第四章算法除特殊说明外使用的数据集为

新闻数据集部分抽样：https://msnews.github.io/

|  |  |  |
| --- | --- | --- |
| 算法名 | 对应文件 | 结果说明 |
| 基于词向量空间模型的文本表示（TF-IDF模型） | tfidfCF.py | precision=0.0226  recall=0.0089  coverage=0.5537  利用tfidf的内容相似度进行协同过滤，效果比朴素item CF要好 |
| 朴素item\_cf | itemCF\_news.py | precision=0.0117  recall=0.0046  coverage=0.2476  朴素item CF没有利用新闻文本的信息，效果较差 |
| 基于语料库的文本相似度（PMI-IR） | PMI\_IR\_CF.py | precision=0.0098  recall=0.0039  coverage=0.1886  基于语料库的文本相似度需要较充足的语料，取样数据较少，效果不佳 |
| 基于约束的推荐 | 基于约束的推荐——MinRelax.ipynb | 示例代码，无评估  数据：笔记本集合.csv |
| 基于效用的推荐 | 基于效用的推荐.ipynb | 示例代码，无评估  数据：笔记本集合.csv |
| 基于实例的推荐 | 基于实例的推荐.ipynb | 示例代码，无评估  数据：笔记本集合.csv |