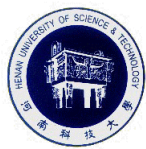


分类号 _____

UDC _____

密 级 _____

编 号 _____



河南科技大学

专业硕士 学位论文

面向中医证候的智能处方推荐方法研究

Research on Intelligent Prescription Recommendation Method for
TCM Syndrome

学位申请人：

刘丹丹

指导教师：

张明川 教授

合作教师：

张虹 主任医师

专业领域：

计算机技术

学位类别：

工程硕士

2022 年 05 月

独创性声明

本人声明，所呈交的论文是我个人在导师指导下完成的研究工作及取得的研究成果。据我所知，文中除了特别加以标注和致谢的地方外，不包含其他人已经发表或撰写过的研究成果，也不包含为获得河南科技大学或其它教育机构的其他学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已 在论文中作了明确的说明并表示了谢意。

研究生签名： 刘丹丹
日 期： 2022.5.23

关于论文使用授权的说明

本人完全了解河南科技大学有关保留、使用学位论文的规定，即：学校拥有对所有学位论文的复制权、传播权、汇编权及其它使用权（特殊情况需要保密的论文应提前说明，但在解密后应遵守此规定）。

需要保密的论文请填写：本学位论文在____年____月至____年____月期间需要保密，解密后适用本授权书。（需要保密的学位论文无须向图书馆提供论文的电子文档）

研究生签名： 刘丹丹 导师签名： 张明
日 期： 2022.5.23

摘 要

中医是我国独具优势和特色的传统医学体系，是中华优秀传统文化不可或缺的一部分。利用现代互联网技术发展中医行业，探索中医与人工智能融合的新模式，是传承中医药精华、守正创新的重要内容，对中医在治未病、重大疾病治疗、疾病康复中具有技术支撑和保障作用。中医师问诊具有以下步骤：通过望、闻、问、切等模式确定患者的证候（即症状和体征）；通过辨别证候给出治法，治法是针对临床证候采取的具体治疗方法；最终依据治法开出个性化处方。因此，本文拟设计一种模拟中医问诊思路，实现以患者证候作为输入，治法作为桥梁，处方信息作为输出的过程，辅助医生提高问诊效率的处方推荐系统。该系统将中医与人工智能相结合，使用深度学习技术从海量中医诊疗数据里学习其规律并发现处方新知识。本研究主要工作如下。

1. 提出文本卷积神经网络中医论治算法，采用文本卷积神经网络构造一个非线性的映射关系，通过训练和学习网络，实现了从中药数据中自动抽取中医证候的内在规律，把治法的多分类结果以概率的形式展现出来，选取最大的概率作为治法的结果进行输出，从而模拟证候到治法的过程，并将其应用于新病例的诊断。经实验结果分析，基于文本卷积神经网络的中医论治模型训练速度快、并且具有更高的准确率。

2. 提出注意力分解机处方推荐算法，采用因子分解机的二阶交叉项将患者和处方的多类别属性进行交叉组合，在因子分解机上添加注意力网络，对处方中有效特征赋予更高的权重，从而区分处方中不同组合特征的重要程度，提高模型的推荐性能。经实验结果分析，基于注意力分解机处方推荐模型不需要依赖人工特征，具有较好的推荐性能，实现了一定程度的个体化推荐效果。

3. 智慧中医辅助诊疗系统的设计与实现。本系统依据中医师选法定方的问诊思路，采用 MVC 架构，实现了系统界面与处方推荐流程的分离，协助医生问诊确认患者的证候，通过本系统进行开方。该系统主要为医院以及中医馆的医生提供方便，提高医生的诊断效率和正确率。

关 键 词： 中医信息化；中医论治；处方推荐；深度学习

论文类型： 设计类

选题来源： 国家自然科学基金(62002102, 62102134)

ABSTRACT

Traditional Chinese Medical(TCM) is a Traditional Medical System with Unique Advantages and Characteristics. It is an indispensable part of China's excellent traditional culture. Using modern Internet technology to develop the TCM industry and explore a new mode of integration of TCM and artificial intelligence is an important content of inheriting the essence of TCM, upholding integrity and innovation, and plays a role of technical support and guarantee for TCM in the prevention of diseases, treatment of major diseases and recovery from diseases. TCM physicians' consultation includes the following steps: to determine the syndromes (symptoms and signs) of patients through the modes of seeing, hearing, asking and cutting; to give therapeutic methods by differentiating syndromes, which are the specific therapeutic methods for clinical syndromes; and to finally make personalized prescriptions according to the therapeutic methods. Therefore, this paper designs a kind of TCM consultation system, which is based on patient syndrome as input, cure as bridge, prescription information as output, and assists doctors to improve consultation efficiency. The system combines traditional Chinese medicine with artificial intelligence, uses deep learning technology to learn its rules and discover new prescriptions from massive traditional Chinese medical data. The main work of this study is as follows.

1. This paper convolutional neural network TCM therapeutic methods algorithm is proposed, and a non-linear mapping relationship is constructed by using the text convolutional neural network, and through the training and learning network, the inherent law of automatic extraction of TCM symptoms from Chinese medicine data is realized, the multi-classification results of the treatment method are displayed in the form of probability, and the maximum probability is selected as the result of the therapeutic for output, so as to simulate the process of waiting for therapeutic and applying it to the diagnosis of new cases. According to the analysis of experimental results, the TCM therapeutic model based on text convolutional neural network has fast training speed and higher accuracy.

2. The prescription recommendation algorithm of the attention factorization machines is proposed, which uses the pair-wise interaction of the factorization machines to interaction the multi-category attributes of the patient and the prescription,

adds the attention network on the factorization machine, and gives higher weight to the effective features in the prescription, so as to distinguish the importance of different combination features in the prescription and improve the recommended performance of the model. According to the analysis of experimental results, the prescription recommendation model based on the attention decomposition machine does not need to rely on manual features, has good recommendation performance, and achieves a certain degree of individualized recommendation effect.

3. Design and Implementation of Intelligent TCM Assistant Diagnosis and Treatment System. According to the idea of TCM doctors selecting legal prescriptions, this system adopts MVC framework, and realizes the separation of system interface and prescription recommendation process. The system provides convenience for doctors in hospitals and Chinese medicine centers, and improves the diagnosis efficiency and accuracy of doctors.

KEY WORDS: Informatization of Traditional Chinese Medicine; Therapeutic Methods; Prescription recommendations; Neural networks

Dissertation type: Design class

Subject source: National natural science foundation of China(62002102,62102134)

目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.3 论文主要内容	4
1.4 论文组织结构	5
第 2 章 相关理论及技术	7
2.1 注意力机制	7
2.2 推荐算法概述	10
2.2.1 协同过滤推荐算法	10
2.2.2 Factorization Machine	11
2.2.3 Neural Factorization Machine	12
2.3 数据挖掘	13
2.3.1 数据挖掘算法	13
2.3.2 数据挖掘过程	14
2.3.3 中医药数据挖掘分析	15
2.4 本章总结	16
第 3 章 基于文本卷积神经网络的中医论治算法研究	17
3.1 引言	17
3.2 系统模型	18
3.2.1 模型应用场景	18
3.2.2 文本卷积神经网络	18
3.2.3 中医论治模型	19
3.3 算法设计与分析	23
3.4 实验	25
3.4.1 实验数据	25
3.4.2 实验设置及分析	25
3.5 本章小结	29
第 4 章 基于注意力分解机的中医处方推荐算法研究	31
4.1 引言	31
4.2 系统模型	33

4.2.1 模型应用场景	33
4.2.2 注意力分解机	33
4.2.3 中医处方推荐模型	34
4.3 算法设计与分析	37
4.4 实验	38
4.4.1 实验数据	38
4.4.2 实验设置及分析	39
4.5 本章小结	42
第 5 章 系统设计与实现	43
5.1 系统需求分析	43
5.1.1 系统目标	43
5.1.2 功能需求分析	43
5.1.3 性能需求分析	44
5.1.4 可行性分析	44
5.2 系统概要设计	45
5.2.1 总体设计	45
5.2.2 架构设计	45
5.2.3 功能模块设计	46
5.2.4 数据库设计	48
5.3 系统实现与测试	52
5.3.1 运行环境与开发平台	52
5.3.2 功能模块实现	52
5.3.3 功能测试	55
5.4 本章小结	56
第 6 章 结论	57
6.1 总结	57
6.2 展望	58
参考文献	59
致 谢	63
攻读学位期间的研究成果	64

