



报告正文

参照以下提纲撰写，要求内容翔实、清晰，层次分明，标题突出。
请勿删除或改动下述提纲标题及括号中的文字。

（一）立项依据与研究内容（5000-10000 字）：

1. 项目的立项依据（研究意义、国内外研究现状及发展动态分析，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）：

中医药学具有十分悠久的历史，几千年来为保障民众的健康发挥了不可替代的重要作用。进入二十一世纪以来，人类医学的目的正在从“疾病医学”向“健康医学”转变，中医学整体观念和辨证论治的优势得到了世界医学和各级政府的高度重视。尤其是在信息技术高速发展的今天，大数据、云计算和人工智能给医学带来了颠覆性的革命。应用人工智能技术为中医诊疗服务，成为业内外高度关注的问题，但是，由于对中医的思维特点及其复杂性、模糊性缺乏有效的把握，成为制约本领域发展的瓶颈，也限制了中医药的时代表达和优势的发挥。

本项目申请者基于长期中医诊断客观化、标准化、智能化研究的实践，拟以整体健康状态的动态、个性化把握为核心，应用人工智能技术，研究中医诊断、治疗、疗效评价系统，实现中医诊疗的规范化、可测量、可评价，为大数据背景下中医药的发展和诊疗仪器装备的研发提供技术支撑和方法学平台。

1.1 研究意义

1.1.1 服务健康产业发展的重大需求

中医的整体思维是中医学的灵魂，中医的思维特色和优势集中体现在：一是把人看作一个有机整体，强调形神统一以及人与自然的统一；二是把生命的全过程当作一个整体，涵盖出生、生长、发育、疾病、衰老、死亡；三是整体观念指导下的辨证论治。伴随着健康观念的转变，现代医学也已从单纯的“生物医学模式”转向“环境—社会—心理—生物医学模式”，已从“预防为主”转变为“促进健康为主”。人们对健康的需求不断增加。早在几年前，健康产业就已被认为是信息产业之后下一个超万亿美元产业^[1]。面对这一重大需求，中医药的优势日益凸显出来，正如《中医药健康服务发展规划（2015-2020 年）》指出：



中医药（含民族医药）强调整体把握健康状态，注重个体化，突出治未病，临床疗效确切，治疗方式灵活，养生保健作用突出，是我国独具特色的健康服务资源，必将在健康产业的发展中发挥重要的基础和引领作用。但是，目前能够真正满足中医健康服务需求的中医诊疗设备还十分匮乏，也无法满足中医健康大数据对数据源的可靠性的需求。因此，要加强中医诊疗设备研发，重点研发中医健康识别系统、智能中医体检系统、经络健康辨识仪等中医健康辨识、干预设备^[2]。

本项目正是在这样的背景和需求下提出的，计划立足中医原创思维，以中医健康状态辨识理论为核心，充分应用现代类人智能技术，研发中医诊疗系统，可以与移动互联网融合，充分发挥其自动化、智能化的优势，在大数据支撑下，融合中医诊疗的共性技术，构建起中医健康服务的通用技术平台，满足中医健康产业发展的需求。

1.1.2 提升中医药服务能力的需要

四诊信息的可靠性和诊断辨证的准确性是提升中医药服务能力的前提和重要抓手。当前我国发展中医药健康服务面临的主要问题是：中医从业人员数较少，并且基层中医的服务能力参差不齐，服务品质不高，不能满足人民群众的健康需要。一项关于社区中医药服务能力调查显示：社区中医师最需要提升的能力当中，中医技术的操作技能（78%）和中医辨证论治的临床能力（72%）高居前两位^[3]。造成这种局面的核心原因是：中医从业人员的中医望闻问切四诊技能不规范，辨证思维不熟练，临床诊断准确性不高，从而难以有效地指导中药、针灸等中医治疗服务，这必然严重限制了中医药整体优势的发挥。因此，发展中医人工智能诊疗系统势在必行。我国《中医药发展战略规划纲要（2016-2030年）》指出：到2030年，大力发展中医远程医疗、移动医疗、智慧医疗等新型医疗服务模式。综合运用现代科技手段，开发一批基于中医理论的诊疗仪器与设备，使中医药健康服务能力显著增强，使中医在治未病中的主导作用、在重大疾病治疗中的协同作用得到充分发挥^[4]。

本项目基于人工智能和数据挖掘，从技术角度解决中医信息采集的客观需求，全面、规范、准确地获取四诊信息，实现四诊合参和中医健康状态的自动辨识，还可以依靠互联网实现中医远程诊疗，从而有效弥补基层中医的技术短板、提升基层中医辨证论治水平。充分发挥中医简、便、廉、验的特点和在“治未病”方面独有的优势，实现医疗卫生工作“中心下移、关卡前移”和提升中医药服



务能力和水平的目标。

1.1.3 解决中医理论研究的瓶颈问题

证是中医理论和临床研究的核心，由于中医思维的模糊性和复杂性特点，长期以来，四诊信息的采集依靠传统的望闻问切，辨证方法繁多且难以形成共识，疗效评价仅限于定性而不能实现量化。中医诊疗过程的客观化、规范化和标准化成为制约中医临床研究的障碍，因而也是理论研究的瓶颈。因此，借鉴系统工程的原理和方法，应用人工智能技术和大数据技术，建立以状态为核心的中医健康认知的方法和体系，能够实现四诊信息的采集、存储、分类和融合；实现健康状态的人工智能诊断；实现干预方案的匹配和临床疗效的合理客观的评价，解决传统四诊与辨证方法的不可测量、不可重复、不可评价等问题，为中医理论与临床研究提供可靠的技术平台，从而解决中医理论与临床研究的瓶颈问题。

1.2 国内外研究现状及发展动态分析

中医诊疗系统由四诊信息采集、健康状态辨识（诊断）、干预（治疗）及效果评价等几个部分构成。上世纪 50 年代以来，部分中医药工作者和医药工程专家开始了中医诊疗仪器的研究与探索。特别是近 10 年来，随着人工智能、大数据、云计算等技术的飞速发展，大量的新技术、新方法开始被广泛应用于中医诊疗系统的研究，推动了本领域的发展。

1.2.1 中医诊断技术研究

随着科学技术的不断发展，很多新技术、新材料也不断被运用于中医诊断设备的研制中，形成多学科交叉的态势，使得传统的中医四诊也有了新的发展。从技术特点看，经历了三个阶段。

第一阶段从上世纪 50 年代到 90 年代，这一阶段的特点是，各类传感器技术运用于四诊信息的采集当中，并且不断更新和成熟，使得信息获取方式更加客观和便捷。不足之处在于，对仪器信息的分析和解读不够成熟，如何将仪器获取的信息与中医临床意义相结合，缺乏统一的认识。例如 50 年代，朱颜首次将杠杆式脉搏描记器引用到中医脉诊的研究中^[5]。80 年代，安徽中医学院和中国科技大学合作，率先进行了舌色的量化分析实验。湖南中医学院利用显微分光光度计、原子分光光度计及计算机自动分析系统等设备，对舌象、脉象进行显微图像及其他信息的观察和分析^[6]。张迺华提出 5 种声诊的现代诊断方法^[7]。90 年代，北京工业大学沈兰荪教授、哈尔滨工业大学、香港理工大学、解放军 211



医院、上海中医药大学许家佗、台湾中原大学胡威志、逢甲大学邱创干等专家和机构，对中医舌象分析的关键技术进行了研究，设定了标准化采集条件，建立了分类标准与数据库，实现了舌象特征的提取算法，建立了舌影像与脏腑对应关系的参数资料库^[6]。

第二个阶段是本世纪初的十年，这一阶段的特点是基于新技术的中医诊断仪器被广泛运用于研究中，如：舌诊仪、脉诊仪、面诊仪、问诊仪等。压电陶瓷式医用微压力传感器、光电容积传感器、触力传感器、红外传感器、驻极体传声器等运用于脉诊仪的研究^[5]。色差计、光电血管容积仪、红外热象仪等检测方法也运用于望诊研究^[6]。上海中医药大学王忆勤等总结了声诊研究中几种主要技术和方法。高也陶利用自行研制的二十五音分析仪进行临床研究^[7]。但是受限于信息分析方法和信息融合的不成熟，中医的诊断仪器仍呈现四诊“各自为战”的特点，未能有效发挥中医“四诊合参”的优势。

第三阶段，2010 年后至今，随着大数据的存储和分析技术的不断成熟，中医诊断仪器开始呈现出四诊结合的趋势。例如上海道生公司与上海中医药大学合作、天津天堰公司与天津中医药大学合作，中国中医科学院等企业和科研院所先后开发了中医四诊仪，并尝试着在教学和临床研究中运用。但研究显示，此类设备的不足之处在于，望闻问切的设备成熟度不同，分类标准不统一，影响了四诊信息的融合，同时由于缺少统一的理论支持，不同信息来源的诊断结论存在很大的差异，甚至相互矛盾，不能真正实现四诊合参。

中医诊断的正确性除了单一信息的准确性外，更重要的是四诊合参，而四诊合参的实现必须做到四诊信息的有效融合。因此，有效的信息融合是中医诊断仪器研发的基础。由于理论体系的差异和思维模式的不同，中医诊断不可能也不应该像西医诊断一样追求单一的“特异性指标”，如果不能四诊合参，单一的仪器或指标很难体现中医诊断的优势，这也是到目前为止，中医诊断仪器在临床中运用较少的原因。

1.2.2 中医健康状态研究

人是一个开放的复杂巨系统。人的一生从出生、生长、发育、疾病、衰老、死亡是一个时序的连续过程，在这一过程中对于健康状态整体、动态、个性化的把握是中医诊断和维护健康的核心，也是干预效果评价的依据。既往研究显示，由于对中医的思维特点及整体健康状态缺乏有效的把握，制约了中医诊疗设备的研发和应用。如：单一的信息来源难以为诊断辨证提供可靠、全面的依



