

二分网络推荐算法与协同过滤算法的关系研究

周 波

(中国原子能科学研究院 北京 102413)

摘 要 文中介绍了协同过滤算法与二分网络推荐算法的原理,分析了协同过滤算法与二分网络推荐算法之间的内在关系,推导出协同过滤算法是二分网络推荐算法的一种特例,并证明了基于二分网络推荐算法的推荐结果优于协同过滤算法。将基于二分网络的推荐算法理论进一步系统化、统一化,以推动推荐算法的进一步发展。

关键词 协同过滤,二分网络,推荐算法

中图法分类号 TP391 **文献标识码** A

Research on Relationship Between Bipartite Network Recommendation Algorithm and Collaborative Filtering Algorithm

ZHOU Bo

(China Institute of Atomic Energy, Beijing 102413, China)

Abstract This paper introduced the basic principle of collaborative filtering algorithm and bipartite network recommendation algorithm, and proposed the general bipartite network recommendation algorithm. The internal relationship between the two algorithms was analyzed. The results show that collaborative filtering algorithm is a special case of the bipartite network recommendation algorithm, and bipartite network algorithm is proved to perform better than collaborative recommendation algorithm. This research systematizes and unifies the bipartite recommendation algorithm theory and promotes the further development of recommendation algorithm.

Keywords Collaborative filtering, Bipartite network, Recommendation algorithm

1 引言

在大数据时代,推荐系统尤为重要。在数据社会中,信息如同空气、水,是人类生存不可或缺的重要资源,然而,面对着过载的海量信息资源,如何寻找和提供有效的信息资源是当前面临的重要的问题。搜索引擎和推荐系统是解决数据爆炸的两种重要方式,搜索引擎的技术已经相对成熟,而推荐系统的相关技术还在不断完善之中。

协同过滤算法是商业推荐中应用最成熟和最广泛的推荐技术。施乐公司 Palo Alto 研究中心于 1992 年首先提出该算法并实现了具有协同过滤特性的邮件推荐系统 Taestry,但该系统的缺点是自动化程度较低。1994 年,GroupLens 的研究团队将协同过滤推荐算法应用在 Usenet 新闻推荐服务中,为协同过滤算法的商业应用打下了基础,2003 亚马逊将协同过滤算法应用到商品销售系统中,标志着协同过滤算法的商业化应用成熟。此后,协同过滤算法作为最主要的推荐算法广泛应用于各种推荐系统中^[1-2]。

协同过滤算法在实际应用中存在数据稀疏性、冷启动、可扩展性差、历史数据依赖性强和实时性差等问题。针对这些问题,众多学者提出了相应的解决办法。李聪等提出基于 n 序访问逻辑的冷启动问题解决办法^[3];李小浩着重研究了协同过滤算法的可扩展性问题^[4];孙小华对协同过滤算法中的数据稀疏性问题进行了研究^[5];徐键针对协同过滤算法中推荐的实时性进行了研究^[6]。这些研究为解决协同过滤算法在

实际应用中的问题提供了思路和方法。

基于二分网络的推荐算法由周涛于 2007 年基于复杂网络理论而提出,该算法提出后受到广泛关注。基于二分网络的推荐算法较传统的协同推荐算法有着更高的推荐精度,此后衍生出一系列基于二分网络的推荐算法以及各种优化算法^[7]。目前,基于二分网络的推荐算法有 6 种,分别是物质扩散算法、热传导算法、偏热传导算法、基于接受者能力的算法、混合算法和基于发送者能力和接受者能力的算法(TAA 算法)。周波等总结了 5 种基于二分网络的推荐算法,提出了基于发送者能力和接受者能力的二分网络算法(TAA)。该算法建立了基于二分网络推荐算法的统一框架,可将其他 5 种推荐算法纳入该框架之中,并通过实验证明该算法优于其他的 5 种推荐算法^[8]。但是该算法主要针对 Top-N 推荐,没有针对评分预测的模型。

大量的文献已经证明基于二分网络推荐算法的推荐效果要远好于协同过滤推荐算法,但现有的研究没有解释基于二分网络的推荐算法优于协同过滤推荐算法的原因,对于两者之间的关系还缺乏相关研究,因此本文将对这两个问题进行研究。