第1章 绪论

1.1 研究背景与意义

中医是中华五千年文明智慧的结晶，蕴藏着深厚的哲学智慧、在促进文明交流、维护人民健康等领域有着举足轻重的地位。中医行业在经历了建国以来的震荡恢复期、改革开放的探索发展期之后，当前已经进入了全面行业复苏时期。在时间与历史的验证下，中医将其独特的生命力彰显出来。中医依靠自身的特点和优势，使其受到世界各地关注和青睐。中医独特的发展潜力，必定对社会经济产生巨大影响力。党的十八大以来，习近平总书记对中医的特殊优势和功能给予了充分的肯定，指出了中医药在新时期的新发展。2020 年新冠疫情期间，中医药在疫情防控、临床救治等方面做出了非常巨大的贡献^[1,2]。习近平总书记也发表重要讲话，他从人民安全和身体健康的战略高度出发，充分肯定中医在疫情防控中做出的贡献。中医证候是疾病在某个阶段发展过程中包括病因、病势等一系列病理变化的总称，是通过望闻问切四诊合参得到的。因此，中医证候对中医诊疗而言至关重要^[3-5]。

伴随着近年来我国促进各方面中医事业发展的政策颁布，同时，结合我国国情，人口老龄化、慢性病增长速度快、医疗资源的供需不平衡、区域分布不均衡，使得医疗人工智能成为一个庞大的市场。所以，人工智能是解决医疗效率问题的关键。近年来，“人工智能+医疗”的发展迅猛，是古老中医在科技时代再次焕发光彩的必要途径^[6-9]，为中医药发展注入了活力，也为中医药发挥了更大价值提供了有力支撑。因此，我国积极推进中医药信息化建设，正确把握中医药科技在中医药传承和创新发展中的大目标和方向，对促进健康中国的建设具有重要作用^[10,11]。

中医的发展模式与西医截然不同，西医的诊疗思维是规范化的、标准化的，西医从实验数据中归纳概括规律，再通过修改发展成诊断理论。中医的临床思维则是多元化的、模糊化的，是在实践中不断提炼、探索、归纳出极具个性化的理论^[12]。而中医这种临床思维的特性，更适合信息技术特别是人工智能技术发挥其辅助诊疗的优势。目前传统中医学习路径复杂、成长过程漫长，加上中医的道理更加深奥，并缺乏令其他领域和专家普遍接受的证据，这些使得中医的传承与发展受到一定限制。因此中医问诊和人工智能技术相结合是中医信息化建设的核心^[13,14]。近些来，很多专家辅助诊断系统的技术和方法都在进行着探索，这类系统在某种程度上可以帮助医生提升诊疗效率，为患者也提供了方便^[15]。然而这

类医学专家系统是依据专家的先验知识，通过先验知识生成了相关联的规则，为医生提供的决策价值也不高。此外，这些中医信息和资源也不能被共享，在一定程度上制约着中医信息化发展。

通过物联网技术以及大数据分析能力，让计算机担当起“名医”的角色，将会在很大程度上提高中医证候开处方的诊断效率^[16,17]。“人工智能+辅助诊疗”，就是利用人工智能技术进行辅助诊疗，通过计算机来学习医学知识，模仿医生的判断，提供正确的诊断和治疗方法^[18]。目前，中医问诊中，选法定方是中医问诊最重要的一个环节，首先医生要辨别患者的证候，然后从病人的临床实践中总结出治疗方法，然后再给患者配药，这是中医诊断的一系列流程。治法作为证候和处方之间的桥梁，治法一旦确定，处方也就相对应。然而患者证候和中医治法间存在复杂非线性关系，使“证候”、“治法”与“处方”的彼此对应变得极其繁杂。“人工智能+辅助诊疗”的应用，将患者全部的数据，包括患者个人信息、证候、治法和方药等结构化数据组织在云端数据库里。利用深度学习，借助标准化的数据采集手段分析中医处方数据文本的特征，构建具有多个隐藏层的学习模型，自动学习更高层次的特征，实现“证候”到“治法”再到“方剂”的对应，从而增强模型分类和推荐的准确率。通过辅助诊疗系统对用户的治疗和用药进行记录，该系统根据患者的疗效反馈，持续学习，促进中医药信息化的发展^[19]。与此同时，随着人工智能仪器地问世，中医人才培养周期得到了极大地缩短，提高了医疗人员的工作效率^[20]。

本课题的研究内容是利用现代计算机技术以及大数据分析能力，运用中医药处方数据，提取归纳中医证候数据的特征信息，结合专家经验、中医理论知识，将治法作为桥梁从而推荐出智能的个性化中医处方。中医处方辅助系统可以辅助医生为患者提供健康管理服务，让中医数据不再是一座座孤岛，打破传统中医派系壁垒，实现数据共享共融，促进高质量中医经验传承发展，对打造健康中国、构建人类卫生健康共同体，具有战略性的意义。

1.2 国内外研究现状

随着人工智能技术和计算机技术的不断发展，知识工程得到了迅速的发展。在知识工程方面，最常用和最有成果的就是专家系统，其中最受瞩目的，就是医疗专家系统，它能够模仿人类的心理思维活动。医学诊断是专家系统中的一个重要领域。

医疗专家系统在技术上经历了孕育、产生、成熟和发展等阶段。在人工智能的影响下，国内外先后出现了诸多专家系统。MYCIN 是斯坦福在 70 年代早期

研发的世界上首个医疗专家系统，它是首个在医学领域中运用人工智能来诊断和治疗传染病的系统。MYCIN 系统的问世对医疗专家系统的发展起着举足轻重的作用，同时也是人工智能医疗中一个重要的里程碑。但是由于该专家系统只能在昂贵大型计算机上执行，因此，第一个人工智能专家系统并没有被真正应用到实践中。20 世纪 80 年代，几个具有庞大知识库和可以直接在计算机界面上操作的专家系统也被研制出来，比如美国犹他大学 1985 年的医疗专家系统，就是以大量的医院数据为基础，通过大量的数据库来做出决定。1986 年匹兹堡大学花费 10000 人时的工作量，通过各个著名科室专家经过不断地更新改进，构建了包含 3000 多种疾病症状，可以治疗 500 多种内科疾病的专家系统，美国国家卫生院曾对该专家系统进行严格临床试验，证明这些系统至少具有一般内科住院医师的诊断能力。

20 世纪 90 年代，因为专家系统地运用领域比较局限，导致先验知识获取困难，以及不确定的常识推理有障碍。专家系统发展比较迟缓，进入 21 世纪以来，医学专家系统虽然发展缓慢，但是也有一些显著的成果。2016 年 9 月，微软宣布和克里夫兰诊所合作，帮助这家医疗中心为接受 ICU 治疗的患者确定潜在的风险。微软开发者利用微软人工智能数字助手对病人的生命体征和实验数据等信息进行了预测性分析的探索，他们利用从受到监控的 ICU 收集到的病人生命体征和实验数据等信息数据，融合了机器学习算法去探索先进的预测性分析。同年，美国休斯顿卫理公会医院的研究者们在《癌症》杂志上写到，配备人工智能工具的放射科医生审查和解读乳腺 X 线照片的速度比正常情况下快 30 倍，准确率高达 99%。2018 年 Cognoa 公司设计开发一款 CognoaAPP，这是首个为自闭症患儿设计的智能辅助系统，利用人工智能技术，对孩子是否以正确的成长速度进行评估，从而判断孩子发育是否正常。据报道，这个 APP 在临床上的应用程序被用来诊断儿童的孤独症，其正确性在 95% 以上。

相对于世界先进国家，我国的医疗专家系统起步较晚，但发展很快。80 年代初期，我国开展了计算机辅助诊疗的研究工作。1987 关幼波教授带领北京中医院的研究小组，利用计算机技术，将传统中医药和现代计算机技术相结合，共同研发了关幼波肝脏病的诊断程序，关老的肝病辨证论治诊疗程序于 1979 年 1 月 23 日正式投入使用。1986 年，福建中医药大学联合省立计算机中心，将林如高教授的医学理论和理论成果录入到计算机中，并联合其他科研单位，研制出了林如高骨伤计算机系统，并且成功的应用于临床。到 21 世纪，各种各样的专家诊疗系统不断出现，比如模糊口腔癌症诊疗专家系统，胃癌诊断专家系统，结核病诊断专家系统等。新冠疫情期间，四川研发“肺部多病变 CT 影像 AI 筛查与

