

DOI: 10.11992/tis.201711036

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180408.0930.002.html>

## 面对智能导诊的个性化推荐算法

马钰, 张岩, 王宏志, 张义策

(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

**摘 要:** 针对电子医疗信息过载和医疗资源严重不足的问题, 本文以辅助诊疗的结果为基础, 将 Skyline 查询和局部范围内基于协同过滤的评分方式有机结合, 提出了一种面向智能导诊的个性化推荐算法。实验结果表明, 本文提出的算法能为用户提供个性化的合理推荐结果。该方法对合理分配和使用医疗资源有很大的促进作用, 能从一定程度上缓解就诊压力, 提高就诊质量, 具有重要的实用价值和社会意义。

**关键词:** 辅助诊疗; 智能导诊; Skyline 查询; 医疗推荐; 协同过滤; 推荐系统; 大数据; 个性化

**中图分类号:** TP311 **文献标志码:** A **文章编号:** 1673-4785(2018)03-0352-07

中文引用格式: 马钰, 张岩, 王宏志, 等. 面对智能导诊的个性化推荐算法[J]. 智能系统学报, 2018, 13(3): 352-358.

英文引用格式: MA Yu, ZHANG Yan, WANG Hongzhi, et al. A personalized recommendation algorithm for intelligent guidance [J]. CAAI transactions on intelligent systems, 2018, 13(3): 352-358.

## A personalized recommendation algorithm for intelligent guidance

MA Yu, ZHANG Yan, WANG Hongzhi, ZHANG Yice

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**Abstract:** Because of the overload in the electronic medical data and serious shortage of medical resources, we combined Skyline queries and the scoring method based on collaborative filtering in local areas and proposed a recommendation algorithm for intelligent personalized guidance. Our experimental results show that the proposed algorithm can provide users with personalized and reasonable recommendations. Thus, the proposed algorithm was found to considerable promote rational allocation and the use of medical resources. Additionally, our algorithm improved the quality of medical treatments to a certain extent.

**Keywords:** auxiliary diagnosis; intelligent guide; skyline query; medical recommendation; collaborative filtering; recommendation system; big data; personality

中国人口数量与医疗资源之间的巨大反差使得医疗资源日趋不足, 短期内增加医疗资源的总量几乎是不可能的, 因此有效整合和合理分配现有的医疗资源, 缓解就诊压力, 提升就医质量, 有着很大的实用价值与社会意义。

智能导诊根据患者的主诉为患者自动推荐相应的医院和医生, 对合理分诊起着重要作用。从数据的角度看, 智能导诊是针对患者主诉和基本信息对医院和医生的推荐。

推荐技术与医疗导诊的结合意义重大, 但是传统的推荐技术在导诊中的应用并不好, 因为导诊和每个用户的特征息息相关: 由于地理位置等因素的影响, 不同用户在选择医院时差别很大, 即使同一个用户, 在所患疾病不同时也会去不同的医院就诊。因此, 针对用户和疾病的不同推荐与导诊的结果也应该随之变化, 这样的推荐才是有意义的。也就是说, 医疗导诊领域的推荐必须体现出用户和疾病的特征, 以个性化推荐为主。因此本文提出了一种面向智能导诊的个性化推荐算法。

### 1 相关工作及分析

近年来国内外与医疗领域推荐相关的研究很

收稿日期: 2017-11-29. 网络出版日期: 2018-04-08.

基金项目: 国家科技支撑计划项目 (2015BAH10F01); 国家自然科学基金项目 (U1509216, 61472099, 61602129); 国家重点研发计划项目 (2016YFB1000703); 黑龙江省留学回国人员基金项目 (LC2016026).

通信作者: 王宏志. E-mail: [wangzh@hit.edu.cn](mailto:wangzh@hit.edu.cn).

