

基于改进 K-NN 和 SVM 的多学科协作诊疗决策支持系统^①



李晓峰¹, 王妍玮², 李 东³

¹(黑龙江外国语学院 信息工程系, 哈尔滨 150025)

²(普渡大学 机械工程系, 西拉法叶市 IN47906)

³(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要: 由于当前的诊疗决策支持系统采用单一学科的决策方法, 导致诊疗精度不高, 获取的数据分类结果准确率较低, 提出并设计一种基于改进 K-NN (K-Nearest Neighbour) 分类算法和 SVM (Support Vector Machine) 的多学科协作诊疗决策支持系统. 在构建系统总体框架的基础上, 对数据库系统模块、人机交互模块和诊疗推理模块进行设计, 其中诊疗推理模块是系统的软件核心, 通过改进 K-NN 分类算法和 SVM 建立推理引擎, 在计算机的辅助下, 搜索与患者病症信息相似的医疗案例, 并进行相似度匹配, 根据匹配结果与患者症状集构建一个新的临床案例, 引入 CDA (Clinical Document Architecture) 概念, 实现改进 K-NN 分类算法和 SVM 算法的有效融合, 完成多学科协作诊疗决策. 实验结果表明, 与传统系统相比, 该系统的诊疗决策精度高, 评价指标测试平均值达到 95.98%, 分类结果准确率较高, 在该系统辅助下能提高医生诊断正确性, 降低误诊率, 且运算复杂度较低.

关键词: 改进 K-NN 分类算法; SVM; 多学科协作; 诊疗决策支持系统

引用格式: 李晓峰, 王妍玮, 李东. 基于改进 K-NN 和 SVM 的多学科协作诊疗决策支持系统. 计算机系统应用, 2020, 29(6): 80-88. <http://www.c-s-a.org.cn/1003-3254/7466.html>

Multidisciplinary Collaborative Diagnosis and Treatment Decision Support System Based on Improved K-NN and SVM

LI Xiao-Feng¹, WANG Yan-Wei², LI Dong³

¹(Department of Information Engineering, Heilongjiang International University, Harbin 150025, China)

²(Department of Mechanical Engineering, Purdue University, West Lafayette, IN47906, USA)

³(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Because the current diagnosis and treatment decision support system adopts a single subject decision-making method, resulting in low diagnosis and treatment accuracy and low accuracy of the obtained data classification results, a multi-disciplinary collaborative diagnosis and treatment decision support system based on improved K-Nearest Neighbor (K-NN) classification algorithm and Support Vector Machine (SVM) is proposed and designed. Based on the overall framework of the system, the database system module, human-computer interaction module, and diagnosis and treatment reasoning module are designed. The diagnosis and treatment reasoning module is the software core of the system. The reasoning engine is established by improved K-NN classification algorithm and SVM. With the help of computer, medical cases similar to the patient's disease information are searched, and similarity matching is carried out. According to the matching results, a new clinical case is constructed based on patient symptom set. The concept of Clinical Document

① 基金项目: 国家自然科学基金 (61803117); 教育部科技发展中心产学研创新基金 (2018A01002); 国家科技部创新方法专项 (2017IM010500)

Foundation item: National Natural Science Foundation of China (61803117); Industry-University-Research Innovation Fund of Science and Technology Development Center, MOE (2018A01002); Special Fund for Innovation Methodology of Ministry of Science and Technology (2017IM010500)

收稿时间: 2019-11-17; 修改时间: 2019-12-01, 2019-12-25; 采用时间: 2020-01-02; csa 在线出版时间: 2020-06-10

Architecture (CDA) is introduced to realize the effective fusion of improved K-NN classification algorithm and SVM algorithm, and to complete the multi-disciplinary collaborative diagnosis and treatment decision. The experimental results show that, compared with the traditional system, the system has high accuracy in diagnosis and treatment decision-making, the average value of the evaluation index is 95.98%, and the accuracy rate of classification results is high. With the help of the system, it can improve the diagnosis accuracy of doctors and reduce the misdiagnosis rate, and the operation complexity is low.

Key words: improved K-Nearest Neighbor (K-NN) classification algorithm; Support Vector Mechine (SVM); multidisciplinary collaboration; diagnosis and treatment decision support system

1 引言

医疗机构在面对现今如此大量医患数量和复杂的人体结构时,诊疗工作面临巨大的挑战,误诊现象时有发生,因为病人病因有时并不会如临床表现信息那样准确,不仅耽误了患者病情,也给医生自身和所在医疗机构形象带来损害,由此造成了非常严重的后果.所以医生做出有效的诊断和治疗措施是一个复杂的决策和思维过程.在此背景下,利用计算机的处理能力辅助医生做出更为准确的诊断判断^[1].为此,相关专家研究了一种多学科协作诊疗决策支持系统.该系统的出现,在一定程度上提高了医生的诊疗精度和治疗效果,并减少了人为疏忽、降低医疗成本.