辅助诊疗系统”，这个辅助诊疗系统可以对数百张 CT 影像进行筛查，并将该影像的“肺炎”标注出来，同时，该辅助诊疗系统可以对患者的胸部 X 线影像进行量化分析，并与之进行对照，为医师提供诊断依据。这套医疗专家系统对于保存、继承、研究、推广名老中医的实践、培养青年中医、推进中医药现代化建设，意义重大。

文本分类是目前国内外研究的热点之一，它在自然语言处理中占有重要的地位。虽然在以往的研究中，已经提出了许多有效的算法来对英文文本进行分类，并取得了良好的效果，但中文文本分类由于在处理过程中遇到了巨大的困难而没有受到足够的重视。在大数据技术和运算速度不断加快的今天，机器学习已经被越来越多地使用。Li 等^[21]提出了一种基于（Latent Dirichlet Allocation, LDA）的新闻文本分类模型，采用主题模型来降低文本维度并获取特征，以解决生活中的多分类文本；Zhang 等^[22]使用 n-gram 技术将自然文本句子转换为单词待的特征，然后将特征输入到双向长短时记忆网络对电影评论数据进行分类；Gao 等^[23]提出利用词嵌入和长短时记忆网络相结合的方法，将学习到的词嵌入作为模型的初始特征，并且使用类别因子和 TF-IDF 生成权值来增加特征，最后对短文本集进行分类。

在人工智能飞速发展的今天，将现有的计算机技术与中医理论知识相结合，使计算机智能化地通过选法定方为病人开出个性化处方，将会在很大程度上提高中医证候开处方的诊断效率。目前研究中，对于中医治法规律的研究主要采用频次分析和主题模型方法进行治法规律分析。2017 张颖等^[24]通过对病案中证候与处方的分析，构建隐语义模型，发现隐含病机，通过隐含病机建立联系从而根据证候推荐药物。2018 年 Li 等^[25]通过收集大量数据，采用神经网络模型，以中医证候作为输入，输出结果为端到端的中草药生成处方。2020 年李鹏飞等^[26]对患者的体征和药物的功效相融合进而推荐处方。虽然都在一定程度上促进了中医个性化诊疗，但是并没有真正建立起中医证候和处方间关于治法的联系，并且可解释性还不强。

从整体上看，目前对中医药的推荐问题的研究相对较少。多数是针对中医证候和处方间的关联关系之间的挖掘，以中医证候为依据，由药物的配伍组成。从某种意义上来说，这与中医选法定方的思维方式不符。因此本文将中医的诊断思路考虑进去。

1.3 论文主要内容

人工智能在医疗领域的应用非常重要，也是医疗的核心领域，“人工智能+

辅助诊疗”的市场潜力很大。本文根据中医师选法定方的问诊思想设计并实现了中医智能处方推荐系统。本文提出利用深度学习和神经网络解决患者证候、治法和处方间存在的复杂关系，建立了中医论治模型和处方推荐模型，从而提高医疗诊治水平。中医辅助诊疗系统在一定程度上节省了医疗人力。本文的研究主要分为以下部分。

1. 基于文本卷积神经网络的中医论治模型。高质量的数据是建模的基础，中医证候与治法间存在非线性的关系，本文对从古典书籍中收集的“半结构化”数据进行预处理，通过 Glove 嵌入方式将中医数据自然语言数值化处理，采用文本卷积神经网络隐藏层提取中医药数据的关键信息，最终通过输出层输出预测分类。文本卷积神经网络模型训练速度快、稳定性好。

2. 基于注意力分解机的处方推荐模型。该算法融合处方药材的治疗信息以及患者的不同体征得到更加智能化、个性化的推荐。本文对目前存在的推荐算法进行学习，根据治法开处方的特点，提出因子分解机和注意力相结合的推荐算法。注意力分解机模型更加简单，不需要依赖人工特征，效果更优。通过在注意力分解机上添加注意力网络，从而对有用目标赋予更高的权重，提取占据重要作用的交叉特征，并且可以根据不同患者的不同体征进行个性化推荐。

3. 智慧中医辅助诊疗系统的设计与实现。根据中医师选法定方的问诊思路，完成系统的需求分析和总体设计。智慧中医辅助诊疗系统主要为医院以及中医馆的医生提供方便，协助医生确认患者的证候和治法，进而给出个性化处方，提高医生的诊断效率和正确率。

1.4 论文组织结构

根据论文的内容要求，全文共分六章。

第一章，绪论。本研究首先分析了我国中医药发展与信息化技术结合的背景以及意义，接下来具体介绍了中医智能诊断开方在国内外发展的现状。通过对以上分析，提出智能处方设计的必要性。最后介绍了本文的主要研究内容与论文的组织结构。

第二章，相关理论及技术。本章对本论文所需要的技术路线以及算法的理论、发展历程、优缺点等进行了详细地介绍。主要包括数据挖掘技术、注意力机制以及常见的推荐算法。

第三章，基于文本卷积神经网络的中医论治算法。为了处理中医证候到治法间复杂的非线性关系。本文基于权重共享、局部连接等特点的卷积神经网络，构建了基于卷积神经网络的中医论治模式。对患者进行论治，并通过实验进行对比

分析,结果显示本研究的算法在提高分类精度地基础上,也能够快速的进行模型的训练。

第四章,基于注意力分解机的处方推荐算法。首先介绍了中医处方数据的特点,并对该数据进行数值化处理,通过分解机对患者信息和中医处方信息进行二阶特征交互,在此基础上加入注意力机制,从而对有用目标赋予更高的权重,提取占据重要作用的交叉特征。从而对患者进行个性化处方推荐。

第五章,智慧中医辅助诊疗系统的设计与实现。本系统依据中医师选法定方的问诊思路,采用 MVC 架构,实现了系统界面与处方推荐流程的分离,协助医生问诊确认患者的证候,通过本系统进行开方。该系统主要为医院以及中医馆的医生提供方便,提高医生的诊断效率和正确率。

第六章,总结与展望。本章是对全文的总结,并概括了本论文的中医论治模型以及处方推荐模型,最终在此系统的基础上提出本论文中的不足以及尚未研究的工作。最终对中医智能化发展为人们带来了福利进行感慨以及对未来的工作进行了展望。

第2章 相关理论及技术

2.1 注意力机制

注意力机制（Attention Mechanism）^[27]是利用神经网络来进行数据的加工，在各个领域都有广泛的应用。比如，机器翻译。换句话说，注意力机制就是当你在某一时刻对某一事物进行观察时，让你把注意力放到某些具体的物体上，对具体事物的重点信息进行关注。在大数据时代，模型将会接收和处理海量的数据，但是对于某一时间段，我们只需要重点关注一些局部关键信息，因此，引入注意力机制就显得至关重要。

大部分注意力模式都是依附于编码-解码（Encoder-Decoder）的结构，而编码-解码器架构则是一个用于研究深度学习的模型，有着广阔的应用场景。在文字处理中，最抽象的 Encoder-Decoder 框架如图 2-1 所示。

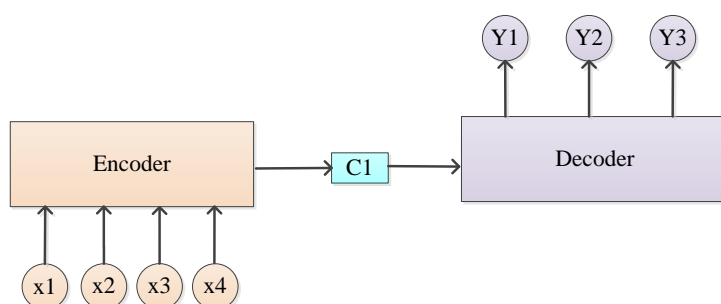


图 2-1 Encoder-Decoder 结构^[27]

Fig. 2-1 Encoder-Decoder structure^[27]