多, M. López-Nores 等<sup>[1]</sup>引入了一种新的过滤策略, 对于有特定疾病的患者, 这种推荐方法的效率较高。P. Pattaraintakorn 等<sup>[2]</sup>提出了一种使用粗糙集和规则分析的医疗推荐系统, 主要目标是将病人的自身的身体数据作为条件属性, 推荐临床检查方式。宫继兵等<sup>[3]</sup>提出了一种医疗社交网络中的多层混合医生推荐结构, 通过挖掘网络中医生与病人之间的关系为患者推荐医生。徐守坤等<sup>[4]</sup>提出过一种医生资源均衡推荐算法, 使用最佳结果优先的推荐算法, 解决了医疗资源的使用过度集中而产生的问题。

由上述分析可以看出, 现行系统主要有两方面的问题: 1) 没有考虑现在医疗推荐方面最迫切需要解决的问题; 2) 没有考虑导诊与医疗推荐和传统的推荐在本质上的不同。

针对这些问题, 本文提出了一种面向智能导诊的个性化推荐算法, 其主要有 3 个特点:

1) 算法提出了辅助诊疗机制, 通过辅助诊疗能根据用户的症状表现, 帮助用户大致确定其所患疾病, 从而明确用户偏好, 提升推荐结果的满意度。

2) 算法获得用户评分及反馈信息的方式以显式方法为主, 并且在辅助诊疗时根据用户的反馈信息对诊断结果进行优化, 从而提高诊断结果的准确性。

3) 提出了一种在有地域约束时的推荐算法。对 Skyline<sup>[5-10]</sup>查询进行改进, 并将其结果作为推荐的候选集合, 之后在局部范围内使用基于协同过滤的评分算法, 对候选集合进行评分并排序。

## 2 整体框架

为了对本文所述系统有一个全局性的了解进而更好地了解算法细节, 本部分将以框架图为基础, 阐述算法中各部分的作用及大致思路。本文的整体框架如图 1 所示。

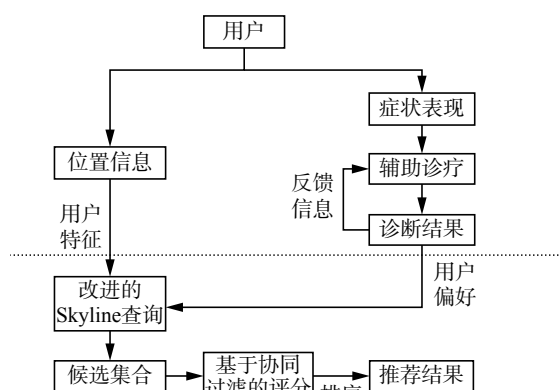


图 1 整体框架

Fig. 1 The overall framework

在图 1 中, 虚线将框架图分为两大部分: 第 1 部分是获取用户偏好的过程, 首先根据用户的症状表现, 使用辅助诊疗算法给出初步诊断结果, 再利用反馈信息和症状联想机制进一步确定所患疾病, 从而完成症状到用户偏好的转化过程; 第 2 部分是生成推荐结果的过程, 由用户的偏好结合其地理位置信息, 采用改进的 Skyline 查询算法, 找出一定范围内符合要求的医院作为候选集合, 最后在局部范围内由基于协同过滤的评分方式对候选集合中的医院进行评分、排序, 从而生成最终的推荐结果。

## 3 基于医疗知识库的辅助诊疗技术

在本节中, 我们主要介绍辅助诊疗技术的具体细节, 用以解决医疗导诊与推荐时用户偏好未知的问题, 通过辅助诊疗帮助用户根据症状表现确定其所患疾病, 从而明确用户偏好。

### 3.1 分词与症状索引表的建立

为了提高分词速度, 本文通过对《同义词林》进行适当的改造, 构建了一个新的字典。首先, 部分无用词汇被剔除; 其次, 按照词语的字数, 字典被分为 5 个部分, 这样每次匹配时词典最多被扫描一次。使用改造后的词典, 诊断时症状描述中遇到的同义词和相关词等问题得到了一定程度的缓解。为了提高分词准确度, 本文选用了逆向最大匹配法。为了加快诊断速度, 本文采取了在症状信息上建立倒排索引的方法。该索引包含 3 个属性, 分别保存症状分词结果对应的同义词林中的编码、疾病 ICD 编码和疾病中症状出现的次数, 以及症状对应的疾病类型数。后两个属性在进行诊断时使用。