但是,由于我国在这方面研究起步较晚,技术还不是很成熟,诊断精度距离标准还有一段距离.文献[2]提出并设计了一种具有辨证论治内涵的智能中医诊疗决策系统,通过病证临床诊断、治疗、疗效评价决策方法,建立现代中医智能诊疗系统,利用人机结合优势,为中医临床诊疗提供智能决策辅助支持,探索创新中医病证诊疗模式.但是该系统的诊疗效果不理想;文献[3]提出并设计了基于心脏团队模式的多学科诊疗实施系统,通过多学科协作,对部分心脏疾病的诊疗策略、患者选择以及患者随访及管理均存在有益影响,目前多见于冠状动脉血运重建的诊疗决策.实施中存在误诊率较高的现象;文献[4]提出并设计了一种国内分级诊疗现状的评价系统,以“分级诊疗”为关键词检索自各数据库中的相关文献,全面收集国内分级诊疗现状的研究,从结局评价指标和面临的问题两方面进行系统评价;文献[5]提出并设计了基于系统复杂性的中医诊疗信息分层可视化系统,将中医诊疗过程分层,并结合现代数据可视化技术为不同层次的诊疗过程选

取可视化方法.但是上述两种系统的诊疗精度较低.一些发达国家经过长时间的分析与探讨,对医疗诊疗决策支持系统也进行了探索和研究.文献[6]提出并设计了一种基于 Web 的心血管分析、诊断和治疗系统,为用户提供完整的数据管理和动态血压监测,在医学报告中,便于用户理解,该系统证明了其在心血管临床研究中的巨大潜力.文献[7]提出并设计了基于大数据挖掘和云计算的疾病诊疗推荐系统,为了更准确有效地识别疾病症状,提出了一种基于密度峰值聚类分析(DPCA)的疾病症状聚类算法,利用 Apriori 算法分别对疾病诊断(DD)规则和疾病治疗(D-T)规则进行关联分析,为了达到高性能和低延迟响应的目标,使用 Apache Spark 云平台为 DDTRS 实现了一个并行解决的方案.但是以上两种系统在进行运算时,运算时间较长,分类效率较差.

针对上述问题,将改进 K-NN 分类算法和 SVM 应用其中,提出一种新的基于改进 K-NN 分类算法和 SVM 的多学科协作诊疗决策支持系统.该系统主要包括数据库系统模块、人机交互模块和诊疗推理模块,其中诊疗推理模块是系统的软件核心,通过改进 K-NN 分类算法和 SVM 算法建立推理引擎,根据推理结果构建一个新的案例,引入 CDA 概念,对改进 K-NN 分类算法和 SVM 算法进行有效融合,完成多学科协作诊疗决策.为测试本系统的诊疗精度,与传统诊疗决策支持系统进行仿真测试,测试结果表明:改进 K-NN 分类算法和 SVM 应用下的多学科协作诊疗决策支持系统的诊疗精度更高,由此说明本系统要比传统系统性能要好,更能帮助医生做出正确诊断,给患者更为准确的及时治疗建议,有利于保障人民身体健康,更有利于提高国民健康水平.

2 基于改进 K-NN 和 SVM 的多学科协作诊疗决策支持系统

由于人体结构的复杂性,所以患病原因也具有多样性、多变性、复杂性和不确定性,又因为人体结构的统一性,所以病理特征还具有动态演化性,因此在进行诊疗时,需要多学科协作的进行^[8,9].为此本次设计的基于改进 K-NN 和 SVM 的多学科协作诊疗决策支持系统要满足以下几点需求:

(1) 多方面

在对患者进行诊疗时,要多方面的考察患者生理信息,如病症表现信息(症状、生命体征等)、病症隐藏信息(心电图、彩超、CT等),才能提升诊疗精度,为此,系统设计需要支持多学科协作诊断.

(2) 及时、准确

当面对发病快的患者,并不允许医生缓慢诊断,所以为辅助医生在较短时间内做出正确的诊断决策,基于改进 K-NN 分类算法和 SVM 的多学科协作诊疗决策支持系统需要快速、及时、准确的提供多方面信息^[10,11].

(3) 完整性

医生做出诊断是一个严肃的过程,所以在完整详细诊疗记录的基础上,给出的诊断不仅是一个结果,更是一个详细的诊断推理过程,所以系统设计要满足信息完整这一需求.

(4) 动态更新

病症、医疗方法等并不是一成不变的,所以为满足需求,系统需要拥有动态更新作用,才能根据患者病情变化以及不断进步的医疗技术,帮助医生给出更为科学、合理的诊断结果以及治疗建议^[12].

2.1 系统总体框架

基于改进 K-NN 分类算法和 SVM 的多学科协作诊疗决策支持系统设计,分为数据库系统模块、人机交互模块和诊疗推理模块^[13].多学科协作诊疗决策支持系统总体框架如图 1 所示.

