

文章编号: 1671-4742(2015)03-0248-06

基于模糊联合聚类方法的针灸处方主穴挖掘研究

郭艳珍, 舒红平, 郑皎凌

(成都信息工程大学软件工程学院 四川 成都 610225)

摘要: 分析针灸处方数据,从治疗某种疾病的处方中找到一类医生用到的主要腧穴,方便不同的医生做出相应的决策。使用模糊联合聚类方法对处方和腧穴进行同时聚类,分别找出集群中具有相似腧穴的处方和具有相似处方的腧穴。但是当针灸数据库中处方集合比较大且分类更多时,这种算法不足以把处方和腧穴更好聚类。文中提出使用改进的 FCR-PA 算法进一步的细粒度分类,可以清晰地辨别处方和腧穴的所属类。实验对疾病数据库中典型的 18 种疾病进行分析,其中有 15 种与临床诊疗经验相符合,准确率达 83.3%。该方法能够有效地挖掘出治疗某一种疾病的主要腧穴,为不同的医生问诊提供有价值的决策支持。

关键词: 计算机科学技术; 数据挖掘; 联合聚类; 针灸处方; 腧穴

中图分类号: TP311.1

文献标志码: A

0 引言

问题源于针灸临床循证决策支持平台系统,中医腧穴是针灸治疗疾病的施术部位,治疗疾病时对针灸产生刺激点与反应,也是中医开具针灸处方的依据。在中医针灸治疗过程中,根据经验和病人的特点,不同的医生针对同一类疾病采用不同的腧穴治疗配方。因此,一些支持针灸决策的系统的数据库中,医生治疗同一种疾病采用的不同腧穴组合就形成集合。研究发现,可以从这些治疗疾病的腧穴集合中发现某种共性^[1],即某些腧穴在治疗针某种疾病时被医生普遍采用。因而考虑挖掘更加可靠的主要腧穴集合问题,揭示腧穴运用特点和配伍规律^[2-3],如何取得合理有效的应用研究证据对提高针灸临床疗效具有非常重要的意义。

针对提出的问题,采用模糊联合聚类算法。分析针灸临床循证决策支持系统中的处方数据库,数据以名老中医针灸处方为基础,将治疗某一种疾病的腧穴字段分解成标准腧穴并对其进行转换成处方和腧穴的共现关系。运用联合聚类方法对处方和腧穴进行同时聚类,从而得出针对某一种疾病的针灸处方相对应的几个主要腧穴集合,主要腧穴集合相对应的名老中医的针灸处方群集。这对针灸临床疗效具有很大的提高。研究重点即是通过改进的模糊联合聚类算法挖掘

主穴为医生提供决策支持。

1 相关工作

网络多媒体的检索和过滤最常用的办法就是聚类。方法根据对象的特性将其聚集成类,从而同类间的相似性尽可能小,不同类之间相似性尽可能大,这其实是无监督的模式识别方法。例如 K-means 聚类法,可以将某个对象(一般以向量表示)集合划分成好几个聚类或者数据子集,这样同一聚类中各个对象之间的相似性比其他聚类中各个对象之间的相似性更高^[4-5]。为提升高维数据的查询效率,通常用聚类算法将相似性高的数据点聚集在一起。因此聚类算法直接或者间接应用于大量高维数据机构。但传统的聚类算法也有不足之处,因为只根据对象与对象之间的相似性进行聚类,却没有考虑对象属性之间的相似性^[6]。在分析基因和文本领域,首先采用同时考虑对象和属性的联合聚类方法^[7]。高维数据集可以用矩阵的形式表示,矩阵有多少行表示有多少对象(向量),有多少列表示每个对象有多少属性。联合聚类法对矩阵的行和列分别进行聚类,然后分别迭代直到收敛,最后可以将矩阵的行跟列分配到行聚类和列聚类里面。文中采用模糊联合聚类的分析方法,可以对腧穴进行聚类,实现对腧穴共性的挖掘。如图 1 表示联合聚类的矩阵 M 。设 u 为联合聚类的行聚类数, w 为列聚类数,经过模糊联合聚类运算可以将矩阵分割成 $u \times w$ 个数据子集。

收稿日期: 2015-05-05

基金项目: 国家自然科学基金资助项目(61501064); 四川省教育厅科技成果转化重大培育资助项目(15CZ0019)