### 3.2 辅助诊断技术

#### 3.2.1 辅助诊疗思路

辅助诊疗技术采用的核心思想是计算用户输入的症状信息与医疗知识库中疾病症状信息之间的相似度, 从而确诊疾病类型, 具体实现过程包括:

1) 对录入的症状信息进行分词, 使用分词结果在索引表中进行查询, 记录所有匹配项, 将对应的疾病作为候选集, 并将其中的疾病总数记为  $N$ 。

2) 利用症状索引表计算输入与候选集中每种疾病症状的相似度, 并依据相似度进行排序。

#### 3.2.2 相似度计算

相似度计算公式借鉴了 TF-IDF 计算相似度的思路<sup>[11]</sup>, 其公式为

$$S = \sum_{i \in q} \left( \frac{\text{num}_i}{\text{num}_{\text{symptom}}} \log_2 \left( \frac{N}{\text{type}_{\text{disease}} + 1} \right) \right) \quad (1)$$

式中:  $S$  为输入  $q$  与疾病症状之间的相似度;  $\text{num}_i$  表示  $q$  中的某一症状  $i$  在该疾病中出现的次数, 当

$q$  值较大时, 意味着此症状为潜在的主要症状, 则提高该症状的权重;  $\text{num}_{\text{symptom}}$  表示疾病的总症状数, 对  $\text{num}_i$  进行归一化, 以防止部分疾病系数偏大;  $\text{type}_{\text{disease}}$  表示症状对应的疾病类型数量, 对数函数则是为了提高潜在主要症状的权重。

### 3.2.3 症状联想机制

算法在用户输入症状信息和得到诊断结果之后均引入症状联想机制, 以此来加强与用户的交互, 提升辅助诊断的准确性。

#### 1) 用户输入时的症状联想

基于分词与症状索引表, 统计症状两两之间共同出现的次数, 并将与该症状共同出现次数前 5 的症状存入症状索引表作为联想症状, 在用户输入完一个症状之后, 将联想症状提供给用户, 用户可以选择与自身相关的症状, 以此来帮助用户提升输入症状的完整性。

#### 2) 诊断之后的症状联想

首先计算疾病中每个症状在该疾病中的权值, 然后选择权重前 3 的症状作为联想症状保存, 在得到诊断结果之后将联想症状提供给用户, 若出现联想症状说明患该种疾病的可能性较高, 从而帮助用户进一步确定所患疾病, 症状在疾病中的权重计算方法与上述相似度计算公式类似, 只不过  $N$  为疾病总数。

辅助诊断部分使用的数据只是疾病与其相关症状信息, 相比于通用检索系统其数据量很小; 而且查询是由症状确定疾病这一种固定模式, 其输入与输出模式不变, 因此引入症状联想机制是合理的, 而且能有效加强系统和用户的交互, 提升推荐结果的满意度。

### 3.3 诊断结果的优化

用户往往会对自己可能患有的疾病进行猜想, 这是非常有价值的信息。系统将考虑来自用户的反馈信息, 对诊断结果进行优化。具体实现过程如下:

- 1) 将反馈疾病的症状加入查询项;
- 2) 增加查询项中已有的症状的权重。

具体而言, 反馈信息与原症状信息的权重比  $\alpha$  的计算过程为

$$\alpha = 0.5 \left( \frac{x}{\max} + 1 \right) \quad (2)$$

式中:  $x$  为某反馈项在诊断结果中的位置,  $\max$  为反馈项在诊断结果中位置的最大值。

## 4 基于地域信息的推荐技术

本节介绍了基于地域信息的推荐技术, 该算法

通过对 Skyline 查询进行改进并将其结果作为推荐结果的候选集合, 而将传统的协同过滤推荐算法加以改进作为局部范围内的一种评分方式, 有效地应对地域约束严格带来的挑战。

### 4.1 kd-tree 索引结构

地理位置信息由经度和纬度组成, 是一个天然的二维数据, 而 kd-tree 常被用来对多维数据结构进行划分, 在 kd-tree 上对指定的点搜索其一定范围内的邻居节点效率很高<sup>[12]</sup>。