在自然语言处理中，对于句子 $\langle Source, Target \rangle$ ，我们将 *Source* 单词序列作为输入目标，通过语义编码 c 进行转化，最后产生目标单词序列 *Target*。*Source* 和 *Target* 的单词序列分别为：

$$Source = (x_1, x_2, \dots, x_m) \quad (2-1)$$

$$Target = (y_1, y_2, \dots, y_n) \quad (2-2)$$

Encoder 是一种编码程序，它把输入的语句用非线性形式转换成一种中间语义 c ，由解码器 Decoder 产生 i 时刻要生成的单词信息 y_i ：

$$c = f(x_1, x_2, \dots, x_m) \quad (2-3)$$

$$y_i = g(c, y_1, y_2, \dots, y_{i-1}) \quad (2-4)$$

因为在 Encoder-Decoder 架构中，目标句子单词 y_i 的产生采用了完全一致的语义编码 c ，没有区别，这就与人在观察某一幅图画时，眼中却没有注意力焦点是一样的。但是在实际应用中，句子 *Source* 对任意输入 x_i 的影响概率是不同的，在图 2-2 中显示了添加了注意力机制的 Encoder-Decoder 框架。

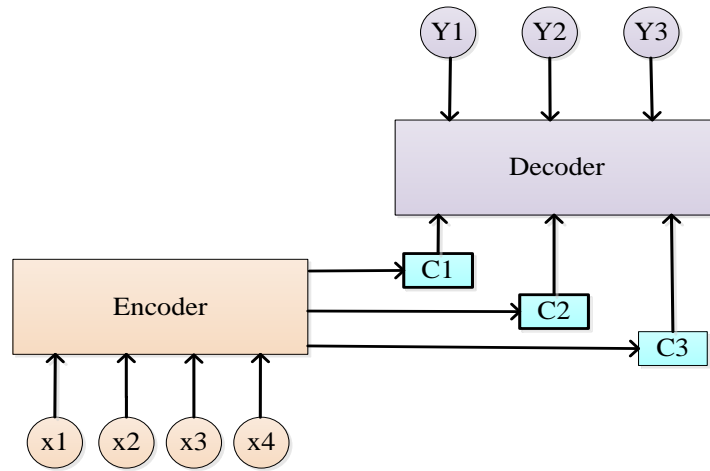


图 2-2 注意力机制的 Encoder-Decoder 结构^[27]

Fig. 2-2 Attention-Encoder-Decoder structure^[27]

引入注意力机制后，输出单词序列 *Source* 的目标函数转化为 y_i 。每个 c_i 表示输入单词相对应的注意力分布概率， G 表示了一个非线性变换函数，下面的方式是 c_i 的权重和运算方法：

$$y_i = G(c_i, y_1, y_2, \dots, y_{i-1}) \quad (2-5)$$

$$c_i = \sum_{j=1}^{L_x} a_{ij} h_j \quad (2-6)$$

其中， a_{ij} 代表 *Target* 输出第 i 个文本单词时，*Source* 在语句中输入第 j 个单词的注意力的概率， h_j 代表了 *Source* 输入语句的第 j 个单词的语义编码， L_x 表示输入句子 *Source* 的长度。

注意力机制的实质是：*Source* 的输出序列是由一些列 *key-value* 数据对构成，是通过对 *Target* 中 *Query* 的判定来获得每个 *Key* 和 *Query* 之间的相似性，从而得到对应的权值。计算出最后的 *Attention* 值。公式如下所示：

$$Attention(Query, Source) = \sum_{i=1}^{L_x} Similarity(Query, Key_i) * Value_i \quad (2-7)$$

注意力机制的权重越大，则 *Attention* 越聚焦于对应的 *Value* 值上。该方法包括以下步骤：首先求出 *Query* 和 *Key* 的权值，第二步是将输入的 *Value* 用权

重系数进行加权求和，*Attention* 的运算过程为如图 2-3 所示。

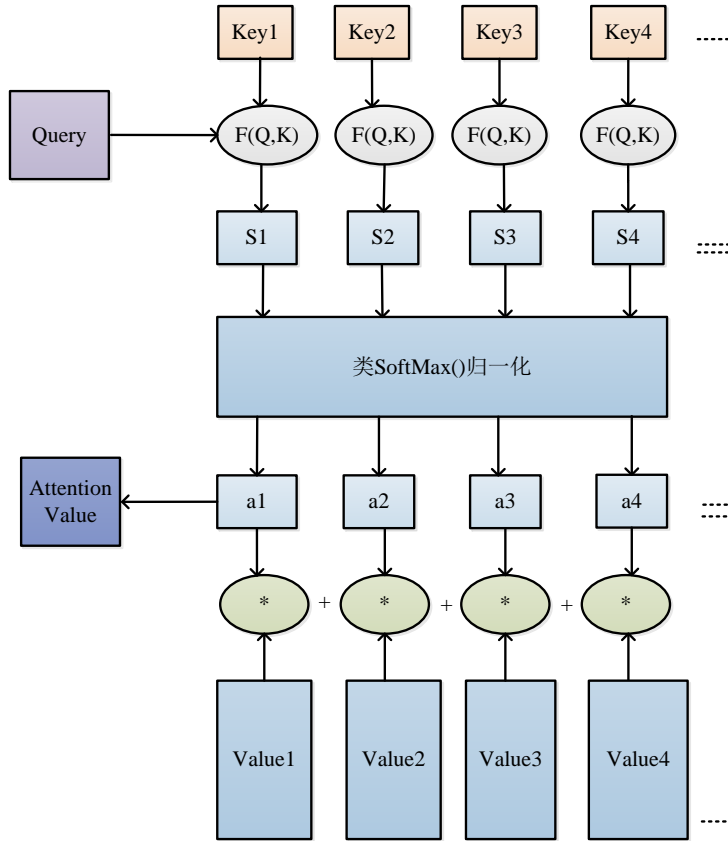


图 2-3 Attention 计算过程^[27]

Fig. 2-3 Attention calculation process^[27]

Q 是输出的目标查询元素 *Query*， K 是关键字 *Key* 的缩写， S 是 *Query* 和 *Key* 的相关性， F 是函数， a 和 *Value* 分别表示系数和权重系数值，*Attention* 值即为注意力机制。

首先，利用各种函数和运算机制实现 *Query* 与 key_i 之间的相似性。大部分的算法都是计算它们的点积、矢量的余弦相似度以及再引进的神经网络的值，它们的表示方式是：

$$\text{点积: } \text{Similarity}(\text{Query}, key_i) = \text{Query} \cdot key_i \quad (2-8)$$

$$\text{Cosine 相似性: } \text{Similarity}(\text{Query}, Key_i) = \frac{\text{Query} \cdot key_i}{\|\text{Query}\| \cdot \|Key_i\|} \quad (2-9)$$

$$\text{MLP 网络: } \text{Similarity}(\text{Query}, Key_i) = \text{MLP}(\text{Query}, Key_i) \quad (2-10)$$

第二个步骤是通过 Softmax 对第一步的得分转化为标准化过程，即 a_i 一般采用如下公式计算：

$$a_i = \text{Softmax}(\text{Sim}_i) = \frac{e^{\text{Sim}_i}}{\sum_{j=1}^{L_x} e^{\text{Sim}_j}} \quad (2-11)$$

第二步的计算结果 a_i 即为 $valu_i$ 相应的权值，再经过加权求和，得到 *Attention* 的值：

$$\text{Attention}(\text{Query}, \text{Source}) = \sum_{i=1}^{L_x} a_i \cdot \text{Value}_i \quad (2-12)$$

由上述三个步骤进行运算，得到 *Query* 和 *Attention* 值。这就是 a 的基本思想。

2.2 推荐算法概述

推荐系统是针对信息过载问题提出的，随着中医药行业的快速发展，中医人才的需求也越来越大，利用人工智能给患者推荐处方是当前各大医院和中医馆的迫切需求。推荐系统能够对大量的中医药材过滤，根据病人描述的证候以及给出的相应治法，对病人建立相应的处方推荐模型，为病人完成数据筛选和处方推送。目前，推荐系统在互联网上已经得到了广泛的应用，研究人员也开展了研究，提出了一系列的推荐算法，效果各有优势。

2.2.1 协同过滤推荐算法

协同过滤推荐^[28]算法主要包括基于用户的协同过滤和基于项目的协同过滤，根据实际情况而定，选择最适合的推荐算法。基于用户的协同过滤推荐机制是一种新的有效的方法。该方法基于用户的喜好，将其分成了预测与建议两类，利用用户的习惯、记录等特征对其进行分析。基于用户偏好找到相似用户的喜好模型，继而进行推荐。基于项目的推荐用于计算项目的相关性，在大多数网站上，项目数量和用户数量相比可以忽略不计，所以，在项目数目上相似度比较稳定，并且，一项目为基础的推荐具有实时性。因而，与基于用户的推荐算法相比，基于物品的推荐方法具有更大的适用性。通常用以下方法来确定相似性 *Sim*。 u 和 v 分别代表用户集合里两个不同的用户。

