教学视频精准推荐算法研究

尹 杭，宋文广，龙 力，吴华超
（长江大学计算机科学学院，湖北荆州 434023）

摘要：通过分析现有精准营销推荐算法，结合学员通过网络学习教学视频的特点，研究传统协同过滤算法的优缺点，提出一种基于教学视频的协同过滤算法。首先设计一种基于教学视频网站的聚类协同过滤算法，使用随机函数在样本中初始化多个聚类中心。再将样本分配给距离其最近的中心向量，由这些样本构造不相交的聚类，用各个聚类的中心向量作为新的中心。最后进行循环计算分组和确定中心的步骤，直至算法收敛或到达确定的迭代步数为止。通过分析实验数据得出，经过聚类后的样本数据在推荐过程中表现出更高的价值。该算法能提高学员搜索相关教学视频的速度，并能提高学习效率。

关键词：算法理论；教学视频；协同过滤；精准推荐
中图分类号：TP301.6 **文献标识码：**A **文章编号：**1674-2850(2017)22-2445-07

Research on accurate recommendation algorithm for teaching video

YIN Hang, SONG Wenguang, LONG Li, WU Huachao
(*School of Computer Science*, *Yangtze University*, *Jingzhou*, *Hubei* 434023, *China*)

Abstract: By analyzing the existing accurate marketing recommendation algorithm, combined with the characteristics of students learning online teaching video, this paper studies the advantages and disadvantages of traditional collaborative filtering algorithm, and proposes a collaborative filtering algorithm based on teaching video. Firstly, a clustering collaborative filtering algorithm based on the teaching video website is designed, which uses the random function to initialize multiple cluster centers in the sample. Then, the samples are assigned to the nearest central vectors, the disjoint clusters are constructed by these samples, and the central vectors of each cluster are taken as the new centers. Finally, the steps of grouping and determining the center are calculated circularly until the algorithm converges or arrives at the determined iteration step. By analyzing the experimental data, the sample data after clustering show higher value in the recommendation process. The algorithm can improve the speed of searching for related teaching video by students, and also improve the learning efficiency.

Key words: algorithm theory; teaching video; collaborative filtering; accurate recommendation

0 引言

突出个性培养，即根据用户自身特征，自主选择适合自己发展的专业知识视频进行在线学习，越来越普及。电商营销中的精准营销策略，同样可以应用到学生个性化学习上。精准营销最早被用于 Amazon 电子商务的商品推荐，因此被称为“个性化推荐技术”。在视频教学网站中，如果能满足用户的个性化学习需求，将会更利于对人才系统全方位的培养，采用个性化推荐算法提高用户在视频教学网站中的学习效率和满意度，对一个视频教学网站来说有极其重要的价值。将传统的协同过滤算法与视频教学网站恰

基金项目：长江大学校级研究项目（JY2014011）；长江大学 2016 创新创业训练计划（2016083）
作者简介：尹杭（1995—），男，本科生，主要研究方向：数据分析与处理及软件研制
通信联系人：宋文广，副教授，主要研究方向：石油行业软件研制. E-mail: whswg1979@126.com

当地结合在一起，形成一种新的教学视频精准推荐学习方法，将极大地方便学员进行在线学习，节省大量网上筛选、寻觅相关教学视频内容的时间，可以很好地提高学员的学习效率和学习质量。

协同过滤算法[1]是目前应用最广泛、最成功的个性化精准推荐算法之一，可以分为基于模型的协同过滤算法和基于记忆的协同过滤算法。基于模型的协同过滤算法是根据用户的喜好数据，训练出推荐模型，再根据实时的信息数据对用户进行预测来推荐相关内容以供用户选择。基于记忆的协同过滤算法，可以分为基于用户的协同过滤算法和基于视频的协同过滤算法[2~3]，前者是找到与用户目标兴趣相似的集合数据，再搜寻该用户集合中目标用户没有浏览过但可能喜欢的对象等相关内容推荐给目标用户。基于用户的协同过滤算法更关注与目标用户相似的群体，更加社会化。这种算法在新闻网站中有所体现，在这种类型的网站中用户更加关注的是当前的热点新闻，给用户推荐相同兴趣爱好的关注内容，同时具备了热点和时效性。基于视频的协同过滤算法是计算视频之间的相似度，根据视频的相似度和用户的历史行为形成为用户推荐列表数据，更关注目标用户的个性化需求。基于视频的协同过滤算法更适用于电子商城等购物网站，根据用户的历史行为，预测其可能感兴趣的商品，提供一种个性化推荐。这种算法在电子商务网站中获得了很好的应用。

