

文章编号: 1002-1566(2012)02-0325-08

奇异值分解法在预测用户页面兴趣度中的应用

刘伟江 王颖

(吉林大学商学院, 吉林 长春 130012)

摘要: 如何挖掘用户感兴趣的内容并向其推荐相应的产品是许多商务网站关注的重点。本文提出一个用户页面兴趣度公式, 据此得到一个稀疏的页面兴趣度矩阵, 考虑到矩阵的稀疏性会影响预测的准确性, 我们首先用稀疏矩阵中每列非零元素的平均值对该列中值为零的元素进行修正, 然后用奇异值分解法对原稀疏矩阵中值为零的元素进行再次修正, 进而根据修正后的矩阵用 Slope One 算法对用户的页面兴趣度进行预测并用某网站的数据验证。结果表明: 与稀疏矩阵相比, 用修正后的矩阵预测的准确性有较大提高。

关键词: 奇异值分解; 页面兴趣度; Slope One 算法**中图分类号:** C939, O212**文献标识码:** A

Predicting User's Page Interest Degree by Using Singular Value Decomposition

LIU Wei-jiang WANG Ying

(Business School, Jilin University, Jilin Changchun 130012, China)

Abstract: How to mine visitors' interest information and recommend corresponding product to them is a goal that many websites pursue. The paper proposes a users' page interest degree formula, then gets a sparse matrix about page interest degree. Considering sparse matrix can affect predicting accuracy, this paper first uses every column's not zero value's average to replace those zeroes, second using singular value decomposition to deal with those zero values in original sparse generator matrix and get a updated matrix, third applying Slope One Algorithm to predict visitors' page interest degree. At last, this paper tests the result by using a website data, experiment result is: comparing with sparse generator matrix, the updated matrix after SVD under Slope One is better in predicting user's page interest degree's accuracy.

Key words: singular value decomposition, page interest degree, slope one algorithm

0 引言

电子商务的发展和普及在给用户带来大量信息的同时也使用户陷入信息过量的困境, 因此如何抓住用户访问网站的机会, 尽可能地向访问者推荐其感兴趣的页面, 从而增加把访问者变成满意的消费者的机会是很多商家追求的目标, 而对用户感兴趣页面的预测是实现此目标的关键。

很多研究表明, 用户对网页的兴趣度与其在该网页上的浏览行为密切相关。用户的很多动作都能暗示其偏好^[1-2], 如鼠标点击、拉动滚动条等。Claypool^[3] 等人认为浏览时间和拉动滚

收稿日期: 2009年12月15日**收到修改稿日期:** 2010年3月18日**基金项目:** 吉林大学“985工程”项目, 吉林省教育厅十一五规划项目。

动条所花的时间可以有效揭示用户的兴趣。Liang^[4] 等人认为网页的平均阅读速度有助于确定用户兴趣的等级等。谭琼^[5]、吕佳^[6] 等人的研究指出, 用户访问页面时的停留时间、访问次数、保存等动作能够揭示用户的兴趣所在。李珊^[7] 等人又考虑了页面发送字节数(页面大小)这一因素对页面兴趣度的影响。此外还有一些学者^[8-10] 也从不同的角度讨论了用户页面兴趣度问题。

本文在总结前人对页面兴趣度的研究工作基础上, 提出一个新的用户页面兴趣度的计算公式, 进而得到用户页面兴趣度矩阵 R_0 , 但它通常是一个稀疏矩阵, 而矩阵的稀疏性影响预测的准确度。考虑到用户对于没有访问过页面的兴趣度实际上不一定为零, 本文提出用平均值填充和奇异值分解(Singular Value Decomposition, 简称 SVD)^[11] 方法对稀疏矩阵中值为零的元素进行修正, 得到一个稠密矩阵, 进而根据这个矩阵运用 Slope One 算法对用户的页面兴趣度进行预测。实验结果表明: 与原稀疏矩阵相比, 预测的准确性有较大提高。