1. 皮尔逊相关系数

皮尔逊相关性是一种常用的方法，它是一种度量变量之间的相关性，它可以在某种意义上帮助我们了解各种因素的关系，而皮尔逊相关性则可以在不太高的情况下得到更好的结论。计算公式如下：

$$\text{Sim}(u, v) = \frac{\sum (u - \bar{u})(v - \bar{v})}{\sqrt{\sum (u - \bar{u})^2 (v - \bar{v})^2}} \quad (2-13)$$

2. 余弦相似度

余弦相似度是由两个向量夹角的余弦值来度量的。 $N(u)$ 是用户 u 已经正反馈过的物品汇合, $N(v)$ 为用户 v 已经正反馈过的物品集合。一般情况下, 余弦值越高说明相关性越大。公式如下:

$$Sim(u, v) = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| * |N(v)|}} \quad (2-14)$$

3. 杰卡德相关系数

杰卡德相关系数是用来测量布尔值与符号测量的个体关联度, 而不能测量特定数值的差别。它们之间的相关性系数越大说明相似度越高。公式如下:

$$Sim(u, v) = \frac{N(u) \cap N(v)}{N(u) \cup N(v)} \quad (2-15)$$

2.2.2 Factorization Machine

(Factorization Machine, FM)^[29]是一种通用的预测方法, 对稀疏数据具有很好的学习能力, 即使数据很少, 也可以组合特征以生成可靠的参数和预测。与传统的线性推荐算法相比, FM 模型特征两两自动组合, 对所有特征向量交叉建模。FM 是推荐系统、广告系统的三大基础模型之一。

首先引入了一种基于线性回归的方法, 该方法的原理和方法都比较简单, 易于学习。线性回归的公式如下:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i \quad (2-16)$$

其中, w_0 表示权重, x_i 表示第 i 个特征的值, n 表示样本特征的数量。线性回归模型的特征是相互独立的, 但是在实际生活特征之间存在一定的内在联系。而 logistic 回归模型虽然在线性回归基础上添加了人工特征组合引入了非线性, 但是人工特征需要耗费大量的时间, 再加上人类的经验具有一定的局限性。因此引入交叉项的公式如下:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} x_i x_j \quad (2-17)$$

该模型在交叉项 $x_i x_j$ 的前边加上交叉项系数 w_{ij} , 直接将两两组合特征引入模型, 可以解决特征组合的问题, 但是组合特征泛化能力弱, 无法对对应的参数进行估算。为了解决这个问题, Steffen Rendle 提出 FM 模型, FM 模型是用两个向量 $\langle v_i, v_j \rangle$ 代替单一的权重系数, 从单纯的用户, 物品隐向量延伸到所有特征上,

具体的方程式如下：

$$\hat{y}_{FM}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \quad (2-18)$$

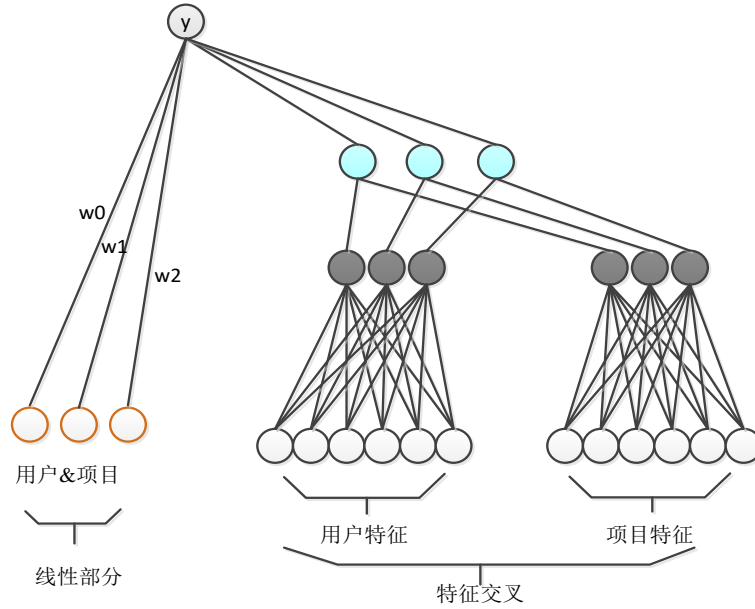


图 2-5 FM 结构图^[29]

Fig. 2-5 FM structure^[29]

其中，参数 $w_0 \in R, w \in R^n, V \in R^{n \times k}$ 。 $\langle v_i, v_j \rangle$ 代表两个不同大小为 k 的向量 v_i 和 v_j 的点积，公式如 (2-19) 所示， v_i 代表系数矩阵 V 的第 i 维向量， $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k})$ ， $K \in N^+$ 称为超参数，FM 结构如图 2-5 所示。

$$\langle v_i, v_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (2-19)$$

传统推荐算法作为一个基础模型，具有实现简单、解释性强、速度快等不可替代的优点，但它具有泛化性能差、表达能力不强、无法对特征进行筛选等局限性。可能会导致有效信息的丢失，甚至会出现错误信息。针对这个问题，推荐算法向更复杂更高级的方向发展，衍生出了深度学习推荐模型。

2.2.3 Neural Factorization Machine

FM 模型将不同特征进行组合，但是实质上仍然是线性模型，由于特征组合仅限于二阶，因此会受到爆炸问题的困扰，也就在所难免的限制 FM 模型的表达。新加坡国立大学的研究者于 2017 年提出将深度神经网络应用于因子分解

机，实现对特征的深层次抽取，这就是模型（Neural Factorization Machine，NFM）。公式如下：

$$\hat{y}_{NFM}(x) = w_0 + \sum_{i=1}^n w_i x_i + f(x) \quad (2-20)$$

与式(2-18)多少相比，将特征的两两组合换成了由深度神经网络充当的 $f(x)$ ，因此，NFM 在底层使用了特征交叉，在高层使用了深度神经网络。

NFM 在嵌入层和多层神经网络之间加入特征交叉池化层，在输入层通常是以 one-hot 稀疏向量作为输入，输入数据通常会变成高维稀疏数据，高维数据不利于模型学习，输入的特征是由输入特征值 x_i 与嵌入层的 v_i 相乘得到：

$$v_x = \{x_1 v_1, \dots, x_n v_n\} \quad (2-21)$$

B-Interaction 层的具体公式如(2-22)所示， \odot 表示两个向量对应元素相乘，对交叉向量特征求和，最终输入 k 维的向量，由池化层输出的向量最终经过多层神经网络输出。

$$f_{BI}(V_x) = \sum_{i=1}^n \sum_{j=i+1}^n x_i v_i \odot x_j v_j \quad (2-22)$$

NFM 中，将不同域的特征向量进行特征交叉后，根据嵌入层的尺寸的维度进行加和，由多层神经网络输出，NFM 的池化加和操作，对所有的特征都均等的被使用，没有探索重要程度对结果的影响程度，这就使得许多有用的信息被忽视。因此研究人员开始在深度学习模型结构上进行探索。

2.3 数据挖掘

在信息技术飞速发展的今天，大量的数据库和数据不断地从各行各业中涌现。信息爆炸时代，人们无法对大量的信息进行处理导致其对生活产生负面影响。对这些数据的价值进行提取并进行决策是当前迫切的需要。因此，数据挖掘技术应运而生。数据挖掘是一种深度数据分析方法，从数据中挖掘知识、提取知识和模式分析。目前，数据挖掘技术在各行各业中得到了广泛的应用。对于中医药行业来说，利用数据挖掘技术可以对相关的中医药数据进行分析 and 处理，有助于中医药行业信息化发展。

2.3.1 数据挖掘算法

信息爆炸时代，各行各业都需要数据挖掘技术。而数据挖掘产生的模式分析算法也多种多样。

1. 决策树

决策树与树形结构类似，决策树以树枝状的层级架构呈现，每一个结点都是一个属性的预测，每一个分支都会产生一个预测结果，每个树叶的节点就是一个预测的类别。决策树可解释性强、可以同时处理对称性和数值型数据，对无关的特征可以进行处理以及适用于有确定属性的样本等优点。但是由于它的产生的模型较复杂、使用能力有限，模型复杂会导致数据的泛化能力差。