在视频教学网站中，用户的知识背景与兴趣爱好均有很大差异，使用传统的协同过滤算法并不能取得很好的推荐效果，故提出使用数据分析中的聚类算法[4~5]，将用户进行分组，增强用户分组内的聚合性。在进行教学视频推荐时可获得更有价值的数据，在经过优化的用户分组中寻找最合适的推荐结果，提高推荐算法的准确性。

1　视频教学数据预处理

在视频教学网站中，不同用户对同一个教学视频的认可程度不尽相同。用户对教学视频的认可程度表现为对视频评分的高低，个性化推荐算法可以根据用户的评分数据，分析用户对教学内容的认可程度，有针对性地为用户推荐其接纳程度比较高的教学视频内容，达到精准推荐教学视频内容的目的。可以采用协同过滤算法，如下。

基于模型的协同过滤算法：先采用相似度计算公式[6]，计算与目标用户 a 最相似的 M 个用户，再根据这 M 个用户的喜好，为用户推荐其可能喜欢的教学视频。设定在系统中有 2 个用户，分别为 u_1 和 u_2，令 $N_{(u_1)}$ 表示用户 u_1 的评分集合，$N_{(u_2)}$ 表示用户 u_2 的评分集合，采用余弦相似度公式计算 2 个用户 u_1 和 u_2 之间兴趣爱好的相似度。用户相似度矩阵如下：

$$\boldsymbol{H}=\begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m-1} & r_{1,m} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,m-1} & r_{2,m} \\ \vdots & \vdots & & \vdots & \vdots \\ r_{n-1,1} & r_{n-1,2} & \cdots & r_{n-1,m-1} & r_{n-1,m} \\ r_{n,1} & r_{n,2} & \cdots & r_{n,m-1} & r_{n,m} \end{bmatrix}.$$

设 $U=\{u_1,u_2,\cdots,u_n\}$ 表示所有用户的集合，$V=\{v_1,v_2,\cdots,v_n\}$ 表示所有教学视频的集合。在矩阵 $\boldsymbol{H}$ 中，任一用户 i 对任一教学视频 j 的评分为 $r_{i,j}$，计算用户 u 与用户 v 的相似度，将用户 u 和用户 v 设成 m 维的向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$，余弦相似度公式为

$$\mathrm{sim}(u,v)=\cos(u,v)=\frac{\boldsymbol{u}\times\boldsymbol{v}}{\|\boldsymbol{u}\|_2\times\|\boldsymbol{v}\|_2}. \tag{1}$$

计算用户 u 和用户 v 的相似度后，同理可以计算他们与其他用户之间的相似度。再计算与目标用户 a

最相似的 M 个用户的相似度，根据这 M 个用户的评分记录，预测用户 a 对从来没有看过的教学视频 q 的喜好程度，计算公式为

$$W(a,q)=\sum_{c\in S(a,M)\cap J(q)}(r_{c,q}\mathrm{sim}(a,c)), \tag{2}$$

其中，$S(a,M)$ 为与用户 a 最相似的 M 个用户的集合；$J(q)$ 为对视频 q 感兴趣的用户集合；$r_{c,q}$ 为在矩阵 $\boldsymbol{H}$ 中用户 c 对视频 q 的评分；$\mathrm{sim}(a,c)$ 为用户 a 与用户 c 的相似度。

基于视频的协同过滤算法：首先根据用户 a 对视频的评分列出视频相似度矩阵，再根据视频相似度矩阵计算出与目标视频 q 最相似的 N 个视频，再根据用户的历史记录数据，预测用户可能喜欢的视频。首先列出视频的相似度矩阵，与用户相似度矩阵类似，任一用户 i 对任一视频 j 的评分为 $R_{j,i}$.