1 用户页面兴趣度矩阵的构成

1.1 用户页面兴趣度的计算公式

用户页面兴趣度是用户对某页面感兴趣的程度。考虑到可操作性和合理性, 本文给出了一个新的页面兴趣度计算公式。如果把用户 i 对页面 j 的兴趣度用 $Insterest_{i,j}$ 表示, 它与用户 i 对页面 j 的访问次数 $N_{i,j}$ 、用户 i 对页面 j 的平均浏览时间 $t_{i,j}$ 以及页面 j 的大小 sb_j 有关。鉴于页面大小 sb_j 对用户的浏览时间有影响, 我们把访问时间对页面兴趣度的影响用 $T_{i,j} = \frac{t_{i,j}}{sb_j}$ 来表示, 而访问次数对页面兴趣度的影响用 $N_{i,j}$ 来表示。由于不知道 $T_{i,j}$ 和 $N_{i,j}$ 哪个影响因素更重要, 本文给这两个影响因素加上了权重, 可得出页面兴趣度的公式为:

$$Insterest_{i,j} = a \times T_{i,j} + b \times N_{i,j}. \quad (1)$$

考虑到 $T_{i,j}$ 和 $N_{i,j}$ 这两个值量纲不同, 本文采用除以最大值的方法来进行归一化处理, 公式分别为:

$$T'_{i,j} = \frac{T_{i,j}}{\max(T_{i,j})}, \quad (2)$$

$$N'_{i,j} = \frac{N_{i,j}}{\max(N_{i,j})}. \quad (3)$$

因而我们得到一个新的可操作的页面兴趣度的计算公式。

$$Insterest_{i,j} = a \times T'_{i,j} + b \times N'_{i,j}. \quad (4)$$

上式中的两个权重因子 a 、 b 可用主成份分析方法来获取。

1.2 用户页面兴趣度矩阵的构成

用户页面兴趣度矩阵是表示每个用户对不同页面的兴趣度的矩阵, 矩阵的行表示用户, 列表示页面(页面排列的顺序按照页面被访问次数的多少排列), 矩阵中的元素就是由公式(4)中计算出的 $Insterest_{i,j}$ 。该矩阵是一个 $m \times n$ 矩阵, m 表示用户的个数, n 表示页面的个数。

2 Slope One 算法在预测用户页面兴趣度中的应用

Daniel Lemire 教授在 2005 年提出的一种新型的类似于 Item-based 的协同过滤算法-Slope One 算法^[12]。算法的主要思想是利用群用户对一组项目评分的平均差值和目标用户对某些项目的评分来预测该用户对其他项目的评分。Slope One 算法中的群用户并不单纯的指“最近邻”

居”, 而是所有对某两个项目给过评分的用户集合。该算法简单, 易于实现。下面我们讨论如何利用 Slope One 算法预测用户页面兴趣度。

在已知用户页面兴趣度矩阵的情况下, 可以定义一个训练集 χ , 对于任意两个页面 i 和 j , 可以通过下式计算出页面 i 和页面 j 的平均差值 $dev_{j,i}$:

$$dev_{j,i} = \sum_{u \in S_{j,i}(\chi)} \frac{u_j - u_i}{\text{card}(S_{j,i}(\chi))}, \quad (5)$$

其中, $S_{j,i}(\chi)$ —同时有页面 i 和 j 兴趣度值的用户的集合; u_i —用户 u 对页面 i 的兴趣度; u_j —用户 u 对页面 j 的兴趣度; $\text{card}(S_{j,i}(\chi))$ —已同时获得页面 i 和 j 兴趣度值的用户个数。