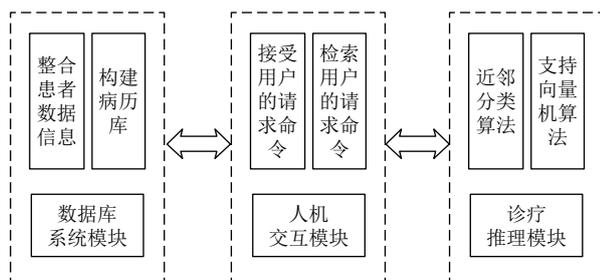


图 1 系统总体框架图

根据图 1 可知,数据库系统模块实际上是一个数据库系统,包括患者数据库和源病历库两个.人机交互模块是系统人机交互界面的窗口,负责接收和检索用户的请求命令,调用系统功能为诊疗决策服务,以及将最终的诊疗结果报告给医生.诊疗推理模块是系统的软件核心,主要作用是通过改进 K-NN 分类算法和 SVM 建立推理引擎,并在计算机的辅助下,在医院源病历库中搜索与患者病症信息相似的医疗案例,并进行相似度匹配,建立患者病症问题与病历库的关联,针对相似案例和患者病症问题相匹配,不需要进行修改;如果两者不匹配,则需要对案例修正,调整诊疗决策方案,最后将得出诊疗结果与患者症状集构建一个新的临床案例,更新源病历库^[14].

通过改进 K-NN 分类算法和 SVM 建立推理引擎过程中,由于 SVM 对于初始参数设置较为敏感,为保证推理引擎建立结果的可靠性,本文利用遗传算法优化初始参数,遗传算法结合了适合生存的思想,通过选择、交换、变异等作用机制实现种群进化^[15].本文利用遗传算法优化初始参数过程中,在解空间同时建立多个初始点,通过建立适应度函数寻找搜索方向,采用实时、同步搜索的方式,能够快速寻找出最优初始参数.

2.2 数据库系统模块

数据库系统模块是系统的基础模块,一方面负责整合患者数据信息,如病人的基本信息、病历信息、病程信息、医嘱信息、检验信息、影像信息、护理信息,以及其他所需要的各类信息等,另一方面负责构建病历库,为医生诊疗决策提供数据支撑.

数据库是为企业进行科学、合理的战略决策创建的数据支持的集合,具有指导业务流程改进、成本、质量以及控制的作用^[16].数据库系统架构主要分为 3 个部分:数据抽取、数据储存以及数据显示.根据实际应用需求,在系统测试后,最后确定相应的配置.

在数据库当中,ETL 是其中的核心和灵魂,负责将大量异构数据从数据源中提取出来,并进行清洗、转换和加工,最后将其整合到一起,按照一定的规则装载到一个大的数据仓库当中,完成数据存储,为软件中的诊疗推理做准备,如图 2 所示.

2.3 人机交互模块

人机交互模块是系统唯一与用户联系的窗口,负责接收和检索用户的请求命令以及将最终的诊疗结果显示给用户.在这里选择 ELAN SMART-PAD 智能触控

面板作为显示与搜索设备. 该设备服务质量很高, 不仅运行速度快, 显示功能也更全面.

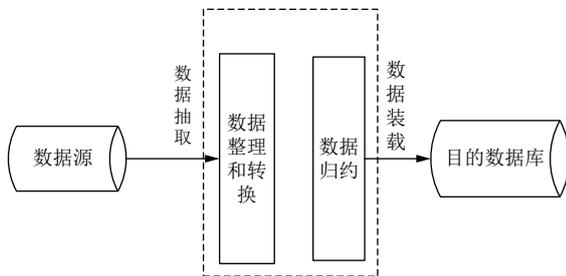


图2 ETL运行架构

2.4 诊疗推理模块

利用得到的患者数据信息和病历仓库中的信息进行匹配, 检索出相似的病历, 并以此推理出目标病历的解, 其实质是计算患者病历与病历仓库临床案例之间的相似度来实现患者病因的诊断, 因此相似度算法的选择直接关系到诊断结果的准确性^[17]. 常见的相似度计算方法主要有最近邻分类算法、归纳推理法以及支持向量机(SVM). 然而, 以上计算方法在实际应用过程中, 人们发现得到的结果与真实结果有时存在一定的差异, 所以在给出的诊疗决策方案后还需要进一步的修正和调整, 付出的计算量将是巨大的, 为解决这一问题, 将近邻分类算法(K-NN)进行改进, 并与支持向量机(SVM)相结合, 弥补各自的缺点, 提高诊断的准确性.

K-近邻(简称K-NN)属于分类算法, 是一个理论上比较成熟的算法, 也是最简单的机器学习算法之一. 改进该算法的思路是: 首先对样本进行预处理, 计算数据属性的信息增益, 确定各数据属性的权重因子, 组成一个属性空间. 在这一属性空间内, 给出一个待分类的样本 x (病人数据), 计算 x 与训练集(病例库)中每个文本的距离, 对距离值进行排序, 找出最近的 k 个最相似病历, 若这些相似病历属于某一个类别, 则该样本也属于这个类别^[18]. 计算过程如下:

计算数据属性的信息增益:

$$Gain(x) = E(S) - E(S|x) \quad (1)$$

式(1)中, $E(S)$ 表示数据整体熵, $E(S|x)$ 表示增加的数据熵.

根据计算得出的信息增益, 计算数据属性的权重因子, 计算公式如下:

$$\xi = \frac{Gain(x)}{\sum_{x=1}^n Gain(x)} \cdot e^{Gain(x)} \quad (2)$$

根据式(2)找出 n 个权重因子最大的数据属性, 构建一个 n 维属性空间, 在 n 维属性空间中, 计算样本与类别之间的相似度和, 如下:

$$y(x, S_i) = \sum_{d_i} sim(x, d_i) y(d_i, S_i) - b_j \quad (3)$$

式(3)中, $y(x, S_i)$ 为样本 x 与类别 S_i 之间的相似度和; $sim(x, d_i)$ 表示 x 与训练文本 d_i 之间的相似度; $y(d_i, S_i)$ 代表训练文本 d_i 与类别 S_i 之间的相似度; b_j 为类别 S_i 的阈值.