视频相似度矩阵如下：

$$\boldsymbol{V}=\begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,m-1} & R_{1,m} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,m-1} & R_{2,m} \\ \vdots & \vdots & & \vdots & \vdots \\ R_{n-1,1} & R_{n-1,2} & \cdots & R_{n-1,m-1} & R_{n-1,m} \\ R_{n,1} & R_{n,2} & \cdots & R_{n,m-1} & R_{n,m} \end{bmatrix}.$$

视频 y 和视频 w 之间的相似度类似于用户之间的相似度，计算公式为

$$\mathrm{sim}(y,w)=\frac{\boldsymbol{y}\times\boldsymbol{w}}{\|\boldsymbol{y}\|_2\times\|\boldsymbol{w}\|_2}. \tag{3}$$

接着，用相似度计算公式计算与视频 q 最相似的 M 个视频，根据用户的操作数据计算，得到用户 a 对视频 q 的感兴趣程度，计算公式为

$$B_r(a,q)=\sum_{p\in S(q,M)\cap J(a)}(\mathrm{sim}(q,p)R_{p,a}), \tag{4}$$

其中，$S(q,M)$ 为与视频 q 相似度最高的 M 个视频的集合；$J(a)$ 为用户 a 看过或感兴趣的视频集合；$\mathrm{sim}(q,p)$ 为视频 q 与视频 p 的相似度；$R_{p,a}$ 为用户 a 对视频 p 的评分。

2 视频精准推荐算法设计

在视频教学网站的应用中，精准推荐教学视频需要考虑用户的兴趣和学习能力。同时考虑到目标用户在一段时间内所属的群体是不变的，因此先采用 K-means 计算方式[7]，将用户进行聚类；进行基于用户的协同过滤时，在已经确定的聚类中心中寻找与目标用户相似的用户组，此时在经过聚类后的用户分组中用户之间的相似度更高，其评分数据对目标用户也更有价值，在该聚类中，计算出与目标用户最相似的 K 个用户，对用户进行精准推荐。

设计一种基于视频教学网站的聚类协同过滤算法：使用随机数随机产生 K 个聚类中心，计算各用户与聚类中心之间的距离，即用户之间的相似度，可采用式（1）计算。采用 K-means 聚类算法[8]将用户集合 $U=\{u_1,u_2,\cdots,u_n\}$ 划分为 K 个聚类 $U_1,U_2,\cdots,U_K$，各聚类中心分别为 $Z_1,Z_2,\cdots,Z_K$. 求出与目标用户 a 最相似的聚类中心 Z，其聚类为 L. 在相似度矩阵 $\boldsymbol{H}$ 中算得与目标用户最相似的 M 个用户的评分矩阵 $\boldsymbol{M}$. 通过式（2）预测目标用户 a 对视频 q 的喜好程度，其计算模型为

$$I_{a,q}=\sum_{k\in L}\boldsymbol{M}_{k,q}\times\boldsymbol{H}_{a,k}, \tag{5}$$

其中，$\boldsymbol{M}_{k,q}$ 为用户 k 对视频 q 的评分；$\boldsymbol{H}_{a,k}$ 为用户 a 与用户 k 的相似度。

3　实验验证与数据结果分析

3.1　算法关键参数设置

评价指标：A 为推荐给用户 a 的视频；T 为测试数据中用户 a 喜欢的视频。准确率表示在测试集合中符合预测的视频数占预测总视频数的比率；召回率表示在测试集合中符合预测的视频数占用户喜欢的视频数的比率；覆盖率表示所有预测视频的并集占视频总集合的比率[9]，公式如下：

$$准确率=\frac{\sum_{a\in U}|A\cap T|}{|A|},$$

$$召回率=\frac{\sum_{a\in U}|A\cap T|}{|T|},$$

$$覆盖率=\frac{|U_{a\in U}A|}{|V|}.$$

参数设置：首先将 M 与 N 的值设置为 10，计算在不使用 K-means 聚类算法情况下的准确率、召回率和覆盖率，再计算使用 K-means 聚类算法情况下的准确率、召回率和覆盖率，并改变 K 值和 step 值，分析最终结果。将对 K-means 划分的聚类数和迭代步数[10]分别定义为 K 和 step，将 K 分别设置为 10，20，…，100，将 step 分别设置为 5，10，…，50.

