

基于约束推荐的网络可视化分析

张超杰¹, 吴果林^{1,2}

(1. 桂林航天工业学院 理学院, 广西 桂林 541004;

2. 桂林航天工业学院 广西航空物流研究中心, 广西 桂林 541004)

摘要:现如今互联网购物成为一种主要的购物方式,而各个购物平台都面临商品数据过量的问题,如何尽快地给出用户需求的商品列表就成为亟待解决的问题。基于约束的推荐系统不依赖用户评分等用户偏好的历史数据,有效解决了“冷启动”的问题。用户需求与用户意向物品的二部图网络包含了许多对推荐有价值隐藏在内部的信息,针对某在线销售记录的案例数据集进行研究,文中构建了二部图网络并对其进行了可视化,采用无标度网络模型(BA模型)分析用户、物品节点的度。由物品节点的度分布结构得出物品节点具有明显的社团属性,因此可以根据社团内部物品节点的度进行排序进行相应的推荐。把这种推荐方法加入到推荐算法中,以增加推荐的精确度,为购物平台的推荐算法设计提供帮助。

关键词:基于约束的推荐;复杂网络;二部图;可视化

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2019)10-0115-05

doi:10.3969/j.issn.1673-629X.2019.10.024

Network Visualization Analysis Based on Constraint Recommendation

ZHANG Chao-jie¹, WU Guo-lin^{1,2}

(1. School of Science, Guilin University of Aerospace Technology, Guilin 541004, China;

2. Guangxi Aviation Logistics Research Center, Guilin University of Aerospace Technology, Guilin 541004, China)

Abstract: At present, shopping on the Internet has become a major way of shopping, and each shopping platform is faced with the problem of excessive product data. How to give a list of products required by users as soon as possible has become an urgent problem to be solved. The constraint-based recommendation system does not rely on historical data such as user ratings and other user preferences, effectively solving the problem of “cold start”. The bipartite network of user needs and user-intentional items contains a lot of information that is valuable to the recommendation. We study the case data set of an online sales record, construct and visualize the bipartite graph network, and adopt a scale-free network model (BA model) to analyze the degree of user and item nodes. Since the degree distribution structure of the item node results in the item node having obvious community attribute, the corresponding recommendation can be performed according to the degree of the item node in the community. This recommendation method is added to the recommendation algorithm to increase the accuracy of the recommendation, which is helpful for the recommended algorithm design of the shopping platform.

Key words: constraint-based recommendation; complex network; bipartite network; visualization

0 引言

随着网络购物的兴起,如何在成千上万的商品中给出用户需求的产品推荐列表,就成了亟待解决的问题。基于知识的推荐系统^[1-3]能够根据用户的明确需求、产品的领域知识、通过推理为用户给出推荐。由于不依赖于用户评分等关于用户偏好的历史数据,故不存在冷启动问题。在已应用的案例中^[4-6],基于知识的推荐技术表现出优异的性能,已成为推荐系统研究

的一个重要分支和热点。一般地,基于知识的推荐系统有两种类型^[1]:基于约束的推荐^[2]和基于案例的推荐^[3]。虽然基于知识的推荐技术不依赖用户的评分,但是从复杂网络的观点来看,以往的推荐记录形成一个用户需求与用户意向物品的二部图网^[7]。显然这个二部图网络隐藏了用户需求与意向物品的内在联系,也包含了用户之间、物品之间的深层联系。为了更好地分析这些深层联系,将这些信息用于推荐系统,提

收稿日期:2018-09-26

修回日期:2019-01-29

网络出版时间:2019-04-24

基金项目:广西中青年教师基础能力提升项目(2017KY0859)

作者简介:张超杰(1990-),男,硕士,研究方向为数值分析;吴果林,博士,副教授,通讯作者,研究方向为非参数统计、推荐系统。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1047.028.html>

高推荐技术的性能,文中对基于约束的推荐过程中形成的用户需求与意向物品的二部图网络进行了可视化分析,主要包括以下几方面的内容:二部图结构可视化,用户、物品节点的度分布,单模网络的社团结构可视化等。

1 基于约束推荐的网络整理与分析

基于约束推荐网络对优质雪茄领域推荐的数据集

表 1 用户偏好与意向物品数据(部分)