据，而单一的辨识结果如体质、经络等难以反映机体的整体状态。

申请者前期致力于以健康状态为核心的中医健康状态的理论、方法技术及应用方面的研究。2011年在《中华中医药杂志》上连续发表12篇健康状态相关论文，初步构建了中医健康状态理论的框架。2016年出版专著《中医状态学》，从健康认知、状态表征、整体状态辨识、状态调整、效果评价等方面系统构建了中医状态学的理论和方法体系，得到了国医大师、中国科学院陈可冀院士和中国中医药科学院院长、中国工程院张伯礼院士的高度评价^[8]，为本项目的人工智能诊疗系统研发奠定了理论基础和方法学依据。

状态是中医健康认知理论的逻辑起点^[9]。健康状态是指人体在一定时间内形态结构、生理功能、心理状态、适应外界环境能力的综合状态，体现的是健康的状况和态势^[10]。健康状态是对生命过程中不同阶段生命特征的概括，包括未病、欲病、已病状态，涵盖了人的各种体质、生理特点、病理特点、病、证等的概念。因此状态辨识包括病、证、体质、生理病理特点等的诊断^[11]。人的健康状态是动态变化的，状态表征参数是辨识状态的主要依据。全面、规范、准确地采集宏观、中观、微观的三观状态表征参数^[12]，通过参数的集合筛选，构建“三观并用”的健康状态表征参数体系，可以对状态进行辨识。健康状态可以用状态要素（已病状态下称为证素）来描述，如程度、部位、性质等，常见的病位如心、肝、肾、胆、胃等约25-30个，常见的性质如寒、热、气虚、血虚等约25-30个，在量化的基础上，通过50-60个要素的组合，形成无穷多的结果，实现了对个体人健康状态测量和个性化把握^[13]。因此，状态辨识是依据中医学理论，“根据表征参数，辨别状态要素，组成状态名称”^[8]。状态辨识是中医诊断的核心，也是治疗和效果评价的依据，适用于各种人群的健康维护、诊断、临床干预和效果评价等。

在大数据时代，以健康状态为核心，应用人工智能技术，能够实现表征参数的采集、存储、分类、融合、状态辨识、干预方案匹配和效果评价的客观化、标准化和智能化，为中医诊疗系统的研发、中医数据库的建立、治未病健康工程和中医养生保健服务产业的发展提供方法学支持^[14-15]。

申请者^[14, 16]前期研发了基于中医原创思维的中医整体健康状态辨识系统、气血津液状态辨识系统，初步构建了常见证型干预方案知识库，建立了基于整体健康状态动态测量的中医药临床疗效评价系统^[17]，并对基于人工智能的中医整体健康状态评估、干预和评价系统进行了有益的探索，能够为本项目的研究提



供理论和方法学依据。

1.2.3 人工智能研究

信息技术的迅猛发展,特别是信息获取技术与物理信息系统等技术的不断突破,引发了网上的数据规模正以前所未有的速度增长。目前,服务业、制造业、科教文化、医疗卫生等领域积累的数据容量已达到 TB、PB 甚至 EB 级别^[18-19]。2011 年 5 月,EMC 公司在美国拉斯维加斯举办以“云计算相遇大数据”为主题的第 11 届 EMC 世界年度大会,正式提出了大数据的概念。

大数据的迅速涌现及其潜在的巨大价值,为社会政治、经济、文化等各个方面的发展注入了新的活力^[20-21],如:通过对社交网络上的大数据进行挖掘,能够跟踪社会动态,以及预见重大和突发性事件;2010 年时代杂志刊载的医学界年度十大突破中,生物科技公司 CardioDx 对 1 亿个基因样本进行分析,识别出 23 个预测冠心病的主要基因。另一方面,伴随着大数据带来的机遇,数据类型多样、数据不确定性、要求及时响应等挑战也随之而来^[22]。其中,数据类型多样要求学习系统不仅能处理结构化数据,还要适应图像、视频、文本、语音等非结构化数据的处理。2011 年《Science》推出关于数据处理的专刊 *Dealing with Data*,指出数据的组织和使用体现了多模态特征;2014 年 973 指南“大数据计算的基础研究”也指明了多源异构大数据的表示、度量、语义理解方法的研究方向;在数据不确定性方面,数据质量是大数据应用的一个关键因素。高质量是对大数据的一项重要要求。数据质量低劣,或者关键数据缺失或陈旧,容易导致基于数据的知识发现与决策产生严重偏差;及时响应则要求系统具有并行处理流式数据的能力。

总之,大数据分析是一项庞大且复杂的系统工程,而人工智能技术为大数据应用提供了有力的支撑^[23]。作为由机器学习、计算机视觉等领域组合而成的技术科学,人工智能的研究目标是使机器能够胜任一些需要人类智能才能完成的工作。目前,人工智能技术已在多个领域取得成功。例如,2016 年 3 月,Google DeepMind 实验室推出的智能机器人 Alpha Go 以绝对优势战胜了世界顶级围棋高手李世石。同年,该公司计划将人工智能用于医疗健康领域。通过采集一系列医疗健康数据,实现对患者病情恶化、死亡以及需要重新治疗的风险评估;2017 年 2 月,《Nature》公布了鉴定皮肤癌的最新研究成果。通过采用人工智能技术分析将近 13 万张临床上的皮肤癌图片,得出机器对皮肤癌的识别能力比大部分参与研究的皮肤科医生都要好。人工智能技术的火热同样为中医健康管理



的发展带来了契机^[24-27]。通过收集被诊断对象的四诊信息，构建以健康状态为核心的中医智能诊疗系统。一方面有利于借助人工智能技术的优势构建高精度的中医诊疗模型算法以辅助临床决策。另一方面，人工智能技术与西医的交叉研究早已开始并取得了令人瞩目的成绩。加快中医信息化进程，促进中医健康产业发展迫在眉睫。因此，立足于中医大数据，研究人工智能技术在中医诊疗中的应用，对于人体健康的智能分析、监测与评估以及扩大中医在世界的影响力具有非凡意义。

1.2.4 人工智能在中医诊断的应用

人工智能与中医诊断信息处理相结合，对于提高中医诊断辨证的客观性、规范性和准确性，降低误诊率有着积极的现实意义。从人工智能技术应用在中医诊断领域的特点来看，可大致分为二个阶段。

第一阶段为20世纪70-80年代，是人工智能在中医诊断领域应用的起步和发展时期。这一阶段的主要特点是研发以专家的经验为知识的专家系统，在技术和方法层面上，多用基于规则的知识表示。1977年，中科院自动化研究所与名中医关幼波合作，成功研制国际上第一个中医专家系统“中医关幼波肝炎诊断治疗程序”。80年代，先后研制了200多个中医专家系统软件^[28]。

第二阶段为20世纪90年代至今，是人工智能在中医诊断领域的深入应用阶段。随着知识发现（KDD，Knowledge Discovery in Database）的提出和在各个行业的广泛应用，国内外的专家学者将人工神经网络、贝叶斯网络、支持向量机、粗糙集等先进的人工智能方法引入中医诊断领域，取得了一定的成果。例如，朱文锋^[29]在模糊定量与综合定性的基础上，提出了证素辨证，这是一种“根据证候，辨别证素，组成证名”的新的辨证方法。宴峻峰^[30]采用支持向量机和粗糙集等算法，对已收集的临床病历数据进行实验分析，探索利用这些方法解决中医证素辨证建模的可行性。王立问^[31]以慢性胃炎患者中医问诊数据为研究样本，采用二类相关和深度置信网络的机器学习方法，建立了慢性胃炎中医证型的分类模型。刘旭龙^[32]等将形式概念分析应用于计算机辅助中医辨证体系研究，把证和证素定为对象与属性，并将它们的概念和层次关系数学化，最后利用概念格将其可视化。褚娜^[33]运用智能技术从中医和西医两个角度对中医证候进行综合性研究，引入混合智能系统理论，为中医辨证过程设计一个具有规范化和客观化的整体框架，并以慢性乙型肝炎为例建立混合智能中医辨证系统，为中医临床实践提供现代化的技术手段。朱咏华等^[34]建立贝叶斯网络中医辨证系统，



进行数据计量分析、推理验证证候—证素—证名间的关系，其结果与中医专家经验有很高的吻合性。白丽娜^[35]根据中医体质辨识的特点，利用BP神经网络理论，构造了适用于中医体质辨识的BP神经网络模型。

上述贝叶斯网络、神经网络、粗糙集理论等数据挖掘的方法，能从大量的数据中发现有关研究对象的新知识，将机理上还不明确的变量关联起来。但是，“人的功能状态”是非线性的复杂系统，在功能状态体系研究中，不能采用还原论的方法，要引进复杂性科学的理念和多学科结合的研究方法，采用类人智能技术以揭示其复杂性的机制和规律。中医健康状态辨识是依据病人的临床表现，重点辨别病位和病性证素并概括为完整证名的意象思维认知过程，其目标是通过在给定的多标记数据上构建中医人工智能诊疗系统以预测测试对象所属类别标记集合^[14, 27]。因此，从症状、病位证素、病性证素及证型等角度出发，构建多尺度的健康状态认知框架，是实现“辨证论治”理论^[36]指导下人工智能应用于中医诊疗的核心步骤。