数据集：训练集中包含 5 603 个用户对 3 468 个视频的评分记录，总计 1 000 000 条，测试集中包含 1 325 个用户对 3 468 个视频的评分记录，总计 250 000 条，将训练集中的数据导入模型中，得出用户对视频的预测评分，再到测试集上检验预测评分，得出最后的评分结果。

3.2　算法处理结果数据分析

传统协同过滤算法的准确率为 0.038 965，召回率为 0.543 267，覆盖率为 0.913 634，在使用传统协同过滤算法产生的预测结果中，预测用户对某视频喜欢，在测试集中命中数目占预测的视频数的比率表示预测的准确率，该数据具有很高的价值，其微量的提高都会带来巨大的经济效益。命中的视频数占测试集中用户喜欢的视频数的比率表示预测的召回率；预测的视频数占总视频数的比率表示预测的覆盖率。有一些冷门的视频缺少用户评分数据，故很难被推荐给用户，因此用覆盖率衡量预测结果对这种情况的处理效果。采用 K-means 聚类算法后的准确率、召回率、覆盖率分别如表 1～表 3 和图 1～图 3 所示。

表 1　教学视频精准推荐准确率

Tab. 1　Accuracy rate of accurate recommendation in teaching video

step	K									
	10	20	30	40	50	60	70	80	90	100
5	0.078 204	0.076 766	0.075 360	0.076 431	0.079 152	0.079 821	0.076 615	0.077 172	0.078 813	0.078 160
10	0.076 716	0.077 684	0.079 731	0.079 852	0.080 133	0.081 342	0.077 471	0.078 776	0.077 587	0.079 119
15	0.077 985	0.078 265	0.078 174	0.078 470	0.080 726	0.080 106	0.080 017	0.080 632	0.080 889	0.078 406
20	0.074 619	0.078 888	0.077 487	0.080 193	0.080 038	0.083 923	0.081 619	0.082 651	0.078 164	0.076 073
25	0.075 731	0.077 241	0.078 028	0.082 911	0.082 947	0.084 692	0.081 711	0.081 547	0.079 477	0.080 470
30	0.074 482	0.079 815	0.079 461	0.080 988	0.081 302	0.082 256	0.082 179	0.080 137	0.081 820	0.082 176
35	0.075 194	0.078 517	0.080 043	0.079 487	0.081 412	0.082 341	0.083 458	0.079 341	0.079 494	0.081 575
40	0.076 477	0.076 691	0.078 648	0.078 159	0.081 274	0.081 942	0.081 410	0.080 517	0.076 641	0.080 127
45	0.073 715	0.075 614	0.079 238	0.077 206	0.080 325	0.081 207	0.078 719	0.078 491	0.077 376	0.078 142
50	0.072 271	0.075 917	0.080 154	0.078 870	0.081 230	0.082 816	0.078 714	0.076 716	0.076 517	0.079 183

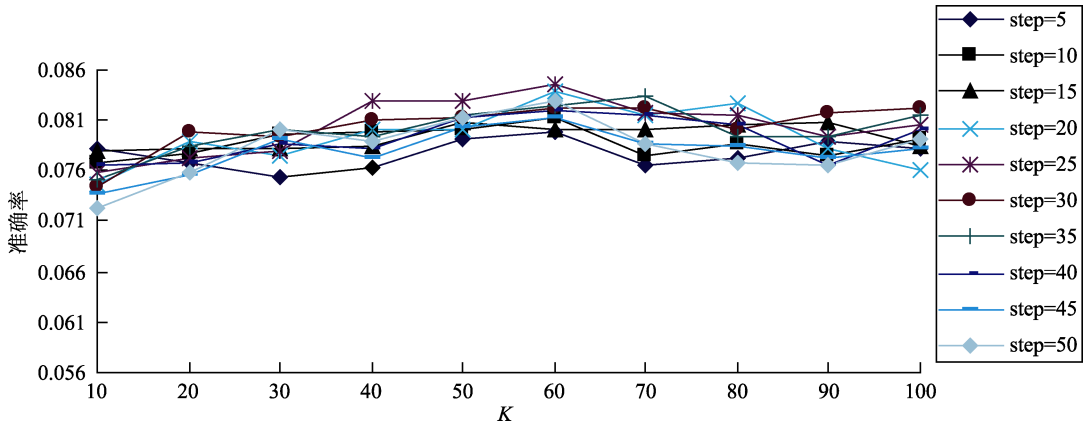


图 1 教学视频精准推荐准确率图

Fig. 1 Accuracy rate chart of accurate recommendation in teaching video

由表 1 和图 1 可知，准确率相对于传统的协同过滤算法有很大提升，当聚类数为 60、迭代步数为 25 时，准确率达到最大值 0.084 692. 当聚类数小于 60 时，随着聚类数的增加，准确率升高；当聚类数大于 60 时，随着聚类数的增加，准确率降低。当聚类数为 60 时，聚类中用户的相似度最好，这时聚类中的用户对目标用户的价值比所有用户集合中挑选出的用户集合中用户的价值更高，因此比普通的协同过滤算法的准确性更高。