用户编号	用户变量	值	用户编号	用户变量	值
12504	UM02	woechentlich	12505	UM08	von_5_bis_10
12504	UM03	selbst	12505	UM09	etui_5
12504	UM04	churchill	12505	UM010	5
12504	UM07	exotisch	12505	UM012	aufmunternd
12504	UM07	vollmundig	12506	UM02	monatlich
12504	UM08	von_10_bis_20	12506	UM03	selbst
12504	UM09	etui_5	12506	UM04	panetela
12504	UM010	117	12506	UM07	mild
12504	UM012	breit	12506	UM08	egal
12504	UM012	aufmunternd	12506	UM09	etui_5
12505	UM02	woechentlich	12506	UM010	4
12505	UM03	selbst	12506	UM010	5
12505	UM04	perfecto	12506	UM010	49
12505	UM07	wuerzig	12506	UM012	mild

表 1 中,用户变量列表示用户偏好提问,值列表示用户偏好回答,其中用户变量 UM010 所对应的值表示用户意向物品回答,关于每个用户变量的意义以及用户变量与物品属性之间的约束关系等其他详细信息可参考文献[9]。这些数据堆砌在一起时,如果仅用数据表格或文字的形式来表示,则理解的内容非常有限。易见,上面数据集包含了用户偏好与意向物品的二元

进行分析研究,以期找到用户和物品之间的一些深层联系。该数据集是由瑞士一个在线雪茄销售商店于 2005 年 10 月至 2009 年 5 月收集,并由 Zanker 等^[8]整理后应用于约束推荐研究的案例数据集,包含了 535 个显示用户偏好与用户意向物品的会话。表 1 显示了其中 3 个用户会话的数据。

关系,为了更直观、更清晰地描述用户偏好与意向物品、用户之间、物品之间的联系,因而将该数据集提炼成一个关于用户偏好与意向物品的二部图网络。为方便计,可以将一个用户的所有偏好看成一个节点(即每个用户看成一个节点),每个物品看成一个节点,这样得到一个由 535 个用户偏好节点、141 个物品节点以及 1 422 条边构成的二部图网络,如图 1 所示。

