格间的联系和差异, 而 CP_e 则计算出所有用户城市针对物品属性-价格的联系和差异。

6.6. 推荐系统非独立同分布

基于上述存在于用户、物品, 以及用户-物品的显式/隐式交互中的非独立同分布, 我们先通过一个总体方程 $\mathbb{RS}^{A+D}(\cdot)$ 来描绘一个完整的包含所有用户-物品的推荐系统非独立同分布, 以便于将显式用户-物品非独立同分布 \mathbb{RS}^A 和隐式用户-物品非独立同分布 \mathbb{RS}^D 结合在一起:

$$\begin{aligned} \mathbb{RS}^{A+D} &= \mathbb{RS}^{A+D}(\mathbb{RS}^A, \mathbb{RS}^D) \\ &= \sum_{i_1, i_2=1}^I \sum_{j_1, j_2=1}^J \mathbb{RS}^{A+D}(A(A_{i_1j_1}, \mathbb{RS}^D(\mathbb{RS}^D_{i_1j_1}))) \odot (A_{i_1j_1}, \mathbb{RS}^D_{i_1j_1}) \end{aligned} \quad (10)$$

式中, $\sum_{j_1, j_2=1}^J \mathbb{RS}^{A+D}(A(A_{i_1j_1}, \mathbb{RS}^D(\mathbb{RS}^D_{i_1j_1}))) \odot$ 表示 \mathbb{RS}^{A+D} 的后续非独立同分布为以 $A(A_{i_1j_1}, \mathbb{RS}^D(\mathbb{RS}^D_{i_1j_1}))$ 形式耦合的 $A_{i_1j_1}$ 和以 $\mathbb{RS}^D(\mathbb{RS}^D_{i_1j_1})$ 形式耦合的 $\mathbb{RS}^D_{i_1j_1}$, 以此类推, 并且具有不确定性。

推荐系统问题中完整的非独立同分布定义如下。

定义1(推荐系统非独立同分布)。一个推荐系统完整的非独立同分布, 即 \mathbb{RS} , 为来自于四个方面的非独立同分布的集合: 表(B)中的用户非独立同分布 \mathbb{RS}^B 、表(C)中的物品非独立同分布 \mathbb{RS}^C 、表(A)中的显式用户-物品非独立同分布 \mathbb{RS}^A , 以及表(D)中的隐式用户-物品非独立同分布 \mathbb{RS}^D 。

$$\mathbb{RS} = \mathbb{RS}(\mathbb{RS}^A, \mathbb{RS}^B, \mathbb{RS}^C, \mathbb{RS}^D) \quad (11)$$

式中, $RS(\cdot)$ 为总体函数。

最后, 我们定义非独立同分布推荐系统。

定义2(非独立同分布推荐系统)。给定一个推荐系统问题 \mathbb{X} , 包含用户信息矩阵 \mathbb{B} 、物品信息矩阵 \mathbb{C} 、评价矩阵 \mathbb{A} 和环境 \mathbb{E} , 一个非独立同分布推荐系统会做到:

(1) 学习完整的非独立同分布 RS , 包括学习评价的非独立同分布 $RS^A(\mathbb{A})$ 、用户非独立同分布 $RS^B(\mathbb{B})$ 、物品非独立同分布 $RS^C(\mathbb{C})$ 、显式用户-物品非独立同分布 $RS^D(\mathbb{B}, \mathbb{C})$, 以及它们的合成方法 $RS(\cdot)$ 。

$$RS = RS(RS^A(\mathbb{A}), RS^B(\mathbb{B}), RS^C(\mathbb{C}), RS^D(\mathbb{B}, \mathbb{C})) \quad (12)$$

(2) 学习环境 \mathbb{E} 条件下的估计函数 $\hat{N}(\cdot)$, 以近似地表达在现实世界中推荐问题 \mathbb{X} 的内在本质 $N(\cdot)$:

$$\hat{N}(\mathbb{A}, \mathbb{B}, \mathbb{C}, RS|\mathbb{E}) \approx N(\mathbb{X}) \quad (13)$$