通过计算 $dev_{j,i}$ 得到的矩阵是一个反对称矩阵, 该矩阵可以在有新页面出现时进行实时更新。

为了预测用户 u 对页面 j 的兴趣度 $P(u)_j$, 我们令: \bar{u} —训练集中用户 u 已有值(非零值)的页面兴趣度的平均值, $\text{card}(R_j)$ —训练集中用户 u 已有页面兴趣度值的页面与页面 j 可以计算平均偏差, $dev_{j,i}$ 的页面的数目, $S(u)$ —训练集 χ 中用户 u 已有兴趣度值的所有页面的集合。

根据 Slope One 算法的基本思想, 我们可以得到用户 u 对页面 j 兴趣度的预测值 $P(u)_j$ 为:

$$P(u)_j = \bar{u} + \frac{1}{\text{card}(R_j)} \sum_{i \in R_j} dev_{j,i}, \quad (6)$$

$$R_j = \{i | i \in S(u), i \neq j, \text{card}(S_{j,i}(\chi)) > 0\}.$$

一般来说, 由于网站中页面数量很多, 而每个用户一般都只访问过少量页面, 这使得整个页面兴趣度矩阵 R_0 非常稀疏。但矩阵稀疏性影响预测的准确度, 考虑到用户没有访问到的页面并不意味着他们对这些页面没有兴趣, 因此我们可以对此进行修正。为此, 我们首先把 R_0 的每一列中用户没有访问过的页面的兴趣度用这列中访问过此页面的所有用户对此页面兴趣度的平均值来代替, 亦即用 R_0 中每列非零元素的平均值来填充 R_0 中相应列的所有空缺值(零值), 进而构成一个稠密矩阵 R , 此时矩阵 R 中每列元素的平均值与矩阵 R_0 中相应列非零元素平均值相同。但实际上各用户对该列中没访问过页面的兴趣度并不都是相同的, 为了减少噪声, 我们在此基础上, 通过奇异值分解对稀疏矩阵 R_0 中的空缺项(现已用稀疏矩阵中的非零元素的平均值替代)作进一步修正。

3 奇异值分解 (SVD)

SVD^[13,p:344-354] 是一种矩阵分解技术, 它可以将一个 $m \times n$ 矩阵 R 分成三个矩阵 T, T, D ; T 和 D 分别是 $m \times m$ 和 $n \times n$ 的正交矩阵 ($TT' = I$, $DD' = I$), 并且:

$$E = TSD', \quad (7)$$

其中 D' 是 D 的转置矩阵, 且

$$S = \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (8)$$

S_1 是一个 $m \times m$ 对角矩阵, $S_1 = \text{diag}(\delta_1, \delta_2, \dots, \delta_m)$, S_1 中的对角元素按照顺序 $\delta_1 \geq \delta_2 \geq \dots \geq \delta_m > 0$ 排列, 数值 $\delta_1, \delta_2, \dots, \delta_m$ 连同 $\delta_{m+1} = 0, \delta_{m+2} = 0, \dots, \delta_n = 0$ 一起称做矩阵 R

的奇异值 (Singular Value)。恰当选择 k 值 ($k < m$): 保留 k 个大的奇异值, 过滤掉小的奇异值 (即用 0 替代小的奇异值) 并删除 S 中值为 0 的行和列, 得到一个新的矩阵 S_k , 它是一个 $k \times k$ 对角矩阵, 相应地把矩阵 T 、 D' 简化为 T_k 、 D'_k , 它们分别是 $m \times k$ 和 $k \times n$ 矩阵, 则重构矩阵 R_k 为:

$$R_k = T_k \times S_k \times D'_k. \quad (9)$$

$R_k \approx R$, R_k 是过滤掉了一定噪声的 $m \times n$ 矩阵, 即当 R 中数据含有噪声时, 这种过滤能减少噪声 [13,p:356], k 过大, 则不能很好的去除矩阵 R 中数据的噪声; 相反 k 太小, 则 R_k 不能保持矩阵 R 的主要特性。

最后, 根据矩阵 R_k 对矩阵 R_0 进行再次修正, 即将矩阵 R_0 中值为零的元素用 R_k 中对应位置的元素值替代, 得到稠密矩阵 R_c , 用它就可对用户页面兴趣度进行预测。

