基于聚类优化算法的中医治疗高血压用药规律分析

宋欣霞 金 卫

（山东中医药大学 济南 250355）

〔摘要〕 在介绍几种聚类算法概念的基础上，提出优化的 K－means 算法，将其用于分析 2 000 条中医治疗高血压临床数据，得到不同的证候所对应的 8 种用药组合。结果显示优化后的 K－means 算法在保证聚类质量的前提下，提高了算法运算速率，在性能上有显著的优越性。

〔关键词〕 K－means 算法；高血压；集群；迭代；用药规律

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2016.11.013

Analysis on the Drug Rule for Traditional Chinese Medicine Treatment of High Blood Pressure Based on Clustering Optimization Algorithm

SONG Xin－xia, JIN Wei, Shandong University of Chinese Medicine, Jinan 250355, China

〔**Abstract**〕 Based on the introduction of the concepts of several clustering algorithms, the paper puts forward the optimized K－means algorithm, applies it to the analysis on 2 000 pieces of clinical data about Traditional Chinese Medicine (TCM) treatment of high blood pressure, and gets 8 corresponding drug combinations of different symptoms. Through the result, on the premise of guaranteeing clustering quality, the optimized K－means algorithm improves the arithmetic speed of the algorithm, and is provided with obvious superiority of performance.

〔**Keywords**〕 K－means algorithm; High blood pressure; Cluster; Iteration; Drug rule

1 引言

近年来聚类分析在数据分析中的应用越来越广泛，其既可以作为独立的工具来分析数据的分布情况，也可以作为其他算法的预处理方法，还可以作为离散点的检测等。本文中阐述一种优化的聚类算法，用于探究中医治疗高血压的用药规律。

〔修回日期〕 2016－08－23

〔作者简介〕 宋欣霞，硕士研究生，发表论文 1 篇；通讯作者：金卫。

2 概念简介

2.1 聚类算法

聚类分析又称群分析，是研究样品或指标分类问题的一种统计方法，同时也是数据挖掘的一个重要算法[1]。聚类分析中一个主要的任务是集群分析，目的是帮助用户了解在一个数据集中的自然分组或结构，减少全凭主观判断所造成的误差，使分析结果更具客观性、合理性。近年来聚类算法在中医药领域得到广泛应用[2]，如将层次聚类应用到中药数据库中[3]，用二元数据的 Jaccard 系数计算两药物间的相异度，对治疗糖尿病的中药的性味进行

聚类，中风病急性期中医证候多元分析，治疗心血管病用药规律的 Varclus 聚类分析，基于聚类分析的银屑病中医证候研究，通过聚类分析探究针灸治疗类风湿性关节炎的主穴运用规律等。聚类分析方法包含许多种，具体使用哪一种应根据研究需要慎重选择。

2.2 K－means 算法

在真实的 D 维空间中给定 n 个数据点，确定一组中心点 k，以便得到从每个数据点到其最近的中心最小均方距离。试图找到邻近的 K－means 算法标准最优解。K－means 算法可以被认为是一种梯度下降程序[4]，它起始于聚类中心，并且迭代地更新这些中心从而减少方程式 $\frac{1}{n}\sum_{i=1}^{n}[\min d^2(x_i, m_j)]$ 中的目标函数。K－means 算法最终收敛到局部最小值[5]。K－means 算法的优势是方便性和可实施性、可扩展性、速度收敛性和稀疏数据的适应性。

2.3 优化的 K－means 算法

为了使得 K－means 算法更有效，尤其是对大数据集进行聚类[6]，提出了改进的 K－means 算法。因为在每次迭代中，K－means 算法都要计算数据点和所有中心之间的距离，对于大数据来说这种计算是相当耗时的。为解决此问题，可以从先前迭代的 K－means 算法中获益[7]。对于每个数据点来说，保留它到最近的簇之间的距离，在下一个迭代中，只计算它到先前的集群的距离。如果新的距离小于或等于先前的距离，该点保持在该集群中，这样就没有必要计算它到其他群集中心的距离了。节省来计算到 K－1 聚类中心距离的时间。

K－means 算法发现球形星团，它的中心点就是所述集群的点的重心，该中心作为被添加或移除的新的点移动。这种运动使中心更接近于一些点，与其他点分开相距得更远、与中心靠得更近的点就留在该集群中，所以也没有必要找到它到其他集群中心的距离。那些距离中心远的点可能会改变集群，所以只计算这些点到其他集群中心的距离，将其分配给最近的中心。在这种思想中，有两个距离函数可以支撑算法功能：第 1 个函数是 K－means 算法的基本函数，即为每个数据点寻找最近的中心点，通过计算到 k 个中心的距离，和为每个数据点保留其到最近的中心点的距离；第 2 个函数是通过比较距离大小分配数据点的归属。