(3) 优化目标函数[如损耗函数 $L(\cdot) \rightarrow 0$]来获得最为合适的近似估计 \hat{N} 。

$$L(\cdot) = \arg \min(N(\mathbb{X}) - \hat{N}(\mathbb{A}, \mathbb{B}, \mathbb{C}, RS|\mathbb{E})) \quad (14)$$

7. 非独立同分布推荐的案例研究

如在6.2部分和6.3部分中所讨论的, 一个非独立同分布推荐的例子是了解用户和物品的耦合关系。对用户/物品耦合建模的原理是:

(1) 学习用户的相似性, 即了解用户自身属性值之间、各用户属性之间和各用户之间的相似性, 从而归纳出值的相似性、属性的相似性和用户的相似性;

(2) 学习物品的相似性, 类似于学习用户的相似性, 了解物品自身属性值之间、各物品属性之间和各物品之间的相似性, 并整合出值的相似性、属性的相似性和物品的相似性;

(3) 集成用户/物品的相似性, 通过考虑不同层次用户的相似性和物品的相似性整合用户/物品的相似性。

一些前期工作已经开始在经典的CF之上探索对用户/物品耦合进行建模, 特别是MF的建模。例如:

(1) 基于耦合物品相似性的协同过滤 [11], 其中, 耦合物品的相似性以耦合对象相似性(文献[39,40])的方式建模, 这里通过合并耦合物品属性的相似性并引入耦合 K -模式算法来预测评价。

(2) 基于耦合物品相似性的MF [27], 根据耦合对象相似性(文献[39,40])来学习耦合物品的相似性, 将这种相似性加入到MF的目标函数中, 以学习潜在的用户和

物品的关系矩阵。

(3) 基于耦合的用户/物品相似性的MF [10], 根据耦合对象相似性(文献[39,40])学习耦合用户相似性与耦合物品相似性, 将两者的相似性纳入MF目标函数用以优化。

表1展示了文献[10]中关于耦合矩阵分解(CMF)与两种CF方法的比较结果: 基于用户的CF(UBCF, 它首先通过在评分矩阵Pearson相关计算用户的相似性, 然后根据这些用户建议的相关物品给定用户谁拥有强关联)[16]和基于物品的CF(IBCF, 它首先通过在评分矩阵Pearson相关考虑物品相似, 之后推荐与给定用户的物品兴趣有强关联的物品)[19]。在CMF的潜在维度100上, Movielens结果表明, CMF获得0.49%和2.42% w.r.t.的平均绝对误差(MAE)和0.18%和19.54% w.r.t.的均方根误差(RMSE)。

Bookcrossing结果表明, CMF获得33.02%和31.03%的平均绝对误差(MAE), 以及24.68%和19.04%的均方根误差(RMSE)。这一结果表明, 考虑到用户/物品的耦合, CMF基本上击败了UBCF和IBCF方法。

CMF的上述结果表明, CMF采用了微观的用户/物品来形成对特定推荐问题更全面的理解, 从而弥补主观评价的不足。因此, CMF拥有通用的和具体的建模能力, 而基本的MF只体现在通用方面。此外, 文献[10,11,27]中的结果(有兴趣的读者可以在文献[10,11,27]中找到详细信息)显示, 在推荐系统, 耦合用户相似性和耦合物品相似性的初步应用, 揭示了通过考虑微观的属性信息和关系可发掘出内在的用户和物品之间微观的互动与影响。

事实上, 如在文献[5]中所述, 这样的耦合关系尚未在经典CF和其他推荐算法中考虑, 这些算法忽视了物品属性、用户属性和物品、用户交互等的完整参与。这就解释了为什么这样的算法对特定工作并不十分适用, 尽管它们提供了通用的并独立于具体应用的解决方案。

表1 数据集Movielens、Bookcrossing的耦合矩阵分解和协同过滤结果对比

Data set	Metrics	UBCF (Improve)	IBCF (Improve)	CMF
Movielens	MAE	0.9027 (0.49%)	0.9220 (2.42%)	0.8978
	RMSE	1.0022 (0.18%)	1.1958 (19.54%)	1.0004
Bookcrossing	MAE	1.8064 (33.02%)	1.7865 (31.03%)	1.4762
	RMSE	3.9847 (24.68%)	3.9283 (19.04%)	3.7379

CMF: coupled matrix factorization; CF: collaborative filtering; UBCF: user-based CF; IBCF: item-based CF; MAE: mean absolute error; RMSE: root mean square error.