图 1 数据处理截图

4 实证分析

本文使用某网站网络日志中 2008 年 8 月份的数据进行分析。

4.1 计算稀疏矩阵

首先用 web log explorer 进行相应的数据处理, 如图 1 所示。

按公式 (4) 进行处理后得到的页面兴趣度的公式为:

$$Interest_{i,j} = 0.7071 \times \frac{t_{i,j}}{sb_j} + 0.8561 \times N_{i,j}. \quad (10)$$

进而得出用户页面兴趣度矩阵 R_0 , 其中用户 103 人, 页面 161 个。用户页面兴趣度矩阵 R_0 的部分结果见图 2。其中, 矩阵 R_0 的稀疏度为 $1 - \frac{882}{103 \times 161} = 0.9468$, 即 94.68% 的数据都是缺失的。

4.2 对 R_0 的第一次修正

用 R_0 里列中的非零元素平均值来填充相应列的所有空缺值 (零值), 得到矩阵 R , R 的部分结果如图 3 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	
2	u1	-0.308																				
3	u2	0.579																				
4	u3	0.258																				
5	u4	0.271	0.418																			
6	u5	0.277	1.431																			
7	u6	0.701	0.96	0.409	0.632	0.351	0.475	0.445	0.316	0.318	0.606											
8	u7	0.315	0.402	0.42																		
9	u8	0.076	0.65	0.769	0.307	0.407	0.445															
10	u9	0.901	0.966	1.141	0.788																	
11	u10	0.007																				
12	u11	0.418	0.886																			
13	u12	0.009																				
14	u13	0.45	0.392																			
15	u14	0.416																				
16	u15	1.043	0.428																			
17	u16	0.023	0.43																			
18	u17	0.377																				
19	u18	0.324	0.302																			
20	u19	0.54	0.42																			
21	u20	0.255	0.338																			
22	u21	0.349	0.309																			
23	u22	0.377	0.629																			
24	u23	0.602	0.524	0.634	0.352																	
25	u24	0.352	0.398	0.42																		
26	u25	0.581	0.39	0.704																		
27	u26	0.566	1.422																			
28	u27	0.327	0.495																			
29	u28	0.77																				