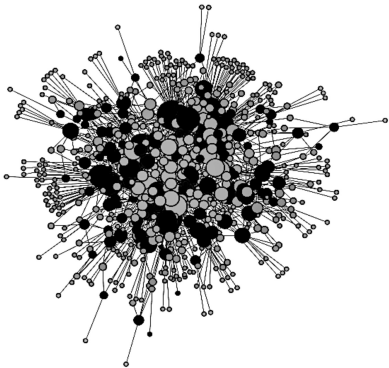


图 1 用户偏好与意向物品网络

图 1 中黑色表示用户节点,灰色表示物品节点,节点的大小由节点的度来决定。不难发现,图中除了一个物品节点与一个用户节点单独相连外,其余的节点

都相互链接且组成一个连通图。另外,从图 1 也可以看出,多数用户选择一、两个物品,少数用户选择多个物品。也即大部分用户节点只有少数几个链接,某些

用户节点却拥有与物品节点的大量链接,在度分布上具有幂律形式。相对而言,物品节点也有度数较大的集散节点,但是度数较少的节点也不是很多,分布相对

比较平均,幂律分布不明显。图 2 显示了用户与物品节点的度分布。

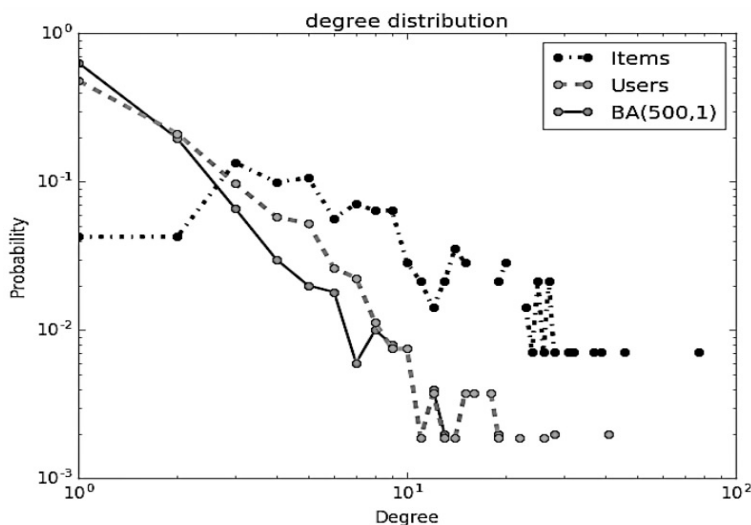


图 2 用户与物品节点的度分布

图 2 中点划线表示用户节点,虚线表示物品节点,实线表示 BA 无标度网络^[10]。图中显示的是 500 个节点,连边为 1 的 BA 无标度网络,记为 BA(500, 1)。BA(500, 1) 模型有两个重要特性:(1) 每次引入一个新的节点,并且连到 1 个已存在的节点上;(2) 一个新节点与一个已经存在的节点 v_i 相连接的概率 p_i 与节点 v_i 的度 k_i 满足如下关系:

$$p_i = \frac{k_i + 1}{\sum_j k_j + 1}$$

易见,用户节点的度分布近似于 BA(500, 1) 的度分布。如果将每次用户偏好调查看成引入一个新的节点,就能很好地解释度分布近似于 BA(500, 1) 的度分布。这是因为,多数用户选择一、两个物品类似于 BA(500, 1) 模型的特性(1);BA(500, 1) 模型随着节点的增加,大度的节点也开始少量出现,而在基于约束推荐的网络中,也有少量用户选择多个意向物品,这与 BA(500, 1) 模型的第二个特性相吻合。从物品节点的度分布来看,尽管多数用户选择一、两个物品,但物品节点的度为 1、2 的概率不是最高的,度为 3 的概率最大,度为 4~9 的概率都要大于度数为 1、2 的节点。此外,物品节点的度分布不像用户节点的度分布那样直线下降,而是缓慢下降,且大度节点的数量也比用户节点的数量要多,这体现了一些物品获得用户的普遍爱好。

2 基于约束推荐的网络社团可视化

从上面物品节点的度分布结构可以看出,一些物品被多数用户所喜好,一些物品被部分用户所喜好,而

另一些物品被少数用户所喜好,物品节点存在明显的社团属性。因此可以对物品节点进行社团可视化,观察它们的社团结构。然而,物品节点处于一个二部图网络,对物品节点进行社团划分不像单模网络那样可以直接应用现有的社团分割算法。当前,针对二部图网络社团划分有两种思路:一种是直接作用在二部图网络上,设计专门针对二部图网络的社团分割算法或者在原有针对单模网络的算法上进行修改使之能够用于二部图网络的算法,诸如 Barber 基于模块的算法^[11]、Barber 与 Clark 的 LPAb 算法^[12]以及 Aikin 与 Francis 的统计模型方法^[13];另一种是将二部图网络映射成单模网络,利用已有的单模网络的社团探测算法进行社团划分。在已知的单模网络的社团探测算法中,比较优秀的算法有 Blondel 等的 Louvain 算法^[14]、Rosvall 与 Bergstrom 的 Infomap 算法^[15]以及 Newman 的基于模块的算法^[16-17]。鉴于 Guimera 发现在二部图网络其中一类节点社团探测时,无论是先映射到该类节点后模块最大化社团划分,还是先二部图网络模块最大化社团划分再映射到这类节点上,两者之间没有差别^[18]。因此,文中采用先将二部图网络映射到物品节点上,形成一个关于物品节点的单模网络,然后再利用单模网络划分社团的方式来划分物品节点的社团。

考虑一个基于约束的推荐网络 $G = (U \cup I, E)$, 其中 U 为用户节点, I 为物品节点, E 为两者之间的连边。 $G_I = (I, E_I)$ 为网络 G 在物品节点上的映射, G_I 中的边由网络 G 确定,当且仅当 I 的节点 v_i, v_j 在 U 中至少有一个共同邻居节点时,则节点 v_i, v_j 之间存在一条边。映射有权重映射与非权重映射两种。在基于约束的推荐网络 G 中, I 中的节点 v_i, v_j 有多个共同邻居

