

一种结合共同邻居和用户评分信息的相似度算法

贺银慧 陈端兵 陈勇 傅彦

(电子科技大学计算机科学与工程学院 成都 610054)

摘要 随着互联网的发展,推荐系统逐步得到广泛应用,协同过滤(CF)是其中运用得最早、最成功的技术之一。CF 首先根据用户间的相似度,找出每个用户的近邻;然后根据目标用户近邻的评分预测目标用户的评分;最后把预测评分较高的项目推荐给目标用户。因此相似度计算方法直接关系到预测结果的准确性,对推荐起着至关重要的作用。目前,学者们已从不同的角度提出了各种各样的相似度计算方法,其中共同邻居算法(common-neighbors)是一种简单有效的方法。但此法仅考虑了两用户间的共同邻居数,忽略了用户的具体评分信息。针对这个问题对共同邻居算法进行了改进,同时考虑了共同邻居数和用户的评分信息。实验结果表明,改进的共同邻居算法在一定程度上可提高评分预测的准确性。

关键词 协同过滤,共同邻居,相似度算法,评分信息

Similarity Algorithm Based on User's Common Neighbors and Grade Information

HE Yin-hui CHEN Duan-bing CHEN Yong FU Yan

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract With the rapid growth of Internet, recommender systems have been used in many fields, and collaborative filtering(CF) is one of the earliest and the most successful ones. CF method usually identifies the neighborhood of each user based on similarity between two users; then predicts items' rating by integrating ratings of target user's neighbors, and lastly those items with higher predicted score are recommended to target user. So similarity plays an important role and affects the accuracy of the prediction. Up to now, various similarity measures have been proposed by researchers from different aspect. And common-neighbor algorithm is a simple and efficient method. However, common-neighbor algorithm just considers the number of common objects scored by two users, doesn't consider the user's grade information. In this paper, an improved algorithm based on common-neighbor and user's grade information was proposed. Experimental results indicate that improved common-neighbor algorithm can obtain rather good predicting results.

Keywords Collaborative filtering, Common-neighbors, Similarity algorithm, Grade information

1 引言

随着 Internet 的迅猛发展,接入 Internet 的服务器数量和 World-Wide-Web 上的网页数目都呈现指数增长。互联网技术的迅猛发展使得大量的信息涌现在我们面前,例如 Netflix 上有数万部电影、Amazon 上有数百万本书、Del.icio.us 上有超过 10 亿的网页收藏。要在这些海量信息中找到自己感兴趣的内容,确实不是一件易事。传统的搜索算法只能呈现给所有用户相同的排序结果,无法针对不同用户的兴趣爱好提供相应的服务。为了有效地利用这些信息,解决信息爆炸问题,个性化推荐技术应运而生。推荐问题从根本上说就是代替用户评估他从未看过的产品^[1-3],这些产品包括书、电影、CD、网页,甚至可以是饭店、音乐、绘画等等。

协同过滤是目前个性化推荐系统中应用得最为广泛的技术之一。协同过滤推荐^[4,5]的基本思想是,根据目标用户与

其他用户的相似性,搜索选取与目标用户最相似的几个(例如 k 个)近邻作为他的最近邻居;然后根据这些最近邻居的评分数据对目标用户进行推荐。因此在协同过滤推荐系统中,最为关键的是用户之间以及产品之间的相似性如何度量,不同的相似性度量方法产生的推荐结果也有所差异。所以,如何定义用户之间以及产品之间的相似性,成为关键问题之一。本文就协同过滤推荐系统中用户相似度的度量问题进行分析 and 讨论。

目前已提出很多相似性度量算法,例如,比较常用的 pearson 相关系数法^[6]和 cosin 相似度法^[7],还有基于结构的 common-neighbor 指标法、jaccard 指标法、salton 指标法、sorenson 指标法^[8]以及 LHN-I 指标法^[9]等。通过实验比较,发现 common-neighbor(简记为 CN)相似性度量指标是上述几种相似性指标中预测效果最佳的一种。与此同时,上述 5 种基于结构的相似性度量指标都没有考虑用户的具体评分信息

到稿日期:2009-10-19 返修日期:2009-12-29 本文受国家高技术研究发展计划(编号:2007AA01Z440),国家自然科学基金(编号:60973069,90924011),四川省应用技术与开发项目支撑计划(编号:2008GZ0009),中国博士后科学基金资助项目(编号:20080431273)资助。