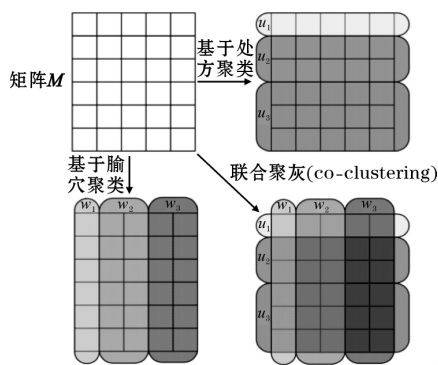


图1 处方和腧穴的联合聚类示意图

2 基于模糊联合聚类的针灸处方主穴挖掘

2.1 模糊联合聚类简介

模糊联合聚类(Fuzzy Co-clustering with Ruspini's condition, FCR) [8-9] 即同时对矩阵的行和列进行聚类, 优势突出, 广泛应用于数据分析和协同过滤等领域。为获得模糊聚类, 给出联合聚类算法聚类准则的定义, 如式(1)所示。在这个定义之下, 那些具有高度彼此相关性的个体和属性将聚集。在这个意义上说, 把聚合度作为聚类准则。模糊集群的聚类标准如下

$$\sum_{i=1}^M \sum_{j=1}^N u_{ci} w_{cj} d_{ij} \quad c = 1, \dots, C \quad (1)$$

式(1)中, 若第 j 个腧穴在第 i 个处方中存在, 则 $d_{ij} = 1$; 否则 $d_{ij} = 0$ 。每个簇的聚合度是由定性变量 d_{ij} 、处方和腧穴的隶属度函数 u_{ci} 和 w_{cj} 的乘积总和。通过给处方和腧穴的隶属度函数不断赋值使聚集度最大化, 从而形成模糊簇。处方和腧穴的隶属度函数如下

$$\sum_{c=1}^C u_{ci} = 1 \quad u_{ci} \in [0, 1] \quad i = 1, \dots, M \quad (2)$$

$$\sum_{j=1}^N w_{cj} = 1 \quad w_{cj} \in [0, 1] \quad c = 1, \dots, C \quad (3)$$

其中 i 和 j 分别指代矩阵的行(处方)与列(腧穴) c 指示聚类数量, u_{ci} 是处方在聚类 c 中的隶属度函数, w_{cj} 是腧穴在聚类 c 中的隶属度函数。

文中提出的联合聚类算法是通过优化目标函数实现聚合度的最大化, 采用熵的最大化和拉格朗日乘法构造目标函数如式(4)。目标函数的解法是已知必备的条件通过迭代处理, 达到局部最小值。通过使用该方法, 可以一次获得全部数据集的聚类结果。模糊聚类后, 可以得到全部处方和腧穴所属的簇。

$$R = \sum_{c=1}^C \sum_{i=1}^M \sum_{j=1}^N u_{ci} w_{cj} d_{ij} - T_u \sum_{c=1}^C \sum_{i=1}^M u_{ci} \log u_{ci} - T_w \sum_{c=1}^C \sum_{j=1}^N w_{cj} \log w_{cj} + \sum_{i=1}^M \lambda_i \left(\sum_{c=1}^C u_{ci} - 1 \right) + \sum_{c=1}^C \gamma_c \left(\sum_{j=1}^N w_{cj} - 1 \right) \quad (4)$$

其中 λ_i, γ_c 是不同的拉格朗日乘数。式(4)中的第2项和第3项把熵的最大化描述成正则表达式, 这个方法在模糊 c-means 算法中运用第一次被 Miyamoto 等 [10] 提出。能获得模糊聚类。 T_u 和 T_w 是指明模糊度的权值参数, 余下描述的是隶属度约束条件, 也就是式(1)和式(2)。从优化目标函数 R 的必需条件, 也就是 $\partial L / \partial u_{ci} = 0$ 和 $\partial L / \partial w_{cj} = 0$, 得到式(5)。

$$u_{ci} = \frac{\exp\left(\sum_{j=1}^N w_{cj} d_{ij} / T_u\right)}{\sum_{c=1}^C \exp\left(\sum_{j=1}^N w_{cj} d_{ij} / T_u\right)}, \quad w_{cj} = \frac{\exp\left(\sum_{i=1}^M u_{ci} d_{ij} / T_w\right)}{\sum_{j=1}^N \exp\left(\sum_{i=1}^M u_{ci} d_{ij} / T_w\right)} \quad (5)$$