图 2 用户页面兴趣度矩阵 R_0 (截图)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	
2	u1	0.689	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.					
3	u2	0.579	0.689	0.42144	0.7	0.547	0.473	0.503	0.537	0.476	0.49777	0.444	0.475	0.828	0.439	0.653	0.435	0.				
4	u3	0.266	0.689	0.791	0.60427	0.547	0.473	0.503	0.537	0.476	0.427791	0.444	0.475	0.828	0.439	0.653	0.435	0.				
5	u4	0.271	0.416204	0.791	0.407702	0.501057	0.473	0.781427	0.424007	0.476	0.548	0.444	0.392005	0.828	0.3431	0.653	0.435	0.				
6	u5	0.277	1.431254	0.791	0.7	0.300134	0.473	0.444605	0.592410	0.476	0.548	0.444	0.373199	0.828	0.439	0.653	0.435	0.				
7	u6	0.701	0.260382	0.4075	0.632285	0.551499	0.4748	0.444652	0.310224	0.310224	0.000214	0.446	0.932016	1.7005	0.439	0.653	0.300320	0.				
8	u7	0.315	0.401007	0.4906	0.7	0.565174	0.37240	0.503	0.527	0.476	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.			
9	u8	0.072	0.950127	0.791	0.769091	0.307220	0.42747	0.444662	0.571	0.300005	0.548	0.444	0.46047	0.828	0.439	0.666019	0.301139	0.61				
10	u9	0.901	0.0565323	1.4141	0.707409	0.547	0.473	0.503	0.537	0.476	0.399093	0.444	0.475	0.309069	0.37556	0.653	0.435	0.				
11	u10	0.007	0.689	0.791	0.36425	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.47471	0.828	0.439	0.653	0.435	0.				
12	u11	0.415	0.669966	0.791	0.7	0.720002	0.59179	0.503	0.537	0.476	0.548	0.444	0.475	0.363024	0.41602	0.653	0.435	0.				
13	u12	0.249	0.6379	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.415				
14	u13	0.45	0.35187	0.791	0.7	0.351874	0.58744	0.394127	0.537	0.476	0.548	0.444	0.552455	0.828	0.439	0.653	0.435	0.				
15	u14	0.419	0.689	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				
16	u15	1.043	0.425262	0.791	0.924562	0.34377	0.4189	0.985867	0.617941	0.476	0.476	0.476	0.476	0.476	0.476	0.476	0.476	0.476	0.476	0.476	0.	
17	u16	0.383	0.689	0.43036	0.7	0.547	0.473	0.503	0.537	0.476	0.373494	0.476	0.548	0.828	0.439	0.653	0.435	0.				
18	u17	0.277	0.689	0.791	0.63425	0.547	0.473	0.503	0.537	0.476	0.382291	0.444	0.475	0.828	0.439	0.653	0.435	0.				
19	u18	0.324	0.361832	0.791	0.7	0.287403	0.54887	0.503	0.537	0.476	0.444	0.475	0.828	0.439	0.653	0.435	0.					
20	u19	0.94	0.689	0.43036	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.423417	0.439	0.653	0.435	0.				
21	u20	0.256	0.357913	0.791	0.7	0.581014	0.59529	0.503	0.537	0.397888	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.54				
22	u21	0.349	0.30988	0.791	0.7	0.301847	0.328008	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.460182	0.				
23	u22	0.377	0.689	0.791	0.7	0.492316	0.43747	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.454				
24	u23	0.532	0.55795	0.63398	0.547	0.473	0.537	0.537	0.614612	0.548	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.	
25	u24	0.382	0.394755	0.43036	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				
26	u25	0.591	0.330425	0.70988	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				
27	u26	0.566	1.425137	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				
28	u27	0.367	0.494934	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				
29	u28	0.77	0.689	0.791	0.7	0.547	0.473	0.503	0.537	0.476	0.548	0.444	0.475	0.828	0.439	0.653	0.435	0.				

4.3 对矩阵 R 进行奇异值分解

为了减少噪声, 我们用 MATLAB 软件对填充后的矩阵 R 进行奇异值分解。分解后, 得到: T 为 103×103 矩阵, S 为 103×161 矩阵, D 为 161×161 方阵。 S 矩阵的部分结果如图 4 所示。

4.4 k 的选取及对矩阵 R_0 的再次修正

根据公式(8), 矩阵 S_1 是一个 103×103 的对角矩阵, 对角线元素是奇异值。在奇异值分解中, 需要对奇异值进行截取, 即要恰当地选择 k ($k < m$), 并根据这个 k 值由公式(9)得到 R_k 。针对本数据集, 我们分别求出当 k 为 10、15、20、25、30 时的重构矩阵 R_k , 进而对矩阵

R_0 进行再次修正得到相应的稠密矩阵 R_c , 并根据 R_c 用下面 5.5 部分所述的 5-fold 交叉验证法计算相应 k 值时的平均绝对误差, 结果如表 1 所示。