2 协同过滤算法的基本原理

从推荐任务的角度来划分,推荐算法可以分为 Top-N 推荐和评分预测两类。评分预测是预测用户对新物品的评分,通过预测评分和实际评分的差别作为判断算法优劣的标准。

Top-N 推荐是推荐用户最感兴趣的 N 个物品,以推荐的准确率作为判断算法优劣的标准。

从协同过滤推荐算法相似度计算的角度来划分,协同过滤算法又可以分为基于用户的推荐算法(user-based)和基于物品的推荐算法(item-based)。user-based 推荐算法是向用户推荐类似用户喜欢的物品,item-based 推荐算法是向用户推荐用户喜欢物品的相似物品,这两种推荐算法的基本原理类似,本文选择 user-based 推荐算法作为主要研究对象,item-based 推荐算法类似,不再赘述。

本文相关的字符含义如下:矩阵 $\mathbf{R}(m,n)$ 是用户对物品的评分矩阵, $r(i,k)$ 表示用户 i 对物品 k 的评分(评分范围为 1—5 分), $\bar{r}(i)$ 表示用户 i 的评分均值。矩阵 $\mathbf{W}(m,n)$ 是用户-物品连接矩阵, $w(i,k)=1$ 表示用户 i 购买过物品 k , $w(i,k)=0$ 表示用户 i 未购买过物品 k 。评分矩阵用于预测评分,连接矩阵用于 Top-N 推荐^[9]。

2.1 基于用户的协同过滤推荐算法的 Top-N 推荐模型

基于用户的协同过滤推荐算法的 Top-N 推荐模型中,用户 i 和用户 j 之间的相似度可表示为:

$$sim_1 = \frac{\sum_{k=1}^n w(i,k) * w(j,k)}{\sqrt{\sum_{k=1}^n w(i,k)^2} * \sqrt{\sum_{k=1}^n w(j,k)^2}} \tag{1}$$

选择与用户 i 相似度最高的 K 个用户构成集合 U' ,则用户 i 对物品 a 的偏好程度可以表示为:

$$p_1(i,a) = \sum_{j \in U'} sim_1(i,j) * w(j,a) \tag{2}$$

将用户对物品的偏好程度降序排列,去掉用户已经购买的物品,取前 Top-N 作为最终的推荐结果^[9]。

2.2 基于用户的协同过滤推荐算法的评分预测模型

基于用户的协同过滤推荐算法的评分预测模型中,用户 i 和用户 j 之间的相似度有两种常用的计算方式,一种是余弦相似度,另一种是皮尔逊相似度。

余弦相似度记为:

$$sim_2 = \frac{\sum_{k=1}^n r(i,k) * r(j,k)}{\sqrt{\sum_{k=1}^n r(i,k)^2} * \sqrt{\sum_{k=1}^n r(j,k)^2}} \tag{3}$$

皮尔逊相似度记为:

$$sim_3 = \frac{\sum_{k=1}^n (r(i,k) - \bar{r}(i)) * (r(j,k) - \bar{r}(j))}{\sqrt{\sum_{k=1}^n (r(i,k) - \bar{r}(i))^2} * \sqrt{\sum_{k=1}^n (r(j,k) - \bar{r}(j))^2}} \tag{4}$$

选择与用户 i 相似度最高的 K 个用户构成集合 U' ,则基于余弦相似度的用户 i 对物品 a 的预测评分可表示为:

$$p_2(i,a) = \frac{\sum_{j \in U'} sim_2(i,j) * r(j,a)}{\sum_{j \in U'} sim_2(i,j)} \tag{5}$$

基于皮尔逊相似度的用户 i 对物品 a 的预测评分可表示为:

$$p_3(i,a) = \bar{r}(i) + \frac{\sum_{j \in U'} sim_3(i,j) * (r(j,a) - \bar{r}(j))}{\sum_{j \in U'} sim_3(i,j)} \tag{6}$$

3 基于二分网络推荐算法的基本原理

基于二分网络推荐算法的基本思想是在用户-物品构成的二分网络中进行资源分配,依据用户对物品分配资源的多

少来进行推荐。基于二分网络的推荐算法有以下 3 类:1)单一的发送者能力或接受者能力的推荐算法,主要是物质扩散算法、热传导算法、偏热传导算法;2)有条件同时考虑发送者能力和接收者能力的算法,主要是混合算法和接收者能力算法;3)同时考虑发送者和接收者能力的算法,主要是 TAA 算法。周波等证明了同时考虑发送者能力和接收者能力的算法的推荐效果最好,有条件同时考虑发送者能力和接收者能力的算法的推荐效果次之,单一考虑发送者能力或接收者能力的推荐效果最差^[8]。