具体过程如下:

步骤1. 计算待分类样本与训练集之间的距离, 计算方法主要有欧几里得、曼哈顿、马氏距离等3种.

(1) 欧几里得距离

$$d(x, y) = \sqrt{(x_1, y_1)^2 + (x_2, y_2)^2 + \dots + (x_n, y_n)^2} \quad (4)$$

转换为:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i, y_i)^2} \quad (5)$$

式(5)中, i 为 (x, y) 坐标上的任意一点, (x_i, y_i) 为任意一点 i 上的坐标, n 表示点个数的总量.

(2) 曼哈顿距离

在平面上, 坐标 (x_1, y_1) 的点 $P1$ 与坐标 (x_2, y_2) 的点 $P2$ 的曼哈顿距离为:

$$d(x, y) = \sum_{i=2}^n |(x_i, y_i)| \quad (6)$$

(3) 马氏距离

假设样本点为:

$$A = (x_1, x_2, \dots, x_N)^T \quad (7)$$

式中, T 表示转置符号. 数据集分布的均值 f 为:

$$f = \frac{(x_1 + x_2 + \dots + x_N)^T}{N} \quad (8)$$

则样本点元素 x_i 与均值 f 之间的马氏距离 $d(x_i, f)$ 为:

$$d(x_i, f) = \sqrt{(x_i - f)^T \sigma (x_i - f)} \quad (9)$$

式中, σ 为均值权重。

上述 3 种距离计算方法中, 欧几里得距离有时不能满足实际需求, 没有考虑到总体样本数据变异对距离远近的影响; 马氏距离在绝大多数情况下是可以顺利计算的, 但由于协方差矩阵的影响, 马氏距离的计算不稳定; 而曼哈顿距离依赖坐标系统的转度进行距离测算, 可以应用于多种类型数据的计算。因此, 本文计算待分类样本与训练集之间的距离时, 更适合使用曼哈顿距离计算。

步骤 2. 按距离递增次序排序;

步骤 3. 选取与当前点距离最小的 k 个病历;

步骤 4. 统计前 k 个病历所在的类别出现的频率;

步骤 5. 返回前 k 个病历出现频率最高的类别作为当前点的预测分类。

由于在预分类过程中, 多学科决策过程面临大量数据, 导致分类过程陷入不收敛的境地, 故采用支持向量机 (SVM) 算法解决该现象。支持向量机 (SVM) 是一种可以训练的机器学习算法, 简单的说是一个分类器, 并且是二类分类器^[19]。其定理描述如下:

将诊疗数据分为一个超平面:

$$(w, x) + b = 0, w, b \in R^j \quad (10)$$

式中, R^j 为 j 维超平面; (w, x) 表示内积; w 为可加权值向量; b 为偏置, 超平面相对原点的偏移。

对式 (8) 进行归一化处理:

$$((w, x)_i + b)r \geq 1, i = 1, 2, \dots, n \quad (11)$$

式 (9) 中, r 表示分类间隔, 其满足条件为:

$$\begin{cases} r_{\min} = \frac{2}{w} \\ r_{\max} = \max\{w^2\} \end{cases} \quad (12)$$

根据分类间隔满足条件可知, 如果分类间隔呈现最小值, 则呈现线性可分状态, 是因为利用非线性映射, 将低维特征空间线性不可分的模式转换为高维特征空间线性可分的模式; 如果分类间隔呈现最大值, 则对高维特征空间进行分类或回归, 但是在高维特征空间运算时会出现多维现象, 致使在分类预处理的过程需要重新进行计算, 存在计算时间长、运算过程复杂等因素。由此, 利用核函数技术可以快速的解决该问题。达到数据降维的效果。对于核函数的选择, 缺乏统一的指导规则, 不同的核函数适用于不同的领域研究, 但多项实验结果表明, 径向基核函数具有较好的适用性, 不会

出现太大偏差, 降维效果好, 且能有效保障有效信息的完整性, 其中高斯核函数是一种非常具有代表性的径向基核函数, 相较于线性核、多项式核函数, 高斯核函数只有一个参数, 容易选择, 决策边界更为多样, 且高斯核函数能够映射到无限维, 优势明显, 因此采用高斯核函数进行数据降维。计算公式如下:

高斯核函数:

$$O(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{a}\right) \quad (13)$$

式中, a 为参数。

数据降维中, 为保证有效信息的完整性, 整个降维过程遵循两个原则: 一个是最近重构性, 使用降维后的数据重构原始数据时误差最小; 二是在低维空间中数据尽量分开。

利用高斯核函数完成数据降维后, 融合改进 K-NN 与 SVM, 来提高诊疗准确性。CDA (HL7 临床文档结构 Clinical Document Architecture) 是以交换文档为目的, 指定结构和语意的文档标记标准。CDA 文档由头和体构成, 头是对文档进行分类, 包括文档信息、服务提供商和服务对象, 包括受访数据、患者等; 体是临床报告, 由可插入的内容框架构成。分为 4 部分: 节, 段, 列表, 表格。HL7 临床文档是一个具有法律效应的临床信息集合^[20]。CDA 能够借助于 XML、HL7 的参考信息模型和词汇编码表, 利用机器处理分析病历文档, 分析结果既能够被电子检索, 也能够被人阅读。引入 CDA 概念, 能够通过机器处理实现改进 K-NN 分类算法和 SVM 算法的有效融合, 以此优化多学科协作诊疗决策支持系统, 并提升该系统的有效性。

设 T 为测试集, Q 为 T 的相应类别集合, 待测分类样本为 x , 给定的改进 K-NN 算法针对每个类别的可信度阈值为 $e_i (e_i > 0), i = 1, 2, \dots, n$, 改进 K-NN 与 SVM 融合算法描述如图 3 所示。

根据图 3 可知, 首先利用改进 K-NN 算法对待分类的样本 (病人数据) 进行预分类, 将改进 K-NN 全部输出类别作为候选, 采用 SVM 分类器对待分类的样本进行分类, 根据模式识别理论会出现两种情况, 一种是线性可分, 另一种是“维数灾难”, 需要返回分类预处理阶段重新开始计算, 但是采用核函数技术可以进行降维处理; 然后将 CDA 概念引入其中; 最后融合改进 K-NN 算法和 SVM 算法, 完成多学科协作诊疗决策。

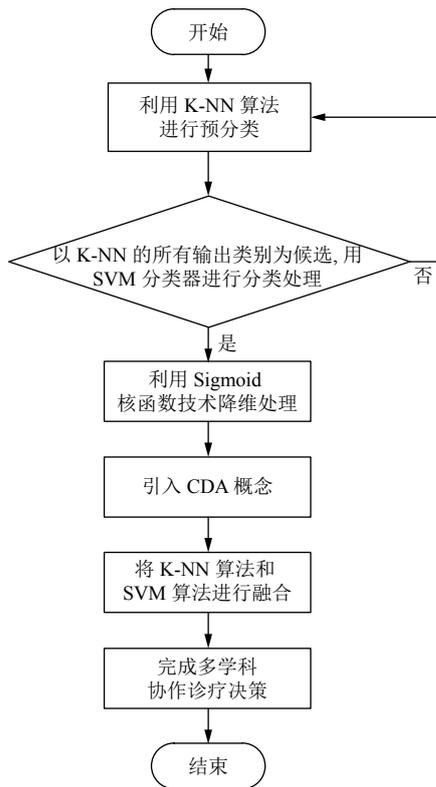


图3 改进 K-NN 分类算法和 SVM 融合推理流程

3 系统仿真测试

传统诊疗决策支持更多的是依赖医生自身的医学知识和多年积累的临床诊断经验,然后在系统帮助下进行医疗诊断的.目前主要传统系统主要有基于辨证论治内涵的智能中医诊疗决策系统、基于心脏团队模式的多学科诊疗实施系统、国内分级诊疗现状的评价系统、基于系统复杂性的中医诊疗信息分层可视化系统、基于 Web 的心血管分析、诊断和治疗系统和基于大数据挖掘和云计算的疾病诊疗推荐系统等.为了验证本文系统的有效性,采用本文系统、文献[2-7]的系统进行对比实验.

3.1 实验环境

本文实验环境配置信息如表 1 所示.

表 1 开发环境配置信息

名称	配置情况
操作系统	Microsoft Windows 10
处理器	Intel(R)Core(TM)i5-2410m
内存	6.0 GB
硬盘	4.0 GB
数据库管理软件	微软 SQL server 2010 R2
JDK	1.6
数学软件	Matlab

3.2 实验数据集

实验数据来源于药度数据库 (<https://data.pharmacodia.com/>),共采集数据 5000 个,进行 50 组实验,每次使用 100 个数据.

本实验选用 587 患者,其中男患者 355 人,女患者 232 人,包括儿童、青年、中年、老年等各个年龄段,进行测试.

3.3 实验指标

(1) 诊疗精度:以往精度主要体现在准确率、召回率,但是这是两个分裂质量的两个不同方面,所以更为好的精度评价就是将二者集合考虑,不可偏颇,所以产生一种新的评价指标 $F1$ 测试值, $F1$ 测试值是用来衡量二分类模型精确度的一种指标, $F1$ 测试值可以看作是模型准确率和召回率的一种加权平均数,公式如下:

$$F1 = \frac{N_1 \cdot N_2}{N_1 + N_2} \quad (14)$$

式中, N_1 为准确率, N_2 为召回率.

(2) 数据分类结果准确率:当多学科决策过程面临多数据时,分类过程会完全陷入不收敛的境地,严重降低了分类结果,所以本文采用支持向量机 (SVM) 算法解决此问题,能够提高分类结果准确率.

(3) 高维特征空间运算时间:高维特征空间运算是多学科协作诊疗决策过程中最重要的运算过程,能够对分类效率产生影响,运算时间越快,分类效率越高.本文使用正文中核函数技术,与文献[2-5]系统对高维特征空间运算时间进行对比.