综上所述，近几十年的研究在中医四诊仪器和中医专家系统研发方面已经积累了一定的基础，尤其是现代科学技术的发展给中医传感器技术带来了革命性的进步，但是，依然存在以下几个问题：（1）研究比较分散，缺乏统一的理论支撑，对与健康疾病相关的“天、时、地”等因素关注不够，难以体现中医的整体思维特征；与人工智能结合不够，多学科交叉的优势难以发挥。（2）由于过分强调指标的特异性，四诊信息采集仍立足于单一信息，如脉诊仪、舌诊仪、经络仪等，虽然传感器技术已经十分发达，但仍存在结果不一致甚至相互矛盾的现象，无法体现四诊合参。（3）由于对中医健康认知核心缺乏深刻的认识，导致中医思维的偏差，加上辨证方法不一致，结果难以形成共识。在信息分类、模型识别方面缺乏统一的标准，成为制约中医药诊疗设备研发与健康大数据接轨的瓶颈。

1.3 科学意义

构建中医智能诊疗模型是一项极富挑战性的工作。原因在于：其一，中医健康数据的不确定性。诊断有赖于未来高质量、大规模的临床研究和观察结果，而且中医健康数据不可避免地存在大量缺失，要求开发的学习算法适用于极端稀疏数据的处理；其二，中医健康数据多模态多标记的表示形式对学习算法的设计提出了更高的要求。这意味着采集的病人信息具有多源异构的特性，故而，多模态数据的融合是数据分析有效性的重要环节。另外，由于需要辨识病人的



病位证素、病性证素与证型，辨识过程可看作一个典型的多标记学习问题，而基于标记之间的关联信息来构建模型，是获取兼具准确性与鲁棒性的辨识结果的关键。因此，以健康状态为核心，应用人工智能实现四诊合参和构建中医智能诊疗模型，紧扣数据不确定性、多模态、多标记等科学问题对大数据分析的影响，是实现基于大数据的中医智能诊疗的必由之路。

申请者前期的工作基础表明：对整体健康状态动态、个性化的把握是中医诊疗系统研发和应用的核心，前期研发的中医健康管理太空舱（四诊采集仪器）、电子鼻、健康状态辨识系统和疗效评价系统，初步实现了中医临床诊断与疗效的可测量、可评价。大数据时代，将人工智能先进理论、方法和技术与中医状态学理论和方法结合起来，研究人工智能在中医健康状态评估、干预、疗效评价中的关键问题，能够实现中医人工智能诊疗系统的研发，对于推动中医现代化发展具有重要意义。

通过系统的文献梳理和理论研究，结合前期的工作基础，我们提出假说：状态的测量与评估是中医诊断、治疗和疗效评价的依据。在中医健康状态理论指导下，应用人工智能技术实现人体健康状态表征信息的采集、分类、融合，进行状态辨识模型识别、干预方案匹配和效果评价，是中医诊疗系统研究的核心。

本课题的科学意义：（1）遵循中医整体思维，提出状态是中医健康认知的逻辑起点，整体健康状态的动态、个性化测量与评估是中医诊断、治疗和疗效评价的依据；（2）应用人工智能技术，建立状态表征参数的采集、存储、分类、融合的标准，构建状态辨识的算法模型，实现干预方案的自动匹配和疗效的可测量、可重复、可评价；（3）应用云计算和大数据技术，构建中医人工智能诊疗系统，实现中医诊疗的客观化、规范化、智能化，为中医的理论和临床研究提供科学基础和方法学平台，为中医数据库建立和健康产业发展提供技术平台。

【参考文献】

- [1]胡鞍钢，鄢一龙.中国：走向 2015[M].杭州：浙江人民出版社，2010：36.
- [2]国务院办公厅关于印发中医药健康服务发展规划（2015-2020 年）的通知[J].中华人民共和国国务院公报，2015，（14）：32-39.
- [3]龚鹏，余小萍，王彦华，等.关于提升社区中医药服务能力的调查[J].中医药导报，2015，21（9）：41-43.
- [4]国务院关于印发中医药发展战略规划纲要（2016-2030 年）的通知[J].中华人民共和国国务院公报，2016，（8）：21-29.



- [5]蒋颖, 刘聪颖, 张亚丹, 等.脉诊检测分析仪的研究进展与新思路[J].中华中医药志, 2017, 32 (1): 218-221.
- [6]邸丹, 周敏, 秦鹏飞, 等.中医舌诊、面诊客观化研究进展[J].上海中医药杂志, 2012, 46 (4): 89-92.
- [7]鄢彬, 王忆勤.中医闻诊客观化研究进展[J].中华中医药学刊, 2014, 32 (2): 243-246.
- [8]李灿东.中医状态学[M].北京: 中国中医药出版社, 2016: 序一, 序二, 60.
- [9]李灿东, 纪立金, 鲁玉辉, 等.论中医健康认知理论的逻辑起点.中华中医药杂志, 2011, 26 (1): 109-111.
- [10]王琦.中医健康状态概念及内涵研究[J].中华健康管理学杂志, 2012, 12 (6): 419-420.
- [11]李灿东, 杨朝阳, 林雪娟, 等.体质、病理特点、证与健康状态的关系[J].中华中医药杂志, 2011, 26 (8): 1770-1773.
- [12]李灿东, 杨雪梅, 纪立金, 等.健康状态表征参数体系的建立与集合分析[J].中华中医药杂志, 2011, 26 (3): 525-528.
- [13]李灿东, 纪立金, 杨朝阳, 等.以状态为中心的健康认知理论的构建[J].中华中医药杂志, 2011, 26 (2): 313-316.
- [14]李灿东, 杨雪梅, 甘慧娟, 等.健康状态辨识模型算法的探讨[J].中华中医药杂志, 2011, 26 (6): 1351-1355.
- [15]许家佗.基于四诊信息决策支持的中医健康评价体系研究述评与展望[J].中国中西医结合杂志, 2012, 32 (3): 307-310.
- [16]杨雪梅, 甘慧娟, 赖新梅, 等.基于证素辨证模型的中医健康管理系统研发[J].中华中医药杂志.2015, 30 (8): 2681-2683.
- [17]王洋, 王昌恩, 林雪娟, 等.疾病诊断与疗效评价相关参数差异性理论探讨[J].中华中医药杂志, 2016, 31 (8): 2939-2941.
- [18]H.V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, et al. Big data and its technical challenges [J]. Communications of the ACM, 2014, 57 (7): 86-94.
- [19]李建中, 刘显敏.大数据的一个重要方面: 数据可用性[J].计算机研究与发展, 2013, 50 (6): 1147-1162.
- [20]王元卓, 靳小龙, 程学旗.网络大数据: 现状与展望[J].计算机学报, 2013, 36 (6): 1125-1138.
- [21]李国杰, 程学旗.大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J].中国科学院院刊, 2012, 27 (6): 647-657.



- [22]程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25 (9): 1889-1908.
- [23]C.L. Philip Chen, Chunyang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data [J]. Information Sciences, 2014, 275 (11): 314-347.
- [24]Peiqin Gu, Huajun Chen n. Modern bioinformatics meets traditional Chinese medicine [J]. Briefings in Bioinformatics, 2014, 15 (6): 984-1003.
- [25]Xuezhong Zhou, Shibo Chen, Baoyan Liu, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support [J]. Artificial Intelligence in Medicine, 2010, 48 (2-3): 139-152.
- [26]Wenhsiang Wu, Jiayou Liu, Henhong Chang. Chang. Latent class model based diagnostic system utilizing traditional Chinese medicine for patients with systemic lupus erythematosus [J]. Expert Systems with Applications, 2011, 38 (1): 281-287.
- [27]徐玮斐, 刘国萍, 王忆勤, 等. 近 5 中医证候诊断客观化研究述评[J]. 中医杂志, 2016, 57 (5): 442-445.
- [28]余江维, 马利庄, 杨华元. 中医智能化诊断的研究现状与展望 [J]. 辽宁中医杂志, 2010, 37 (1): 50-53.
- [29]朱文锋. 创立以证素为核心的辨证新体系[J]. 湖南中医学院学报, 2004, 24 (6): 38-39.
- [30]晏峻峰. 基于数据挖掘技术的证素辨证方法研究[D]. 湖南中医药大学: 2007.
- [31]王立问. 基于深度学习与条件随机场的多标记学习方法的中医问诊建模研究[D]. 华东理工大学: 2013.
- [32]刘旭龙, 洪文学, 张涛, 等. 基于形式概念分析的中医辨证可视化方法[J]. 燕山大学学报, 2010, 34 (2): 162-164.
- [33]褚娜. 基于混合智能的中医辨证系统研究[D]. 上海交通大学: 2012.
- [34]朱咏华, 朱文锋. 基于贝叶斯网络的中医辨证系统[J]. 湖南大学学报 (自然科学版), 2006, 33 (4): 123-125.
- [35]白丽娜. 基于 BP 神经网络的中医体质辨识研究[D]. 天津理工大学: 2014.
- [36]邓铁涛. 辨证论治是中医临床医学的灵魂[J]. 中华中医药学刊, 2002, 20 (4): 394-395.