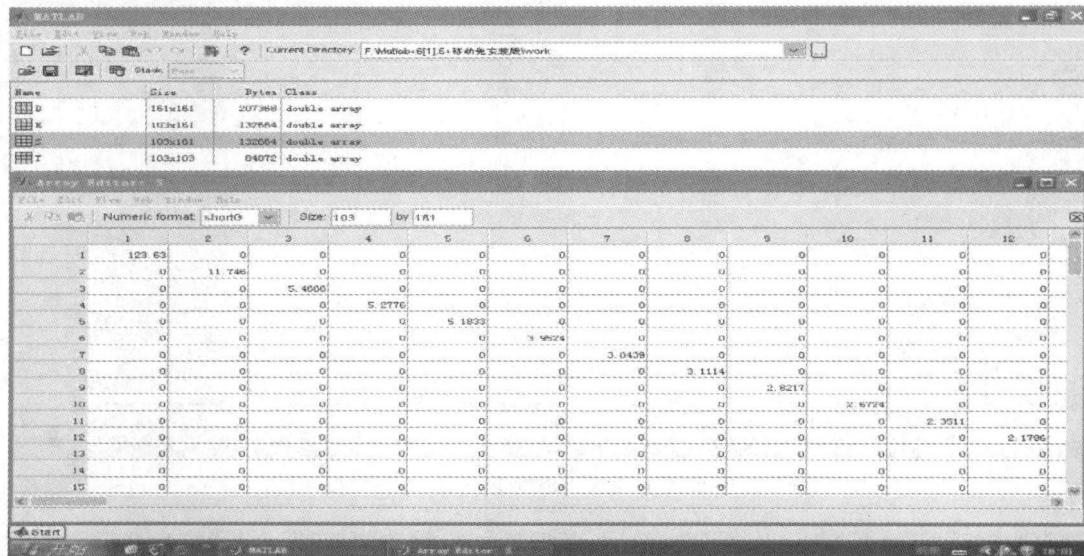


图 4 奇异值分解后的 S 矩阵(截图)

表 1 不同 k 值时预测结果的平均绝对误差

组号	训练集 页面 序号	验证集 页面 序号	$k = 10$ 时的平均 绝对误差	$k = 15$ 时的平均 绝对误差	$k = 20$ 时的平均 绝对误差	$k = 25$ 时的平均 绝对误差	$k = 30$ 时的平均 绝对误差
1	33-161	1-32	0.2850	0.28509	0.28594	0.28626	0.28651
2	1-32, 65-161	33-64	0.3740	0.3741	0.3739	0.3729	0.3729
3	1-64, 97-161	65-96	0.3383	0.3380	0.3384	0.3376	0.3380
4	1-96, 129-161	97-128	0.12701	0.12641	0.12533	0.12532	0.12554
5	1-128	129-161	0.02690	0.02586	0.02583	0.02540	0.02560
总的平均绝对误差			0.23024	0.22989	0.22988	0.22950	0.22971