因此, 针对医院的经纬度信息, 将所有的医院以 kd-tree 的结构组织起来, 每个节点保存医院的经纬度以及医院在数据库中对应的编号。在后面的推荐算法中, 查询指定点一定范围内的医院时就可以在该 kd-tree 上进行, 返回医院的 idhospital 集合, 而根据 idhospital 返回医院的其他信息的时间可忽略不计。

### 4.2 改进的 Skyline 查询算法

医疗推荐可以抽象为一个多目标优化问题, 可以通过将 Skyline 查询引入到推荐算法中解决。

Skyline 查询的目的是找到不被其他点支配的点集合作为 Skyline 集合, 本文中支配的定义如下。

支配: 一个医院节点主要考虑距离和评分两方面, 因此可以抽象地表示为  $h = \langle s, d \rangle$ , 其中  $s$  表示该医院治疗某种疾病的评分,  $d$  表示该医院与该用户的距离; 若  $h_1$  支配  $h_2$ , 则  $(h_1.s \geq h_2.s \text{ 且 } h_1.d \leq h_2.d)$  并且  $(h_1.s > h_2.s \text{ 或 } h_1.d < h_2.d)$  为真。

该算法是针对传统 Skyline 查询中结果集合小于推荐结果集的最小阈值而提出的, 此时推荐结果候选集合由两部分组成: 传统的 Skyline 集合和优先级队列中的元素集合。其中 Skyline 集合中元素的优先度高于优先队列中的元素, 只有当 Skyline 集合中的元素少于推荐的最小阈值时才从优先级队列选取元素进行补充。

在算法中, 维护一个优先级队列  $Q$ , 长度为  $k+1$ , 以便在队列中的元素达到推荐结果的最小阈值时还能进行一次插入操作, 队列  $Q$  的优先级定义为: 若  $h_1$  支配  $h_2$ , 则  $h_1$  的优先级大于  $h_2$ ; 否则,  $h_2$  的优先级大于  $h_1$ 。  $Q$  中从头到尾元素的优先级依次升高, 也就是说,  $Q$  的头元素  $Q.\text{front}$  的优先级最低。对于在 BNL 中淘汰的节点  $p$ , 进行如下操作:

```

if ( $Q.\text{size}() < k$ )
     $Q = Q + p$ ; /*若  $Q$  还没有满, 则将  $p$  插入  $Q$ */
else if ( $Q.\text{size}() = k$ ) /*若优先级队列  $Q$  满了*/
    if ( $Q.\text{front}()$  支配  $p$ )
        pass  $p$ ; /*若  $Q$  的头节点支配  $p$ , 淘汰  $p$ */

```



if ( $p$  支配  $Q.front()$ )

$Q=Q-Q.front()$ ;

$Q=Q+p$ ;

/\*若  $p$  支配  $Q$  的头节点, 将  $Q$  头节点删除, 将  $p$  插入  $Q^*$ /\*

#### 4.3 局部范围内基于协同过滤的评分算法

这一步要解决的问题就是结合用户信息与医院信息, 对 Skyline 中的医院集合进行评分与排序, 算法提出了局部范围内基于协同过滤的评分方式。该评分算法能够充分利用用户数据以及用户之间的相似性, 而且针对该领域中数据分布的局部稠密性, 提出了近邻用户的筛选机制, 有效降低了算法的时间复杂度。

##### 4.3.1 近邻用户的选择

传统的基于用户的协同过滤算法局限性在于不能很好地适应大规模用户和物品数据<sup>[13]</sup>, 假定  $M$  个用户和  $N$  个物品, 在最坏的情况下, 评估最多包含这  $N$  个物品的所有  $M$  个用户的记录, 因此复杂度较高, 而且评分矩阵一般非常稀疏。

针对数据分布局部范围内稠密这一特性, 近邻用户选择算法将用户信息同样采用 kd-tree 索引结构组织起来, 可以快速找到距离该医院一定范围内的用户集合, 该集合是所有用户集合的一个很小的子集。选取一定范围内的用户能够覆盖大多数的评分记录。

