

文章编号:1672-058X(2012)07-0103-05

# 基于协同过滤的在线教学视频推荐方法\*

贾志洋, 石宜金, 李 丁, 马瑞英

(云南大学 旅游文化学院, 云南 丽江 674100)

**摘 要:**在目前的在线教学系统中,用户对教学视频的选择具有一定的盲目性,根据这一特点,提出了一种基于协同过滤的在线教学视频推荐方法,可以将用户可能感兴趣的的教学视频“推”给用户;首先将用户的观看视频纪录整理并保存至数据库中,依据各用户历史播放纪录以及用户的基本信息的兴趣差异来查询邻居用户,然后利用这些邻居用户的视频观看记录基于协同过滤的方法进行教学视频的推荐;改进了传统协同过滤推荐方法中普遍存在的稀疏性(Sparse)和冷启动(Cold Start)等问题,因此能使推荐更为精确;另外,通过用户是否观看所推荐的视频,可以对系统做出隐性评价以修正系统的参数,以提高推荐的准确性。

**关键词:**推荐系统;协同过滤;在线教学;教学视频

**中图分类号:**TP391.7

**文献标志码:**A

视频教学已经越来越广泛地被应用于在线交互学习<sup>[1]</sup>中,如何让用户快速有效地获取到自己需要的教学视频是一项非常重要的工作。传统的方式更多地是采用“拉”的方式,即由用户进行基于关键字或基于自然语言的视频检索<sup>[2]</sup>。在此根据电子商务个性化推荐系统的相关成熟的技术,采用“推”的方式向用户推荐其可能感兴趣的的教学视频。协同过滤(collaborative filtering)<sup>[3]</sup>是推荐系统中应用最广泛、最成功的个性化推荐技术。推荐系统以个性化方式向用户推荐商品,帮助用户找到所需要的商品,并便捷地完成购买过程。协同过滤技术一般采用最近邻技术,利用用户的历史喜好信息计算用户之间的距离,然后利用目标用户的最近邻居对商品评价的加权平均值来预测用户对特定商品的喜好程度,从而根据喜好程度对目标用户进行推荐。

## 1 视频推荐流程

协同过滤的出发点是兴趣相近的用户可能会对同样的东西感兴趣。所以,只要维护关于用户喜好的数据,从中分析得出具有相似兴趣的用户(邻居用户),然后就以邻居用户的兴趣来向其进行推荐。

根据协同过滤的基本思想<sup>[4]</sup>,教学视频推荐方法的推荐流程分为 3 个步骤:

- (1) 用户信息表示。对用户的视频观看记录进行建模,从而可以有效度量用户之间的相似性;
- (2) 邻居用户的形成。根据用户信息计算用户之间的相似度,从而形成邻居用户;
- (3) 产生推荐。根据当前用户最近邻居的视频观看记录预测当前用户的观看兴趣,产生 Top-N 视频

收稿日期:2011-12-27;修回日期:2012-01-21.

\* 基金项目:云南省教育厅科学研究基金资助项目(2011y345).

作者简介:贾志洋(1980-),男,吉林人,硕士,讲师,从事机器学习研究.

推荐。

根据这 3 个步骤,可以利用用户的基本信息及过去的视频观看历史记录进行查询,并形成邻居用户名单并记录在用户数据库的对应记录中,当用户登录系统后即可根据其邻居用户的观看记录进行推荐。