4 广义二分网络的推荐算法的基本原理

基于二分网络推荐算法根据网络节点之间的联系和节点的网络属性在节点之间进行初始资源分配,因此,可据此推出网络中两个节点之间的资源广义分配算法。设有两网络节点 i 和节点 k ,节点 i 和节点 k 之间的联系用函数 $G(i,k)$ 表示,节点 i 和节点 k 的网络属性分别是 $N(i)$ 和 $N(k)$,则广义二分网络推荐算法中,每一步资源扩散的分配方案是:

$$f(i,k) = f(i) * G(i,k) * N(i) * N(k) \tag{7}$$

进一步地,可以考虑有记忆的资源分配方案,设第 $n(n > 2)$ 步资源扩散是从节点 i 向节点 k 扩散,则扩散路径已经过的 $n-2$ 集合为 E ,因此在进行第 n 步资源扩散时可考虑已经过的节点集合 E 的影响,其资源分配方案是:

$$f(i,k) = f(i) * G(i,k,E) * N(i) * N(k) \tag{8}$$

仅考虑两个节点是否相连和两个节点的度,令 $G(i,k) = r(i,k)$, $N(i) = 1/d(i)$, $N(k) = 1/d(k)$,可有 $f(i,k) = f(i) * r(i,k)/(d(i) * d(k))$,即为 TAA 算法的资源分配方案。

该资源分配方法不仅可用于二分网络节点之间的分配,对于其他类型的网络节点之间的资源分配同样适用;不仅可用于有直接联系的两个节点之间的资源分配,也可用于两个存在间接联系的两个节点之间的资源分配。

该资源分配方案可对二分网络推荐算法进行推广,形成二分网络推荐算法簇。目前的基于二分网络推荐算法在节点之间的联系方面仅考虑了节点间的直接连接关系,即 $G(i,k) = r(i,k)=1$,其实节点之间的联系有很多种,如共同邻居、路径联系、边介数等^[11]。节点的网络属性除节点的度外还有节点的簇系数、点介数等^[12],可以基于这些节点之间的联系和节点间的网络属性来构造新的推荐算法,并对新的推荐算法进行研究。在具体的推荐系统和推荐业务中,节点具有社会属性,也可将这些社会属性融入二分网络推荐算法中。

4.1 基于用户的 TAA 算法 Top-N 推荐模型

基于二分网络模型的资源传递过程如图 2 所示。利用用户 i 向物品 a 分配的资源量预测用户 i 对物品 a 的偏好的过程可分为 3 步,其中 $f(x,y)$ 表示节点 x 向节点 y 扩散的资源量。

第 1 步 由用户 i 向物品 k 分配资源, $f(i)$ 表示用户 i 的初始资源量,物品 k 分配得到的资源量为:

$$f(i,k) = \frac{f(i) * w(i,k)}{d(i)^{\lambda_1} d(k)^{\lambda_2}} \tag{9}$$

第 2 步 由物品 k 向用户 j 分配资源,用户 j 分配得到的资源量为:

$$f(k,j) = \frac{\sum_{k=1}^n f(i,k) * w(j,k)}{d(k)^{\lambda_3} d(j)^{\lambda_4}} \tag{10}$$

第 3 步 由用户 j 向物品 a 分配资源,物品 a 分配得到的资源量为:

$$f(j, a) = \sum_{j=1}^m \frac{f(k, j)w(j, a)}{d(j)^{\lambda_5} d(a)^{\lambda_6}} \quad (11)$$

设用户 i 的初始资源量 $f(i) = 1$, 可得基于 TAA 算法时, 用户 i 向物品 a 的资源扩散量为 $M(i, a)$, 按 $M(i, a)$ 值的大小降序排列取前 Top-N 即为推荐结果。

$$\begin{aligned} M(i, a) &= f(j, a) \\ &= \sum_{j=1}^m \frac{f(k, j) * w(j, a)}{d(j)^{\lambda_5} * d(a)^{\lambda_6}} \\ &= \sum_{j=1}^m \left(\frac{w(j, a)}{d(j)^{\lambda_5} * d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{f(i, k) * w(j, k)}{d(k)^{\lambda_3} * d(j)^{\lambda_4}} \right) \\ &= \sum_{j=1}^m \left(\frac{w(j, a)}{d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{f(i, k) * w(j, k)}{d(k)^{\lambda_3} * d(j)^{\lambda_4 + \lambda_5}} \right) \\ &= \sum_{j=1}^m \left(\frac{w(j, a)}{d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{f(i) * w(j, k) * w(i, k)}{d(i)^{\lambda_1} * d(k)^{\lambda_2 + \lambda_3} * d(j)^{\lambda_4 + \lambda_5}} \right) \\ &= \sum_{j=1}^m \left(\frac{w(j, a)}{d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{w(j, k) * w(i, k)}{d(i)^{\lambda_1} * d(k)^{\lambda_2 + \lambda_3} * d(j)^{\lambda_4 + \lambda_5}} \right) \end{aligned} \quad (12)$$