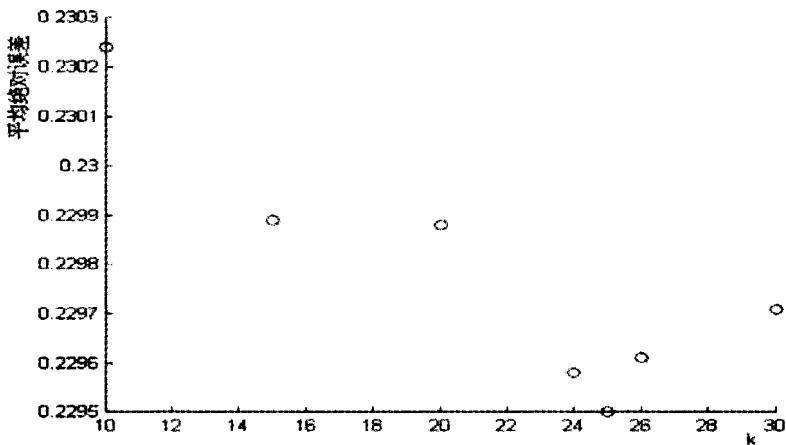
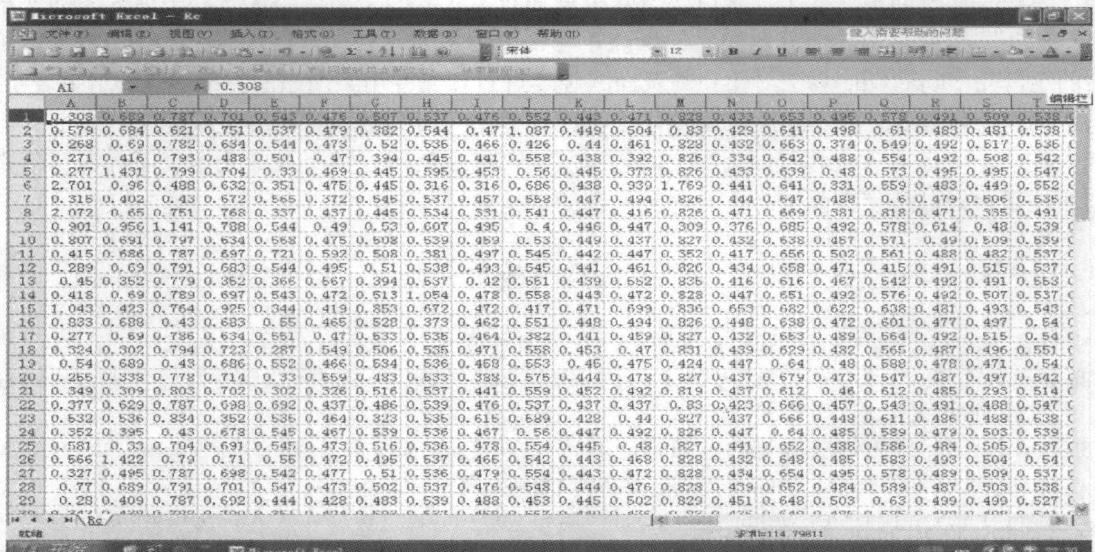
显然 k 为 25 时总的平均绝对误差最小, 为了检验这一结果的正确性, 本文又分别对 k 为 24 和 26 的平均绝对误差进行了计算, 结果如表 2 所示。

表 2 $k = 24, 26$ 时预测结果的平均绝对误差

组号	训练集页面序号	验证集页面序号	$k = 24$ 时的平均绝对误差	$k = 26$ 时的平均绝对误差
1	33-161	1-32	0.28624	0.28633
2	1-32, 65-161	33-64	0.3731	0.3729
3	1-64, 97-161	65-96	0.3377	0.3378
4	1-96, 129-161	97-128	0.1254	0.12545
5	1-128	129-161	0.02546	0.02546
总的平均绝对误差			0.22958	0.22961

由表 1 和表 2 的数据得到 k 与总的平均绝对误差的关系图, 如图 5 所示。

从表 1、表 2 和图 5 中均可看出, 对本数据集, $k = 25$ 的结果确实是最佳的, 其对应的稠密矩阵为 R_c , 如图 6 所示。

图 5 k 与平均绝对误差的关系图 6 $k = 25$ 时的稠密矩阵 R_c (截图)

4.5 5-fold 交叉验证

采用 5-fold 交叉验证法 (将数据集分成五份, 轮流将其中 4 份做训练集 1 份做验证集, 5 次的结果的均值作为对算法精度的估计) 对 Slope one 算法预测出的稀疏矩阵 R_0 和稠密矩阵 R_c (与 $k = 25$ 相对应) 的页面兴趣度的结果, 通过平均绝对误差 (Mean Absolute Error, 简称 MAE) 值的大小来进行比较。MAE 的表达式如下:

$$MAE = \frac{\sum_{u,i} |r_{(u,i)} - \hat{r}_{(u,i)}|}{N}, \quad (11)$$

其中, $r_{(u,i)}$ 是通过公式 (6) 计算得出的用户 u 对页面 i 的兴趣度, $\hat{r}_{(u,i)}$ 是用户 u 对页面 i 的实际兴趣度, N 为验证集中用户 u 有实际兴趣度值的项的个数。结果如表 3 所示。

