|  |  |
| --- | --- |
| 1，模式挖掘 | **完成**  **成果**：  形成一种新的基于模式挖掘的评分填充算法。并在填充后的矩阵的基础上提出了改进的相似度计算方法。  **新问题**：  由于原始矩阵的稀疏，传统的计算相似度的复杂度比较小。  如果直接在填充后的相似度矩阵上使用传统的在线计算，则**复杂度大大增加。**  **解决办法**：填充后的矩阵，对于填充的部分使用**离线计算，（填充评分属于代替评分，由于无法做到太高的精确性，因此不需要太高的实时性，离线计算比较合适）** |
| 2，（模糊）聚类 | 在论文中加入聚类的方法（创新点，k-means算法初始选中心点的方法） |
| Idea1  1,选取打分普遍高的用户Uo（积极用户），打分普遍低的用户Up（消极用户），计算出中立用户Un（是计算出来的，实际上不存在）。得到K-means算法初始的三个中心点。  2，执行k-means算法，最终收敛后，得出三个分组。  3，执行模糊k-means算法，得出三个模糊分组  4，计算MAE |
| Idea2  1，计算每个用户的熵，根据熵的分布，选择初始的几个用户作为中心点。  2，进行k-means聚类。  3，计算MAE |
| 3，用户分组 |  |
| 4，信息熵（暂不研究） | 参考：基于加权信息熵相似度的协同过滤算法研究--刘文龙 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

记录结果SML1M

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PCC | HybirdSim3 | HybirdSim3  +  K-means |  |  |  |  |  |  |
| 5 | 0.90834 | 0.89441 | 0.91066 |  |  |  |  |  |  |
| 10 | 0.89563 | 0.88030 | 0.90367 |  |  |  |  |  |  |
| 15 | 0.89247 | 0.87718 | 0.89795 |  |  |  |  |  |  |
| 20 | 0.89157 | 0.87633 | 0.89785 |  |  |  |  |  |  |
| 25 | 0.89115 | 0.87601 | 0.89406 |  |  |  |  |  |  |
| 30 | 0.89112 | 0.87593 | 0.90019 |  |  |  |  |  |  |
| 35 | 0.89116 | 0.87612 | 0.89176 |  |  |  |  |  |  |
| 40 | 0.89122 | 0.87640 | 0.89809 |  |  |  |  |  |  |
| 45 | 0.89125 | 0.87648 | 0.90622 |  |  |  |  |  |  |
| 50 | 0.89127 | 0.87653 | 0.89998 |  |  |  |  |  |  |

ML100K

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PCC | HybirdSim3 | HybirdSim3  +  K-means |  |  |  |  |  |  |
| 5 | 0.81401 | 0.89441 | 0.91066 |  |  |  |  |  |  |
| 10 | 0.78190 | 0.88030 | 0.90367 |  |  |  |  |  |  |
| 15 | 0.77053 | 0.87718 | 0.89795 |  |  |  |  |  |  |
| 20 | 0.76463 | 0.87633 | 0.89785 |  |  |  |  |  |  |
| 25 | 0.76180 | 0.87601 | 0.89406 |  |  |  |  |  |  |
| 30 | 0.76045 | 0.87593 | 0.90019 |  |  |  |  |  |  |
| 35 | 0.75921 | 0.87612 | 0.89176 |  |  |  |  |  |  |
| 40 | 0.75847 | 0.87640 | 0.89809 |  |  |  |  |  |  |
| 45 | 0.75798 | 0.87648 | 0.90622 |  |  |  |  |  |  |
| 50 | 0.75774 | 0.87653 | 0.89998 |  |  |  |  |  |  |