节点表示有多个用户都喜好物品 v_i, v_j , 也即权重映射更能清楚地表达节点之间链接的信息^[19]。因此文中选择权重映射作为图 G 到图 G_l 的映射。设 f 为图 G 到图 G_l 的映射, 图 G_l 的边 (v_i, v_j) 的权重 w_{ij} 定义如下:

$$w_{ij} = |n(v_i) \cap n(v_j)|, i \neq j$$

其中, $n(v_i)$ 表示节点 v_i 的邻居节点。

利用上述映射方法, 将基于约束的优质雪茄推荐网络映射到物品节点上, 得到一个由 141 个物品节点以及 2 300 条边构成的单模网络, 并测得网络的模块度为 0.321, 是一个具有明显社团结构的网络。应用 Louvain 算法^[14], 计算出物品单模网络可分成 7 个社团, 每个社团的节点数量分布如图 3 所示。

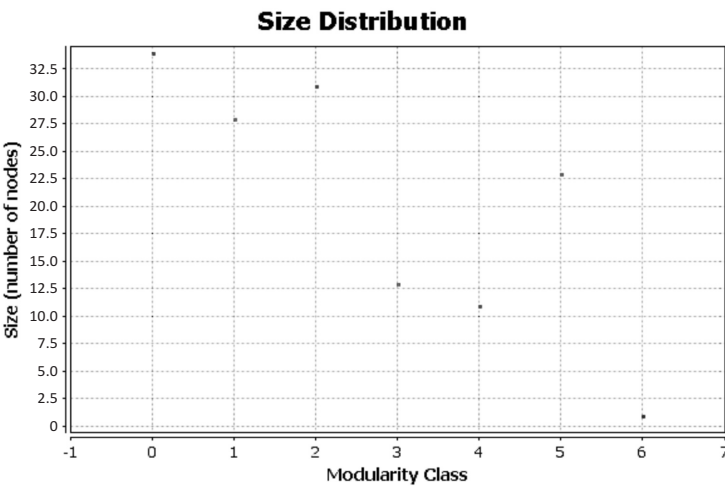


图 3 物品网络的社团数量分布

从图 3 可知, 有 1 个社团只有 1 个物品节点, 这与二部图网络中只有 1 个物品和 1 个用户组成的连通子集相对应, 可见 Louvain 算法能够很好地区分连通分支。除此之外, 在其他的社团中, 最大的社团含有 34 个物品节点, 次大的社团有 31 个节点, 次小的社团也有 12 个节点。整个网络除了一个只含单个节点的社

团外, 聚集程度还是很大的, 表现了大社团的特征。图 4 可视化了社团分类的结果。由图 4 不难看出, 权重较大的边所对应的节点大都归属于同一类, 这与定义的节点间权重越大, 两个节点包含更丰富的关联信息一致。

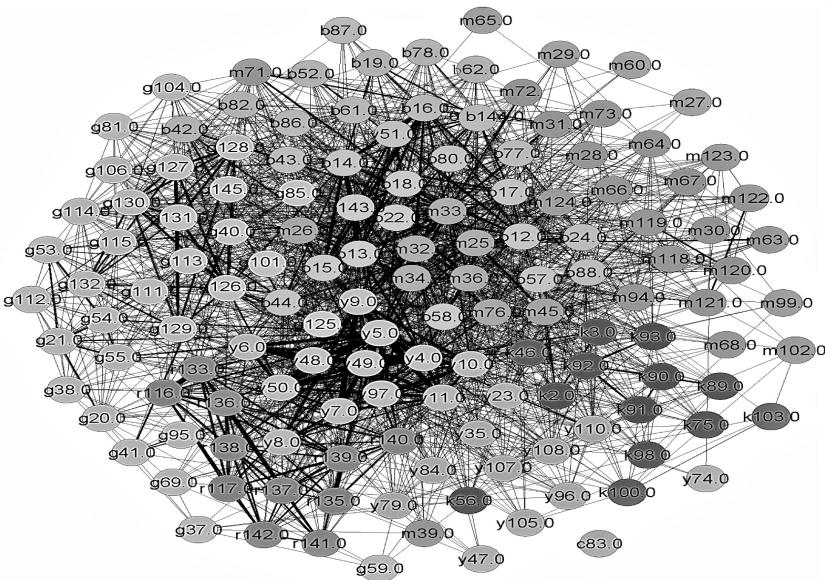


图 4 物品网络的社团结构

3 结束语

基于约束的推荐系统能够根据用户指定的需求进行个性化的推荐, 在 Felfernig 和 Zanker 等^[2-6,8-9] 许多

学者的努力下, 基于约束的推荐已成为基于知识的推荐系统一个主要的研究方向, 并成为 一个研究热点。尽管基于约束的推荐不依赖于用户评分等关于用户偏好的历史数据, 但以往的推荐记录形成的用户需求与