本篇文章所执行 K－means 算法是基于上面提到的两种函数的。对于 K－means 算法有两种实现[8]。在第 1 种实现中，这两个函数都涉及执行的次数，它们之间存在着重叠，称为“重叠 K－means”算法；在第 2 种实现中，distance（）函数被执行两次，而 distance_ new（）函数执行迭代的提醒。被称为是“改进的 K－means”算法。

3 试验分析

3.1 数据来源

山东中医药大学附属医院心血管门诊 2 000 例高血压医案（其中单纯高血压 1 571 份，合并其他疾病 429 份），患者共 1 378 例，男患者 516 例，女患者 862 例，年龄在 2～86 周岁，就诊次数在 1～14 次。

3.2 试验方法

将医案信息输入电子病历，包括症状、证候、疾病的中医诊断、疾病的西医诊断等。导入中医医案分析系统，根据数据的规范化原则（高血压、高血压病、血压高、原发性高血压、一级高血压、早期高血压等统一为高血压），对数据进行预处理，转化为计算机处理的数据单元，使之规范、准确和有序，实现数据的正确表达和合理组织。分别对数据进行 K－means 聚类、改进 K－means 算法聚类，然后分析聚类结果。

3.3 试验结果（表 1 – 表 3）

表 1 聚类结果统计

症型	例数	用药组合
肝阳上亢或热毒型	621	（1）钩藤，黄连，黄答，黄柏，丹皮，栀子，泽泻，荻荃，稀签草，野葛根
肝肾阴虚型	237	（2）杜仲，牛膝，制首乌，女贞子
高血压病兼有血癖	119	（3）当归，川芍，元胡，丹参，生地
高血压病伴有气阴两虚型冠心病	489	（4）黄茂，麦冬，五味子，三七粉，冰片
高血压病伴有失眠	518	（5）炒枣仁，夜交藤，紫石英
高血压病伴有高血脂病	355	（6）草决明
高血压病伴有快速性心律失常	23	（7）青蒿
高血压病伴有腹痛、腹胀	35	（8）木香

表 2 不同算法产生最终簇执行时间

名称 \ 例数 / 时间（s）	50	100	200	400	600	800	1 000	1 200	1 400	1 600	1 800	2 000
K – means 算法	0	5	7	21	13	18	27	30	40	60	55	50
重叠的 K – means 算法	0	5	6	11	10	7	12	18	20	21	22	20
改进的 K – means 算法	0	4	5	5	6	6	7	8	10	10	9	10

表 3 不同算法的实现产生簇的质量

名称 \ 例数 / 时间（s）	50	100	200	400	600	800	1 000	1 200	1 400	1 600	1 800	2 000
K – means 算法	2	5	10	13	15	15	17	19	22	22	23	25
重叠的 K – means 算法	2	5	9	11	15	16	17	18	20	21	22	25
改进的 K – means 算法	2	4	10	10	15	14	15	20	20	22	22	15

4 结论

4.1 聚类结果分析

从表 1 可以看出聚类最终产生 8 种辩证用药规律组合。药组（1），这一组合的药均有降压的功效，其中黄连可清热燥湿、泻火解毒；丹皮、泽泻均可清热凉血、活血散瘀，因此此类药组可治疗肝阳上亢、热毒型高血压病。药组（2），其中杜仲有补益肝肾、强筋壮骨、调理冲任、固经安胎的功效；牛膝入药有逐瘀通经、补肝肾、强筋骨、利尿通淋等效用；制首乌可补益精血、养肝安神、强筋骨；女贞子是一味补肾滋阴、养肝明目的中药，可治肝肾不足、头晕耳鸣、头发早白及两目昏糊等病症。因此此类药物组合可治疗肝肾阴虚型高血压病。药组（3），此类药组可治疗血虚诸证、月经不调、经闭、痛经、症瘕结聚等。因此用此药组治疗高血压病兼有血癖。药组（4），此药组可直接扩张外周血管，降低外周阻力，从而降低血压，同时可养阴生津、润肺清心，用于治疗内热消渴、心烦失眠、肠燥便秘。因此此药组可治疗气阴两虚型高血压病。药组（5），此药组具有镇心、安神、降逆气的功效，可用于缓解失眠。药组（6），草决明可润肠通便、降脂明目，治疗便秘及高血脂、高血压。清肝明目、利水通便，有缓泻作用，用于治疗高血压高血脂症。药组（7），青蒿具有清热解暑、除蒸、截疟的功效，用于暑邪发热、阴虚发热等。因此加入此药可治疗高血压病伴随心率过快的症状。药组（8），木香具有行气止痛、调中导滞的功效，可用于治疗胞胁胀满足、脘腹胀痛。因此此药组可用于治疗高血压病伴随腹胀的症状。