TAA 算法的三次资源扩散过程如图 1 所示。

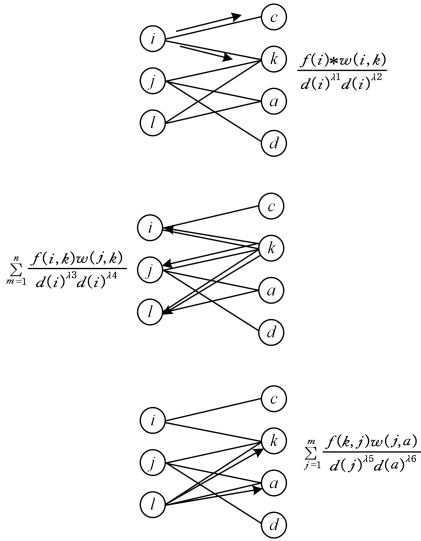


图 1 TAA 算法的 3 次资源扩散过程

4.2 基于用户的 TAA 算法的评分预测模型

基于二分网络的评分预测模型就是在二分网络中加入边的权重来进行评分预测, TAA 算法可基于用户或物品的节点强度进行资源分配。在加权网络中, 节点的度可以很自然地推广为节点的强度。有 3 种方式来计算节点的强度 $s(i)$:

第 1 种 用户 i 节点的强度定义为^[13]:

$$s_1(i) = \sum_{k=1}^n r(i, k) \quad (13)$$

第 2 种 为降低高权重边对推荐结果带来的马太效应, 可将用户 i 节点的强度定义为:

$$s_2(i) = \sum_{k=1}^n r(i, k)^2 \quad (14)$$

第 3 种 为降低用户打分偏好对推荐结果的影响, 可减去用户 i 的平均评分作为用户 i 的新评分, 用户新的评分为:

$$r'(i, j) = r(i, k) - r(i) \quad (15)$$

结合第 2 种考虑权重影响的节点强度计算方法, 第 3 种用户强度的计算方法为:

$$s_3(i) = \sum_{k=1}^n r'(i, j)^2 = \sum_{k=1}^n (r(i, k) - r(i))^2 \quad (16)$$

同 Top-N 推荐类似, 评分预测模型的资源扩散路径为:

第 1 步 由用户 i 向物品 k 分配资源, $f(i)$ 表示用户 i 的初始资源量, 物品 k 分配得到的资源量为:

$$f(i, k) = \frac{f(i) * r(i, k)}{s(i)^{\lambda_1} s(k)^{\lambda_2}} \quad (17)$$

第 2 步 由物品 k 向用户 j 分配资源, 用户 j 分配得到的资源量为:

$$\begin{aligned} f(k, j) &= \sum_{k=1}^n \frac{f(i, k) * r(j, k)}{s(k)^{\lambda_3} s(j)^{\lambda_4}} \\ &= \sum_{k=1}^n \frac{f(i) * r(i, k) * r(j, k)}{s(i)^{\lambda_1} s(k)^{\lambda_2 + \lambda_3} s(j)^{\lambda_4}} \end{aligned} \quad (18)$$

第 3 步 由用户 j 向物品 a 分配资源, 物品 a 分配得到的资源量为:

$$f(j, a) = \sum_{j=1}^m \frac{f(k, j) r(j, a)}{s(j)^{\lambda_5} s(a)^{\lambda_6}} \quad (19)$$

令 $f(i) = 1$, 将物品 a 分配得到的资源量进行归一化可得用户 i 对物品 a 的评分, 采用第 2 种节点强度, 得到用户 i 对物品 a 的评分为:

$$\begin{aligned} p_4(i, a) &= \frac{1}{\sum_{j=1}^m f(k, j)} \sum_{j=1}^m \frac{f(k, j) r(j, a)}{s_2(j)^{\lambda_5} s_2(a)^{\lambda_6}} \\ &= \frac{1}{\sum_{j=1}^m f(k, j)} \sum_{j=1}^m \left(\frac{r(j, a)}{s_2(a)^{\lambda_6}} * \sum_{k=1}^n \frac{r(j, k) * r(i, k)}{s_2(i)^{\lambda_1} * s_2(k)^{\lambda_2 + \lambda_3} * s_2(j)^{\lambda_4 + \lambda_5}} \right) \end{aligned} \quad (20)$$