2. 项目的研究内容、研究目标, 以及拟解决的关键科学问题 (此部分为重点阐述内容);

2.1 研究内容

本项目围绕中医诊疗系统研究中存在的基础科学问题和关键技术展开研究,



以中医健康状态理论为指导,应用系统工程的原理,基于人工智能技术,构建中医整体健康状态评估、干预和评价系统。研究内容如下:

2.1.1 四诊信息的采集、存储

四诊信息采集系统从中医“望闻问切”的四个方面获取信息,系统包括望诊(包括舌诊)、闻诊(声音、气味)、问诊、脉诊等主要模块,信息类型和获取方式均在本课题组前期研究的基础上,以成熟的技术为支持,实现信息的采集、存储、识别。

2.1.2 四诊信息的分类、融合

基于预调查的参数分类:在对四诊信息预调查的基础上,对四诊参数进行分类,以确定参数等级。

四诊参数的集合筛选:采用德尔菲专家咨询法对参数集合进行分析、筛选并优化,为建立诊断模型打下基础。

参数集成:采用优化设计构建多源异构数据的融合方法以及决策机制。

2.1.3 中医诊断算法模型

根据证素辨证和中医状态学的原理,依据从表征参数判断状态要素(证素),最后形成状态名称(证名)的辨证思维过程,构建中医诊断的模型算法。

(1) 研究融合多源异构大数据的中医辨证论治分析模型和方法,构建类人认知体系架构和思维机理。

(2) 借鉴多种深度学习机制,研究多标记学习框架下基于深度学习的中医智能诊断算法。

(3) 研究基于标记局部相关性的多标记分类学习算法和多标记特征选择算法,实现基于人工智能的中医诊断机制。

2.1.4 干预方案推荐与效果评价

(1) 干预(治疗)方案推荐

在课题组前期建设的干预方案数据库,包括《中医方剂大辞典》、《方剂学》等近10万首方剂的基础上,系统整理干预方案适应的证型、不适应的证型等相关信息,在中医治则、治法理论指导下,研究基于证型匹配的干预方案自动推荐算法模型。

(2) 干预效果评价

在前期制定的基于证素辨证原理的证的疗效评价方法基础上,对治疗前后的病位、病性证素及其积分进行测量,实现对证素与证型的双重疗效评价。通过



构建时间序列并对其进行分析，建立证的疗效评价的统计模型。

2.1.5 人工智能诊疗系统

中医人工智能诊疗系统的主要功能有四个：四诊信息采集、人工智能辅助诊疗、随访和疗效评价以及病案的存储等管理工作。根据这四个主要功能，提出基于云计算和大数据的人工智能诊疗系统架构，描述相关诊疗服务，展示诊疗的处理流程，完成人工智能诊疗系统的设计与实现。

(1) 针对中医临床大数据异构、多模态的特点，借鉴 lambda 的混合架构体系，研究基于 Hadoop 的中医临床数据收集、存储、融合等问题。

(2) 基于 Hadoop 等分布式计算框架，设计基于云计算的中医大数据人工智能诊疗系统。

2.1.6 验证性实验

(1) 选取 2010-2015 年福建中医药大学附属人民医院、附属第二人民医院、附属厦门中医院的 10 万份住院病例为本底作为标准数据集以验证中医人工智能诊疗系统的学习性能。

(2) 实际收集福建中医药大学附属人民医院、附属第二人民医院、附属晋江中医院、附属国医堂的门诊病人共 1000 例对系统进行测试，采用调查问卷的方式对医生和患者进行随访，以进一步考察系统的学习性能、功能的全面性、使用满意度、使用流畅性、操作方便性等等。

研究思路：

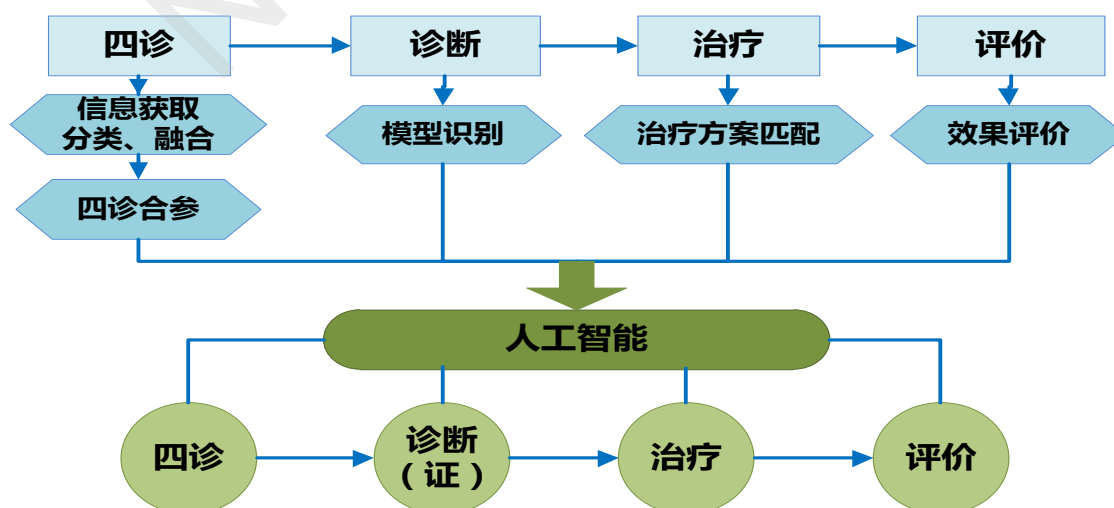


图 1 总体研究思路



2.2 研究目标

(1) 遵循中医思维规律，完善以整体健康状态动态、个性化测量与评估为核心的中医诊断、治疗和效果评价的理论体系。

(2) 应用多源信息融合和深度学习的方法，建立四诊信息的采集、存储、分类、融合的标准，构建四诊合参和四诊信息采集的共性技术平台。

(3) 应用人工智能技术，建立中医诊断、治疗和效果评价模型算法，为实现中医诊疗的智能化提供技术支撑。

(4) 应用系统工程的原理和人工智能技术，基于云计算和大数据，研发以健康状态为核心的中医人工智能诊疗系统，实现中医诊疗的客观化、规范化和智能化。

2.3 拟解决的关键科学问题

(1) 中医人工智能诊疗系统的内涵

中医人工智能诊疗系统是遵循中医思维规律，以健康状态为核心，深度融合人工智能、云计算、大数据技术，形成三位一体的具有自主获取新知识、自我学习能力的辅助诊疗系统。在系统整理中医的科学知识和经验知识，建立符合计算机语言的数据结构及知识规则库的基础上，全面、客观、准确地采集表征参数，通过多源信息融合实现真正的四诊合参，进而建立多标记学习框架下基于深度学习的辨证算法，从而构建一个精准的人工智能诊疗系统，实现中医诊疗的规范化、可测量、可评价。

(2) 四诊合参与多源异构信息处理

对于四诊设备采集的包括症状、舌像、脉象、声音、气味等具有多模态特性的数据资料，设计多源异构信息的处理方法，是实现四诊合参的关键。另外，数据源之间关联模式的不确定性为设计中医四诊合参理论指导下的人工智能诊断算法带来了严峻挑战。因此，构建融合四诊信息的优化模型，是中医诊疗系统研究的关键科学问题之一。

(3) 中医思维与辨证诊断算法模型

中医辨证诊断是一个典型的多标记学习方法的应用。以证素辨证原理为基础，建立多模态多标记数据下的分类学习算法，有利于构建中医类人认知体系和思维机制。因此，面对复杂繁多而又相互矛盾的中医健康数据，以及中医病位证素、病性证素兼夹及证型匹配，实现中医辨证诊断算法的有效性、准确性与可计算性是中医诊疗系统研究的另一个关键科学问题。



3. 拟采取的研究方案及可行性分析(包括研究方法、技术路线、实验手段、关键技术等说明);

3.1 研究方法

3.1.1 四诊信息采集系统研究

依据全国中医药行业高等教育“十三五”规划教材《中医诊断学》(李灿东. 中医诊断学[M]. 北京: 中国中医药出版社, 2016), 采用分模块研究, 通过“多源信息集成”的方法构建四诊信息采集系统。

望诊模块采集面部的色泽、形态, 及舌质和舌苔的相关信息, 由高清数码照相机、光源系统、伺服软件构成。闻诊模块采集声音信息和气味信息。声音诊断模块采集音调、音频等特征, 由声音采集装置、存储装置和分析软件构成。气味诊断模块采集人体呼出气体的特征, 由阵列式传感器、储气袋及分析软件构成。问诊模块采集问诊信息, 由中医健康状态通用问诊量表构成。脉诊模块采集脉象的频率、压力、节律、长度、宽度、流利度、紧张度等脉象要素, 由柔性压力传感器和控制系统构成。

3.1.2 四诊合参与信息融合研究

(1) 表征参数的分类计算

由于四诊采集的方式方法有所差异, 所采集的参数有精确的、模糊的、计量的、计数的、形态学的、影像的等多种数据类型, 因此需采用预调查的方式对不同类型的参数进行分类以及等级划分, 以便根据四诊信息生成健康状态的辨识结果。

(2) 四诊参数的集合筛选

采用德尔菲专家咨询法对参数集合进行分析、筛选并优化, 确定参数之间的关系。

(3) 参数集成

采用多源融合的方法对参数进行集成, 为建立诊断模型奠定基础。

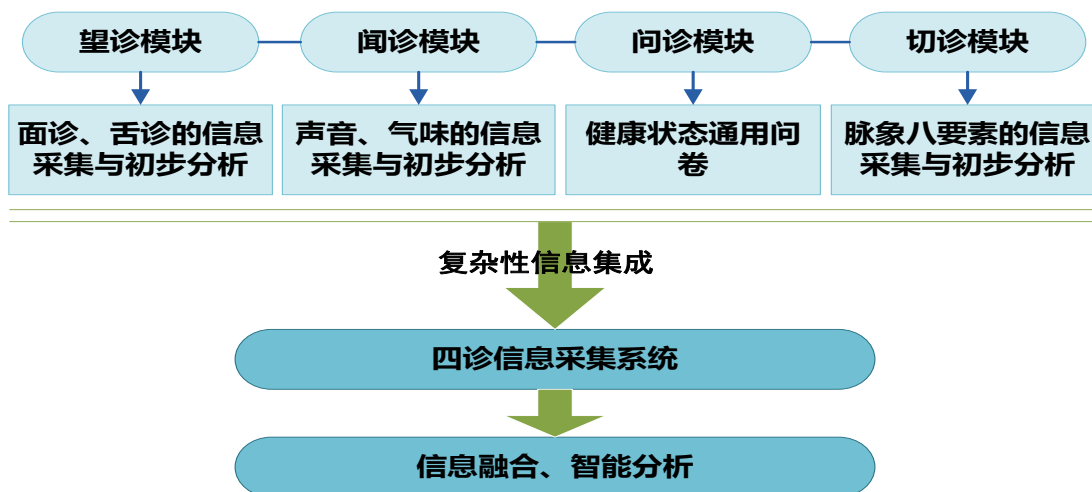


图 2 四诊合参与信息融合研究技术路线图

按照中医四诊合参的诊断原则，首先利用望、闻、问、切的方式生成四诊信息，然后融合四诊信息用于智能分析，是实现基于人工智能的辨证诊断算法的必要步骤。因此，对于多源异构数据融合方法的设计，我们主要关注不同模态数据类型的处理，多源异构数据的融合，以及流式数据的并行处理。