2. 粗糙集

粗糙集理论由波兰科学家提出，它被用于数据分析中对模糊和不确定的信息进行处理。当前我们处于大数据时代，我们希望能从大量的决策属性中找到被隐藏的、有用的条件属性，筛选过滤掉可有可无的属性，因此粗糙集就派上了用场，粗糙集在各个领域的研究近年来有了很大发展。它的主要思想是可以在信息不完整和不一致下，对数据进行约减，从而获取模型的分类规则。

3. 关联规则

关联法则的应用是从大量的资料中发现项目的内在联系。就拿超市来说，可以看到啤酒和纸尿裤放在一块，因为一般在晚上年轻的爸爸会去给孩子买尿不湿，当他看到啤酒时，就会顺手买一瓶啤酒。这就是通过关联规则挖掘背后的商机出现的。关联规则算法简单，易于实现，适合稀疏数据集的关联规则挖掘。但是由于它没有考虑属性程度的重要性导致该算法的适应面窄。

4. 神经网络

人工神经网络大多为 BP 网络，在输入层输入的数据，通过隐藏层对网络权重进行学习，以适应环境的要求，最终在输出层输出正确的分类。神经网络可以具有从复杂或者不精确的数据获取有价值的信息的能力，因此能提取出人类和计算机技术都无法识别的有用信息或者检测趋势。神经网络具有应用方便，能更快更好的获得结果等优点。

2.3.2 数据挖掘过程

日常生活中，会产生大量的结构化或半结构化两种表现数据。数据挖掘技术是预测和描述这些潜在的数据表现形式。它一般由多个步骤组成，通常分为明确挖掘的主题和界定挖掘的任务。首先，根据我们研究的方向确定数据集的知识类型范围，当我们明确了挖掘的主要内容时，接下来需要了解有关领域的情况，熟悉我们研究的领域背景和界定业务问题。这一步的工作包括以下几个方面：熟悉研究对象，并确定研究的数据集，采用某种方法对研究的数据集进行抽样，最终确定数据表达形式。本文挖掘的对象是中医药中证候、治法以及处方等数据，目的是从挖掘出的数据中对患者进行选法定方，最终达到给病人推荐处方的需求。数据挖掘通常有以下步骤。

数据收集：根据已经查明的数据对对象进行分析，选择适当的数据收集方式将其存储在数据库中。

数据预处理：数据库中收集的数据可能缺少属性值、包含错误的属性值或者格式不规范等，因此需要对数据进行填充缺失值、更改错误信息以及转变格式等操作，数据预处理是数据挖掘中至关重要的一步。

数据转换：由于收集到数据的渠道很多，因此数据的格式存在一定的区别，需要进行转换成统一格式，模型选择的不同会使数据的表达不一样，也需要转化成适合不同模型分析的数据。

数据挖掘：运用特定的知识发现算法，根据不同的需求建立不同的模型提取数据集隐藏的规律。

知识评估：将模型的结果进行评估，以可视化的方式进行呈现，方便医生用户理解。若该模型不能满足用户需求，可以重新进行挖掘过程，直到能满足用户的预测需求。

2.3.3 中医药数据挖掘分析

在分析中医药数据时，数据集收集在预测算法中也显得至关重要，由于数据的质量对试验的结果有很大的影响，所以对数据的质量也有一定的要求。尽管国内外已经出现了医药数据集，但是已经公开的中医药行业的数据集并不常见，极少数的数据集也不能很好地应对本文特定的场景，并且已经出现的数据集和本文所运用的处方推荐场景的有一定的差异，不能满足实际生活需求。因此，本文以全国中医药行业高等教育“十三五计划”中的中医古典书籍进行分析。在中医院、中医馆和诊所之间进行多方研究和调查，结合医生给病人诊断的数据参考，并将研究对象的数据信息进行整理和完善，构建符合事实要求的数据集。

中医药数据来源：经过对国内外公开医药数据集的学习，结合实际生活需求，对于中医论治分类预测，首先保证的丰富且真实的数据来源，主要包括中医学古典书籍等非结构化数据。

首先利用爬虫等技术对非结构化数据进行提取，其中包括：患者年龄，性别，证候，治法和处方等数据，前期的工作是对获取的中医药数据整合和完善，进行存储，为下一步数据预处理做准备。

本文的中医药数据来源多，数据量大，获取的数据信息有些不完善或者有错误，需要对数据进行预处理，预处理是数据挖掘中非常重要的一步，预处理结果的好坏直接影响模型的输出结果。本文数据预处理的前期工作主要是包括数据清洗和数据变换。

数据清洗：对缺少的局部空值、异常值以及重复发生值，可以通过删除不正

确的数据、手工填写的方式完成，或利用算法进行补足。对于患者来说，如果有多个属性缺失，则该数据基本属于无用数据，则直接将该数据剔除，数据集中针对缺失性不多的数据，科技逆行多重插值法为该数据填充，并且对于属性值超过50%以上缺失时，通过调研可知，该数据对实验没有意义，不具备研究价值。比如：职业，日期等，对相关数据进行删除。

数据转换：整理的数据集来源渠道不同，因此数据的格式存在一定的区别，比如性别、年龄等属性。这些数据对实验的结果有一定的影响。需要对字符型数据进行数值化操作来满足实验数据的要求。

将以上预处理整合的中医药数据作为神经网络的输入，神经网络可以从大量的、复杂的数据中获取人们无法得出的有价值的信息，通过模型的训练最终得出预测的分类以及处方的推荐结果。

2.4 本章总结

本章对注意力机制、推荐算法和数据挖掘进行了主要介绍，首先介绍了注意力机制的原理以及他们在不同场景下的优缺点，然后对主流的推荐算法做了详细介绍，阐述了推荐算法的发展历史以及意义。最后介绍目前处于大数据时代所需要的数据挖掘技术，主要介绍了数据挖掘的算法以及过程，并针对本文所需要的中医药数据，介绍了数据的来源以及对数据的处理过程。

第3章 基于文本卷积神经网络的中医论治算法研究

3.1 引言

作为一个古老的整体治疗体系，中医药以其独特的优势和特点，受到越来越多的人重视和使用。中医理论的基础是整体论思想，着重于人体的整体和与自然环境之间的联系。当前，全球新冠肺炎仍处于大流行状态，而中医药在新冠疫情中发挥了它独特的优势。据资料显示，本次疫情中医治愈率普遍超过 80%，治愈率已达 90%^[30,31]。未来十年将是中医发展的黄金时期。中医诊疗三步过程如下，首先采集患者的证候，证候主要包括患者的症状和体征；其次通过对证候进行全面分析，得到相应的治法名称；最后医生通过治法名称结合患者的个人情况给出中医处方。正如我们所见，治法名称是医生诊断的第二步，通过对证候进行总结，对于最终的处方推荐至关重要。

中医的治疗方法是人们在长期医学诊断中认识疾病的基础，并逐渐总结出来的归纳性理论。中医治法是治疗疾病必须遵循的基本方法，是中医证候和处方之间的桥梁^[32,33]，对临床诊治具有指导意义。深入研究中医证候到治法，是突出中医特色、提高临床诊疗水平的有效措施之一^[34]。中医师对患者进行治法推荐时，由于主观意识、环境不同等原因，这导致了经验传承碰壁，使得中医治法的推荐结果不是特别准确。

近年来，随着社会生活的需要和国家政策的支持，人工智能和医疗正在迅速发展。利用图像处理、模式识别、深度学习等人工智能技术辅助医学诊断已成为研究热点^[35-37]。这也是让古代中医在科技时代重新焕发光芒的必要途径。目前，针对证候得出的治法在中医师中更为普遍。但治法的信息化研究较少，导致中医推荐治法的适用性越来越低。因此，将人工智能与中医治疗方法相结合，对中医的传承具有重要意义。本文设计适合证候与治法之间复杂非线性关系的算法，以完成证候与治法的准确对应。最终实现中医问诊过程的规范化和标准化。

目前对于中医治法规律的研究主要采用频次分析和主题模型方法进行治法规律分析。例如，杨玉凤等^[38]使用频次统计方法进行关于乳腺癌的中医治法规律分析，结果发现了 21 种常用治法；Yao 等^[39]以中医临床医案为文档集，他们将患者证候视为显变量、治法视为隐变量，应用主题模型方法来挖掘隐含的治法分布，生成基于证候的治法预测主题模型。现有的使用频次分析和主题模型等算法从临床医案中挖掘治法规律，传统的主题模型采用无监督的算法，因此对模型的

