**1. 协同过滤的简介**

   关于协同过滤的一个最经典的例子就是看电影，有时候不知道哪一部电影是我们喜欢的或者评分比较高的，那

   么通常的做法就是问问周围的朋友，看看最近有什么好的电影推荐。在问的时候，都习惯于问跟自己口味差不

   多的朋友，这就是协同过滤的核心思想。

   协同过滤是在海量数据中挖掘出小部分与你品味类似的用户，在协同过滤中，这些用户成为邻居，然后根据他

   们喜欢的东西组织成一个排序的目录推荐给你。所以就有如下两个核心问题

   （1）如何确定一个用户是否与你有相似的品味？

   （2）如何将邻居们的喜好组织成一个排序目录？

   协同过滤算法的出现标志着推荐系统的产生，协同过滤算法包括基于用户和基于物品的协同过滤算法。

**2. 协同过滤的核心**

   要实现协同过滤，需要进行如下几个步骤

   （1）收集用户偏好

   （2）找到相似的用户或者物品

   （3）计算并推荐

实例：最简单的在线电影推荐系统

根据电影的评分来推荐，特征只有一个：评分

**3. 数据来源**

数据集下载ml-latest-small(1MB): <http://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

**4. 算法流程**：

1 找到目标用户看过的每一部电影

2 目标用户看过的每一部电影， 找到和目标用户看过相同电影的初始邻居

3 遍历初始邻居，找到初始邻居看过的所有电影

4 计算目标用户 和 所有初始邻居的余弦距离，按相似度由高到低建立邻居用户倒排表

5 遍历邻居用户倒排表，对邻居用户看过的电影，计算累积余弦距离，建立电影倒排表

6 按电影倒排表，输出前20个电影名字