针对上述问题相应的解决方案如下：

问题 1：如何处理不同模态的数据类型？

针对不同类型的数据，采用线性核、高斯核及直方图交叉核等适用于不同类型的核函数将特征空间映射到核空间。具体地，给定二个对象 x 和 y ，每个对象由离散型数据、连续型数据或图像等形式的混合数据来表示其属性特征。若对象 a 和 b 的属性特征由一组离散值表示，则采用线性核函数 $s(a, b) = a^t b$ 来计算对象之间的相似性大小。

若 a 和 b 的属性特征由一组连续值表示，则采用高斯核函数计算对象之间的相似性大小。假设 σ 为其参数，高斯核函数的计算公式如下：

$$s(a, b) = \exp\left(-\frac{\|a - b\|^2}{\sigma^2}\right)$$

对于图像类型的数据，应用直方图交叉核函数计算图像直方图之间的相似性大小。假设 a_i 和 b_i 表示 a 和 b 直方图中的第 i 个 bin ， P 表示直方图 bin 的个数，那么有 $\sum_{i=1}^P a_i = 1$ ， $\sum_{i=1}^P b_i = 1$ 。直方图交叉核函数的计算公式为：

$$s(a, b) = \sum_{i=1}^P \min(a_i, b_i)$$

另外，考虑利用相似性度量方法与损失函数以生成不同对象之间的相似度矩阵。如常用的 Jaccard 相似系数、余弦相似性度量方法、0-1 损失及平均绝对误差损失等。

**问题 2：如何进行多源异构数据的融合？**

针对多个数据源对于预测结果具有不同的贡献值，构建基于多数据源的融合方法。通过构建的核空间和相似度矩阵，从知识层面进行相似性融合：

$$s(a, b) = \sum_{k=1}^K w_k \times s^k(a, b)$$

其中， K 为数据源总数， w_k 为第 k 个数据源的权重， $s^k(a, b)$ 为对象 a 和 b 基于第 k 个数据源的相似度。根据 $s(a, b)$ ，定义分类器或回归函数 $f: X \rightarrow Y$ ，其中 $a \in X$ 为对象 a 的特征属性而 $y \in Y$ 表示其所属类别标记，设计优化算法以获取最终优化结果。

$$\operatorname{argmin}_W \sum_{i=1}^n \sum_{l=1}^L \left(y_{il} - f_{il} \left(\sum_{p=1}^n \sum_{q=1}^n s(a_p, b_q) \right) \right)^2$$

其中， n 为对象个数， L 为标记个数。 y_{il} 为第 i 个对象在第 l 个类别标记上的实际值。 W 为数据源的权重分布 $\{w_1, w_2, \dots, w_K\}$ ，其满足正则化函数 $\delta(W) = 1$ 。

类似地，考虑从决策层面设计优化模型实现多源异构数据的融合。

$$\operatorname{argmin}_W \sum_{i=1}^n \sum_{l=1}^L \left(y_{il} - \sum_{k=1}^K w_k \times f_{il} \left(\sum_{p=1}^n \sum_{q=1}^n s^k(a_p, b_q) \right) \right)^2$$

该模型综合考虑了多数据源对于同一类别标记的预测结果。通过调节多个预测结果的权重来降低训练误差，从而实现融合多数据源的中医智能诊断算法。

问题 3：如何实现流式数据的并行处理？

被诊断对象不同时间阶段的跟踪随访数据决定学习系统应具有增量更新的能力。针对上述情况，探索数据源权重更新依赖对象增量信息，从而实现并行处理流式数据的中医智能诊疗系统。

3.1.3 辨证诊断算法模型研究

根据“根据表征参数，辨别状态要素，组成状态名称”的原则，以病位证素和病性证素为核心，构建基于朱文锋教授证素辨证原理（见图3）的中医诊断算法模型。根据前期基础，提取常见的病位证素：心、肝、脾、肺、肾等20个，病性证素：风、寒、暑、气滞、血瘀、痰等22个，最后匹配成规范证名。以600个常见症状为本底资料，通过文献调研和专家咨询（Delphi Method）归纳出650个表征参数。

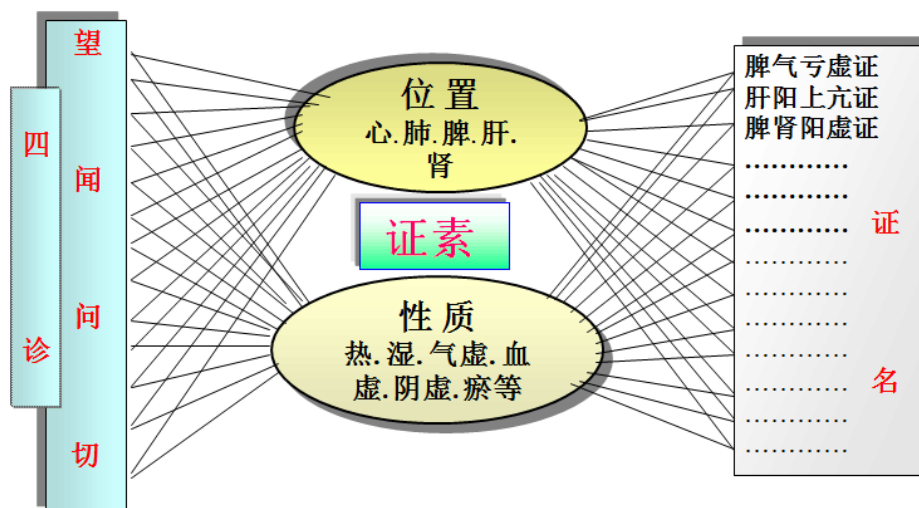


图 3 证素辨证原理图

辨证诊断算法模型研究的重点在于根据生成的四诊信息,对被诊断对象的病位证素、病性证素及证型进行辨识。基于此,一方面借鉴深度学习机制设计辨证诊断算法。另一方面,针对辨证结果的局部相关性展开研究。

(1) 研究多标记学习框架下基于深度学习的辨证诊断算法

在本课题中,对于辨证诊断算法的设计,关键在于如何根据选择的深度学习模型构建多标记分类学习机制。基于此,我们主要关注基于深度置信网络和卷积神经网络的多标记分类学习算法的研究。

针对上述问题相应的解决方案如下:

问题 1: 如何构建基于深度置信网络的多标记分类学习算法?

拟采用深度置信网络设计多标记学习框架下的中医智能诊断算法。深度置信网络(Deep Belief Networks, DBNs)由多层限制玻尔兹曼机(Restricted Boltzmann Machines)组成,其目标在于对原始数据进行高层特征表示,从而构建多标记训练集 $\{(z_i, y_i) | 1 \leq i \leq n\}$ 用于训练多标记分类器 h ,其中 z_i 表示训练对象 x_i 的高层特征表示。

基于深度置信网络的多标记分类学习算法的训练过程是:首先不考虑最顶构成的联想记忆(Associative Memory)层,采用非监督贪婪逐层方法预训练多层限制玻尔兹曼机的权值;然后添加联想记忆层,整个过程采用反向传播算法对 DBNs 网络的权重进行计算。对于任意一个训练对象 x_i ,通过预训练的 DBNs 网络获取其预测结果 \hat{y}_i 。因此, x_i 用于训练的误差可以表示为 $\epsilon_i = y_i - \hat{y}_i$ 。

通过训练多个周期使得 DBNs 网络全局误差最小化,从而生成多标记分类器 h 。对于新来测试对象 \tilde{x} ,获取该对象的高层特征表示 \tilde{z} ,然后预测该对象所属类别标记集合 \hat{y} ,表示



为: $\hat{y} = h(\bar{z})$ 。

问题 2: 如何构建基于卷积神经网络的多标记分类学习算法?

卷积神经网络 (Convolutional Neural Network, CNN) 在中医智能诊断中的应用受到多标记兼夹这一问题的制约。为构建基于卷积神经网络的多标记分类学习算法, 拟采用二进制编码等方式将多标记学习问题转化为多分类问题。如: 考虑对象 x_i 所属类别标记集合 $y_i = \{l_1, l_2, \dots, l_L\}$, 设计 x_i 新的标记表示形式:

$$y_i = l_1 \times 2^{L-1} + l_2 \times 2^{L-2} + \dots + l_L \times 2^0$$

至此, 设计多分类学习框架下的卷积神经网络模型。对于测试对象 \tilde{x} , 通过将预测结果 \hat{y} 转换为二进制形式, 从而实现多标记分类学习。上述方法在一定程度上考虑了利用标记相关性建模, 然而, 由于问题转换造成的类别多样性, 算法的鲁棒性有待提高。基于此, 探索基于卷积神经网络实现多数据源融合的集成学习。

(2) 研究基于标记局部相关性的多标记分类学习和特征选择算法

在本课题中, 对于辨证诊断算法的设计, 我们主要关注**如何考虑标记局部相关性, 如何实现多标记特征选择, 以及如何进行多标记分类学习**。

针对上述问题相应的解决方案如下:

问题 1: 如何考虑标记局部相关性?