8. 展望

现阶段，基于推荐系统的研究正面临着重大挑战。一方面，许多经典问题尚未得到解决；另一方面，重大研究突破与系统创新越来越少。对所研究的推荐系统内在的复杂性及其本质的深入了解显得越发重要。为此，我们必须掌握微观数据的特征(特别是所推荐的用户之间以及物品之间存在的复杂的耦合和异构关系)，进而着眼于新一代非独立同分布推荐系统的研究。

基于以上观点，尽管推荐系统已有众多深入的研究，但未来仍面临着各类重大挑战，并同时具有很多在理论上突破的机遇。以下是对非独立同分布推荐系统研究的五大基本原则的简单总结。

(1) 原则1。将微观和分层变量以及值到对象(value-to-object)的映射(如用户、物品、环境)的耦合关系整合到基于模型的方法中，通过创建数据和模型驱动的推荐系统获取合理的推荐。

(2) 原则2。抓取推荐模型中显式/隐式变量、关系以及特定的异构性特征，以便应对推荐系统问题可能存在的各个环节、各类特征以及复杂性。

(3) 原则3。针对用户、物品和用户-物品交互，对非独立同分布性进行学习。除了对复杂交互的耦合学习[6]之外，非独立同分布性的学习[5]还涉及多要素、类型、形式、层次、结构、分布、关系及其协同(更多的相关讨论请参考引用的文献[5,6,38])，而更具有挑战性的是耦合与异构学习的融合。

(4) 原则4。对图2中的表(D)所示的用户-物品的非独立同分布性的建模。表(D)所展示的非独立同分布性尚未在相关的领域中得到研究。对表(D)中隐式且复杂的用户-物品交互的学习十分重要，因为它们是实现评分行为和偏好及其整个过程的主要动力。为此，我们必须学习多个矩阵的耦合以及矩阵间耦合的层次性。

(5) 原则5。将图2中所有表格中的非独立同分布性进行整合。这包括主观与客观的耦合以及显式与隐式的耦合，四个表格中包含了这些耦合且表现形式是不同的。

第6部分提出的非独立同分布推荐系统框架在理论和系统创新，解决现存典型难题，提供合理的、相关的、个性化的以及可执行的推荐等方面都具有重大潜力。此外，现有大部分的工作[41](包括对社会关系的引入、跨域以及跨群组)只讨论了一些特定的案例或仅仅解决了所述的非独立同分布推荐系统框架的某些特殊的议题。

特别地，以下对非独立同分布推荐系统研究的一些延伸与实例化的可能性展开讨论。

(1) 物品耦合关系建模。对物品内部以及物品间的耦合关系进行建模，将它们的相似性整合到已有的学习模型中，可以大幅提升基本模型的功能以满足物品间的交互与关联挖掘的需求。

(2) 提高对用户的分析与建模的能力。现有模型往往仅侧重某类用户信息或某方面的建模分析，如所谓的用户本体分析、社交用户分析、隐式分析、显式分析、基于CF的分析以及基于数据挖掘和机器学习应用的分析等，在分析与建模过程中利用完整的用户信息[42,43]可以有效地弥补现有模型的这些不足。

(3) 社会关系建模。把用户耦合关系作为独立主题(如用户好友关系或用户间的推特发布及转发)或多个主题(如同时模拟用户好友关系和用户形象)进行建模，这类在推荐系统的社交媒体中的社会关系的建模是我们所提出的非独立同分布推荐系统的一个特例。事实上，针对单个用户属性或多个用户属性的用户耦合关系的建模可以提出许多不同算法。这里的主要区别在于不仅对用户内在的属性耦合而且对用户之间关系属性的耦合关系进行建模。

(4) 处理关注度偏差。在大量长尾物品中经常只有一个小部分的人气物品，这将引起数据稀疏和覆盖不足等问题[44]。从非独立同分布的视角，对较有人气和冷门(rare)的用户/商品的相似性进行建模，可以补充有效打分信息的缺失。这就在现有数据降维和基于图数据传递关系之上建立了一个新的视角来连接人气物品评分和冷门商品评分。对此感兴趣的读者可以参考文献[45]，从文档分析中关键词的间接联系可以得到有用的启发。