贺银慧 硕士生,主要研究方向为数据挖掘,E-mail:niqui_1986@163.com;陈端兵 博士,副教授,主要研究方向为数据挖掘、社会计算、NP 难问题高效求解等;陈勇 硕士生,主要研究方向为数据挖掘;傅彦 教授,博士生导师,主要研究方向为数据挖掘、信息安全、模式识别等。

如 CN 指标只计算用户共同评分的项目数。评分值代表用户的喜爱程度,所以丢失评分信息,很有可能把原本不太相似的用户误判成很相似的用户,进而影响预测评分的准确性。而另外 4 种基于结构的相似性度量指标都是在 CN 指标的基础上经过改造而得到的,因此本文仅讨论如何在 CN 指标的基础上加入用户的评分信息。通过试验比较分析发现,以某种函数形式加入用户评分信息,能有效地改善 CN 方法的推荐效果。

2 方法

2.1 基本理论与技术

2.1.1 网络描述

推荐系统常常借助一个二部图网络进行研究。二部图网络可用数据集 $S=(U, I, R)$ 表示,其中 $U=\{User_1, User_2, \dots, User_m\}$ 表示二部图网络的用户集, $|U|=m$; $I=\{Item_1, Item_2, \dots, Item_n\}$ 表示二部图网络的项目集, $|I|=n$; $m \times n$ 阶矩阵 R 是用户对各个项目评分的邻接矩阵。其中 R 的第 i 行第 j 列元素 r_{ij} 表示 U 中第 i 个用户对 I 中第 j 个项目的评分值,反映了 i 对 j 的喜欢程度。 r_{ij} 越大表示 i 越喜欢 j , r_{ij} 的值可以是任意实数,不过通常取某个区间上的整数。对于 u 和 v 两个用户,即 $u, v \in U$,用 S_{uv} 表示 u 和 v 的相似度。 $T_u=\{v_i \in U | v_1, v_2, \dots, v_k\}$ 表示用户 u 的 k 个最近邻居集,即与用户 u 最相似的 k 个邻居。 $R_u=\{r_{u1}, r_{u2}, \dots, r_{uk}, \dots\}$ 表示用户 u 评过分的项目集。

2.1.2 评分预测方法

通过相似度计算,建立起用户间的相似关系网。若要预测目标用户 u 对项目 $i \in I$ 的评分(该项目是 u 未评过分的项),则要先找出 u 的最近邻居集 T_u ,然后根据 T_u 中已对 i 评过分的用户的评分值,预估 u 对 i 的评分,即

$$r_{ui}' = r_u + \frac{\sum_{v \in T_u, r_{vi} > 0} S_{uv}(r_{vi} - \bar{r}_v)}{\sum_{v \in T_u} S_{uv}} \quad (1)$$

式中, r_{ui}' 为用户 u 对项目 i 的预估评分, \bar{r}_v 表示用户 v 评分的平均值。

2.1.3 评价标准

不同的相似性度量指标会得到不同的预测评分结果。平均绝对误差^[10](MAE)和均方根误差^[10](RMSE)是两个衡量相似性度量方法预测准确性的常用指标,它们的值越小说明预测准确性越好。通常的做法是将原始评分数据划分成训练集(trainset)和测试集(testset)两部分,利用训练集得到预测模型,然后利用测试集对预测结果进行评价,即可得到预测结果和实际评分的 MAE 和 RMSE 值,其定义为

1) 平均绝对误差^[10](MAE): 评分测试集中用户预测评分 r_{ui}' 与真实评分 r_{ui} 的绝对误差的平均值。

$$MAE = \frac{1}{|p|} \sum_{(u,i) \in p} |r_{ui} - r_{ui}'| \quad (2)$$

2) 均方根误差^[10](RMSE): 测试集中用户对项目的真实评分值 r_{ui} 与预测评分值 r_{ui}' 的平均差方根。

$$RMSE = \sqrt{\frac{1}{|p|} \sum_{(u,i) \in p} (r_{ui} - r_{ui}')^2} \quad (3)$$