2.2 基于改进的 FCR-PA 方法的主穴挖掘应用模型

2.2.1 模糊联合聚类的挖掘模型

处方数据库记录了名中医对某一种疾病的处方及其对应的腧穴集合, 利用模糊联合聚类算法对处方数据库进行腧穴挖掘, 目的是找出治疗每一种疾病的主要腧穴集合。不仅挖掘治疗某种疾病的主要处方, 同时挖掘在处方数据库中某种疾病的名老中医处方对应的主要腧穴。如表1所示。

表1 数据库中的名老中医感冒处方

处方 ID	处方名	腧穴
处方 1	王氏治感冒方	风池、风门、合谷
处方 2	玉龙经感冒方	函会、风门
处方 3	刘氏治感冒方	风池、大椎、合谷
处方 4	肖福庵感冒方	大杼、肝俞、膈俞、陶道
处方 5	肖氏治感冒方	大椎、风门、肾俞、肺俞、合谷
处方 6	王执中感冒方	百会、天府、曲池、列缺
处方 7	王登旗治感冒方	曲池、足三里、合谷、大椎、少商

由表1可以看出针对治疗感冒有7个处方, 每个处方可以看作是一个腧穴集合, 7个腧穴集合共有合谷、大椎、风池等17个不同的腧穴。由于每个处方对应的腧穴集合不等长, 为方便计算机处理需要将其转换成等长集合, 即处方和腧穴的共现关系(二进制矩阵 d_{ij}) 如表2所示。

表 2 处方和腧穴的矩阵表示

处方 ID	足三里	囟会	膈俞	合谷	大椎	大杼	肝俞	风池	少商	百会	肾俞	风门	陶道	天府	列缺	曲池	肺俞
处方 1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
处方 2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
处方 3	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
处方 4	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
处方 5	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	1
处方 6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
处方 7	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0

在处方数据中,由表 2 可以看出处方和腧穴的二进制矩阵 d_{ij} ,设置参数聚类数 C 为 2 簇,聚类模糊度 Tu 、 Tw 分别为 0.1、1.5,控制算法达到最优结果的参数 ε 为 0.0001。通过执行和迭代处理目标函数(5),并且给处方和腧穴的隶属度函数不断赋值达到局部最小值,从而使聚集度最大化形成处方和腧穴的模糊簇如表 3 和表 4 所示。

表 3 感冒处方簇

处方编号	A 类	B 类
1	0.9652	0.0348
2	0.5323	0.4677
3	0.0724	0.9276
4	0.9867	0.0133
5	0.1755	0.8245
6	0.9668	0.0332
7	0.9816	0.0184

表 4 感冒腧穴簇

腧穴编号	A 类	B 类
1	0.0438	0.0519
2	0.0328	0.0692
3	0.0242	0.0941
4	0.3101	0.0542
5	0.1628	0.053
6	0.0242	0.0941
7	0.0242	0.0941
8	0.0844	0.0525
9	0.0438	0.0519
10	0.0259	0.0878
11	0.0445	0.0511
12	0.1208	0.0714
13	0.0017	0.047
14	0.004	0.0418
15	0.004	0.0418
16	0.0263	0.0436
17	0.0227	0.0007

表 5 中风处方簇

处方编号	A 类	B 类	C 类	D 类
1	0.4325	0.0636	0.4315	0.0724
2	0.2444	0.2442	0.2443	0.2672
3	0.0958	0.6857	0.0956	0.1229
4	0.3054	0.3132	0.3098	0.0715
5	0.2472	0.2457	0.2463	0.2608
6	0.3052	0.3174	0.3024	0.065
7	0.2396	0.2393	0.2392	0.2819
8	0.4326	0.0618	0.4244	0.0813
9	0.4222	0.0599	0.4281	0.0898
10	0.2486	0.2485	0.2486	0.2543
⋮	⋮	⋮	⋮	⋮
98	0.2486	0.2486	0.2486	0.2542
99	0.1098	0.0991	0.6785	0.1126
100	0.0674	0.4402	0.4179	0.0746
101	0.0963	0.6968	0.0967	0.1102
102	0.1091	0.0961	0.6655	0.1293