(5) 跨域推荐建模。跨域推荐用来向本域的用户推荐另一个域的物品[46]。当关注物品的域信息(如品类和子品类、产品类型和使用目的)时，对物品耦合性的学习可被认为特指对跨域要素的学习。与用户耦合建模类似，针对跨域物品耦合模型，不管是单个还是多个物品的属性，可以提出多种不同的算法。物品内部以及物品之间属性的相似性都需要进行学习。例如，文献[24,25]就对跨域推荐进行了建模。部分相关工作用到了迁移学习[46,47]，这些其实属于上文所述跨域建模的一些特殊例子。当源和目标域具有差异性时，当前的迁移学习可能并不有效。非独立同分布推荐则可以通过源和目标域之间的非独立同分布性的建模发挥良好的作用。

(6) 群组推荐建模。这是通过用户分组建模对用户

耦合性学习的另一个特殊例子。例如,文献[23]中所涉及的工作即对群的偏好进行了建模。在文献[27]中提出了基于群的矩阵分解算法(CGMF),这种算法在考虑用户组内组外耦合关系的同时,增加了社交媒体中的用户分组以适应某类特定群组的特征。更为复杂的是,当一个推荐问题涉及系统嵌套的现象时[48],如何对跨组的偏好和差异进行合理的建模。此时我们需要学习群之间的非独立同分布性。

(7) 冷启动问题。这主要涉及预测新物品评分或向新用户推荐现有物品[49-53]。需要解决长尾以及新的用户/物品由于得到的反馈远低于(甚至没有)人气用户/物品而无法进行准确建模的问题。根据非独立同分布推荐的原则,这一问题可能通过对非独立同分布的用户和(或)物品的相似性进行建模得到解决,依据这些相似性,向新用户或新物品提供相应的推荐。

(8) 欺诈攻击问题。出于一些不正当的目的,会产生许多虚假评分[54]。通过对用户和物品真正的非独立同分布性的建模,以识别出与模型结果不一致、相偏离的评分,虚假评分问题可能得到解决(感兴趣的读者可以参考文献[55]中介绍的耦合孤立点检测方法)。

(9) 情境感知推荐。对情境推荐[56]的一般理解可以认为是在一定的用户/物品限制条件下进行推荐或者假设,如表(B)和表(C)中所列在某个特定用户/物品集条件下进行推荐,因此从非独立同分布用户/物品的视角对相应的非独立同分布性进行建模,就可以有效解决此类问题。此外,当等式(14)中的环境E不包含用户/物品信息时,如当E代表推荐问题中可能涉及的季节、经济或社会文化因素时,我们需要通过加入第五个表格(E)来获取情境信息,以考虑环境和用户/物品之间的相互作用。对于这种情况,推荐理论需要处理定义2中所列的目标。

(10) 面向推荐的人机交互。主要涉及:①对表(B)中所列用户信息的高效全面的采集与应用;以及②将人类定性的智慧作为计算/决策组件的一部分,与推荐系统相融合。对图1中表(B)中用户非独立同分布性的学习可以解决目标1。例如,基于用户对物品的偏好与观点的建模是图2中表(D)所示用户-物品耦合学习的一个特例,它仅着眼于对用户评论的理解。目标2涵盖了多方面人类定性的智慧[48],这些内容无法通过数据源直接获取,但对高质量的推荐十分重要。许多跨学科的研究机会可能涌现,如对人类感知、认知、心理和社会文化等方面的认识,以及对人类智慧如何影响某个推荐的采纳与作用的理解;对人类决策与选择的学习;对人的个性与偏好的表达与了解;对信用和隐私及它们

在某个群组 and 群体场景下的推论的模拟;将真实的用户需求与期望模型化;以及对个体或群组在推荐和决策过程中涉及的情感、人际交往、经验、态度和动机等因素进行建模[2]。本研究将导向基于人机交互的人机合作式的或者以人为本的推荐系统[57],并促进人类与机器智能的融合[58]。