评估是非常困难的。此外，对于主题模型不适宜用于主题数多的情况，主题数若大于 10 会使观测结果不直观。主题模型适合文本长，而中医文本普遍偏短，导致传统的机器学习不适合证候到治法的匹配。

因此，本研究将文本卷积神经网络（TextCNN）应用在中医论治模型中，用于解决上述问题。Kim^[40]提出卷积神经网络对文本进行分类算法，并且在自然语言处理中取得了较好效果。目前在自然语言处理中，较多使用循环神经网络（RNN）模型，它可以较好的处理时序问题。在本研究中，通过使用稀疏的连接和权重共享，TextCNN 在某种意义上减小了参数数目，从而减小了网络的过拟合^[41]。因此本文采用 TextCNN 对中医治法分类进行了研究。在实验中采用了双向长短时记忆网络（Bi_LSTM）和结合注意力机制的双向长短时记忆网络（Bi_LSTM Attention）的与本研究的中医学论治模型进行了比较。

3.2 系统模型

3.2.1 模型应用场景

目前，中医药在疫情防控、临床救治方面做出了巨大的贡献，习近平总书记也发表重要讲话，从人民安全和身体健康的战略高度出发，充分肯定中医在疫情防控中做出的贡献，近几年，人工智能和医疗发展迅猛，是古老中医在科技时代再次焕发光彩的必然途径，因此，将人工智能+辅助诊疗相结合，利用人工智能技术对患者进行论治，通过计算机来学习医学知识，模仿医生的判断，提供正确的诊断和治疗方法。本章节的中医论治模型通过深度学习模型，可以根据患者的各个证候体征，给出相应的治法，为中医医师提供决策进行帮助，减少了中医诊断的时间和提高效率。

3.2.2 文本卷积神经网络

随着计算机硬件技术的不断进步和机器学习的深入发展，深度学习已经成为文本分类的重要方法，与传统的文本分类模型相比，深度学习算法的主要工作就是通过神经网络模型来学习单词向量的表达，然后将所学到的词汇向量进行融合再进行分类。该方法利用隐藏层将单词向量从稀疏编码转化为低维词向量空间，其实质是一种特征抽取，用于对词语的维数进行编码。在词向量的稠密表达中，具有相似意义的词语，其欧几里得或余弦距离是相同的。

卷积神经网络是深度学习算法中应用最广泛的模型之一。卷积神经网络最开始是对计算机视觉进行发明的，后来被证明该方法在语义分析、查询检索、句子建模等方面得到了广泛的运用^[40]。在自然语言处理的文本分类中，文本的语

义表达依赖于关键字的选择，所以对各个段落之间的关联性进行识别显得十分重要。文本卷积神经网络首先通过嵌入层对输入的数据进行预处理，转换成词向量，然后，采用卷积操作对文字进行高层次的特征抽取，使用了池化层对其进行降维处理，并将其在全连接层进行特征的集和，使用 Softmax 层进行分类。对于具有复杂特征的数据集，深度特征提取 CNN 可以通过反复叠加卷积层和池化层来挖掘更丰富的文本语义。TextCNN 可以更好地处理高维度数据，在建模过程中，不需要人工选择特征，卷积层采用了一种具有非线性激活函数的卷积运算，能够高效地对资料进行压缩，从而减小了参数数目，并减小了过分拟合。

文本卷积神经网络主要是在电影评论以及客户对各种产品的评论，他们预测的分类都是预测正面或负面的评论。文本卷积神经网络通过无监督语言模型获得的单词向量上加上一层卷积，保持最初单词向量不变，保持模型的其他参数不变，也取得了较好的效果。文本卷积神经网络通过嵌入层对输入的数据进行处理，主要使用 Word2Vec 无监督学习模型，Word2Vec 是一种训练中文单词的有效表示方法，该模型克服了传统 CBOW 模型的不足，揭示了汉语词汇之间的语义关联，也就是说中文单词通过 Word2Vec 方法生成相似的单词向量，将词汇信息映射到语义空间，最终将使用将预处理好的词向量模型作为卷积神经网络的输入，通过卷积层和池化层提取文本特征，可以得到一个适用于多场景的文本分类模型。

3.2.3 中医论治模型

在中医论治模型中，使用单词的分布式表达作为输入句子分类的模型，其优点在于可以在语句的初始化过程中，这种体系结构能够灵活地交换不同的预先训练好的单词向量。在中医论治模型中，通过对中医证候分类的敏感性，以及使用的输入表征的探讨，本研究通过使用 GloVe 模型取代了 Word2Vec，Word2Vec 使用从谷歌新闻中训练的 1000 亿个单词的本地上下文窗口模型，而 glove 是一个基于全局单词共现统计的模型，本研究使用的 GloVe 模型是在 8400 亿标记的网络数据语料上训练的。利用文本卷积神经网络，对中医治法文本诊断进行分类，中医证候都有相对应的治法标签，在对中医治法诊断的过程中，也和用文本卷积神经网络处理文本的思路是一样的。将中医证候作为文本进行向量化操作，首先对嵌入矩阵中的证候在分词之后进行向量投射，经过 TextCNN 之后，得到相应的治法标签。

基于TextCNN的中医论治模型，使用单通道和不同大小的卷积核提取中医证候治法句子中的重要信息，对中医证候句子的局部相关性进一步捕捉。针对中医

的证候到治法名称之间具有非线性和复杂性等特点，因此将TextCNN算法应用于复杂中医治法诊断的特征提取和分类。中医论治模型的流程步骤如下：首先，采用中药文本数据进行预处理，将传统的中医文本数据转换成一个统一的格式，同时对中医的证候进行分词，停用词等数据处理，最终形成预处理的语料库；第二步，利用语料库建立词汇共现矩阵，然后根据词共现矩阵学习词向量矩阵；第三步，构建基于中医词向量矩阵训练中医论治神经网络模型。第四步，基于中医测试数据集验证文本卷积神经网络模型并优化，并将其作为最终的中医诊断论治模型。

如图 3-1 所示，将一个包含 x_1, x_2, \dots, x_m 的证候治法文本输入到中医论治模型中，其中， $x_i (i \in (1, m))$ 表示证候治法句子。本研究对证候治法文本进行输入操作，使用卷积核对输入的证候治法文本进行关键特征筛选，通过激活函数的非线性能力对特征进行保留并映射，得到证候治法数据集的关键特征，利用最大池化层对最大特征进行提取，最后经过中医论治模型的全连接层预测中医治法。