用户意向物品的二部图网络显然包含了许多对推荐有价值的信息。且混合的推荐方法也是当前推荐技术发展的一种趋势^[1,20-21]。通过可视化客户购物的历史记录形成的二部图网络,以期发现对推荐有价值的信息,为开发或构造混合的推荐算法提供直观的帮助。分析某在线雪茄销售记录表明:在基于用户需求与用户意向物品的二部图网络中,用户节点的度表现为类似BA(500, 1)的分布;物品节点里含有许多大度的节点,呈现出集结性、社团性特点;对物品节点的社团分析显示,物品节点可以分为7个网络社团,这些社团具有很强的层次性。

社团可视化表明,社团中的节点通过共同用户的多少聚集在一起,因此对于给定物品,可以根据物品社团内给定节点间的度进行排序,然后进行相应的推荐,将此推荐方法加入到基于约束的推荐方法中,构造一个混合的推荐算法,增加推荐的精确度,为推荐算法的设计提供帮助。

参考文献:

- [1] RICCI F, ROKACH L, SHAPIRA B. Recommender systems handbook[M]. 2nd ed. Berlin: Springer, 2015.
- [2] FELFERNIG A, BURKE R. Constraint-based recommender systems: technologies and research issues[C]//Proceedings of the 10th international conference on electronic commerce. Innsbruck, Austria: ACM, 2008.
- [3] BRIDGE D, GOEKER M, MCGINTY L, et al. Case-based recommender systems[J]. Knowledge Engineering Review, 2005, 20(3): 315-320.
- [4] ZANKER M, BRICMAN M, GORDEA S, et al. Persuasive online selling in quality & taste domains[C]//Proceedings of the 7th international conference on electronic commerce and web technologies. Krakow, Poland: Springer-Verlag, 2006: 51-60.
- [5] FELFERNIG A, FRIEDRICH G, JANNACH D, et al. An integrated environment for the development of knowledge-based recommender applications[J]. International Journal of Electronic Commerce, 2007, 11(2): 11-34.
- [6] FELFERNIG A, ISAK K, SZABO K, et al. The VITA financial services sales support environment[C]//Proceedings of the 19th national conference on Innovative applications of artificial intelligence. Vancouver, British Columbia, Canada: ACM, 2007: 1692-1699.
- [7] 何绯娟, 缪相林, 许大伟, 等. 基于“读者—图书”二部图的个性化图书推荐方法[J]. 计算机技术与发展, 2015, 25(5): 25-28.
- [8] ZANKER M, JESSENITSCHNIG M, SCHMID W. Preference reasoning with soft constraints in constraint-based recommender systems[J]. Constraints, 2010, 15(4): 574-595.
- [9] ZANKER M, JESSENITSCHNIG M. Case-studies on exploiting explicit customer requirements in recommender systems[J]. User Modeling and User-Adapted Interaction, 2009, 19(1-2): 133-166.
- [10] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [11] BARBER M J. Modularity and community detection in bipartite networks[J]. Physical Review E, 2007, 76(6): 066102.
- [12] BARBER M J, CLARK J W. Detecting network communities by propagating labels under constraints[J]. Physical Review E, 2009, 80(2): 026129.
- [13] AITKIN M, VU D, FRANCIS B. Statistical modelling of the group structure of social networks[J]. Social Network, 2014, 38(1): 74-87.
- [14] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): P10008.
- [15] ROSVALL M, BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.
- [16] NEWMAN M E. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, 2006, 74(3): 036104.
- [17] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [18] GUIMERA R, SALES-PARDO M, AMARAL L S A N. Module identification in bipartite and directed networks[J]. Physical Review E, 2007, 76(3): 036102.
- [19] 王 茜, 段双艳. 一种改进的基于二部图网络结构的推荐算法[J]. 计算机应用研究, 2013, 30(3): 771-774.
- [20] 朱斯丹. 个性化推荐算法与应用[J]. 电子世界, 2018(15): 21-23.
- [21] 王显飞, 陈 梅, 李小天. 基于约束的旅游推荐系统的研究与设计[J]. 计算机技术与发展, 2012, 22(2): 141-145.