##### 4.3.2 带“分级诊疗”的医院累积评分

每次用户就诊后对就诊的医院的治疗效果、收费水平、服务态度这 3 个方面进行评分, 系统根据 3 项各自的权重计算出评分的平均值, 均值计算函数为

$$\text{avgScore} = \frac{5\text{effect} + 3\text{charge} + 2\text{attitude}}{10} \quad (3)$$

其中治疗效果、收费水平、服务态度 3 者之间的权重暂时定为 5:3:2, 这 3 者的权重比是通过查阅相关资料与调查问卷的方式相结合得出的<sup>[14]</sup>。

计算医院治疗每种疾病的累积评分时采用动态的惩罚机制, 惩罚系数设为  $\alpha$ :

$$\alpha = \frac{1}{|Ld - Lh| + 1} \quad (4)$$

式中:  $Ld$  表示疾病标注的默认等级,  $Lh$  为实际所去的医院。当  $Ld=Lh$  时  $\alpha=1$ , 对该评分不惩罚; 当  $|Ld-Lh|$  越大时  $\alpha$  越小, 对该评分的惩罚越严重,  $\alpha$  在计算每项均值时作为系数使用。

完整的医院累积评分计算公式为

$$\text{accScore} = \frac{5\text{effect} + 3\text{charge} + 2\text{attitude}}{10} \log \text{userCounter} \quad (5)$$

式中:  $\text{effect}$ 、 $\text{charge}$ 、 $\text{attitude}$  为各项评分的均值,  $\text{userCounter}$  为评分的总人数, 对评分总人数取对数是用来平衡规模不同的医院之间评价人数的差别引起的偏斜。

##### 4.3.3 计算预测评分

###### 1) 确定相似用户集

在经过筛选的用户子集上计算用户相似度时采用的是 Pearson 相关系数, 因为 Pearson 相关系数能在计算中不考虑平均值的差异, 用户相似度计算公式为

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad (6)$$

相似度系数取值在  $-1 \sim 1$  之间,  $-1$  表示完全负相关,  $1$  表示完全正相关,  $0$  表示不相关。

###### 2) 计算预测值

预测值计算公式为

$$\text{pre}(a, b) = \bar{r}_a + \frac{\sum_{b \in N} \text{sim}(a, b)(r_{b,p} - \bar{r}_b)}{\sum_{b \in N} \text{sim}(a, b)} \quad (7)$$

针对空白评分数据, 采用的是缺省投票机制, 即用该医院的平均评分值填充空白值。

## 5 实验验证

### 5.1 辅助诊疗实验

本文所使用的测试数据来自超星医疗知识库。为了模拟症状表述不全的情况, 每种疾病的症状信息都被拆分为两部分。实验在医疗知识库中疾病的种类数为 100 和 1 000 时分别进行了测试。

统计实验结果的策略是: 如果诊断结果的前 10 位中出现该疾病, 则认为成功召回; 如果诊断结果中的前 3 位中出现该疾病, 则认为诊断结果准确。

1) 数据库中的疾病种类为 100 时, 选取了 10 种疾病, 共 20 个测试用例, 实验结果如表 1 所示。

表 1 100 种疾病时的辅助诊断结果

Table 1 The results of auxiliary diagnoses on 100 diseases

位置	初步	优化	位置	初步	优化
1	9	14	7	1	0
2	5	4	8	0	0
3	3	1	9	0	0
4	0	1	10	0	0
5	1	0	其他	1	0
6	0	0			

2) 数据库中疾病种类为 1 000 时, 选取 100 种, 共 200 个测试用例, 实验结果如表 2 所示。

表2 1 000种疾病时辅助诊断结果

Table 2 The results of auxiliary diagnoses on 1 000 diseases

位置	初步	优化	位置	初步	优化
1	105	116	7	2	3
2	43	39	8	3	1
3	14	22	9	3	0
4	7	8	10	1	1
5	1	2	其他	17	8
6	4	0			