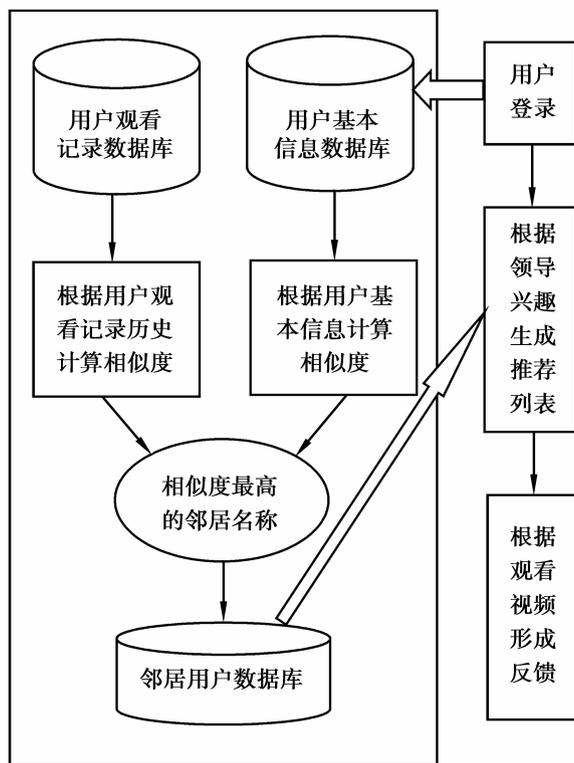


图 1 推荐流程图

主要的推荐流程如图 1 所示。其中,为了增加系统的执行效率,在此采用离线的方式计算每个用户的邻居用户名单,每隔一定时间(假设时间周期  $T=1\text{ d}$ )计算每个用户彼此观看视频的相似度或用户基本信息的相似度,并将最相似的邻居名单存储于邻居用户数据库中,并将用户每次的视频观看记录视为一次反馈,以便精确计算推荐的效度。

## 2 视频推荐方法

### 2.1 邻居用户的生成

邻居用户的生成主要依据用户彼此间的相似度来决定。但是用户的数量较大,而且大量的用户不具有代表性或已经失去时效性,所以在生成邻居用户名单前,必须先将用户数据做预处理,包括数据清洗、数据整合、数据转换以及数据精简,最后得到较精简且具有代表性的用户名单,这些通常是最近观看记录的用户,即活跃用户。因此,每个用户就可以与这些活跃用户做相似度比较,并从中找出相似度较高的活跃用户并加入其邻居用户名单。对于已经有视频观看记录的用户,可以利用观看过的视频记录作为其计算相似度的依据;对于尚无观看视频记录的用户,可以利用用户的基本信息(如登陆次数、年级、专业、班级等)作为其计算相似度的依据。假设所有用户的集合  $U = \{U_1, U_2, \dots, U_n\}$ , 每一个用户  $U_i (i=1, 2, \dots, n)$  通过计算相似度找出相似度高于阈值  $\theta$  的  $N$  个邻居用户的名单。

用户相似度的计算是整个基于用户的协同过滤算法的核心部分。度量用户  $U_i$  和用户  $U_j$  之间相似度的

方法如下:首先得到用户  $U_i$  和用户  $U_j$  评分过的所有项,然后通过不同的相似性度量方法计算用户  $U_i$  和用户  $U_j$  之间的相似度,记为  $\text{sim}(U_i, U_j)$ 。度量用户间相似度的方法有许多种,主要包括3种方法:余弦相似度(Cosine)、相关相似度(Correlation)和修正的余弦相似度(Adjusted Cosine)<sup>[5]</sup>。采用了余弦相似度计算方法,其原理为用户评分看作为  $n$  维项空间上的向量,如果用户没有进行评分(没有观看过视频),则将用户对该项(视频)的评分设为0,否则为1,用户间的相似度通过向量间的余弦夹角度量。设用户  $U_i$  和  $U_j$  在  $n$  维项空间上的评分分别表示为向量  $R_{U_i}$ 、 $R_{U_j}$ ,则用户  $U_i$  和用户  $U_j$  之间的相似度  $\text{sim}(U_i, U_j)$ :

$$\text{sim}(U_i, U_j) = \frac{R_{U_i} \cdot R_{U_j}}{\|R_{U_i}\| * \|R_{U_j}\|} \quad (1)$$

式(1)中分子为两个用户评分向量的内积,分母为两个用户向量模的乘积。

在此以表1为例(假设活跃用户共5个)说明用户之间相似度计算的方法。表1中的数字0表示用户  $U_i$  没有观看过此视频,数字1表示用户  $U_i$  观看过此视频。