式(2)和式(3)中的 p 表示测试集。

2.2 传统的相似性度量方法

cosine 相似性和 pearson 相关相似性是两种较常用的相似性度量方法。本小节的主要内容是将这两种相似性指标与另外 5 种基于结构的相似性指标在 movielens 数据集上进行试验对比分析。

1) cosine 相似性^[7]:

$$S_{xy} = \frac{\vec{R}_x \cdot \vec{R}_y}{\|\vec{R}_x\| \times \|\vec{R}_y\|} \quad (4)$$

式中, \vec{R}_x 和 \vec{R}_y 分别表示用户 x 和 y 的评分向量。

2) pearson 相似性^[6]:

$$S_{xy} = \frac{\sum_{\alpha} (r_{x\alpha} - \bar{r}_x)(r_{y\alpha} - \bar{r}_y)}{\sqrt{\sum_{\alpha} (r_{x\alpha} - \bar{r}_x)^2} \sqrt{\sum_{\alpha} (r_{y\alpha} - \bar{r}_y)^2}} \quad (5)$$

式中, \bar{r}_x 表示用户 x 的评分平均值, α 表示 x 和 y 共同评分的项目集。

3) 共同邻居(common-neighbors)指标: 把两个用户 x 和 y 的共同邻居数量(在二部图网络中即为共同评过分的项数)定义为他们之间的相似值,即

$$s_{xy} = |R_x \cap R_y| \quad (6)$$

4) Salton 指标: 其定义为

$$S_{xy} = \frac{|R_x \cap R_y|}{\sqrt{k(x) \times k(y)}} \quad (7)$$

式中, $k(x) = \|R_x\|$, 表示节点 x 的度。

5) Jaccard 指标: 其定义为

$$S_{xy} = \frac{|R_x \cap R_y|}{|R_x \cup R_y|} \quad (8)$$

6) Sorensen 指标^[8]: 这个指标常用于研究生态学数据, 定义为

$$S_{xy} = \frac{2 \times |R_x \cap R_y|}{k(x) + k(y)} \quad (9)$$

7) LHN-I 指标^[9]: 它是 Leicht, Holme 和 Newman 提出的一种相似性度量方法, 定义为

$$S_{xy} = \frac{|R_x \cap R_y|}{k(x) \times k(y)} \quad (10)$$

表 1 给出了 7 种相似性度量方法在 movielens 上的 RMSE 值和 MAE 值的比较, 其中黑色字体标注的是效果最好的方法。表 1 中的每组数据都是在 80% 的训练集上(数据的划分方法在第 3 节有详述)分别运算 5 次, 取其平均值。

表 1 几种相似性度量指标比较

	RMSE	MAE
Cosine	0.836667	0.7846
Pearson	0.8581	0.828568
Salton	0.836767	0.787
Jaccard	0.837867	0.789267
CN	0.831833	0.779333
LHN-I	0.846163	0.802377
Sorensen	0.837967	0.789367

由表 1 可看出, CN 指标不论其 RMSE 值还是 MAE 值都比其它几种度量指标小, 表明 CN 指标的整体平均评分预测误差最小, 推荐效果较好; 而 pearson 相似性的评分预测误差在几种指标中最大, cosine 相似性的评分预测误差仅次于 CN 指标。

2.3 改进的共同邻居相似性度量方法

2.3.1 问题分析

由表 1 可看出, CN 方法是所列几种指标中推荐效果最

佳的方法。但是 CN 指标在计算两用户的相似度时,只统计两用户共同评过分的的项目数,并未考虑用户对这些项目的具体评分值。而不同的评分值代表用户不同的喜爱程度,CN 指标法有时会遇到一些尴尬的场景。

例如,一个数据集中总共有 10 部电影,现有用户 A, B, C 和 D,他们对自己看过的电影进行了评分,评分值为 1~5 区间上的整数,1 分表示不喜欢,5 分表示非常喜好,0 分表示他还没看过这部电影,评分的邻接矩阵 R 如表 2 所列。

表 2 用户评分矩阵

电影 \ 用户	1	2	3	4	5	6	7	8	9	10
A	1	1	0	5	5	5	1	5	1	1
B	5	5	1	1	1	1	5	0	5	5
C	1	1	5	4	5	4	0	5	2	1
D	1	0	1	5	0	0	1	5	0	0

由 CN 方法计算相似度,分别得 $S_{AB} = S_{AC} = 8, S_{AD} = 4$ 。但观察实际的用户评分信息可知:

1) 虽然 A 与 B, A 与 C 都对 8 部电影评过,但他们的评分差异却不同:A 和 B 的评分差值很大(只要是 A 评 1 分的电影, B 都评 5 分;而 A 评 5 分的电影, B 反而都评 1 分),说明他们的兴趣差异很大(只要是 A 喜欢的电影, B 都不喜欢;而 A 不喜欢的电影, B 都喜欢),也可以说他们兴趣不相投;而 A 和 C 的评分差值却很小(对同一部电影,他们的评分值大多相同,或者只差一分),说明他们的兴趣很相投,因此应该是 $S_{AB} < S_{AC}$ 。

2) 虽然 A 和 D 共同评过分的电影只有 4 部,但他们的评分值都相同,说明他们目前是兴趣相同的用户,他们间的相似值不应该比 S_{AB} 小很多,有可能是 $S_{AD} \geq S_{AB}$ 。

通过上例分析,清楚地显示了评分信息对用户相似度计算的影响,因此保留这部分信息是必要的。这里我们考虑如何在 CN 算法基础上加上用户的评分信息,以充分利用网络提供的数据信息来提高预测的准确性。

2.3.2 改进的相似度计算方法

$\alpha = R_x \cap R_y$ 代表用户 x 和 y 共同评分的项目集。因为不同评分值表示他们的喜好程度不一样,又因为在比较两用户的喜好差异时,没有同时评过分的项在此不具有可比性,因此可以用他们对 α 中项目评分差的绝对值之和来表示他们之间的品味差异。将这个差异值定义为用户间的距离,记为 D_{xy} ,则

$$D_{xy} = \sum_{i \in \alpha} |r_{xi} - r_{yi}| \quad (11)$$

式(11)表明, D_{xy} 越大,则表示 x 和 y 的喜好差异越大。所以当 $|\alpha|$ 一定时, D_{xy} 越大,他们间的相似度就越小,即 S_{xy} 值越小。而当 $|\alpha|$ 越大时, x 和 y 相似的可能性就越大。因此 D_{xy} 与 S_{xy} 是负相关的, $|\alpha|$ 与 S_{xy} 是正相关的,考虑如下 3 种情况:

1) 当 $|R_x \cap R_y| = 0$ 时,因为用户 x 和 y 没有共同的评分项目,认为 x 和 y 没有相似性,即 $S_{xy} = 0$;

2) 当 $D_{xy} = 0, |R_x \cap R_y| \neq 0$ 时,表示这两个用户目前的喜好相同,它们间的相似度由共同邻居数决定。

3) 当 $D_{xy} \neq 0, |R_x \cap R_y| \neq 0$ 时, S_{xy} 的值就由共同邻居数和他们的距离共同决定。

如果直接用 D_{xy} 的倒数乘以 $|\alpha|$,其虽然也满足 D_{xy} 和 S_{xy} 负相关的性质,但 $D_{xy} = 0$ 时,要单独处理。而且试验发现,直接加入 D_{xy} 因素,会对 S_{xy} 值产生过大的影响,使最终的预测误差变大。所以要减小 D_{xy} 对 S_{xy} 的影响,以适当的形式加入 D_{xy} 因素。通过多次试验总结比较发现, D_{xy} 以函数 $f(x) = 1 + e^{(-x/2)}$ 的形式加入,可以适当减小 D_{xy} 对 S_{xy} 值的影响,减小预测误差。同时还满足上述的 3 个要求,因此我们的相似度方法定义为

$$S_{xy} = (1 + e^{-D_{xy}/2}) \times ||R_x \cap R_y|| \quad (12)$$

对表 2 所列的评分表, A, B 的距离 $D_{AB} = 32$, A, C 的距离 $D_{AC} = 3$, A, D 的距离 $D_{AD} = 0$,所以 $S_{AB} \approx 8, S_{AC} = 9.7850, S_{AD} = 8$ 。这时例子中提及的尴尬局面就得到了较为圆满的解决。