表 2 教学视频精准推荐召回率

Tab. 2 Recall rate of accurate recommendation in teaching video

step	*K*									
	10	20	30	40	50	60	70	80	90	100
5	0.412 721	0.451 572	0.473 512	0.464 862	0.481 612	0.479 135	0.491 264	0.464 283	0.459 513	0.471 612
10	0.392 372	0.443 808	0.462 171	0.457 816	0.478 972	0.491 643	0.487 824	0.483 713	0.462 487	0.472 512
15	0.414 631	0.443 715	0.459 106	0.493 178	0.489 714	0.471 349	0.481 539	0.461 481	0.461 249	0.452 414
20	0.439 173	0.440 245	0.447 513	0.457 951	0.462 664	0.453 811	0.481 417	0.473 612	0.483 148	0.491 372
25	0.385 313	0.425 741	0.458 514	0.481 624	0.504 153	0.476 851	0.447 821	0.462 678	0.497 372	0.482 471
30	0.431 461	0.437 162	0.433 826	0.448 619	0.455 932	0.443 751	0.459 106	0.489 625	0.490 101	0.493 517
35	0.403 272	0.423 616	0.458 718	0.477 916	0.482 006	0.490 619	0.485 606	0.482 618	0.463 419	0.516 428
40	0.399 327	0.443 721	0.472 816	0.475 100	0.489 535	0.490 157	0.496 617	0.480 731	0.478 291	0.485 169
45	0.405 737	0.426 921	0.471 536	0.476 268	0.462 868	0.472 622	0.481 650	0.479 761	0.481 633	0.507 527
50	0.394 260	0.445 746	0.449 357	0.473 825	0.492 361	0.472 647	0.497 240	0.481 934	0.478 845	0.488 974

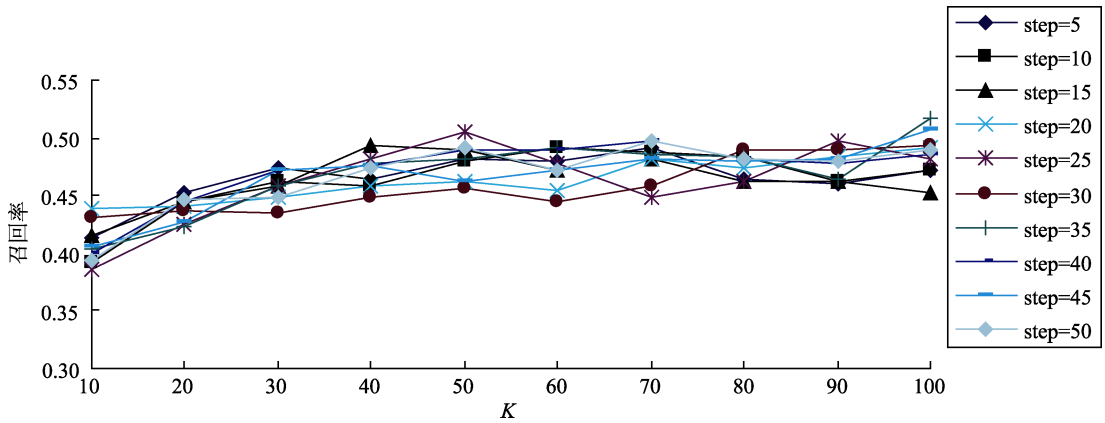


图 2 教学视频精准推荐召回率图

Fig. 2 Recall rate chart of accurate recommendation in teaching video

由表 2 和图 2 可知，当划分的聚类数和迭代步数不同时，随着聚类数和迭代步数的增加，召回率有微量递增趋势，但召回率相对于传统的协同过滤算法有所降低，最低值为 0.385 313. 将预测结果看作集合 A，测试集中用户的测试数据看作集合 B，则命中的视频集合为 $A \cap B$，使用 K-means 聚类算法后，将用户划分为一个个聚类，在聚类中用户的相似度更高，获得更有价值的评分，获得的结果集 A 相对于传统协同过滤算法的结果集更小，$A \cap B$ 即命中的视频集合占集合 A 的比率上升，即推荐的准确性提高，使用 K-means 后集合比传统协同过滤算法变小，而在 C 不变的情况下，召回率降低。由此可见，K-means 算法提高了预测结果的聚合性，聚合的同时失去了全面性，导致召回率下降，但通过提高聚类数可以在一定范围内极微小地提高召回率，但并不能改变最终聚合的结果。