在利用第 3 种用户的强度时, 应采用用户更新的评分 $r'(j, a)$, 得到用户 i 对物品 a 的评分:

$$\begin{aligned} p_5(i, a) &= r(i) + \frac{1}{\sum_{j=1}^m f(k, j)} \sum_{j=1}^m \left(\frac{r'(j, a)}{s_3(a)^{\lambda_6}} * \sum_{k=1, s_3(i)^{\lambda_1}}^n \frac{r'(j, k) * r'(i, k)}{s_3(k)^{\lambda_2 + \lambda_3} * s_3(j)^{\lambda_4 + \lambda_5}} \right) \end{aligned} \quad (21)$$

5 协同过滤算法的相关证明

5.1 协同过滤算法 Top-N 推荐是二分网络推荐算法一种特例的证明

基于二分网络的 Top-N 推荐算法的预测结果为:

$$M(i, a) = \sum_{j=1}^m \left(\frac{w(j, a)}{d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{w(j, k) * w(i, k)}{d(i)^{\lambda_1} * d(k)^{\lambda_2 + \lambda_3} * d(j)^{\lambda_4 + \lambda_5}} \right) \quad (22)$$

令 $\lambda_2 = \lambda_3 = \lambda_5 = \lambda_6 = 0, \lambda_1 = \lambda_4 = 0.5$, 可得:

$$\begin{aligned} M(i, a) &= \sum_{j=1}^m \left(\frac{w(j, a)}{d(a)^{\lambda_6}} * \sum_{k=1}^n \frac{w(j, k) * w(i, k)}{d(i)^{\lambda_1} * d(k)^{\lambda_2 + \lambda_3} * d(j)^{\lambda_4 + \lambda_5}} \right) \\ &= \sum_{j=1}^m \left(w(j, a) * \sum_{k=1}^n \frac{w(j, k) * w(i, k)}{\sqrt{d(i)} \sqrt{d(j)}} \right) \\ &= \sum_{j=1}^m \left(w(j, a) * \sum_{k=1}^n \frac{w(j, k) * w(i, k)}{\sqrt{\sum_{k=1}^n w(i, k)^2} * \sqrt{\sum_{k=1}^n w(j, k)^2}} \right) \\ &= \sum_{j=1}^m (w(j, a) * sim_1(i, j)) \end{aligned} \quad (23)$$

式(23)与式(2)的结果相同, 唯一不同的是基于用户的协同过滤推荐算法的 Top-N 推荐模型, 在第 3 步资源扩散之前选择与用户 i 相似度最高的 K 个用户构成集合 U' , 通过这 K 个用户来进行资源扩散, 而基于 user-based 的二分网络推荐算法的 Top-N 推荐模型是选择所有 m 个用户来进行资源扩散。

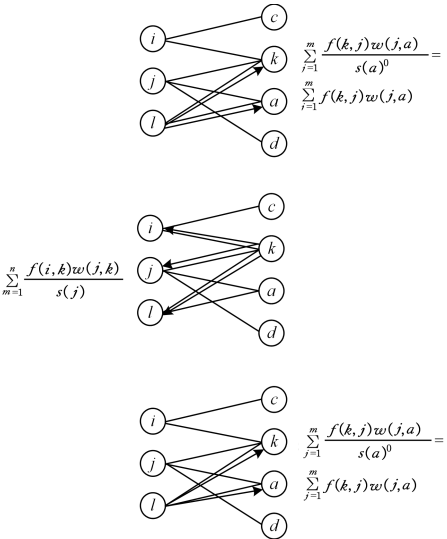


图 2 基于二分网络模型的协同过滤算法的 3 次资源扩散过程

5.2 协同过滤的评分预测推荐算法是二分网络推荐算法一种特例的证明

令基于用户的 TAA 算法的评分预测模型中, $\lambda_2=\lambda_3=\lambda_5=\lambda_6=0, \lambda_1=\lambda_4=0.5$ 。