标记局部相关性是指标记之间的关联性仅由部分对象共有, 而不是所有对象都具有的性质。针对上述情况, 从标记局部相关性这一角度出发, 首先对所有对象进行聚类分析, 然后设计优化算法以学习不同对象集合的标记局部相关性特征。

给定一个多标记训练集 $\{(x_i, y_i) | 1 \leq i \leq n\}$, $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$ 为一个 d 维的特征向量, $y_i = [y_{i1}, y_{i2}, \dots, y_{iL}]$ 为 x_i 的标记集合。假设 $c = [c_1, c_2, \dots, c_n]$, 则标记 y_l ($1 \leq l \leq L$) 可以由函数 $f_l(x, c, w_l)$ 生成预测结果。其中, c 表示生成的标记局部相关性特征矩阵, $w_l = [w_l^x, w_l^c]$ 。基于此, 构建以下优化目标函数 $f(W, c)$:

$$\operatorname{argmin}_{W, c} f(W, c) = \sum_{i=1}^n \sum_{l=1}^L \operatorname{loss}(x_i, c_i, y_{il}, f_l) + \alpha \sum_{i=1}^n \sum_{j=1}^n s(i, j) \times \|c_i - c_j\|^2$$

其中, $\operatorname{loss}(x_i, c_i, y_{il}, f_l)$ 为定义的损失函数。 $s(i, j)$ 表示对象 i 和 j 类别标记的相似性大小, α 为优化参数。通过增加优化约束项, 增强标记局部相关性特征对于分类学习的影响, 使得学得特征矩阵 c 逼近标记局部之间的相关关系。

由于上述优化目标函数具有二个变量 W 和 c , 采用迭代优化的方式对 W 和 c 进行学习。首先固定 c 的值, 初始化任意一个对象 x_i 的标记局部相关性特征向量 c_i 为 x_i 所属聚类的聚类中心。通过最小化优化目标函数 $f(W, c)$, 实现 W 值的更新:

$$W \leftarrow \operatorname{argmin}_W f(W, c) \text{ s.t. } \delta(W) = 1$$



其中, $\delta(W)$ 为正则项。依据更新的 W 值, 计算 c 的值:

$$c \leftarrow \underset{c}{\operatorname{argmin}} f(W, c)$$

以此反复, 得到优化的标记局部相关性特征矩阵 c , 与此同时, 优化目标函数 $f(W, c)$ 的值趋于稳定。

问题 2: 如何实现多标记特征选择?

在上述基础上, 构建多标记特征选择算法。一般来说, 冗余特征和无关特征是不提供任何信息的。在最小化优化目标函数时考虑这些特征, 尽管能获取更小的训练误差, 但在预测新来对象时, 冗余特征和无关特征反而会干扰学习系统对中医病位、病性及证型的预测。基于此, 在上述优化目标函数的基础上增加 L_1 正则项作为稀疏规则算子, 从而实现具有决策性特征的自动选择。假设 β 为优化参数, 优化目标函数更新如下:

$$\underset{W, c}{\operatorname{argmin}} \sum_{i=1}^n \sum_{l=1}^L \operatorname{loss}(x_i, c_i, y_{il}, f_l) + \alpha \sum_{i=1}^n \sum_{j=1}^n s(i, j) \times \|c_i - y_j\|^2 + \beta \|w_l\|_1$$

另外, 考虑联合原始特征空间 x 与生成的特征矩阵 c , 设计基于模糊互信息度量方法, 设计贪婪式多标记特征选择算法以获取具有最小冗余性且与标记相关性最大的特征排序结果。具体的特征选择过程如下:

$$\underset{f_j \in T}{\operatorname{argmax}} FMI(f_j, Y) - \frac{1}{|S|} \sum_{f_i \in S} FMI(f_j, f_i)$$

其中, T 为待选的特征集合, S 为已选的特征集合。 $FMI(f_j, Y)$ 表示特征 f_j 与标记集合 Y 的模糊互信息, $FMI(f_j, f_i)$ 表示特征 f_j 和 f_i 的模糊互信息。

问题 3: 如何进行多标记分类学习?

对于新来测试对象, 首先用回归函数 (如支持向量机和线性回归函数) 预测该测试对象的标记局部相关性特征向量, 然后根据特征选择结果采用多标记分类器或回归函数对测试对象所属标记集合进行预测。

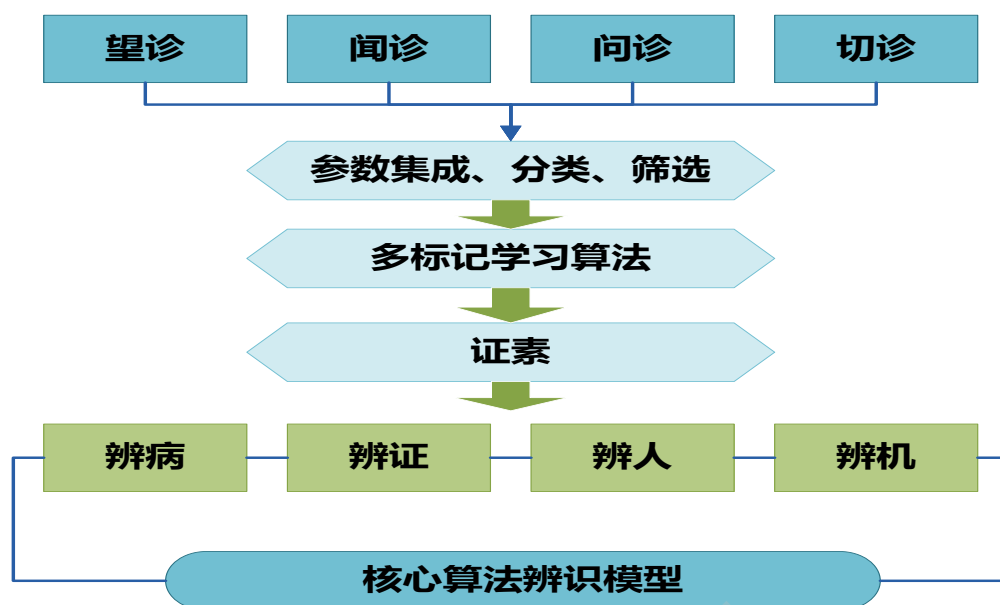


图 4 辨证诊断算法模型研究技术路线图

3.1.4 干预方案推荐与效果评价研究

3.1.4.1 干预（治疗）方案推荐

（1）干预方案数据库的建设

在课题组前期建设的包括《中医方剂大辞典》（彭怀仁. 中医方剂大辞典[M]. 北京：人民卫生出版社，1993）、《方剂学》（邓中甲. 方剂学[M]. 北京：中国中医药出版社，2003）等近 10 万首方剂的基础上，系统整理干预方案适应的证型、不适应的证型等相关信息，为干预方案与诊断结果基于证型的自动关联提供依据，并在“寒者热之、热者寒之、虚则补之、实则泻之”等中医治则、治法理论指导下，实现基于证型匹配的干预方案自动推荐。

（2）干预方案自动推荐算法模型的研究

拟采用协同过滤等推荐算法实现基于证型的干预方案自动推荐。通过采用相似性度量方法对被诊断对象之间的相似度进行计算，然后采用推荐技术预测病位或病性证素及其组合。在此基础上，按照先病性后病位、先主证后次证的规则，分析预测得到的证型组合的主次、轻重以及缓急关系，依据中医的治则治法理论，主动提供个性化的干预方案推荐服务。

3.1.4.2 干预（治疗）效果评价

（1）基于证素辨证原理的证的疗效评价方法研究

以病位、病性证素及其轻重程度为主要内容的证的动态测量，作为证的疗效评价的特色指标，对证素与证型进行双重疗效评价。



①制定中医证的疗效评价量表

根据前期工作基础，以 600 个常见症状（朱文锋. 中医主症鉴别诊断学[M]. 长沙：湖南科学技术出版社，2000）为本底资料，通过文献调研和专家咨询（Delphi Method），制定基于证素辨证原理的中医证的疗效评价量表，对治疗前后的病位、病性证素及其积分进行测量，实现对证素的疗效评价。同时参照《中药新药临床研究指导原则》，对证型进行疗效评价。

②确定证的疗效评价的时间点

根据不同疾病的种类选择不同的评价时间点：如慢性病，采用分级的评价方式，病情不太稳定时，五天之内评定一次；病情稳定后，十五天评定一次；完全缓解，三个月评定一次。急性病，可每小时或每天评价一次。实现对干预效果的动态评价。

③确定证的疗效评价结果

通过整理临床真实案例，研究并确定治疗前后证素积分变化阈值，作为证的转归与变化的指标，结果分为加重、无明显变化、减轻、消失 4 个等级。

（2）证的疗效评价的算法模型研究

拟采用应用时间序列分析技术，用以建立证的疗效评价的算法模型。通过采用自相关分析、谱分析等方法对历史数据资料进行统计分析，从而获取参与疗效评价的影响因素（如状态要素与证型）的变化情况。