表1 代表性用户的视频观看记录表

会员	视频1	视频2	视频3	视频4	视频5
$U_1$	1	0	1	1	1
$U_2$	0	1	1	1	1
$U_3$	1	0	1	1	0
$U_4$	0	1	1	0	1
$U_5$	0	1	0	1	0

假设  $S_i$  为用户  $U_i$  所有观看过的视频的集合,那么用户  $U_i$  与用户  $U_j$  的相似度可以定义:

$$\text{sim}(U_i, U_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \quad (2)$$

那么可以根据式(2)分别计算得出用户  $U_1$  与用户  $U_2$ 、用户  $U_3$ 、用户  $U_4$  和用户  $U_5$  的相似度:

$$\text{sim}(U_1, U_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{3}{5} = 0.6$$

$$\text{sim}(U_1, U_3) = \frac{|S_1 \cap S_3|}{|S_1 \cup S_3|} = \frac{3}{5} = 0.6$$

$$\text{sim}(U_1, U_4) = \frac{|S_1 \cap S_4|}{|S_1 \cup S_4|} = \frac{2}{5} = 0.4$$

$$\text{sim}(U_1, U_5) = \frac{|S_1 \cap S_5|}{|S_1 \cup S_5|} = \frac{1}{5} = 0.2$$

如果令阈值  $\theta = 0.5$ ,则用户  $U_1$  的邻居用户包括用户  $U_2$  及用户  $U_3$ 。另外,对于尚无观看记录的用户,可以将用户基本信息作为计算相似度的依据,其计算方式与有观看记录的用户类似,只是把表1的视频改成用每一个不同的用户的信息特征项(如有登陆次数、年级、班级、专业等)来取代,若某一信息特征项相同则赋值1 否则赋值0。如可计算出相似度并求出邻居用户。

## 2.2 产生推荐

推荐列表是根据邻居用户的观看人数排序得出。以表2例,假设某用户的邻居用户包括  $U_1$ 、 $U_2$  及  $U_3$ ,当表2中某个视频最近被  $U_i$  观看过,则相应用户  $U_i$  的值就会被设定为1,否则设定为0。因此,只要将每

个视频所观看的人数累加到最后一行中,就可以根据视频观看的总人数进行排序并决定那些视频应该被推荐。在此例中,当用户  $U_i$  登陆系统后,系统从邻居用户数据库中查询其邻居有  $U_1$ 、 $U_2$  和  $U_3$ ,于是由表 2 就可以得知视频 3 以及视频 4 的观看人数较高,因此,视频 3 以及视频 4 就会被列入推荐列表。

表 2 邻居用户的视频观看记录统计表

会员	视频 1	视频 2	视频 3	视频 4	视频 5
$U_1$	1	0	1	1	1
$U_2$	0	1	1	1	1
$U_3$	1	0	1	1	0
合计	2	1	3	3	2

教学视频的推荐不同于一般商品推荐,用户在不同的时间点对于购物的需求因人而异,然而,对于在线教学视频,用户有可能每次登陆后观看的都是相同的视频。所以,观看的次数也相当重要,研究目前只针对视频是否有被观看过作为计算的依据,将来的研究也将考虑视频被观看过的频率,使得每个用户对某个视频的兴趣取向更为明显。

### 2.3 隐式评价

用户信息的获取主要获得用户对给定信息的评价。评价分为显式评价<sup>[6]</sup>和隐式评价<sup>[7]</sup>两种。显式评价需要用户有意识地表达自己对某一信息的认同程度,一般用整数值来表示喜欢的程度。协同过滤系统向新用户提供一个信息列表,要求用户对其中全部或部分信息进行评价,系统获得用户的这些初始信息后,就能将用户加入到用户库中,随着用户不断使用推荐系统,用户信息不断积累,从而为用户提供越来越符合其个人需要的推荐。隐式评价希望从用户的行为中获得用户信息,目前已做的研究有通过分析用户网上购物记录、阅读文章时间的长短、URL 的连接次数等数据记录获取。就目前来说,隐式评价对数据分析的难度较大,准确性和有效性有待进一步提高,所以协同过滤领域的研究主要以显式评价为主,从发展潜力上来看,随着隐式评价数据分析、处理能力的不断提高,运用隐式评价的推荐系统将会占越来越大的比重。