由第 2 种节点强度计算公式得到的评分预测模型为:

$$p_4(i,a)=\frac{1}{\sum_{j=1}^m f(k,j)}\sum_{j=1}^m (r(j,a)*\sum_{k=1}^n \frac{r(j,k)*r(i,k)}{s_2(i)^{0.5}*s_2(j)^{0.5}})$$
$$=\frac{1}{\sum_{j=1}^m f(k,j)}\sum_{j=1}^m (r(j,a)*\sum_{k=1}^n \frac{r(j,k)*r(i,k)}{\sqrt{\sum_{k=1}^n r(i,k)^2}\sqrt{\sum_{k=1}^n r(j,k)^2}})$$
$$=\frac{1}{\sum_{j=1}^m sim_2(i,j)}\sum_{j=1}^m (r(j,a)*sim_2(i,j)) \tag{24}$$

$$p_2(i,a)=\frac{\sum_{j\in U'} sim_2(i,j)*r(j,a)}{\sum_{j\in U'} sim_2(i,j)} \tag{25}$$

由第 3 种节点强度计算公式得到的评分预测模型为:

$$p_5(i,a)=r(i)+\frac{1}{\sum_{j=1}^m f(k,j)}\sum_{j=1}^m (r'(j,a)*\sum_{k=1}^n \frac{r'(j,k)*r'(i,k)}{s_3(i)^{0.5}*s_3(j)^{0.5}})$$
$$=r(i)+\frac{1}{\sum_{j=1}^m f(k,j)}\sum_{j=1}^m (r(j,a)*\sum_{k=1}^n \frac{(r(i,k)-r(i))*(r(j,k)-r(i))}{\sqrt{\sum_{k=1}^n ((r(i,k)-r(i))^2)*\sum_{k=1}^n ((r(j,k)-r(i))^2}})$$
$$=r(i)+\frac{1}{\sum_{j=1}^m sim_2(i,j)}\sum_{j=1}^m (r'(j,a)*sim_3(i,j)) \tag{26}$$

$$p_3(i,a)=r(i)+\frac{\sum_{j\in U'} sim_3(i,j)*(r(j,a)-r(j))}{\sum_{j\in U'} sim_3(i,j)} \tag{27}$$

比较 $p_2(i,a)$ 与 $p_4(i,a)$ 、 $p_3(i,a)$ 与 $p_5(i,a)$, 唯一不同的是基于用户的协同过滤推荐算法的评分预测推荐模型, 在第 3 步资源扩散之前选择与用户 i 相似度最高的 K 个用户构成集合 U' , 通过这 K 个用户来进行资源扩散; 而基于用户的二分网络推荐算法的 Top-N 推荐模型是选择所有 m 个用户来进行资源扩散。可发现基于余弦相似度的基于用户的协同过滤推荐算法的评分预测模型是基于第 2 种节点强度进行资源

分配的二分网络推荐算法的评分预测模型的特例。可发现, 基于皮尔逊相似度的基于用户的协同过滤推荐算法的评分预测模型是基于第 3 种节点强度进行资源分配的二分网络推荐算法的评分预测模型的特例。

因此, 不论是协同过滤推荐算法的评分预测模型或 Top-N 推荐模型都是基于二分网络模型的特例。

6 基于二分网络的 TAA 算恒优于协同过滤算法的证明

由于不同的参数取值对应唯一确定的推荐结果, 则可用函数 $H(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ 表示参数 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$ 和推荐结果之间的函数关系。 $H(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ 表示基于二分网络算法的推荐结果, $H(0.5, 0, 0, 0.5, 0, 0)$ 表示基于协同过滤推荐算法的推荐结果。

极值定理表述如下: 如果函数 $f(x)$ 在闭区间 $[a, b]$ 上是连续函数, 则它一定存在至少一个最大值, 即任取 $x' \in [a, b]$, 都存在 c 属于 $[a, b]$, 有 $f(x') \leq f(c)$ 成立。

由于 $H(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ 在 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6 \in [-1, 1]$ 上是连续函数, 即任取 $\lambda_1' \in [-1, 1]$ 都存在 $\lambda_1'' \in [-1, 1]$, 有 $H(\lambda_1', \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) \leq H(\lambda_1'', \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ 成立。