(11) 众包的推荐系统。这一系统涉及多角色,包括任务/服务请求者、工作人员和供应商;而且是多目标的[59,60]:包括奖励、配送的成本效益、技能匹配、总体任务完成率和累计佣金等。推荐可以在多目标的优化过程中发挥作用。如果我们能获取不同用户的角色、任务描述以及优化目标等信息,这一问题就可以被映射为一个多视角问题:一个多类型用户信息表、一个任务信息表以及一个优化目标表,分别可以对应为图1中的表(A)~(C)。由此,本文所提出的非独立同分布推荐可以应用于优化众包。

9. 结论

在社交媒体、电子商务、移动服务以及广告等数字经济与商业活动中,推荐系统扮演了日益重要的角色。当前的推荐理论和系统主要建立在被推荐的物品和接收推荐信息的用户呈现独立而且同分布的假设之上。本研究分析了围绕独立同分布理论的若干问题,并基于对用户与用户、物品与物品以及用户与物品之间的耦合性和异构性的考量,将非独立同分布性引入推荐功能中来。介绍了非独立同分布推荐框架用来整合显式与隐式、主观与客观、本地与全局的非独立同分布性。这一非独立同分布框架对现有推荐理论和系统提出了挑战,并将带来重大的理论突破,同时大大激发了下一代推荐研究与应用的创新机会。

非独立同分布性的学习是数据科学和大数据分析领域的一个重大挑战。它提出了在遇到数据分析、信息处理、统计、模式识别以及学习系统等领域的经典理论和工具时所面临的一些关键问题。希望本研究对非独立同分布推荐的一些探索,以及其他议题的讨论有所启发,形成一个从独立同分布学习向非独立同分布学习思维的转变,以期能为理论的突破与实践的提升做出贡献。

References

- [1] Jannach D, Zanker M, Felfernig A, Friedrich G. *Recommender systems: an introduction*. Cambridge: Cambridge University Press; 2010.