4.2 通过不同聚类方法均能得到可靠结果

通过上述对药组的分析[9]得知，针对证型所用的中药是合理的，说明通过不同的聚类算法得到的最终结果也是可靠的。通过表 2 可知，在个体数相同的前提下来说，改进的 K－means 算法所用的时间较少。表 3 是在相同聚类个体数的前提下，比较均方误差的不同值来判断聚类结果的质量，可以看出 3 种聚类算法在聚类质量上是相差不大的。所以综合来看，重叠 K－means 算法和改进的 K－means 算法在时间上都比原始算法快，在聚类质量上和原始算法相差不大，但是改进的 K－means 算法在时间上更有优势，因此说改进的 K－means 算法是很有效的。

5 结语

本文通过对聚类算法的描述与概括，简单了解 K－means 算法的原理，从而提出优化的 K－means 算法；并且通过探究中医辩证治疗高血压病用药规律的试验，证实改进后的算法是切实可行的。在不影响聚类质量的案例中，所提出的这种优化方案可以改善 K－means 算法的执行时间，从而提高聚类效率。

参考文献

1 郭军华. 数据挖掘中聚类分析的研究［D］. 武汉：武汉理工大学，2003.

2 唐东明. 聚类分析及其应用研究［D］. 成都：电子科技大学，2010.

3 杨小兵. 聚类分析中若干关键技术的研究［D］. 杭州：浙江大学，2005.

4 杨淑莹. 模式识别与智能计算：Matlab 技术实现［M］. 北京：电子工业出版社，2008.

5 雷小锋，谢昆青，林帆，等. 一种基于 K－means 局部最优性的高效聚类算法［J］. 软件学报，2008，19（7）：1683－1692.

6 Hansen P，Jaumard B. Cluster Analysis Andmathematieal Programming［J］. Math Progrmmaing，1997，（79）：191－215.

7 张建萍，刘希玉. 基于计算智能技术的聚类分析研究与应用［D］. 济南：山东师范大学，2014.

8 Berkhin P. Survey of Clustering Data Mining Techniques［EB/OL］.［2016－01－10］. http：//www. ee. ucr. edu/～barth/EE242/clustering_ survey. pdf.

9 钱海燕. 关于降压中药的研究［J］. 中华实用中西医杂志，2007，20（18）：1569－1571.

（上接第 37 页）

4 Steel DH，Parkes C，Papastavrou VT，et al. Predicting Macular Hole Closure with Ocriplasmin Based on Spectral Domain Optical Coherence Tomography［J］. Eye，2016，30（5）：740－745.

5 Rahimy E，Mccannel C A. Impact of Internal Limiting Membrane Peeling on Macular Hole Reopening：a systematic review and Meta－analysis［J］. Retina，2016，36（4）：679－687.

6 Schneider EW，Todorich B，Kelly MP，et al. Effect of Optical Coherence Tomography Scan Pattern and Density on the Detection of Full－Thickness Macular Holes［J］. American Journal of Ophthalmology，2014，157（5）：978－984.

7 Matet A，Savastano MC，Rispoli M，et al. En Face Optical Coherence Tomography of Foveal Microstructure in Full－Thickness Macular Hole：a model to study perifoveal müller cells［J］. American Journal of Ophthalmology，2015，159（6）：1142－1151.

8 Dilraj SG，Varun R，Tamer HM. Assessment of Foveal Microstructure and Foveal Lucencies Using Optical Coherence Tomography Radial Scans Following Macular Hole Surgery［J］. American Journal of Ophthalmology，2015，160（5）：990－999.

9 孟娟娜. 基于 Android 平台的移动电子商务系统设计与实现［J］. 电子设计工程，2016，24（8）：27－29，33.

10 Lee Y，Yoo H. Three－dimensional Visualization of Objects in Sscattering Medium Using Integral Imaging and Spectral Analysis［J］. Optics & Lasers in Engineering，2016，77（7）：31－38.

11 West S，Wagner M，Engelke C，et al. Optical Coherence Tomography for the in Situ Three－dimensional Visualization and Quantification of Feed Spacer Channel Fouling in Reverse Osmosis Membrane Modules［J］. Journal of Membrane Science，2015，498（9）：345－352.

12 Cho CW，Hong CP，Piao JC，et al. Performance Optimization of 3D Applications by OpenGL ES Library Hooking in Mobile Devices［C］. International Conference on Computer and Information Science，2014：471－476.