表 3　教学视频精准推荐覆盖率

Tab. 3　Coverage rate of accurate recommendation in teaching video

step	K 值									
	10	20	30	40	50	60	70	80	90	100
5	0.887 263	0.889 104	0.903 745	0.895 217	0.894 317	0.887 072	0.898 428	0.904 483	0.904 752	0.915 374
10	0.879 655	0.883 521	0.889 356	0.890 346	0.894 293	0.890 357	0.903 681	0.895 626	0.885 789	0.891 752
15	0.865 715	0.882 155	0.886 864	0.891 642	0.889 217	0.889 563	0.892 156	0.904 513	0.900 532	0.900 872
20	0.856 424	0.867 342	0.893 475	0.882 125	0.874 529	0.898 345	0.899 006	0.906 648	0.907 342	0.905 333
25	0.877 462	0.878 463	0.880 461	0.874 111	0.889 745	0.889 777	0.894 527	0.909 745	0.902 622	0.909 528
30	0.869 362	0.879 302	0.888 455	0.887 846	0.890 365	0.893 644	0.899 574	0.910 731	0.898 936	0.908 754
35	0.889 345	0.863 714	0.888 956	0.889 462	0.890 646	0.893 342	0.897 444	0.902 559	0.903 461	0.893 572
40	0.872 433	0.893 623	0.894 643	0.890 454	0.897 453	0.904 653	0.904 652	0.902 364	0.910 352	0.905 412
45	0.844 672	0.879 392	0.881 538	0.874 991	0.899 645	0.904 345	0.900 455	0.904 112	0.903 520	0.907 431
50	0.871 242	0.885 278	0.889 535	0.883 647	0.893 248	0.894 228	0.904 644	0.905 361	0.907 460	0.894 626

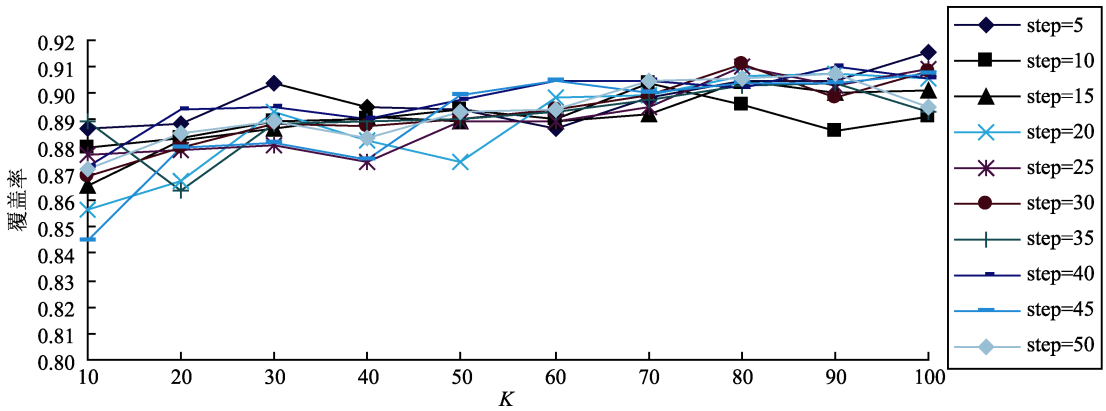


图 3　教学视频精准推荐覆盖率图

Fig. 3　Coverage rate chart of accurate recommendation in teaching video

由表 3 和图 3 可知，覆盖率在聚类数小于 60 时比传统的协同过滤算法略低，但随着聚类数的增加，覆盖率逐步上升，聚类数在 60～100 时，与传统的协同过滤算法持平。当聚类数过小时，经过聚合的预测结果的聚合性不够，存在较多的重合结果，导致覆盖率降低；当聚类数增加时，预测结果的聚合性增加，重合的结果减少，故覆盖率增加。