3 实验结果分析

3.1 测试数据

本文利用 Movielens 数据集对提出的相似性度量方法的预测性能进行了测试。Movielens 包括 943 个用户、1682 部电影、100000 条评分信息,其中每个用户至少有 20 条评分信息。为了便于评估各个相似度指标的性能,将原始数据集划分成训练集和测试集。从每个用户评分集中随机选取一部分构成训练集,剩下的那部分评分数据构成测试集。训练集用于模型训练,测试集用于测试方法性能的优劣。为了便于比较分析,把数据集分别划分成 90%, 80%, 60%, 40%, 20% 的训练集和对应的测试集。每个训练集都用相同的方法划分 5 次,得到 5 组不同的数据集。这样每种相似度指标在每个百分比的训练集上分别进行 5 次运算,然后取其平均值进行比较。

3.2 实验分析

由式(7)一式(10)知, salton, jaccard, sorensen 和 LHN-I 等 4 种基于结构的相似度指标,通过对用户度的不同处理,对 CN 指标进行改造。但由表 1 显示它们并没有改进预测效果,反而使 RMSE 值和 MAE 值都增大了。而我们在 CN 基础上考虑用户的距离因素所得到的改进算法却使 RMSE 值和 MAE 值都减小了。表 3 显示了我们的方法与 CN 及 LHN-I 的预测结果的对比情况(DCN 代表我们的方法,其中 D 表示距离, DCN 表示加入距离因素的 common-neighbors 算法)。表 3 是在不同训练集上分别做 5 次实验,取其平均值作比较,其中黑色字体标注的是最佳的结果。表 3 显示了当改变训练集大小时,用我们方法得到的 RMSE 和 MAE 的平均值都比 common-neighbors 方法所得的 RMSE 和 MAE 要小。图 1 为在 90% 的训练集上, 5 次实验所得 MAE 值的对比情况。表 3 和图 1 的对比结果说明在 CN 基础上引入距离因素对预测准确度有一定程度的提高。

表 3 CN, LHN-I 和 DCN 方法在不同训练集上的比较

训练集	RMSE			MAE		
	LHN-I	CN	DCN	LHN-I	CN	DCN
90%	0.717242	0.699833	0.699449	0.798246	0.773976	0.773282
80%	0.842828	0.824165	0.824079	0.807898	0.781544	0.781232
60%	0.933186	0.920397	0.920289	0.811274	0.793063	0.792686
40%	0.989992	0.981316	0.981041	0.822314	0.809384	0.80862
20%	1.0321	1.028505	1.028234	0.84317	0.837096	0.836762

(下转第 204 页)

GA, EDAs 和 ACO 取得更高的效率。CEA 的优势来自于可以将一次目标评价所产生的知识更好地进行利用,通过复杂的思维过程代替了评价操作,从而提高了优化效率。

(2)CEA 的演化过程是可以解释的,因为 CEA 是对人类认知行为的直接模拟,符合人类的认知习惯,这是其他智能算法所不具备的。

但是 CEA 在求解中也存在以下的局限性。首先,需要目标函数评价是计算密集型操作才能体现 CEA 的优势,否则复杂的思维操作将抵消由于减少目标评价次数所带来的收益;其次,必须能够从目标评价活动中发现知识,否则反而增加了学习等操作的开销;最后,需要合理地设置 CEA 的参数。

结束语 创造性思维是人类社会进步的动力。本文借鉴认知心理学的研究成果,通过对人类认知系统和创造性思维过程进行建模,提出了一种以知识为基础(学习和记忆),综合两类思维技巧(发散和收敛),以价值为导向(价值体系)的新的复杂问题求解算法——认知演化算法。通过实验证实了本算法可以大大减少目标评价次数,提高优化效率。

本文的研究成果已成功应用于基于仿真的作战方案生成的研究中,可以有效地减少计算密集型仿真的运行次数,提高仿真优化效率。下一步研究需要对结论进行严格的理论推导,同时后续研究可将以非逻辑思维为主的创造性思维以及群体创新思维纳入算法框架中。

参考文献

[1] Goldberg D E. Genetic algorithms in search, optimization and machine learning[M]. Inc. Boston, MA, USA; Addison-Wesley Longman Publishing Co., 1989