令任取 $\lambda_1'=0.5$, 都存在 $\lambda_1'' \in [-1, 1]$ 使得下式成立: $H(0.5, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) \leq H(\lambda_1'', \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ (28) 以此类推, 同理可得: 必然存在 $\lambda_1'', \lambda_2'', \lambda_3'', \lambda_4'', \lambda_5'', \lambda_6'' \in [-1, 1]$ 使得下式成立

$$H(0.5, 0, 0, 0.5, 0, 0) \leq H(0.5, 0, 0, 0.5, 0, \lambda_6'') \leq H(0.5, 0, 0, 0.5, \lambda_5'', \lambda_6'') \leq H(0.5, 0, 0, \lambda_4'', \lambda_5'', \lambda_6'') \leq H(0.5, 0, \lambda_3'', \lambda_4'', \lambda_5'', \lambda_6'') \leq H(0.5, \lambda_2'', \lambda_3'', \lambda_4'', \lambda_5'', \lambda_6'') \leq H(\lambda_1'', \lambda_2'', \lambda_3'', \lambda_4'', \lambda_5'', \lambda_6'') \tag{29}$$

即证, 基于 TAA 算法的二分网络推荐算法恒优于协同过滤推荐算法。

结束语 协同过滤算法和基于二分网络推荐算法分别于 1992 年和 2007 年提出, 并已在实际中得到了广泛应用, 这两种算法看似无关, 然而经过详细分析发现它们之间存在密切联系, 具体结论如下:

协同过滤算法是基于二分网络推荐算法的一种特例, 即协同过滤算法只是 TAA 算法取特定参数下的一种算法, 在二分网络中进行特殊的资源分配。本文从理论证明了基于二分网络的推荐算法(TAA 算法)要优于协同过滤算法。

基于二分网络的推荐算法在应用上会面临着数据稀疏性、冷启动、可扩展性和实时性等问题。而本文证明协同过滤算法是二分网络推荐算法的特例, 那么, 可直接使用协同过滤算法在解决这些问题时的成果来解决二分网络的这些问题。

本文提出广义的二分网络推荐算法中, 节点之间的资源分配不再仅仅基于节点的度和节点的初始资源, 还可以基于节点的簇系数、点介数等和用户的信用、消费能力和社会地位等社会属性, 这将对二分网络推荐算法应用的广度和深度进行扩充, 同时提供更好的推荐效果。

参 考 文 献

[1] 裴中佑. 基于随机游走的推荐技术研究及应用[D]. 成都: 西南交通大学, 2014.

练数据的增大,学生的平均得分出现波动,但是始终稳定在 5 分左右,表明推荐的试题相对较难,对于学生的做题心理产生较大负担。当推荐试题难度较大,而学生在进行练习时,保持较低的准确率时,会严重打击学生学习的自信心,不利于学习的进步。对于 NMF 推荐的试题,学生得分一直处于较高水平,虽然可以提高学生信心,但是从教育角度看,推荐的试题较为容易,并不能达到学习知识点的目的。而使用本文提出的个性化试题推荐算法进行试题推荐时,学生的平均得分保持在 7 分左右,既能对学生起到激励作用,又能使学生在训练过程中学习和巩固所学知识。

表 2 平均得分 AR

比例	0.6	0.7	0.8
本文方法	6.9	6.7	7.1
ItemCF	5.7	5.5	5.2
KNN	5.3	4.6	4.9
NMF	8.9	9.2	8.8

综上所述,本文提出了一种新的基于深度自编码器和二次协同过滤的个性化试题推荐方法,其在进行试题推荐时有较好的准确性和稳定性,且符合教育理念,对提升学生能力具较好的表现。

结束语 本文在大数据背景下,对智慧教学中的适应性测试的方法进行了研究,不同于传统的 TOP-N 试题推荐,本文进行了基于知识点的二次协同过滤试题推荐,并将项目反应理论与深度自编码评分预测融入推荐算法中,进行了基于深度自编码器和二次协同过滤的个性化试题推荐方法的研究。与其他传统的单一的推荐方法相比,本文所提方法在推荐试题的难度控制上更为合理,且在各难度范围的试题推荐上取得了较好的推荐结果。

参 考 文 献

[1] 杨现民,唐斯斯,李冀红. 发展教育大数据:内涵、价值和挑战[J]. 现代远程教育研究,2016(1):50-61.

[2] 杜婧敏,方海光,李维杨,等. 教育大数据研究综述[J]. 中国教育信息化,2016(19):1-3.

[3] OXMAN S, WONG W, INNOVATIONS D V X. WhitePaper: Adaptive Learning Systems[EB/OL]. [2015-01-02]. <http://www.integratededsolution.com/wp-content/uploads/2014/10/DVx-Adaptive-Learning-White-Paper-February-20131.pdf>.

[4] 徐立芳,莫宏伟,李金,等. 智能教育与教育智能化技术研究[J].

教育现代化,2018,4(3):115-117,119.