- [2] Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommender systems handbook. 2nd ed. New York: Springer; 2015.
- [3] Cao L. Data science: a comprehensive overview. Technical report. Sydney: University of Technology Sydney; 2016.
- [4] McKinsey Global Institute; Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: the next frontier for innovation, competition, and productivity. New York: McKinsey Global Institute; 2011.
- [5] Cao L. Non-IIDness learning in behavioral and social data. *Comput J* 2014;57(9):1358–70.
- [6] Cao L. Coupling learning of complex interactions. *Inform Process Manag* 2015;51(2):167–86.
- [7] Cao L. In-depth behavior understanding and use: the behavior informatics approach. *Inform Sciences* 2010;180(17):3067–85.
- [8] Cao L, Yu PS, editors. Behavior computing: modeling, analysis, mining and decision. London: Springer; 2012.
- [9] Fu B, Xu G, Cao L, Wang Z, Wu Z. Coupling multiple views of relations for recommendation. In: Cao T, Lim EP, Zhou ZH, Ho TB, Cheung D, Motoda H, editors *Advances in Knowledge Discovery and Data Mining: 19th Pacific-Asia Conference, Part II*; 2015 May 19–22; Ho Chi Minh City, Vietnam. Switzerland: Springer International Publishing; 2015. p. 723–43.
- [10] Li T, Lu J, López LM. Preface: intelligent techniques for data science. *Int J Intell Syst* 2015;30(8):851–3.
- [11] Yu Y, Wang C, Gao Y, Cao L, Chen X. A coupled clustering approach for items recommendation. In: Pei J, Tseng VS, Cao L, Motoda H, Xu G, editors *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, Part II*; 2013 Apr 14–17; Gold Coast, Australia. Heidelberg: Springer; 2013. p. 365–76.
- [12] Cao L, Yu PS. Non-IID recommendation theories and systems. *IEEE Intell Syst* 2016;31(2):81–4.
- [13] Cao L. Data science and analytics: a new era. *Int J Data Sci Analyt* 2016;1(1):1–2.
- [14] Cao L. Data science: intrinsic challenges and directions. Technical report. Sydney: University of Technology Sydney; 2016.
- [15] Cao L. Data science: nature and pitfalls. Technical report. Sydney: University of Technology Sydney; 2016.
- [16] Su X, Khoshgoftaar TM. A survey of collaborative filtering techniques. *Adv Artif Intell* 2009;2009(4):1–19.
- [17] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24–27; New York, USA; 2008. p. 426–34.
- [18] Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on the World Wide Web; 2001 May 1–5; Hong Kong, China; 2001. p. 285–95.
- [19] Deshpande M, Karypis G. Item-based top-N recommendation algorithms. *ACM Trans Inform Syst* 2004;22(1):143–77.
- [20] Ma H, Yang H, Lyu MR, King I. SoRec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management; 2008 Oct 26–30; Napa Valley, CA, USA; 2008. p. 931–40.
- [21] Ma H. An experimental study on implicit social recommendation. In: Proceedings of the 36th International ACM SIGIR conference on Research and Development in Information Retrieval; 2013 Jul 28–Aug 1; Dublin, Ireland; 2013. p. 73–82.
- [22] Singh AP, Gordon GJ. Relational learning via collective matrix factorization. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24–27; Las Vegas, NV, USA; 2008. p. 650–8.
- [23] Hu L, Cao J, Xu G, Cao L, Gu Z, Cao W. Deep modeling of group preferences for group-based recommendation. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence; 2014 Jul 27–31; Québec City, Canada; 2014. p. 1861–7.
- [24] Hu L, Cao J, Xu G, Wang J, Gu Z, Cao L. Cross-domain collaborative filtering via bilinear multilevel analysis. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence; 2013 Aug 3–9; Beijing, China; 2013. p. 1–7.
- [25] Hu L, Cao J, Xu G, Cao L, Gu Z, Zhu C. Personalized recommendation via cross-domain triadic factorization. In: Proceedings of the 22nd International Conference on World Wide Web; 2013 May 13–17; Rio de Janeiro, Brazil; 2013. p. 595–606.
- [26] Yang X, Steck H, Liu Y. Circle-based recommendation in online social networks. In: Proceedings of the 18th ACM SIGKDD Knowledge Discovery and Data Mining; 2012 Aug 12–16; Beijing, China; 2012. p. 1267–75.
- [27] Li F, Xu G, Cao L, Fan X, Niu Z. CGMF: coupled group-based matrix factorization for recommender system. In: Lin X, Manolopoulos Y, Srivastava D, Huang G, editors *Web Information Systems Engineering—WISE 2013: 14th International Conference, Part I*; 2013 Oct 13–15; Nanjing, China. Heidelberg: Springer. 2013. p. 189–98.
- [28] Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence; 1998 Jul 24–26; Madison, WI, USA. San Francisco: Morgan Kaufmann Publishers Inc.; 1998. p. 43–52.
- [29] Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work; 1994 Oct 22–26; Chapel Hill, NC, USA; 1994. p. 175–86.
- [30] Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* 2000;97(18):10101–6.
- [31] Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In: Platt JC, Koller D, Singer Y, Roweis ST, editors *Proceedings of the 21st Annual Conference on Neural Information Processing Systems 2007*; 2007 Dec 3–6; Vancouver, Canada; 2007. p. 1257–64.
