

权利要求书

1、一种基于 TextRank 的应用偏好文本分类方法，其特征在于，包括如下步骤：

S1、根据 TextRank 算法，生成每个应用的关键词字段，构成第一关键词库；

S2、根据多个二级分类，为每个二级分类标记一个种子关键词；

S3、根据种子关键词，在第一关键词库中模糊检索包含所述种子关键词的应用，并将所述包含种子关键词的应用打上二级分类；

S4、再次使用 TextRank 算法，对所有二级分类下的所有应用的种子关键词进行全量计算，生成所述多个二级分类下的第二关键词库；

S5、再次遍历应用表，对每一个关键词字段中的内容与第二关键词库进行字符串相似度匹配，如果相似度低于预设阈值，则删除所述关键词字段对应的应用与当前二级分类之间的关联。

2、根据权利要求 1 所述的一种基于 TextRank 的应用偏好文本分类方法，其特征在于，

所述多个二级分类为应用分类领域公认的 75 个分类。

3、根据权利要求 1 所述的一种基于 TextRank 的应用偏好文本分类方法，其特征在于，

所述预设阈值为 70%或 75%。

4、根据权利要求 1 所述的一种基于 TextRank 的应用偏好文本分类方法，其特征在于，所述方法进一步包括：

S6、遍历完所述应用表后，重新生成第二关键词库，重复步骤 S1-S5。

5、根据权利要求 4 所述的一种基于 TextRank 的应用偏好文本分类方法，其特征在于，所述方法进一步包括：

S7、根据最终的生成结果，人工抽查准确度情况，如果效果不理想，继续再次迭代步骤 S1-S5。

6、一种电子设备，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，其特征在于，所述处理器运行所述计算机程序时执行以实现如权利要求 1-5 任一项所述的方法。

7、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述程序被处理器执行时实现如权利要求 1-5 中任一项所述的方法。