表 6 中风腧穴簇

腧穴编号	A 类	B 类	C 类	D 类
1	0.9103	0.0002	0.0153	0.0303
2	0.0002	0	0	0.0071
3	0.0001	0	0	0.0043
4	0.0044	0	0.0014	0.0148
5	0.0051	0.0001	0.0019	0.0215
6	0.0001	0	0	0.0105
7	0	0	0	0.0047
8	0.0014	0	0.0006	0.0126
9	0	0	0	0.0047
10	0.0001	0	0	0.0061
⋮	⋮	⋮	⋮	⋮
149	0	0	0	0.0047
150	0	0	0	0.0047
151	0	0	0	0.0047
152	0	0	0	0.0042
153	0	0	0	0.0047

如表 3 和表 4 的结果所示,若感冒处方簇和腧穴簇中隶属度比较大,则其属于相对应的类。由表 3 可知处方{1,2,4,6,7}属于 A 类,处方{3,5}属于 B 类。相应的由表 4 可知腧穴{1,4,5,8,9,11,12,16,17}属

于 A 类 腧穴{2,3,6,7,10,13,14,15} 属于 B 类。虽然由此方法可以得出感冒的处方簇和腧穴簇,再通过处方和腧穴的映射关系即可挖掘治疗感冒的主要腧穴。但是当针灸数据库中治疗某一疾病的处方集合比较大的时候,这种算法存在不足以把处方和腧穴更好的聚类。例如治疗中风有 102 个处方数据及对应的 153 个腧穴,将其划分为 4 类,由模糊联合聚类算法得出处方簇和腧穴簇分别如表 5 和表 6 所示。

为解决这类问题,在以前的模糊联合聚类基础上进行更进一步的无监督细粒度分类。

2.2.2 改进的 FCR-PA 方法的挖掘模型

根据前面的论述分析,主要改进是在模糊联合聚类的基础上将处方和腧穴作进一步的细粒度分类。将表 5 中处方簇看作是处方集合{ $P_{ij} \mid i = 1, \cdots, M, j = 1, \cdots, N$ },即每个处方在所有的簇中的权重之和为 1。其中 M 为处方数, N 为聚类数,定义第 i 行所属的类为 $C(i)$ 。由此提出改进的模型:

$$j = C(i) = \arg \max \left(\frac{P_{ij}}{\sum_{j=1}^N P_{ij}} \right) \quad (1 \leq i \leq M, 1 \leq j \leq N)$$

其中先求出处方集合中的每一个元素与每类中所有处方元素总和的比率,然后如果当第 i 行取得元素的最大值时,则返回其所属的类 $C(i) = j$,即可得到处方的具体分类结果。例 1 给出一个对 FCR-PA 进行解释的简单例子。

例 1: 基于治疗中风表格 5 和表格 6 的数据对 FRC-PA 的解释。

针对表 6 的腧穴集簇可以看做腧穴集合{ $A_{ij} \mid i = 1, \cdots, M, j = 1, \cdots, N$ },即所有的腧穴在每一个类中的权重之和为 1: 其中 M 为腧穴数, N 为聚类数,首先定义腧穴簇中每列元素值的最低限制为 θ ,如果 $A_{ij} \geq \theta$,则取其值,从而得出属于每类新的腧穴集合: $a_{ij} = \{A_{ij} \mid A_{ij} \geq \theta, i = 1, \cdots, M, j = 1, \cdots, N\}$ 。通过进一步的规范,即可得到腧穴的具体分类结果。如下实例 1 是对该模型的进一步阐述,由表 5 的处方集合可以转化成矩阵所示。

$$\begin{bmatrix} \frac{P_{11}}{\sum_{i=1}^{105} P_{11}} & \frac{P_{12}}{\sum_{i=1}^{105} P_{12}} & \frac{P_{13}}{\sum_{i=1}^{105} P_{13}} & \frac{P_{14}}{\sum_{i=1}^{105} P_{14}} \\ \frac{P_{21}}{\sum_{i=1}^{105} P_{21}} & \frac{P_{22}}{\sum_{i=1}^{105} P_{22}} & \frac{P_{23}}{\sum_{i=1}^{105} P_{23}} & \frac{P_{24}}{\sum_{i=1}^{105} P_{24}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{P_{105,1}}{\sum_{i=1}^{105} P_{105,1}} & \frac{P_{105,2}}{\sum_{i=1}^{105} P_{105,2}} & \frac{P_{105,3}}{\sum_{i=1}^{105} P_{105,3}} & \frac{P_{105,4}}{\sum_{i=1}^{105} P_{105,4}} \end{bmatrix}_{105 \times 4}$$