[2] Rudolph G. Convergence analysis of canonical genetic algorithms [J]. IEEE Transactions on Neural Networks, 1994, 5(1):96-101
 [3] 马永杰, 马义德, 蒋兆远, 等. 一种快速遗传算法及其收敛性[J]. 系统工程与电子技术, 2009, 31(3):714-718
 [4] Bonissone P P, Subbu R, Eklund N, et al. Evolutionary algorithms + domain knowledge = real-world evolutionary computation [J]. IEEE Transactions on Evolutionary Computation, 2006, 10(3):256-280
 [5] 董占球, 刘岩. 一类智能计算思想: 按自然法则计算[J]. 计算机科学, 1996, 23(03):14-16
 [6] Turchin V F, Frenzt B. The phenomenon of science[M]. New York: Columbia University Press, 1977
 [7] 钱学森. 创建系统学[M]. 太原: 山西科技出版社, 2001
 [8] Boden M A. The creative mind: myths and mechanisms[M]. 1991
 [9] Gabora L. Cognitive mechanisms underlying the creative process [C]//Proceedings of the 4th Conference on Creativity & Cognition. Loughborough, UK; ACM, 2002: 126-133
 [10] 廖海林. 关于记忆与遗忘定量研究的两个数学模型[J]. 江西广播电视大学学报, 2003, 20(4):41-42
 [11] Kirton M J. Adaptors and innovators: A description and measure [J]. Journal of Applied Psychology, 1976, 61(5):622-629
 [12] Higgins L F. A comparison of scales for assessing personal creativity in IS[C]//Proceedings of the 29th Hawaii International Conference on System Sciences. Hawaii, US; IEEE, 1996: 13-19
 [13] Higgins L F, Couger J D. Comparison of KAI and ISP instruments for determining style of creativity of IS professionals[C]//Proceedings of the 28th Hawaii International Conference on System Sciences. Hawaii, US; IEEE, 1995:566-570

(上接第 186 页)

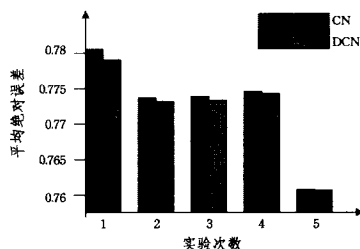


图 1 CN 和 DCN 算法的 MAE 值在 5 次实验上的结果比较

在时间复杂度方面,我们可以在查找共同评分项目的同时,计算它们的评分差,因此本方法与 CN 方法有相同的时间复杂度。

结束语 本文在 CN 算法的基础上考虑用户的评分信息,得到改进的 CN 算法——DCN 方法。并在 movielens 数据集上,对 CN 方法和本文提出的改进算法进行了实验对比分析。实验结果表明,本文算法 DCN 在一定程度上提高了评分预测准确度。未来的研究工作将进一步在其他网络数据集上进行试验,检测距离因素对推荐效果的影响。同时研究距离因素是否与最近邻居数有关系,以及随着最近邻居数目的改变,它对相似度的影响情况有何变化趋势。

参考文献

[1] Adomavicius G, Tuzhilin A. Expert-driven validation of rule-based user models in personalization applications[J]. Data Mining and Knowledge Discovery, 2001, 5(1/2):33-58

[2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transaction on Knowledge and Data Engineering, 2005, 17(6):734-749
 [3] 梅田望夫. 网络巨变元年——你必须参加的大未来[M]. 先觉: 先觉出版社, 2006
 [4] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]//The 14th Conference on Uncertainty in Artificial Intelligence. 1998: 43-52
 [5] Liu Jie, Shang M S, Chen D B. Personal recommendation based on weighted bipartite networks[C]//The 6th International Conference on Fuzzy Systems and Knowledge Discovery. 2009, 8: 134-137
 [6] Resnick P, Iacovou N, Suchak M, et al. GroupLens: An open architecture for collaborative filtering of Netnews[C]//Proceeding of ACM 1994 Conference on Computer Supported Cooperative Work. 1994:175-186
 [7] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithm [C] // The 10th Interling World Wide Web Conference. 2001:285-295
 [8] Sørensen T. A method of establishing groups of equal amplitude in a plant based on similarity of species content and its applications to analysis of vegetation on Danish commons[J]. Biologiske Skrifter, 1948, 5:1-34
 [9] Leicht E A, Holme P, Newman M E J. Vertex similarity in networks[J]. Physical Review E73, 2006:026120
 [10] Shang M S, Jin C H, Zhou T, et al. Collaborative filtering based on multi-channel diffusion[J]. Physica A, 2009, 388(23):4867-4871