[5] 刘邦奇. 智慧教育:新时代的教育变革与转型[N]. 中国教育报,2018-01-27(3).

[6] LORD F. Some test theory for tailored testing[C]// Computer assisted instruction, testing, and guidance. New York: Harper&Row,1970:139-183.

[7] BATMAZ Z, YUREKLI A, BILGE A, et al. A review on deep learning for recommender systems: challenges and remedies[J]. Artificial Intelligence Review, 2018:1-37.

[8] 蒋一君,邱飞岳,刘迎春,等. 基于错题库的个性化练习生成模型研究[J]. 中国教育信息化,2011(8):73-74.

[9] 郭辰. 高中英语诊断性练习系统的需求分析、模块设计和知识体系构建[D]. 合肥:中国科学技术大学,2013:22.

[10] 王文泉. 错题管理系统中个性化推荐练习算法的设计与实现[J]. 中国教育信息化,2016(11):67-70.

[11] 申瑞民,汤轶阳,韩鹏. 基于概念图的教学内容智能调整模型及算法实现[J]. 上海交通大学学报,2002(5):102-105.

[12] 牛文娟. 基于协同过滤的学习资源个性化推荐研究[D]. 北京:北京理工大学,2014.

[13] 杨超. 基于粒子群优化算法的学习资源推荐方法[J]. 计算机应用,2014,34(5):1350-1353.

[14] ADACHE I, FOURNIER S, CHIFU A G. Harnessing Ratings and Aspect-Sentiment to Estimate Contradiction Intensity in Temporal-Related Reviews[J]. Procedia Computer Science, 2017,112(C):1711-1720.

[15] DASCALU M I, BODEA C N, MOLDOVEANU A, et al. A recommender agent based on learning styles for better virtual collaborative learning experiences[J]. Computers in Human Behavior, 2015,45(45):243-253.

[16] 黄立威,江碧涛,吕守业,等. 基于深度学习的推荐系统研究综述[J]. 计算机学报,2018,41(7):1619-1647.

[17] 孙琳琳. 认知诊断评估对实现有效诊断教学的促进作用[J]. 世界最新医学信息文摘,2018,18(20):243.

[18] RAMI S, BENNANI S, IDRISSE M K. Towards a method for analyzing learning style using item response theory[C]// International Conference on Information Technology Based Higher Education and Training. IEEE,2017:1-5.

[19] SIGKDD:KDD Cup 2010:Student performance evaluation [OL]. <http://www.kdd.org/kdd-cup/view/kdd-cup-2010-student-performance-evaluation/Data>.

[20] KIRK J. jfkirk/tensorrec[OL]. <https://github.com/jfkirk/tensorrec>.

(上接第 166 页)

[2] MASSA P, AVESANI P. Trust-Aware Collaborative Filtering for Recommender Systems[M]// On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE. Springer Berlin Heidelberg,2004:492-508.

[3] 李聪,梁昌勇. 基于 n 序访问解析逻辑的协同过滤冷启动消除方法[J]. 系统工程理论与实践,2012(7):1537-1545.

[4] 李小浩. 协同过滤推荐算法稀疏性与可扩展性问题研究[D]. 重庆:重庆大学,2015.

[5] 孙小华. 协同过滤系统的稀疏性与冷启动问题研究[D]. 杭州:浙江大学,2005.

[6] 徐键. 协同过滤中数据稀疏问题与推荐实时性的研究[D]. 兰州:兰州大学,2016.

[7] TAO Z, JIE R, MATÜS M, et al. Bipartite network projection

and personal recommendation[J]. Physical Review E, 2007, 76(4):70-80.

[8] 周波,杨朝峰. 发送者和接受者能力的二分网络推荐算法研究[J]. 情报工程,2016,2(2):71-80.

[9] 何平凡. 基于排序学习的 Top-N 推荐算法研究[D]. 北京:北京理工大学,2016.

[10] 赵向宇. Top-N 协同过滤推荐技术研究[D]. 北京:北京理工大学,2014.

[11] 陈嘉颖,于炯,杨兴耀,等. 基于复杂网络节点重要性的链路预测算法[J]. 计算机应用,2016(12):3251-3255,3268.

[12] 荣莉莉,郭天柱,王建伟. 复杂网络节点中心性[J]. 上海理工大学学报,2008,30(3):227-230.

[13] 姚尊强,尚可,许可,等. 加权网络的常用统计量[J]. 上海理工大学学报,2012,34(1):18-26.