实验结果表明,所选测试数据在此统计策略下,当数据库中疾病种类为100时,召回率为95%,准确率为85%,在优化诊断之后,准确率达到95%,召回率达到100%;当数据库中疾病种类为1 000时,召回率为91.5%,准确率为81%,在优化诊断之后,准确率达到88.5%,召回率达到96%。

## 5.2 推荐技术实验

推荐算法的质量衡量方法一直是存在争议的,甚至有人认为推荐系统的质量根本就不可能直接衡量,因为太多的目标函数。目前最主流的评估方案是根据电影领域用户历史评分来估计不同算法的实验方法,EachMovie数据集和Netflix数据集是其中的典型代表,但是很显然,这些数据不可能应用到本文中的推荐算法上,因为文中提出的推荐算法是和医疗与导诊密切相关的。基于以上原因,本文中使用合成的数据集,并采用定量和定性结合的方式评估算法质量。

### 5.2.1 生成实验数据

#### 1) 合理性分析

经过查阅相关医学资料以及调查问卷的方式,得出在选择医院时人们主要关注的有两点:医院与自己的距离,一般对于常见疾病人们倾向于选择距离自己近的医院就诊;在医院质量方面,患者主要关心的因素是诊疗效果、收费水平、服务态度,对这三者的关注度权重接近于5:3:2。既然人们在就诊选择时有这样的倾向性,实际的数据分布也会体现出这样的特点,那么我们就按照上面的原则模拟生成实验数据,并且这样的数据具有一定的合理性。

#### 2) 数据规模

假定地域范围是 $44 \times 44 \text{ km}^2$ 的一个矩形,有50家医院,医院的等级共有10级,有50种疾病,疾病的严重等级共有10级,有5 000个病人,每人有10条就诊评价记录,共有50 000条记录。

#### 3) 生成数据

① 医院和用户的地域信息使用经纬度表示,经

纬度数值随机生成,只要将其限制在上述的矩形区域即可,医院和疾病的等级信息也随机生成。

② 医院的诊疗效果(effect)、收费水平(charge)、服务态度(attitude)都是医院的固有属性,因此先生成医院擅长治疗的疾病与医院之间的对应集合 $S_1$ 、收费较低的医院集合 $S_2$ 、服务态度好的医院集合 $S_3$ 。

③ 对每条评价记录, effect、charge、attitude 初始的评价都是5分。

④ 按式(8)~(10)更新 effect、charge、attitude 的值。

$$\text{effect} = \begin{cases} (1 + \Delta_1)\text{effect}, & \langle \text{disease}, \text{hospital} \rangle \in S_1 \\ (1 - \Delta_1)\text{effect}, & \langle \text{disease}, \text{hospital} \rangle \notin S_1 \end{cases} \quad (8)$$

$$\text{charge} = \begin{cases} (1 + \Delta_2)\text{charge}, & \text{hospital} \in S_2 \\ (1 - \Delta_2)\text{charge}, & \text{hospital} \notin S_2 \end{cases} \quad (9)$$

$$\text{attitude} = \begin{cases} (1 + \Delta_3)\text{attitude}, & \text{hospital} \in S_3 \\ (1 - \Delta_3)\text{attitude}, & \text{hospital} \notin S_3 \end{cases} \quad (10)$$

⑤ 给每一项再加一个小的正负随机的扰动 $\Delta_0$ ,以模拟评分时的噪音。

### 5.2.2 实验方案

使用上面生成的模拟数据作为训练数据,再随机选择100名用户,每人2种疾病,距离用户 $R=15 \text{ km}$ 的医院作为候选医院,在该测试数据集上运行推荐算法,并记录推荐结果中前3位。

前面已经说过,就诊时的关注点主要有距离和医院的质量,因此,一方面对距离用户范围 $R$ 内的医院按照距离和治疗对应疾病的综合评分分别排序,统计推荐的医院在两个序列中的具体分布情况;另一方面查看并统计推荐的医院在上述 $S_1$ 、 $S_2$ 、 $S_3$ 这3个集合中分布情况,以此来定性确定推荐的医院的质量。最终,通过对以上两个指标分布的分析,来半定量地评估推荐算法。