由表 3 可知, 相对于用原稀疏矩阵的进行预测的总的平均绝对误差, 用修正后的矩阵进行预测时总的平均绝对误差减少了 20.9%, 即预测的准确性有很大的提高。

表3 不同训练集和验证集取值范围及平均绝对误差

组号	训练集页面序号	验证集页面序号	MAE_{R_0}	MAE_{R_c}
1	33~161	1~32	0.43833	0.28626
2	1~32, 65~161	33~64	0.4219	0.3729
3	1~64, 97~161	65~96	0.425	0.3376
4	1~96, 129~161	97~128	0.1533	0.12532
5	1~128	129~161	0.012	0.02540
总的平均绝对误差			0.29011	0.2295

5 结论

本文提出一个新的易于操作的用户页面兴趣度的计算公式, 根据这个公式得到用户页面兴趣度矩阵, 但该矩阵通常是稀疏矩阵, 这种稀疏性使用户对某些页面兴趣度的预测不够准确。为解决这一问题, 本文首先用该矩阵中每列非零元素的平均值替代此列中值为零的元素进行修正, 然后用奇异值分解法对原稀疏矩阵中值为零的元素进行再次修正, 进而根据修正后的矩阵用 Slope One 算法对用户的页面兴趣度进行预测并用某网站的数据验证。结果表明: 与原稀疏矩阵相比, 用 Slope One 算法根据两次修正后得到的稠密矩阵在预测用户页面兴趣度的准确性方面有较大提高。因而, 以此为基础可根据新用户某个或某几个页面兴趣度值来预测该用户对其他页面的兴趣度并作为推荐的依据是可行的。

[参考文献]

- [1] 曾春, 邢春晓, 周立柱. 个性化服务技术综述 [J]. 软件学报, 2002, 13(10): 1952~1961.
- [2] 孙铁利, 杨凤芹. 根据用户隐式反馈建立和更新用户兴趣模型 [J]. 东北大学学报(自然科学版), 2003, 35(3): 99~104.
- [3] Claypool M, Le P, Waseda M, et al. Implicit interest indicators [A]. In: Proceedings of the ACM Intelligent User Interfaces Conf(IUI) [C]. ACM Press, 2001: 33~41.
- [4] Liang T P, Lai H J. Discovering user internet from web browsing behavior: An application to internet news services [A]. In: Proceedings of the 35th Hawaii International Conference on System Sciences [C]. IEEE Computer Society Press, 2002: 203~212.
- [5] 谭琼, 李晓黎, 史忠植. 一种实现搜索引擎个性化服务的方法 [J]. 计算机科学, 2002, 29(1): 23~25.
- [6] 吕佳. 基于兴趣度的 Web 用户访问模式分析 [J]. 计算机工程与设计, 2007, (5): 2403~2404.
- [7] 李珊, 袁方. 基于 Web 日志挖掘的页面兴趣度方法的改进 [J]. 计算机时代, 2007, (3): 29~31.
- [8] 张新香. WEB 日志挖掘在电子商务中的应用研究 [J]. 计算机系统应用, 2006, (1): 52~55.
- [9] 张莹. 从电子商务网站用户行为数据提取用户兴趣 [J]. 潍坊学报, 2005, 5(4): 1~23.
- [10] 周晓兰, 王随平. WEB 文本挖掘中用户兴趣模型的建立和更新 [J]. 湘潭师范学院学报(自然科学版), 2006, 28(3): 33~36.
- [11] 田东风, 欧飞, 申维. 矩阵奇异值分解理论在中文文本分类中的应用 [J]. 数学的实践与认识, 2008, 38(24): 132~139.
- [12] Daniel Lemire, Anna Maclachlan. Slope one predictors for online rating-based collaborative filtering [A]. In: Proceedings of the 2005 SIAM International Conference on Data Mining [C]. SIAM Press, 2005: 471~475.
- [13] 张贤达. 矩阵分析与应用 [M]. 北京: 清华大学出版社, 2004.