(4) 误诊率:较多多学科协作诊疗决策支持系统会借助计算机的处理能力辅助医生做出更为准确的诊断判断,同时也会出现误诊率,大大降低了诊疗效果,为此采用本文系统、文献[4-7]系统进行对比分析.

(5) 相似度匹配:利用得到的患者数据信息和病历仓库中的信息进行匹配,检索出相似的病历,并以此推理出目标病历的解,其实质是计算患者病历与病历仓库临床案例之间的相似度,来实现患者病因的诊断,相似度匹配率越高,其诊断结果的准确性就越高.

(6) 算法复杂度分析:改进 K-NN 分类算法和 SVM 算法自身都需一定的运算复杂度,对两者结合以后的整体运算复杂度进行综合分析,能够测试本文算法的实际应用性能.因此,以算法复杂度为指标进行分析,从时间复杂度和空间复杂度两方面入手分析,算法时间复杂度的计算公式如下:

$$t = t_1\alpha_1 + t_2\alpha_2 + t_3\alpha_3 + t_4\alpha_4 + \dots \quad (15)$$

式中, t 表示算法运行总时长, t_1 、 t_2 、 t_3 、 t_4 表示算法运行中各步骤操作所需时间, α_1 、 α_2 、 α_3 、 α_4 表示各步骤对应的运算次数。

空间复杂度通常与算法运行次数相关, 假设算法的平均运算次数为 $\alpha' = \frac{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}{4}$, 则空间复杂度可表示为 $O((\alpha')^2)$, 因此算法复杂度的计算公式可表示为:

$$t' = t + O((\alpha')^2) \quad (16)$$

3.4 测试结果

从表2中可以看出, 利用本系统对587患者进行诊断, 其F1测试平均值为95.98%, 而在文献[4-7]系统应用下, F1测试平均值分别为54.14%、84.4%、75.2%和58.1%。

表2 系统诊疗精度测试结果 (%)

实验次数	改进 K-NN 分类算法和 SVM	文献[4]	文献[5]	文献[6]	文献[7]
10	92.68	50.33	82.4	71.50	62.30
20	96.62	54.35	84.64	72.66	63.00
30	95.30	55.37	86.32	73.55	63.98
40	98.30	54.68	80.00	76.45	50.80
50	93.66	56.77	85.33	78.10	68.12
60	99.29	53.32	88.00	79.00	40.53
平均值	95.98	54.14	84.40	75.20	58.10

为了进一步验证本文系统的有效性, 采用本文系统、文献[2-5]系统对数据分类结果的准确率进行对比分析, 对比结果如图4所示。

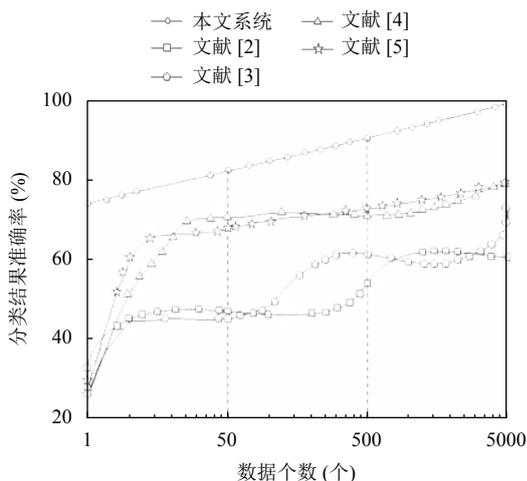


图4 数据分类结果准确率对比

根据图4可知, 本文系统的数据分类结果准确率在72%~95%之间, 呈上升趋势; 文献[2]系统的数据分

类结果准确率在32%~60%之间; 文献[3]系统的数据分类结果准确率在26%~70%之间; 文献[4]系统的数据分类结果准确率在27%~72%之间; 文献[5]系统的数据分类结果准确率在29%~73%之间. 说明本文系统的数据分类结果准确率比文献[2]系统、文献[3]系统、文献[4]系统和文献[5]系统的分类结果准确率高, 是因为多学科决策过程面临多数据时, 分类过程容易陷入不收敛的境地, 导致分类结果准确率较低, 而本文系统在K-NN分类算法的基础上, 利用支持向量机(SVM)算法解决不收敛的问题, 提高了分类结果准确率。

为了验证本文系统的有效性, 采用本文系统、文献[2-5]系统对高维特征空间运算时间进行对比分析, 对比结果如图5所示。

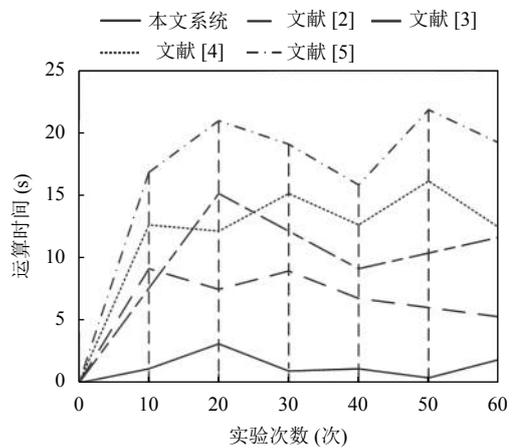


图5 运算时间对比结果

根据图5可知, 采用本文系统对高维特征空间进行计算, 其运算时间在5s以下; 采用文献[2]系统对高维特征空间进行计算, 其运算时间在10s以下; 采用文献[3]系统对高维特征空间进行计算, 其运算时间在15s以下; 采用文献[4]系统对高维特征空间进行计算, 其运算时间在16s以下; 采用文献[5]系统对高维特征空间进行计算, 其运算时间在21s以下. 采用本文系统对高维特征空间进行计算, 其运算时间比传统系统的运算时间短, 是因为本文系统采用中核函数技术做降维处理, 从而提高了分类效率。