$_{105 \times 4}$

然后求出每行的最大值,由模型可知当取得矩阵中每行元素的最大值所对应的变量 j ,既是第 i 行所属的类 $C(i)$ 。处方分类结果如表 7 所示。

表 7 处方分类结果

$C(i)$	处方名 ID	模型	所属类
$C_{(1)}$	处方 1	$\arg \max (P_{11} / \sum_{i=1}^M P_{11}) = 1$, 即 $C_1 = 1$	A 类
$C_{(2)}$	处方 2	$\arg \max (P_{24} / \sum_{i=1}^M P_{24}) = 4$, 即 $C_2 = 4$	D 类
$C_{(3)}$	处方 3	$\arg \max (P_{32} / \sum_{i=1}^M P_{32}) = 2$, 即 $C_3 = 2$	B 类
$C_{(4)}$	处方 4	$\arg \max (P_{43} / \sum_{i=1}^M P_{43}) = 3$, 即 $C_4 = 3$	C 类
$C_{(5)}$	处方 5	$\arg \max (P_{54} / \sum_{i=1}^M P_{54}) = 4$, 即 $C_5 = 4$	D 类
\vdots	\vdots	\vdots	\vdots
$C_{(98)}$	处方 98	$\arg \max (P_{98,4} / \sum_{i=1}^M P_{98,4}) = 4$, 即 $C_{98} = 4$	D 类
$C_{(99)}$	处方 99	$\arg \max (P_{99,3} / \sum_{i=1}^M P_{99,3}) = 3$, 即 $C_{99} = 3$	C 类
$C_{(100)}$	处方 100	$\arg \max (P_{100,3} / \sum_{i=1}^M P_{100,3}) = 3$, 即 $C_{100} = 3$	C 类
$C_{(101)}$	处方 101	$\arg \max (P_{101,2} / \sum_{i=1}^M P_{101,2}) = 2$, 即 $C_{101} = 2$	B 类
$C_{(102)}$	处方 102	$\arg \max (P_{102,3} / \sum_{i=1}^M P_{102,3}) = 3$, 即 $C_{102} = 3$	C 类

由表 7 可知改进的 FCR-PA 可以高效地将治疗中风的高维处方数据聚类。然后通过多变量分析技术将处方和腧穴数据进行分析后可得挖掘治疗中风的主要腧穴。所以治疗中风疾病的聚类结果为属于 A 类的主要腧穴为{足三里,曲池,合谷,绝骨},属于 B 类的主要腧穴为{人中,曲池,合谷,足三里,肩髃},属于 C 类的主要腧穴为{肩髃,曲池,百会,合谷},属于 D 类的主要腧穴为{合谷,曲池,百会,人中,足三里}。

3 实验流程

3.1 数据预处理

具体预处理过程如图 2 所示。

第一步: 数据准备

由于高质量的数据一定决定高质量的决策,所以一个很重要步骤就是对数据进行预处理。而且数据质量的提高对其后挖掘过程的精度和性能的提高也很有帮助^[11]。因此,要使挖掘算法更有效地挖掘准确可靠的知识,就必须对数据进行预处理。数据准备^[12]是对源数据进行预处理的一个步骤,首先保证数据准确,然后滤除冗余以及错误的数据,使数据处理的效果和质量得以提高。同时,因为是从临床获取的数据形成的处方数据库,所以获取到的数据具有很高的重复性和冗余性。为提高数据的可用性,必须在对针灸处方数据进行联合聚类之前对数据进行冗余消除处理。

第二步: 数据转换

用于数据挖掘的数据一般来自多个数据源,而在整合不同数据源的数据时,由于数据定义不同,容易产生数据不一致的问题,包括数据结构不一致、类型不一致、计量单位不一致、随意缩写造成的一致、属性名称不一致等^[13]。数据转换^[12]的目的在于将来自不同数据源中不一致的数据按照一定标准进行转换,使之成为适合数据挖掘的形式。数据转换与数据清理常同时或反复进行。当数据清理完成后,就需要将针灸决策支持系统中的名老中医处方数据的腧穴字段转化为标准腧穴。