4　结论

通过对视频教学网站中个性化推荐技术的研究，分析视频教学网站用户的特征和传统的协同过滤算

法的原理，提出了一种基于传统的协同过滤算法结合 *K*-means 聚合算法的个性化推荐算法。该算法提高了传统协同过滤算法的准确率，实现了传统过滤算法与 *K*-means 算法之间的有效结合，优化了传统过滤算法推荐结果准确性低的问题。创新点是借鉴协同过滤算法与数据分析中的 *K*-means 聚类算法，提出一种教学视频推荐方法，能提高教学视频推荐效率，方便学员快速搜索到感兴趣的教学视频，有利于网络教学。

[参考文献] (References)

[1] 王成，朱志刚，张玉侠，等. 基于用户的协同过滤算法的推荐效率和个性化改进[J]. 小型微型计算机系统，2016，37（3）：428-432.
WANG C, ZHU Z G, ZHANG Y X, et al. Improvement in recommendation efficiency and personalized of user-based collaborative filtering algorithm[J]. Journal of Chinese Computer Systems, 2016, 37(3): 428-432. (in Chinese)

[2] 王威，郑骏. 基于用户相似度的协同过滤算法改进[J]. 华东师范大学学报（自然科学版），2016（3）：60-66.
WANG W, ZHENG J. Improved collaborative filtering algorithm based on user similarity[J]. Journal of East China Normal University (Natural Science Edition), 2016(3): 60-66. (in Chinese)

[3] 姚禹丞，宋玲，鄂驰. 同态加密的分布式 *K* 均值聚类算法研究[J]. 计算机技术与发展，2017，27（2）：81-85.
YAO Y C, SONG L, E C. Investigation on distributed *K*-means clustering algorithm of homomorphic encryption[J]. Computer Technology and Development, 2017, 27(2): 81-85. (in Chinese)

[4] 王行甫，付欢欢，王琳. 基于余弦相似度和实例加权改进的贝叶斯算法[J]. 计算机系统应用，2016，25（8）：166-170.
WANG H F, FU H H, WANG L. Improved Naive Bayes algorithm based on weighted instance with cosine similarity[J]. Computer System & Applications, 2016, 25(8): 166-170. (in Chinese)

[5] 李武，赵娇燕，严太山. 基于平均差异度优选初始聚类中心的改进 *K*-均值聚类算法[J]. 控制与决策，2017，32（4）：759-762.
LI W, ZHAO J Y, YAN T S. Improved *K*-means clustering algorithm optimizing initial clustering centers based on average difference degree[J]. Control and Decision, 2017, 32(4): 759-762. (in Chinese)

[6] 吴夙慧，成颖，郑彦宁，等. *K*-means 算法研究综述[J]. 现代图书情报技术，2011（5）：28-35.
WU S H, CHENG Y, ZHENG Y N, et al. Survey on *K*-means algorithm[J]. Modern Library and Information Technology, 2011(5): 28-35. (in Chinese)

[7] 郭文健，高仲合，段婷婷. 基于信任网络的个性化推荐算法[J]. 电子技术，2016（12）：65-67.
GUO W J, GAO Z H, DUAN T T. Personalized recommendation algorithm based on trust network[J]. Electronic Technology, 2016(12): 65-67. (in Chinese)

[8] 谢月荣，冯璐瑶，闫炳阳，等. 基于个性化混合推荐算法的网络推荐系统[J]. 电脑知识与技术，2016，12（10）：192-193.
XIE Y R, FENG L Y, YAN B Y, et al. Network recommendation system based on hybrid personalized recommendation algorithm[J]. Computer Knowledge and Technology, 2016, 12(10): 192-193. (in Chinese)

[9] 李敬明，程家兴，张伟，等. 基于改进协同过滤算法的图书网站个性化推荐模型构建研究[J]. 长春师范大学学报，2016，35（2）：40-45.
LI J M, CHENG J X, ZHANG W, et al. Research on personalized recommendation model of book website based on improved collaborative filtering algorithm[J]. Journal of Changchun Normal University, 2016, 35(2): 40-45. (in Chinese)

[10] 鲍玉斌，王大玲，于戈. 关联规则和聚类分析在个性化推荐中的应用[J]. 东北大学学报（自然科学版），2003，24（12）：1149-1152.
BAO Y B, WANG D L, YU G. Application in association rules and clustering analysis to personalized recommendation[J]. Journal of Northeastern University (Natural Science Edition), 2003, 24(12): 1149-1152. (in Chinese)