### 5.2.3 实验结果

推荐结果中排前三的医院在按照评分距离和评分排序的两个序列中的具体分布情况如图2所示。

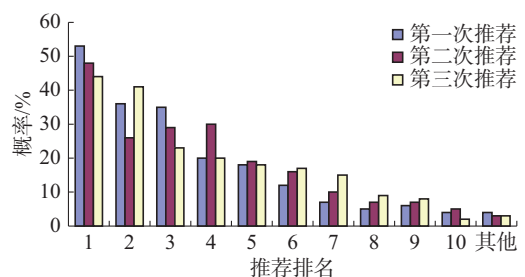


图2 推荐结果在两个序列中的分布

Fig. 2 The distribution of recommended results in two lists

由统计结果可得:排名第一的推荐结果两个序列中前6位出现的概率是87%,在前3位出现的概

率是62%;排名第二的推荐结果在前6位出现的概率是84%,在前3位出现的概率是51.5%;推荐结果排在前6位出现的概率是81.5%,在前3位出现的概率是54%。可以看出,推荐的医院主要出现在上述两个列表中的前6位,其概率大于81%,出现在前3位的概率大于51%,而且出现在前6位的概率随着推荐结果的次序依次从高到低线性排列,这符合预期的结果,推荐的医院至少在距离和质量两方面之一有优势,而且越处于推荐结果前列的医院其占优的可能性应该越大。

推荐结果中前3的医院在上述 $S_1$ 、 $S_2$ 、 $S_3$ 这3个集合中分布情况如图3所示。

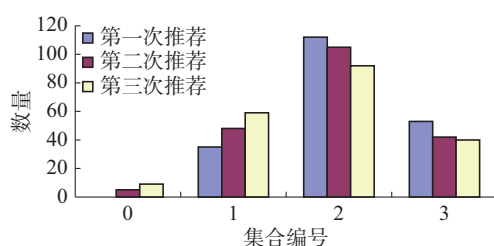


图3 推荐结果在3个集合中的分布

Fig. 3 The distribution of recommended results in three sets

推荐的医院在诊疗效果、收费水平、服务态度这3项中至少有2项占优的概率:结果一为82.5%,结果二为73.5%,结果三为66%,至少有一项占优的概率大于95%。这就说明,推荐的医院从一定程度上来看是较优的,而且推荐结果中位置越靠前的就越优,推荐结果具有一定的合理性。

## 6 结束语

本文通过对医疗与导诊领域的深入研究,发现了医疗领域的推荐与传统推荐在本质上的区别,提出了一种面向智能导诊的个性化推荐算法,以辅助诊疗结果为基础,将Skyline查询和局部范围内基于协同过滤的评分方式相结合。算法能根据用户的症状表现与地理位置等个人信息,为用户提供个性化的推荐结果。并且通过模拟生成实验数据,半定量地验证了推荐算法的合理性和有效性。本文中提出的算法和思路对于有效地利用电子医疗数据,合理分配和使用现有的医疗资源,缓解就诊压力,提升就医质量意义重大,有着很大的实用价值与社会意义。

## 参考文献:

[1] LOPEZ-NORES M, BLANCO-FERNÁNDEZ Y, PAZOS-ARIAS J J, et al. Property-based collaborative filtering for

health-aware recommender systems[C]//Proceedings of 2011 IEEE International Conference on Consumer Electronics. Las Vegas, NV, USA, 2011.

[2] PATTARINTAKORN P, ZAVERUCHA G M, CERCONI N. Web based health recommender system using rough sets, survival analysis and rule-based expert systems [C]//Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Toronto, Canada, 2007: 491-499.

[3] GONG Jibing, WANG Lili, SUN Shengtao, et al. iBole: a hybrid multi-layer architecture for doctor recommendation in medical social networks[J]. Journal of computer science and technology, 2015, 30(5): 1073-1081.

[4] 徐守坤, 吴伟伟. 基于语义的医疗资源均衡推荐算法[J]. 计算机工程, 2015, 41(9): 74-79.

XU Shoukun, WU Weiwei. Balance recommendation algorithm for medical resources based on semantic[J]. Computer engineering, 2015, 41(9): 74-79.