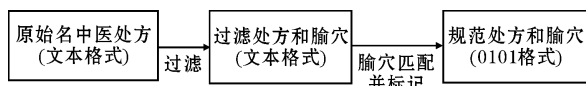


图2 数据预处理流程

3.2 基于改进的 FCR-PA 的算法实现

算法: 基于改进的 FCR-PA

输入: 处方数 M , 聚类数 N , 处方集合 P_{ij} , 腧穴集合 A_{ij} , 参数 ∂

输出: 每个处方所属的类 $C_{(i)}$, 分类后新的腧穴集合 a_{ij}

(1) for $j = 1$ to N do

for $i = 1$ to M do

$sum = \sum_{i=1}^M p_{ij}$ // sum 表示处方矩阵中所有处方在每类中的权重之和

$l_{m \times n} = p_{ij} / sum$ // 矩阵中的每个元素与所有处方在每类中权重之和的比率

(2) end for

(3) for $t = 1$ to N do

$k_{m \times 1} = \max(l_{m \times n})$ // $k_{m \times 1}$ 表示每行元素的最大值

(4) end for

(5) return j // j 表示第 i 行所属的类

(6) for $j = 1$ to N do

if $(A_{ij} \geq \partial)$ // ∂ 表示腧穴簇每列元素中的最低限制值

$a_{ij} = A_{ij}$ // a_{ij} 表示新的腧穴集合

(7) end for

(8) return a_{ij}

在算法中,联合聚类分别用行和列代表实体和属性组成一个矩阵,矩阵内的每个元素代表着实体和属性的共现关系,通过对行和列进行同时聚类揭示实体和属性之间的内在关系。治疗中风疾病的处方数据进

行数据预处理后转换成处方(实体)和腧穴(对象)的共现关系矩阵。但是对这种高维的大量的数据信息进行聚类时出现冷启动的问题,比如有些腧穴很少有医生用到。然而通过上述改进的联合聚类算法可以有效地避免此问题的出现。

4 实验结果与分析

通常,用于临床治疗疾病的主穴为 3 - 10 个^[14],实验通过改进的方法,当治疗某一疾病的处方和腧穴集合比较大的时候,对相对应数据进行分析处理后即可挖掘主要腧穴。而对于集合比较大的情况模糊联合聚类方法不能有效的进行聚类,如表 5 和表 6 所示,不能明显看出处方和腧穴具体属于哪一类。但在该方法的基础上提出的 FCR-PA 算法对前面两表中的数据进行处理后,就可以清晰的辨别处方和腧穴的所属类,具体结果如表 7 所示。

文中提到治疗感冒疾病,挖掘其主穴为合谷、风门、大椎、风池;治疗中风疾病的主穴为肩髃、曲池、足三里以及合谷。此实验结果经成都中医药大学老师鉴定,上述的腧穴为临床治疗感冒和中风最主要的穴位。另外,相继做实验和分析治疗头痛和哮喘的处方数据,针灸治疗头痛疾病的主要腧穴集合为合谷、风池、太阳、上星、足三里;治疗哮喘疾病主要腧穴集合为膻中、乳根、俞府以及璇玑。

实验表明:基于改进的 FCR-PA 算法可以很好地应用于挖掘中医针灸决策系统中的主穴,较好地支持了中医针灸临床决策。

5 结束语

针灸是中国古代治疗各种疾病的主要手段,有疏通经络、扶正祛邪、调和阴阳等功效。中医针灸现在已经成为世界非文化物质遗产,发展到现在,留下文献资料、名师典籍以及世代相传的诊治规律各种珍贵的智慧结晶。因此,从中发现腧穴运用的方法特点以及配伍规律,从而应用于现在的针灸决策支持系统是非常重要的^[15-16]。近年,中医正在走向世界,越来越需要对中医针灸的数据进行挖掘,这必然也要求相关的研究应该客观化、标准化和数量化。所以中医研究引入聚类分析给中医客观化、标准化研究提供了一种新的思路与方法,同时为中医规范化研究注入了新的生机和活力。文中提出的联合聚类改进算法分析是数据挖掘中一门非常有用的技术,可以用于从大量数据中寻找隐含的数据分布和模式^[17]。