为了验证本文方法的误诊率, 采用本文系统、文献[4-7]系统进行对比分析, 对比结果如图6所示。

根据图6可知, 本文系统的误诊率随着实验次数的增长而逐渐降低, 其误诊率在30%以下; 文献[4]系统的误诊率在50%以下; 文献[5]系统的误诊率在

75% 以下; 文献[6]系统的误诊率在 80% 以下; 文献[7]系统的误诊率在 61% 以下. 本文系统的误诊率比传统系统的误诊率低, 说明本文系统具有较高的诊疗效果.

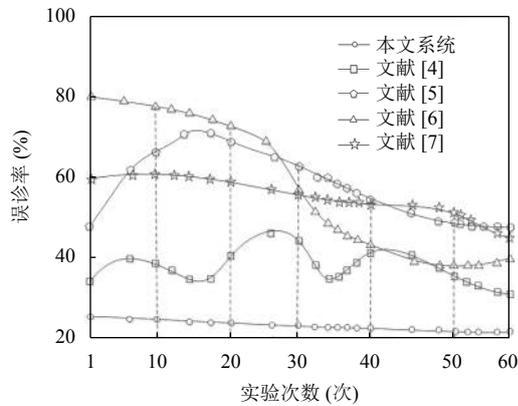


图6 误诊率对比结果

为了进一步验证本文系统的有效性, 采用本文系统、文献[2-5]系统, 对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配结果如图7所示.

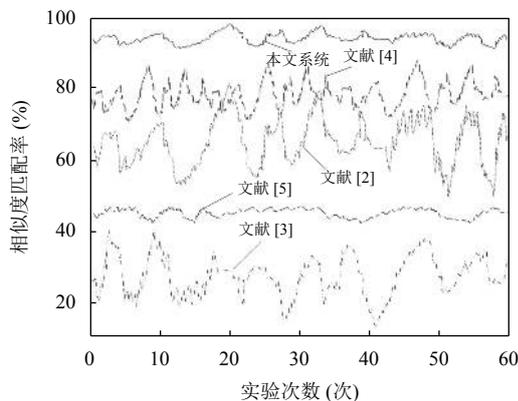


图7 相似度匹配率对比结果

根据图7可知, 采用本文系统对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配率在 90%~100% 之间; 采用文献[2]系统对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配率在 60%~80% 之间; 采用文献[3]系统对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配率在 40% 以下; 采用文献[4]系统对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配率在 70%~90% 之间; 采用文献[5]系统对患者病历与病历仓库临床案例之间的相似度进行匹配, 匹配率在 40%~50% 之间. 本文系统的相似度匹配比传统系统的相似度匹配高, 说明其诊

断结果的准确性高. 由此可知, 本文系统的诊疗精度更好, 能更有效帮助医生做出准确的诊断结果和治疗方案, 保证了患者的生命健康.

依据式 (15) 和式 (16), 计算本文算法的复杂度, 与文献[5-7]进行对比, 具体分析情况如表3所示.

表3 算法复杂度分析 (单位: s)

算法	复杂度
本文算法	10
文献[5]算法	50
文献[6]算法	62
文献[7]算法	74

根据表3对算法复杂度的分析, 可以看出, 本文算法的复杂度相对较低, 用量化数值计算, 具体为 10 s 完成算法运行, 远远低于文献[5-7]算法的复杂度, 能够满足实际需求, 由此验证了本文算法的实际应用性.

4 结语

综上所述, 随着生活水平的提高, 人类健康状况反而越加严重, 因此各种医疗机构每天涌入大量患者进行就医, 并希望医生给出有效的治疗建议. 为此, 辅助医生进行决策的诊疗系统应运而生. 然而, 当前较为常用的几个类型的诊疗决策支持系统, 在精度上还无法完全达到需求, 因此本次研究了基于改进 K-NN 分类算法和 SVM 的多学科协作诊疗决策支持系统. 该系统利用改进 K-NN 算法和 SVM 对患者病情信息进行分类, 从而根据分类结果得出诊断结果. 该系统经过测试, 实现了研究的预期目标, 即降低了误诊率, 诊断精度得到了提高, 且分类结果准确率较高、在分类过程中, 运算时间较短, 对患者病历与病历仓库临床案例之间的相似度进行匹配的结果较好, 说明本系统能更好的辅助医生进行诊断, 提高了医生工作效率, 为人们健康提供了保证. 鉴于本文系统诊疗推理模块的特性, 结合临床知识和术语库, 将知识推理和本文算法相结合是日后的研究重点. 基于改进 K-NN 分类算法和 SVM 的多学科协作诊疗决策支持系统的研究目前还处于初始阶段, 对日后案例的修正和形成是实现多学科协作诊疗决策支持实用性的基础.