[5] PAPADIAS D, TAO Yufei, FU G, et al. An optimal and progressive algorithm for skyline queries[C]//Proceedings of 2003 ACM SIGMOD International Conference on Management of Data. San Diego, USA, 2003: 467-478.

[6] CHAN C Y, JAGADISH H V, TAN K L, et al. Finding k-dominant skylines in high dimensional space[C]//Proceedings of 2006 ACM SIGMOD International Conference on Management of Data. Chicago, IL, USA, 2006: 503-514.

[7] VLACHOU A, DOULKERIDIS C, KOTIDIS Y. Angle-based space partitioning for efficient parallel skyline computation[C]//Proceedings of 2008 ACM SIGMOD International Conference on Management of Data. Vancouver, Canada, 2008: 227-238.

[8] SARKAS N, DAS G, KOUDAS N, et al. Categorical skylines for streaming data[C]//Proceedings of 2008 ACM SIGMOD International Conference on Management of Data. Vancouver, Canada, 2008: 239-250.

[9] BÖRZSÖNYI S, KOSSMANN D, STOCKER K. The skyline operator[C]//Proceedings of the 17th International Conference on Data Engineering. Washington, DC, USA, 2001: 421-430.

[10] CHAN C Y, ENG P K, TAN K L. Efficient processing of skyline queries with partially-ordered domains[C]//Proceedings of the 21st International Conference on Data Engineering. Tokyo, Japan, 2005: 190-191.

[11] GUO Qinglin. The similarity computing of documents based on VSM[C]//Proceedings of the 32nd Annual IEEE International Computer Software and Applications. Turku, Finland, 2008: 585-586.

[12] ROBINSON J T. A search structure for large multidimensional dynamic indexes[C]//Proceedings of 1981 International Conference on Management of Data. Ann Arbor, Michigan, 1981: 10-18.

[13] HERLOCKER J L, KONSTAN J A, RIEDL J. Explaining collaborative filtering recommendations[C]//Proceedings of

2000 ACM Conference on Computer Supported Cooperative Work. Philadelphia, Pennsylvania, USA, 2000: 241–250.

[14] RICCI F, ROKACH L, SHAPIRA B, et al. Recommender systems handbook[M]. Boston, MA: Springer, 2011: 1–35.

#### 作者简介:



马钰, 男, 1993 年生, 硕士研究生, 主要研究方向为数据库、数据挖掘、大数据。发表学术论文 2 篇。



张岩, 男, 1962 年生, 副教授, 主要研究方向为数据质量和生物信息学。发表学术论文 20 余篇。



王宏志, 男, 1978 年生, 教授, 博士生导师, 主要研究方向为大数据管理, 包括数据质量、XML 数据管理和图数据管理。获得省部级科技进步奖 2 项, 发表学术论文 200 余篇。

## 2018 第三届机电一体化和自动化技术国际会议 (ICMAT 2018) 2018 3rd International Conference on Mechatronics and Automation Technology (ICMAT 2018)

ICMAT 2018 是由皇家墨尔本理工大学, AUT 大学, 斯文本科技大学, 澳大利亚国立大学和香港机械工程师协会联合主办。本次会议将于 2018 年 9 月 27—29 日在泰国普吉岛举办。ICMAT2016 和 ICMAT2017 已成功在奥克兰和墨尔本举行。

欢迎在机电一体化和自动化技术, 在处理理论的新发展, 分析建模与仿真、实验、示范和数值, 高级部署和案例研究, 实验室或现场运行试验结果的投稿。此次会议将为世界各地机电一体化和自动化技术这一领域的科研专家们提供面对面的交流机会。

#### 【出版与检索】

所有注册和提交的论文将发表在会议论文集, 并提交 EI Compendex, Scopus, Thomson Reuters (WoS), Inspec 及其他检索机构检索。

被选中的优秀论文将刊发在国际期刊上。

#### 【征稿主题】

智能机电一体化	光电元件和材料
机器人	激光技术与激光加工
仿生	控制系统建模
自动化和控制系统	仿真技术

**Website:** <http://www.icmat.net/cfp.html>