利用模糊联合聚类思想并改进挖掘针对某种疾病所用针灸穴位的主要穴位,实验结果表明,文中提出的算法能够很好的挖掘所用的主穴信息并应用于决策支持系统,经过成都中医药大学老师的鉴定,方法非常实用。接下来,为给用户提供友好的用户界面,可以将挖掘的主穴信息在决策系统中以可视化的方式显示,从而提高决策支持系统的辅助能力。

参考文献:

- [1] 任玉兰,梁繁荣,吴曦,等. 基于数据挖掘的针灸临床循证决策支持系统研究[J]. 中华中医药杂志 2011 26(4):795-797.
- [2] 毛爱民. 古代针灸治疗失眠处方配穴原则及规律[J]. 辽宁中医杂志 2005 32(5):463-464.
- [3] 段颖华,洪文. 原发性高血压针灸治疗规律探讨[J]. 陕西中医 2002 23(8):723-725.
- [4] Kamber J H M. 范明,孟小峰,译. 数据挖掘概念与技术[M]. 北京:机械工业出版社 2001:374.
- [5] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报 2008 19(1):48-61.
- [6] 姜园,张朝阳,仇佩亮,等. 用于数据挖掘的聚类算法[J]. 电子与信息学报 2005 27(4):655-662.
- [7] 叶冲轶. 高维海量数据联合聚类算法的研究与应用[D]. 杭州:浙江工商大学 2010.
- [8] Dumitrescu D, Lazzerini B, Jain L. Fuzzy Sets and Their Applications to Clustering and Training [M]. New york: CRC press 2002: 622. 2000.
- [9] Oh C H, Honda K, Ichihashi H. Fuzzy clustering for categorical multivariate data [C]. IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th. IEEE 2001 4: 2154-2159.
- [10] Miyamoto S, Mukaidono M. Fuzzy c-means as a regularization and maximum entropy approach [J]. IFSA 97 Prague Proceeding 1997 (2): 86-92.
- [11] 彭高辉,王志良. 数据挖掘中的数据预处理方法[J]. 华北水利水电学院学报 2008 (6):61-63.
- [12] 任玉兰,梁繁荣,吴曦,等. 针灸数据挖掘与临床决策[M]. 巴蜀书社 2010.
- [13] 刘明吉,王秀峰,黄亚楼. 数据挖掘中的数据预处理[J]. 计算机科学 2000 27(4):54-57.
- [14] 金君梅. 针刺治疗血管性痴呆选穴规律分析及主穴处方评价[J]. 中国中医药科技 2011, (5):371-372.
- [15] 张旸,李亚敏,董晓燕. 尺泽穴临床应用规律初探[J]. 针灸临床杂志 2007 22(12):40-42.
- [16] 裴景春,冯起国. 晋隋唐时期针灸处方配穴的原则及规律[J]. 中医药学报 2001 29(2):49-50.
- [17] 李锋刚,郝彦,许梁海. 数据挖掘技术在新安医学研究中的应用[J]. 安徽中医学院学报, 2006 24(6):12-14.

The Primary Acupoints Mining in Acupuncture and Moxibustion Prescription based on Co-clustering Algorithm

GUO Yan-zhen, SHU Hong-ping, ZHENG Jiao-ling

(College of Software Engineering ,Chengdu University of Infomation Technology , Chengdu 610225 ,China)

Abstract: In order to support the doctor to make policy decisions, the database of acupuncture prescription is analyzed to find the main acupoint from all the acupoints which is used to cure a disease. This paper clusters prescriptions and acupoints at the same time using the method of fuzzy co-clustering to identify prescriptions of similar acupoints and acupoints having similar prescriptions in the cluster. But this algorithm is insufficient to cluster prescriptions and acupoints better when prescription collection of acupuncture database is relatively large and categories are more. This paper proposes the use of improved FCR-PA algorithm to further fine-grained classification which can clearly identify the class which the prescriptions and acupoints belong to. The experiment analyzes 18 kinds of typical diseases of disease database, of which 15 types of diseases are consistent with clinical diagnosis showing the accuracy rate of 83.3%. Experiment results show that the improved method can achieve the main acupoint used to cure the disease effectively, providing valuable decision support for different doctor.

Key words: technology of computer application; data mining; co-clustering; acupuncture and moxibustion prescription; acupoints