参考文献

- 张寅升, 王瑞, 乔清治, 等. 基于双层建模的知识表达方法在医学知识库构建中的应用. 中国生物医学工程学报,

- 2017, 36(5): 573–579. [doi: [10.3969/j.issn.0258-8021.2017.05.009](https://doi.org/10.3969/j.issn.0258-8021.2017.05.009)]
- 2 崔骥, 许家佗. 人工智能背景下中医诊疗技术的应用与展望. 第二军医大学学报, 2018, 39(8): 846–851.
 - 3 杨玲, 王梦洁, DeBelder A, 等. 基于心脏团队模式的多学科诊疗实施现状及思考. 中华医院管理杂志, 2018, 34(2): 128–132. [doi: [10.3760/cma.j.issn.1000-6672.2018.02.009](https://doi.org/10.3760/cma.j.issn.1000-6672.2018.02.009)]
 - 4 周晓梅, 杨春松, 林芸竹. 国内分级诊疗现状的系统评价. 中国药房, 2017, 28(34): 4763–4766. [doi: [10.6039/j.issn.1001-0408.2017.34.03](https://doi.org/10.6039/j.issn.1001-0408.2017.34.03)]
 - 5 井含光, 孟庆刚. 基于系统复杂性的中医诊疗信息分层可视化研究. 中华中医药学刊, 2017, 35(12): 3081–3083.
 - 6 Novo J, Hermida A, Ortega M, *et al.* Hydra: A web-based system for cardiovascular analysis, diagnosis and treatment. *Computer Methods and Programs in Biomedicine*, 2017, 139: 61–81. [doi: [10.1016/j.cmpb.2016.10.019](https://doi.org/10.1016/j.cmpb.2016.10.019)]
 - 7 Chen JG, Li KL, Rong HG, *et al.* A disease diagnosis and treatment recommendation system based on big data mining and cloud computing. *Information Sciences*, 2018, 435: 124–149. [doi: [10.1016/j.ins.2018.01.001](https://doi.org/10.1016/j.ins.2018.01.001)]
 - 8 夏洪图, 孙静. 宁夏农村基层远程诊疗系统的建设应用现状分析. 中国数字医学, 2018, 13(1): 85–87. [doi: [10.3969/j.issn.1673-7571.2018.01.028](https://doi.org/10.3969/j.issn.1673-7571.2018.01.028)]
 - 9 陈青筱. 临床决策支持系统的研究现状及其在口腔修复领域的应用. 实用口腔医学杂志, 2016, 32(5): 722–726. [doi: [10.3969/j.issn.1001-3733.2016.05.029](https://doi.org/10.3969/j.issn.1001-3733.2016.05.029)]
 - 10 赵学彤, 杨亚东, 渠鸿竹, 等. 组学时代下机器学习方法在临床决策支持中的应用. 遗传, 2018, 40(9): 693–703.
 - 11 王飞跃, 李长贵, 国元元, 等. 平行高特: 基于 ACP 的平行痛风诊疗系统框架. 模式识别与人工智能, 2017, 30(12): 1057–1068.
 - 12 孔鸣, 何前锋, 李兰娟. 人工智能辅助诊疗发展现状与战略研究. 中国工程科学, 2018, 20(2): 86–91.
 - 13 Li XF, Wang YW, Li D. Medical data stream distribution pattern association rule mining algorithm based on density estimation. *IEEE Access*, 2019, 7: 141319–141329. [doi: [10.1109/ACCESS.2019.2943817](https://doi.org/10.1109/ACCESS.2019.2943817)]
 - 14 张颖, 纪文迪, 周毅萍, 等. 基于隐语义模型的中医在线辅助诊疗系统. 计算机应用, 2017, 37(S1): 303–307.
 - 15 赵新苗, 冯向萍, 赵涛. 农业信息分类中 K-means 与 SVM 的混合算法研究. 计算机技术与发展, 2017, 27(6): 178–182. [doi: [10.3969/j.issn.1673-629X.2017.06.037](https://doi.org/10.3969/j.issn.1673-629X.2017.06.037)]
 - 16 梁修荣, 杨正益. 基于聚类和 SVM 的数据分类方法与实验研究. 西南师范大学学报(自然科学版), 2018, 43(3): 91–96.
 - 17 黄杰, 朱广平. 水声目标识别中的 K-D 树 KNN-SVM 分类器研究. 海洋技术学报, 2018, 37(1): 15–22.
 - 18 程凤伟. 一种基于 SVM 的线性分类算法. 太原学院学报, 2018, 36(4): 52–54.
 - 19 龚静, 黄欣阳. 基于 k 最近邻和改进 TF-IDF 的文本分类框架. 计算机工程与设计, 2018, 39(5): 1340–1344, 1349.
 - 20 王雪梅, 刘敏超, 季磊, 等. 基于 FHIR 的病案首页信息化模型建立. 中国数字医学, 2018, 13(6): 43–46. [doi: [10.3969/j.issn.1673-7571.2018.06.015](https://doi.org/10.3969/j.issn.1673-7571.2018.06.015)]