- [32] Burke R. Hybrid web recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors *The adaptive web*. Heidelberg: Springer; 2007. p. 377–408.
- [33] Burke R. Hybrid recommender systems: survey and experiments. *User Model User-Adapt Interac* 2002;12(4):331–70.
- [34] Lv LL, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T. Recommender systems. *Phys Rep* 2012;519(1):1–49.
- [35] Konstan JA, Riedl J. Recommender systems: from algorithms to user experience. *User Model User-Adapt Interact* 2012;22(1):101–23.
- [36] Bobadilla J, Ortega F, Hernando A, Gutiérrez A. Recommender systems survey. *Knowl-Based Syst* 2013;46:109–32.
- [37] Park DH, Kim HK, Choi IY, Kim JK. A literature review and classification of recommender systems research. *Expert Syst Appl* 2012;39(11):10059–72.
- [38] Cao L, Ou Y, Yu PS. Coupled behavior analysis with applications. *IEEE Trans Knowl Data Eng* 2012;24(8):1378–92.
- [39] Wang C, Dong X, Zhou F, Cao L, Chi CH. Coupled attribute similarity learning on categorical data. *IEEE Trans Neural Netw Learn Syst* 2015;26(4):781–97.
- [40] Wang C, Cao L, Wang M, Li J, Wei W, Ou Y. Coupled nominal similarity in unsupervised learning. In: Proceedings of the 20th ACM Conference on Information and Knowledge Management; 2011 Oct 24–28; Glasgow, UK; 2011. p. 973–8.
- [41] Chen L, Zeng W, Yuan Q. A unified framework for recommending items, groups and friends in social media environment via mutual resource fusion. *Expert Syst Appl* 2013;40(8):2889–903.
- [42] Nadee W. Modeling user profiles for recommender systems [dissertation]. Brisbane: Queensland University of Technology; 2016.
- [43] Li R, Wang S, Deng H, Wang R, Chang KCC. Towards social user profiling: unified and discriminative influence model for inferring home locations. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2012 Aug 12–16; Beijing, China; 2012. p. 1023–31.
- [44] Popescul A, Ungar LH, Pennock DM, Lawrence S. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence; 2001 Aug 2–5; Seattle, WA, USA. San Francisco: Morgan Kaufmann Publishers Inc.; 2001. p. 437–44.
- [45] Chen Q, Hu L, Xu J, Liu W, Cao L. Document similarity analysis via involving both explicit and implicit semantic couplings. In: Proceedings of IEEE Data Science and Advanced Analytics 2015; 2015 Oct 19–21; Paris, France; 2015. p. 1–10.
- [46] Jiang M, Cui P, Chen X, Wang F, Zhu W, Yang S. Social recommendation with cross-domain transferable knowledge. *IEEE Trans Knowl Data Eng* 2015;27(11):3084–97.
- [47] Pan W, Yang Q. Transfer learning in heterogeneous collaborative filtering domains. *Artif Intell* 2013;197:39–55.
- [48] Cao L. Metasyntactic computing and engineering of complex systems. London: Springer-Verlag; 2015.
- [49] Son LH. Dealing with the new user cold-start problem in recommender systems: a comparative review. *Inform Syst* 2016;58:87–104.
- [50] Gantner Z, Drumond L, Freudenthaler C, Rendle S, Schmidt-Thieme L. Learning attribute-to-feature mappings for cold-start recommendations. In: Proceedings of the 10th IEEE International Conference on Data Mining; 2010 Dec 13–17; Sydney, Australia; 2010. p. 176–85.
- [51] Mirbakhsh N, Ling CX. Improving top-N recommendation for cold-start users via cross-domain information. *ACM Trans Knowl Discov Data* 2015;9(4):33.
- [52] Lika B, Kolomvatos K, Hadjiethymiades S. Facing the cold start problem in recommender systems. *Expert Syst Appl* 2014;41(4):2065–73.
- [53] Gao H, Tang J, Liu H. Addressing the cold-start problem in location recommendation using geo-social correlations. *Data Min Knowl Disc* 2015;29(2):299–323.
- [54] Gunes I, Kaleli C, Bilge A, Polat H. Shilling attacks against recommender systems: a comprehensive survey. *Artif Intell Rev* 2014;42(4):767–99.
- [55] Pang G, Cao L, Chen L. Outlier detection in complex categorical data by modelling the feature value couplings. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence 2016; 2016 Jul 9–15; New York, NY, USA; 2016. p. 1–7.
- [56] Hidasi B, Tikk D. General factorization framework for context-aware recommendations. *Data Min Knowl Disc* 2016;30(2):342–71.
- [57] Jacko JA, editor. The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. 3rd ed. Boca Raton: CRC Press; 2006.
- [58] Qian XS, Yu JY, Dai RW. A new discipline of science—the study of open complex giant system and its methodology. *Chin J Syst Eng Electron* 1993;4(2):2–12.
- [59] Liu X, Nielek R, Adamska P, Wierzbicki A, Aberer K. Towards a highly effective and robust Web credibility evaluation system. *Decis Support Syst* 2015;79:99–108.
- [60] Aldahhri E, Shandilya V, Shiva S. Towards an effective crowdsourcing recommendation system: a survey of the state-of-the-art. In: Proceedings of the 2015 IEEE Symposium on Service-Oriented System Engineering; 2015 Mar 30–Apr 3; San Francisco Bay, CA, USA; 2015. p. 372–7.