在此采用的评价方式为隐性评价,用户通过推荐视频列表观看或通过搜索的方式进行观看,每一次的观看记录都是对系统的评价,不需要另外要求用户进行评价,减少了用户的困扰。通过用户每次的观看记录,对系统来说即是做了评比,可以得到精确的评比分数。系统接收到用户的对视频观看的请求,即可以依设定的时间重新进行相似度的计算,重新寻找兴趣相似的邻居用户,并且可以重新推荐视频。

## 3 结 论

研究克服了传统的协同过滤的缺点,如针对评价稀疏性(Sparse)<sup>[8]</sup>问题,用户的评价即为用户观看记录,只要有观看,即有评价,每一次的视频的观看都视为一次的隐性评价;针对冷启动(Cold-Start)<sup>[9]</sup>问题,新的用户可以通过用户基本信息<sup>[10]</sup>,依照专业、登陆次数等数据进行相似度计算生成邻居。

研究将协同过滤的方法应用于在线教学视频的推荐,使得在线教学系统具有个性化的特色并提升用户的查找视频的速度。系统定期以离线方式更新每个用户的邻居用户列表,使得用户群不会因为用户兴趣改变而不具代表性。另外,通过每次推荐列表的评价反馈,相关系统参数可以不断地学习使系统越来越成熟,并且能够适应多变化的用户群。

**参考文献:**

- [1] 胡开明. 基于.NET的发卷交卷系统设计[J]. 重庆工商大学学报:自然科学版,2010,27(3):252-256
- [2] 李涛,王国营. 在线学习视频索引与检索方法[J]. 计算机工程与设计,2010,31(3):647-650
- [3] RESNICK P, IACOVOU N, SSUCHAK M, et al. GroupLens: an open architecture for collaborative filtering of netnews [C]. Proceedings of the 1994 ACM conference on Computer supported cooperative work, New York: ACM, 1994
- [4] 邓爱林,朱扬勇,施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报,2003,14(9):1621-1628
- [5] 徐翔,王煦法. 协同过滤算法中的相似度优化方法[J]. 计算机工程,2010,36(6):52-57
- [6] 姚曜,赵洪利,杨海涛,等. 协同过滤技术研究综述[J]. 装备指挥技术学院学报,2011,22(5):81-88
- [7] NICHOLS D. Implicit rating and filtering [C]. Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering, Budapest: ERCIM, 1997
- [8] 曾小波,魏祖宽,金在弘. 协同过滤系统的矩阵稀疏性问题的研究[J]. 计算机应用,2010,30(4):1079-1082
- [9] 陈志敏,李志强. 基于用户特征和项目属性的协同过滤推荐算法[J]. 计算机应用,2011,31(7):1748-1755
- [10] WASFI A. Collecting user access patterns for building user profiles and collaborative filtering [C]. Proceedings of the 4th international conference on intelligent user interfaces, New York: ACM, 1999

## Online Teaching Video Recommendation Method Based on Collaborative Filtering

**JIA Zhi-yang, SHI Yi-jin, LI Ding, MA Rui-ying**

(School of Tourism and Culture, Yunnan University, Yunnan Lijiang 674100, China)

**Abstract:** In current online teaching system, the users have certain blindness on choosing teaching videos, according to this characteristic, online teaching video recommendation method based on collaborative filtering is pointed out, which can recommend the teaching videos in which the users are interested to other users. Firstly, the video-watching records are saved in database, and neighboring users are searched based on historic viewing records of each user and the interest difference between each user in basic information, then teaching video recommendation is conducted according to video-watching records of these neighboring users based on collaborative filtering method. This method improves the Sparse and Cold Start commonly existed in traditional collaborative filtering recommendation method and makes the recommendation more accurate. In addition, whether the users watch the recommended videos can make implicit evaluation on the system to revise the parameters of the system and to improve the accuracy of the recommendation.

**Key words:** recommendation system; collaborative filtering; online teaching; teaching video