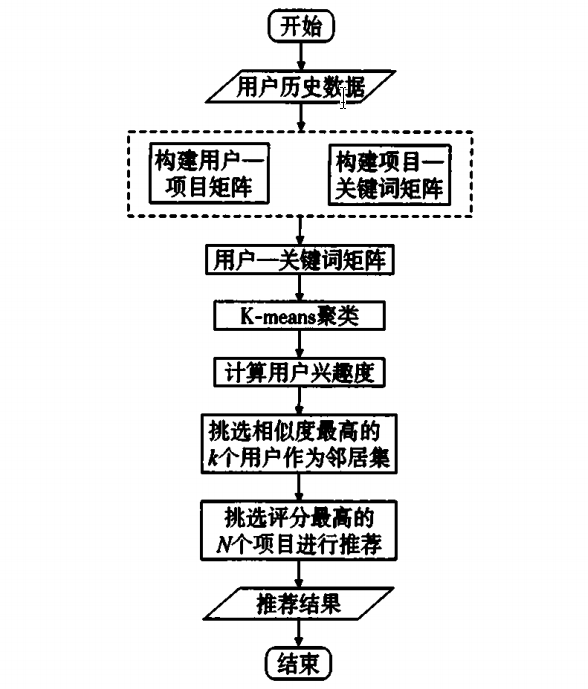
1、RF-IIF算法流程图：



2、数据结构

①项目-关键词矩阵



每一行是项目i的关键词属性集合。

每一列是关键词属性j在每个项目中的集合。

每一个元素是项目i中关键词属性j的值，取1或0，值为1表示项目j具有该关键词属性i，值为0表示该项目i不具有关键词属性j。

②用户-项目评分矩阵



每一行是用户i评价过的电影集合。

每一列是评价过电影j的用户集合。

每一个元素是用户i对电影j的评分，通常取1~5整数，数据越大用户对该项目越满意。

③用户-项目兴趣度矩阵

用logistic函数得到用户-项目兴趣度，用户i对项目j感兴趣程度：取值为（0,1）



其中：为用户i对项目j的评分次数，为用户i对所有项目的平均评分次数。

用户i的兴趣度向量：

所有用户兴趣度向量构成用户-项目兴趣度矩阵。

④用户-关键词评分矩阵

公式如下：



其中：R为用户-项目评分矩阵，K为项目-关键词矩阵。

⑤用户-关键词偏好矩阵

评分频率：某用户对某关键词的评分频率



其中：为用户-关键词评分矩阵中用户i对关键词j的评分次数，T为矩阵中所有关键词的集合，为用户评分总数之和。

反向项目频率：评分该关键词的用户总数



n为用户-关键词评分矩阵中的用户总数，为用户-关键词矩阵中评分过该关键词的用户个数。

用户对关键词偏好度：



用户的偏好向量：

所有用户偏好向量构成用户-关键词偏好矩阵：



⑥用户相似度

用k-mean算法进行聚类，划分用户类簇，采用用户a和b的偏好向量的余弦距离来计算相似度：



⑦最近邻居集

在类簇中寻找与用户相似的Top-N邻居集：



其中：S为用户a和b共同感兴趣项目的集合，和分别表示用户a和用户b对项目j的兴趣度，和分别表示用户a和b对集合i中项目的平均兴趣度。

⑧预测评分

根据Top-N邻居集、用户-关键词评分矩阵和用户-项目兴趣度矩阵，得出用户a对项目q的评分预测值。



其中：为用户a评分过的所有关键词的评分兴趣度，为用户b对项目q的兴趣度。

3、度量方法

①平均绝对误差



其中：系统预测的用户兴趣集合：

用户实际兴趣集合：

②准确率



其中：N为推荐系统产生的推荐总数，C为系统产生的正确推荐总数。