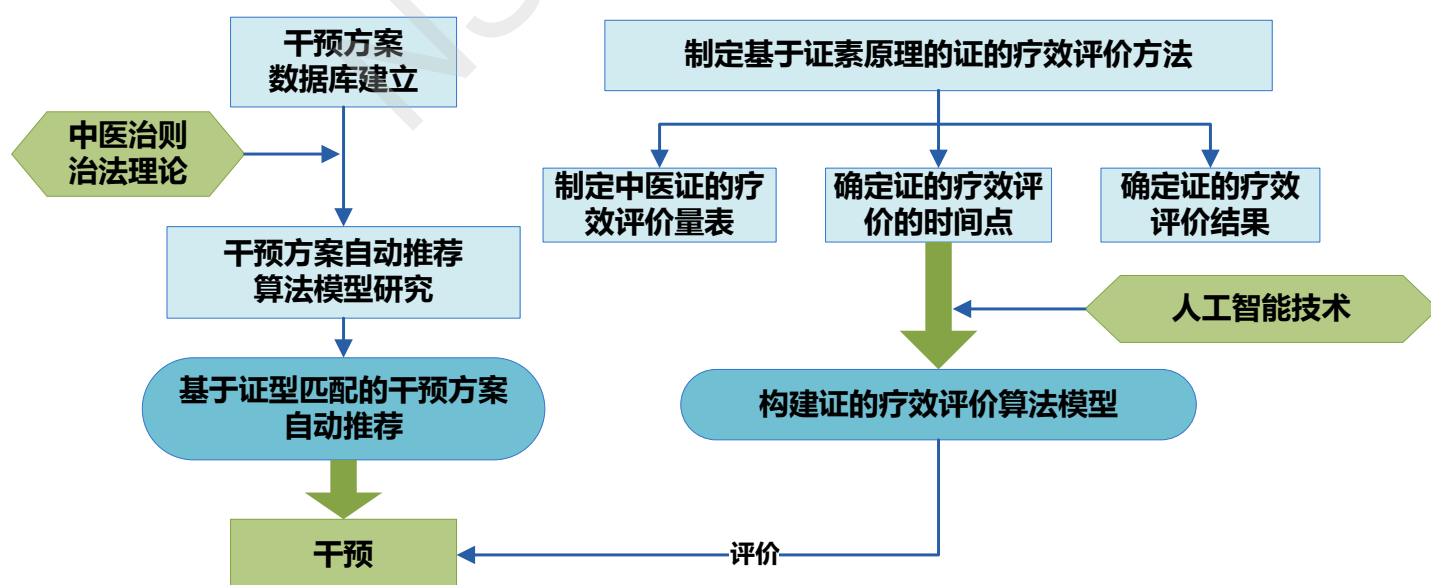


图 5 干预方案推荐与效果评价研究技术路线图



3.1.5 人工智能诊疗系统的集成

采用 Java 语言、计算机软件工程等相关技术，开发基于 B/S 模式的人工智能诊疗系统。基础层基于 Hadoop 集群实现中医四诊大数据的存储及处理。平台层以云计算为基本，实现 SaaS（软件即服务）的模式；以 MapReduce 并行计算引擎为核心，基于 Mahout 搭建并行机器学习引擎，提供多标记学习、聚类、深度学习等多种算法库，为人工智能诊断提供算法支撑。服务层以 Web service 为对外接口，实现实时诊断、治疗、查询、统计报表、硬件接口等功能。



图 6 人工智能诊疗系统的总体技术架构

3.1.6 验证性实验

3.1.6.1 系统性能测试

（1）研究对象来源：2010.01–2015.12 福建中医药大学附属人民医院、附属第二人民医院和附属厦门市中医医院的住院患者 10 万例医疗案例。

（2）医疗数据的整理和导入：对 10 万例医疗案例进行数据预处理和结构化处理，从而生成标准的测试数据。

（3）系统运行结果比对：首先，运用人工智能诊疗系统对导入的测试数据进行辨证分析，然后采用精度、召回率等评价指标分析学习系统的预测结果与实际的医院诊断记录的差异，从而实现系统性能的测试。



3.1.6.2 系统运行试验

(1) 研究对象来源：2020.07-2021.06 福建中医药大学附属人民医院、附属第二人民医院、附属晋江中医院、附属国医堂门诊病人共 1000 例。

(2) 四诊资料采集与辨证：由培训后的中级职称的中医师采用四诊仪对患者的四诊信息进行采集，并运用人工智能诊断系统辨识证素和证型。

(3) 问卷调查：设计《人工智能诊断系统医生调查问卷》和《人工智能诊断系统患者调查问卷》，内容涵盖系统性能是否良好，功能是否全面，使用是否感觉友好流畅等各个方面。根据问卷调查结果，实现从整体上评价中医智能诊疗系统的运行情况。

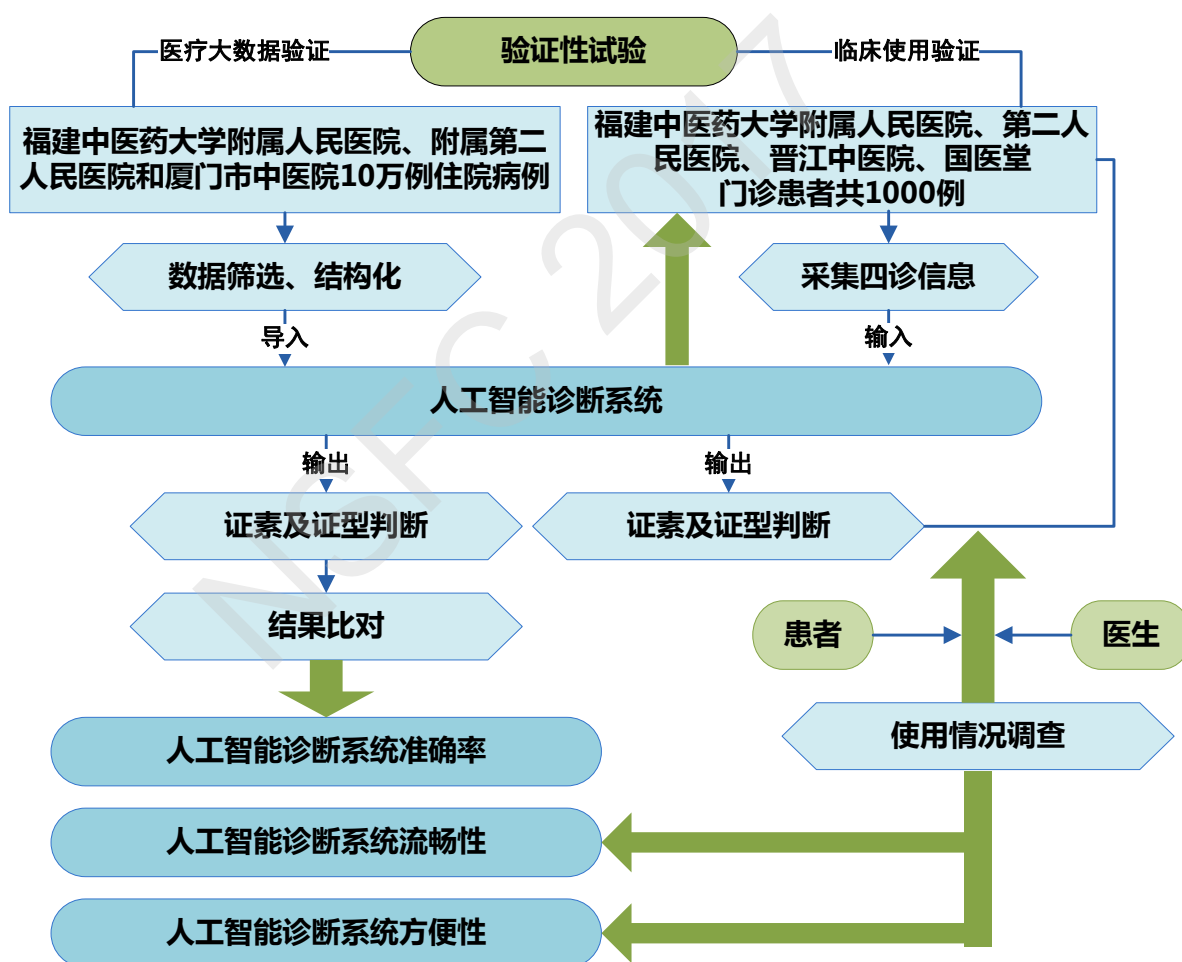


图 7 验证性实验技术路线图



3.2 总体技术路线

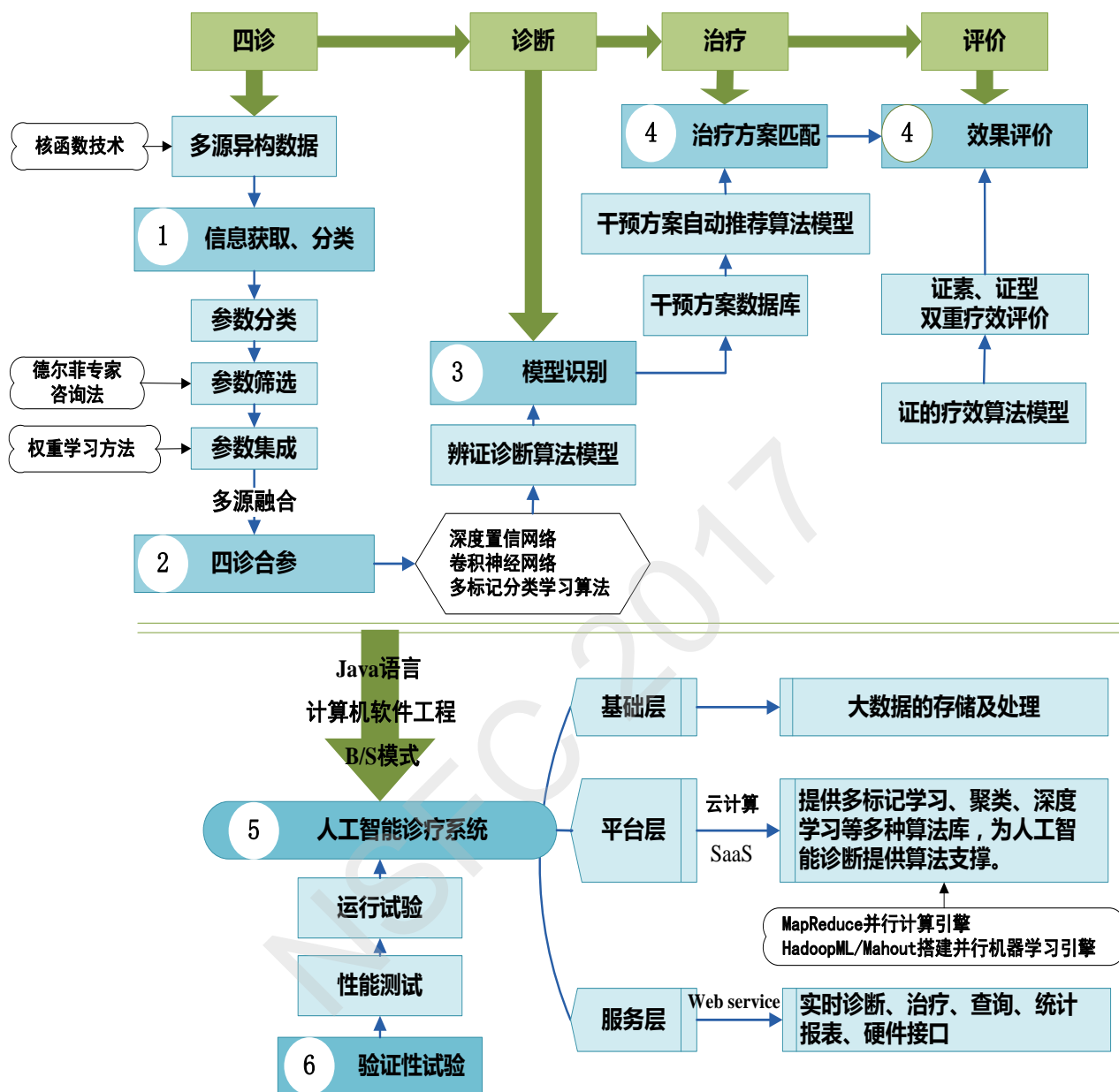


图 8 总体技术路线图



3.3 关键技术

(1) 研究融合多源异构大数据的中医辨证论治分析模型和方法

对于真实世界的被诊断对象,学习系统用多模态数据来定量表示或定性描述对象的四诊信息,以及背景与环境等信息。不同模态数据源之间关联模式的挖掘影响着分类器或回归函数的预测性能,故而,研究融合多源异构数据的中医辨证诊断算法。不同于单模态数据的分类学习过程,多源异构数据的融合不仅需要借助源于不同模态数据的知识发现,更重要的是利用多数据源的互补信息来辅助决策。因此,构建多源异构数据融合方法,采用核函数技术、相似性度量方法或损失函数将特征空间映射到知识层面,在此基础上,构建融合多数据源的权重学习方法。通过利用分类器或回归函数对训练数据进行预测,设计基于知识层和决策层的最优化算法以及基于对象增量信息的流式数据处理机制。

(2) 研究多标记学习框架下基于深度学习的中医辨证诊断算法

中医智能诊疗系统的输出联结被诊断对象的病位、病性与证型等多个类别标记,这要求学习系统具有强大的特征学习能力以应对复杂的分类学习问题。深度学习作为一种流行的机器学习模型,能够生成表示数据本质的特征,从而有效地降低多标记数据分析的复杂性(如标记相关性、类属属性及标记强弱性)以实现高精度泛化模型的构建。因此,以深度置信网络与卷积神经网络二种深度学习模型为例,研究多标记学习框架下基于深度学习的中医辨证诊断算法。其中,深度置信网络对于提升算法的鲁棒性具有积极意义。首先,以无监督的方式逐次训练多个限制玻尔兹曼机层的权值。其次,采用反向传播算法的权值更新方法,利用多标记训练数据调整模型的分类性能。另一方面,考虑对象所属标记集合,采用二进制编码等方式设计多分类学习框架下的卷积神经网络模型以及基于该模型实现融合多源数据的集成学习。

(3) 研究基于标记局部相关性的多标记分类学习和特征选择算法

中医病位、病性及证型的兼夹性导致传统的监督学习框架难以取得很好的效果。这是由于被诊断对象标记之间具有某种程度的联系,如:病患在确诊为性湿的同时,可能存在多名病患兼夹热这一病性。另外,多标记数据的高维度影响多标记分类器的分类性能。因此,相比传统的单标记学习方法,中医辨证诊断系统中的多标记学习算法研究不仅需要避免特征维数灾难,而且有必要充分利用不同病患标记相关性的差异来构造具有强泛化能力的学习系统。因此,研究基于标记局部相关性的多标记分类学习算法和多标记特征选择算法。首先对



所有对象进行聚类分析，然后设计优化算法学习标记相关性局部特征矩阵，在此基础上，一方面考虑采用稀疏规则算子作为约束项去除冗余和无关特征。另外，构建基于模糊互信息度量方法进行特征选择，最后对测试对象所属类别标记进行分类。

(4) 中医人工智能诊疗系统的构建

中医四诊数据由症状、舌苔、脉、声音、气味等多层次的信息组成，具有症状多样性、信息复杂性的特点，属于信息学中典型的“海量数据”特征。因此设计基于 Hadoop 的四诊大数据处理系统，采用 MapReduce 的分布式和并行算法来对数据进行分析。为了提高系统的灵活性和效率，满足其大规模应用中的扩展性、可靠性和按需服务等要求，将云计算的技术和理念应用于该系统，以云计算为大脑，大数据为载体，实现“云服务”式的智能诊疗系统。

3.4 可行性分析

(1) 理论可行：中医诊断学理论，特别是申请者前期关于中医健康状态的理论研究以及中医四诊、辨证及健康状态辨识相关工作基础，为本项目的开展提供了理论依据。

(2) 方法可行：借鉴现代系统工程方法，申请者已研发的四诊采集仪、中医健康管理太空舱、中医健康管理系统以及在数据库的构建、中医医案的知识挖掘、深度学习及其应用、多源异构数据分析、多标记学习等方面的研究基础能够为本项目的研究提供方法学支撑。

(3) 技术可行：课题组由不同学科、专业人员组成，主要成员长期从事中医诊断、诊疗设备、人工智能、数据库构建等相关技术研究，具有丰富的研究经验，能够为本项目在原有研究基础上进一步深入和高质量完成提供技术支持。

(4) 条件可行：申请单位和合作单位是福建省 2011 中医健康管理协同创新中心的协同单位，拥有“国家中医药管理局中医健康状态辨识重点研究室”、“福建省中医健康状态辨识重点实验室”、“福建省中医证高校重点实验室”，同时拥有的“四诊资料标准化采集实验室”和“证素辨证和数据挖掘技术实验室”均为全国同类中唯一的等级实验室，承担了包括国家重点基础研究发展计划（973 计划）项目、国家自然科学基金重点项目及面上项目等基础课题研究工作，能够保证本研究结果正确、可靠。

项目合作单位厦门大学智能科学与技术系，拥有福建省类脑计算技术及应用重点实验室和智能信息技术福建省高校重点实验室两个科研平台，在类脑计算、



机器学习和数据挖掘、计算机视觉与自然语言处理、智能中医信息处理等领域有良好的研究基础和实验条件，可以很好地协助、合作完成相关的研究。

4. 本项目的特色与创新之处；

(1) 理论创新：遵循中医思维规律，以整体健康状态的动态、个性化测量与评估为核心，构建中医诊断、治疗和疗效评价的理论体系，为中医诊疗系统的研发奠定理论基础。

(2) 方法学创新：基于状态辨识的原理，应用人工智能技术，设计多源异构信息的处理方法，实现四诊合参；建立多模态多标记数据下的分类学习算法，构建中医类人认知体系和思维机制，构建四诊构建状态辨识（诊断）的算法模型，为中医诊疗系统的研发提供方法学依据。

(3) 技术创新：借鉴系统工程的原理，结合大数据、云计算和人工智能技术，构建中医人工智能诊疗系统，实现从表征参数采集、分类、融合到模型识别、诊断（状态辨识）、治疗和疗效评价的中医诊疗过程的一体化、客观化、规范化和智能化，为中医的理论和临床研究提供技术支撑。

5. 年度研究计划及预期研究结果（包括拟组织的重要学术交流活动、国际合作与交流计划等）。

5.1 年度研究计划

2018.01-2018.06：完成国内外文献调研，召开课题讨论会，进一步完善课题的顶层设计，制定课题具体实施方案，明确任务分工及要求，完善操作管理规范，培训参加临床、实验、统计分析研究人员。

2018.07-2018.12：四诊信息采集、存储研究，完善各信息采集模块；参加学术交流 1-2 次。

2019.01-2019.06：多源异构数据分类、信息融合方法及决策研究；发表论文 3-4 篇，参加学术交流 1-2 次。

2019.07-2019.12：中医诊断算法模型研究，构建类人认知体系架构和思维机理，多标记学习框架下基于深度学习的中医智能诊断算法研究；发表论文 3-4 篇，参加学术交流 2-3 次。

2020.01-2020.06：实现基于人工智能的中医诊断机制，构建以中医健康状态为核心的人工智能中医诊疗系统，调试系统流畅性；发表论文 3-4 篇，参加学术交流 2-3 次。

2020.07-2020.12：以附属医院住院病历 10 万份为标准数据集，验证人工智