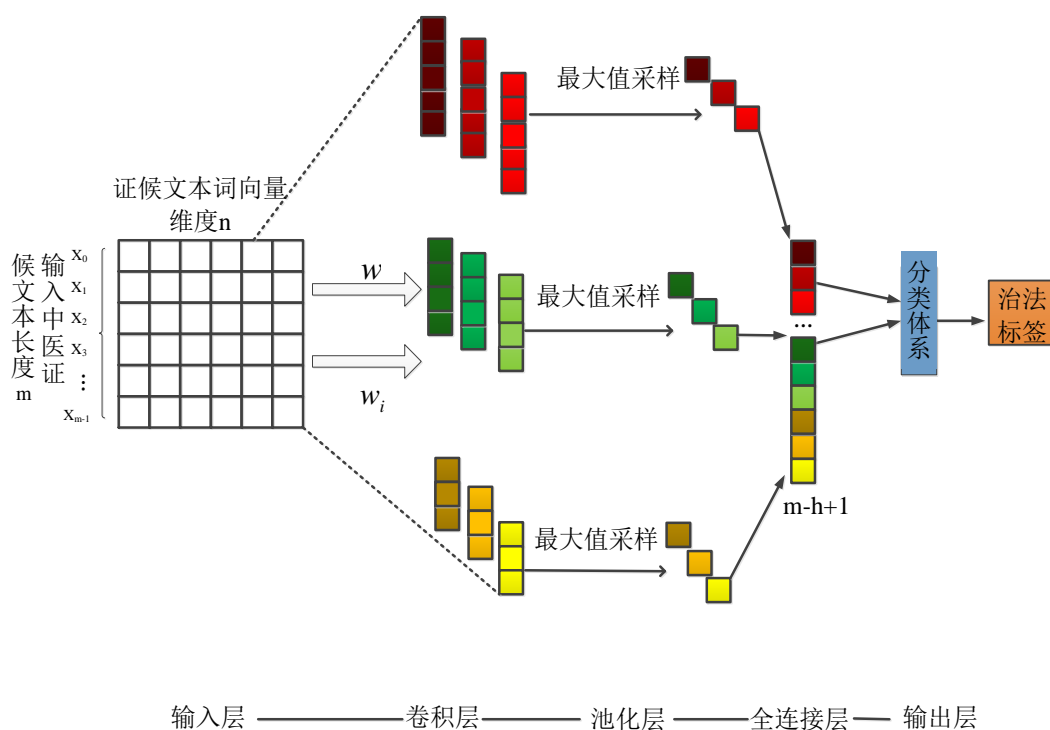


图 3-1 中医论治模型结构图

Fig. 3-1 The structure of a TCM Therapeutic Methods model

在本研究的中医诊断论治模型中，将中医证候治法数据在嵌入层进行预处理，首先，本研究对中医证候文本数据集进行处理：通过读取中医证候治法数据

集，读取数据集，进行切词处理和停用词处理。本研究中采用基于字符串匹配的分析算法，使用了正向最大匹配算法。

正向最大匹配算法，例如中医文本语句为“每天早起后或者饭后咳嗽多并且痰也多”，进行分词处理后为“每天/早起/后/或者/饭后/咳嗽/多/并且/痰/也/多”。接下来利用正则匹配对中医证候治法文本数据集的无效字符进行了剔除，然后用 utf-8 进行了文字的处理，

其次，基于共现矩阵和 GloVe 模型学习词向量矩阵的方法主要包括两个步骤^[18,19]：步骤一，根据两个中医证候数据在上下窗口的距离 d ，构建衰减函数 D 用于计算权重，公式如下：

$$D=1/d \quad (3-1)$$

步骤二，构建 GloVe 模型的损失函数，损失函数公式如(3-2)所示。其中， p 是被统计的中医证候治法句子，它表示统计的词汇表上第 j 个证候治法句子。 s_p 和 s_q 是中医证候 p 和中医证候 q 的词向量。 X 为语料库中构建的共现矩阵， b_p 和 b_q 是两个标量， o 是权重函数， N 表示词汇表的大小，共现矩阵维度为 $N \times N$ 。最后将基于共现矩阵训练好的中医证候治法文本词向量矩阵输出，输入到中医论治模型的卷积层。

$$L = \sum_{p,q=1}^N o(x_{p,q})(s_p^T s_q + b_p + b_q - \log(x_{p,q}))^2 \quad (3-2)$$

通过嵌入层输出的中医文本共有 m 行证候治法词向量，中医文本中证候治法词的词向量一共有 n 维，那么对于证候治法文本句子，总共有 m 行 n 列的证候治法矩阵 M 。本研究可以把证候治法矩阵 M 看成是一副图像，使用文本卷积神经网络去提取中医文本特征。其中， h 为中医论治模型中卷积核的高度。在本研究的中医论治模型中，使用多个高度不同的卷积核 w ，可以得到更丰富的中医证候治法特征表达。中医证候治法词向量经过嵌入层后，就可以输出中医证候治法矩阵 M ， $M[i:j]$ 表示的是中医证候矩阵 M 的第 i 行到第 j 行，那么中医论治模型卷积操作被定义为：

$$a_i = w \cdot M[i:i+h-1] \quad (3-3)$$

其中， i 表示中医证候矩阵中的第 i 行词向量，其中 $i=1, \dots, m-h+1$ ， w 和 M 之间是点乘。对于一个卷积核，可以得到中医治法特征 $a \in m-h+1$ ，共有 $m-h+1$ 个治法特征。本研究使用激活函数 ReLU 激活，并叠加上偏置 b 。本研究的卷积操作 f_i 被定义为：

$$f_i = \text{ReLU}(a_i + b) \quad (3-4)$$

经过中医论治模型的卷积操作之后，可以得到一个 $m-h+1$ 维的中医证候词

向量, 本研究的中医论治模型使用 ReLU 激活函数, 它具有可以使网络训练更快, 并且增加网络的非线性, 防止梯度消失等优点。本研究的中医论治模型最终输出的非线性证候特征词向量 c 被定义为:

$$c = [c_1, c_2, \dots, c_{m-h+1}] \quad (3-5)$$

由于使用不同尺寸的卷积核得到的中医证候治法文本特征大小是不一样的, 因此本研究需要在池化层对每个特征使用池化函数, 利用最大值采样, 使他们的维度相同来提取最关键的特征。本模型使用最大池化层的好处是将一维数组转换成单一数值, 能减少后续卷积层或隐藏层的神经元个数, 有利于降低模型过拟合问题。在中医证候论治模型中, 本研究使用最大池化层来提取中医证候的关键特征 z^{\max} , 公式如下:

$$z^{\max} = \max f_i, i = 1, 2, \dots, m - h + 1 \quad (3-6)$$

全连接层在中医论治模型中起到分类器的作用, 在全连接层中所有的神经元都有权重连接。将卷积操作和非线性激活函数得到的不同特征输入到池化层利用最大值采样将其拼接得到的一个一维的中医证候向量以此抓取足以用来识别中医证候文本特征, 并将得到的一维证候特征平摊成一个长的向量, 最终送入全连接层配合输出层进行分类, 最大池化层得到的中医证候向量 z^{\max} 放入全连接层来预测治法类别的概率分布, 公式如下:

$$y_{z^{\max}} = \frac{e^{z^{\max}}}{\sum_{j=1}^n e^j} \quad (3-7)$$

中医论治模型的全连接层利用上式函数输出 0 到 1 之间的数值, 数值的大小表示输入中医证候属于某个治法的概率。最终选择输出概率最大的作为本研究的中医论治模型预测的输出治法结果。对于中医论治数据来说, 通过对中医数据的分析, 找出中医文本的最长序列作为中医论治模型输入的文本序列长度 n , 其他不足的进行自动补充, 把中医证候的输入 n 转化成定长的文本序列作为中医论治神经网络的输入的好处是方便设计神经网络的结构。通过对中医证候的诊断治法适配策略进行分类研究, 确定证候的合理性。本文设计适合证候与治法之间复杂非线性关系的算法, 以完成证候与治法的准确对应。最终实现中医问诊过程的规范化和标准化。

基于文本卷积神经网络的中医治法诊断模型, 通过使用 GloVe 模型, 它是一个基于全局单词-单词共现统计的模型, 对中医证候数据文本进行分词和停用词处理, 并将处理后的语料库送入统计共现矩阵, 进行训练, 得到词向量并进行保存, 将得到的中医证候词向量矩阵, 作为中医论治模型的嵌入层的输入进神经

网络进行中医治法诊断并进行分类,给出分类结果。GloVe 模型对其他的模型如 Word2Vec、CBOW、Skip-Gram、SVD 等进行了优化,做到了充分兼顾全局信息和局部信息,使得效果有了一定的提升,算法复杂度也得到一定的优化。

3.3 算法设计与分析

基于文本卷积神经网络的中医论治模型算法步骤流程算法如下表 1 所示,本研究将中医证候文本分为训练集和测试集,主要包含了患者的证候和治法名称。根据中医师的诊断结果,将中医证候对应的治法标签信息分为训练标签和测试标签集。

首先从中医论治模型中读取一个表示成向量的样本,通过卷积操作,最大值采样并通过全连接层得出中医治法诊断结果,采用梯度下降优化卷积神经网络,通过训练得到最终的中医论治模型网络结构。同样,本研究使用中医测试集和对应的中医治法诊断结果测试中医论治算法的准确度。

算法 1 中医论治分析算法

输入: 训练样本, 中医证候文本词向量 X
 训练样本中中医证候对应的治法名称标签

输出: 中医论治模型中参数值和治法大类以及治法小类名称 Y

```

1: while 输入证候治法文本向量  $X$  do
2:   中医证候文本序列进行卷积核大小  $h$  的卷积操作
3:   for  $i < X$  do
4:     中医证候治法文本进行卷积操作
5:      $a_i \leftarrow w \cdot M[i:i+h-1]$ 
6:     经过卷积操作得到  $m-h+1$  维中医证候词向量  $c$ 
7:      $c \leftarrow [c_1, c_2, \dots, c_{m-h+1}]$ 
8:   end for
9:   if 证候治法文本特征 not null
10:    使用非线性激活函数保留并映射证候治法文本特征
11:     $f_i \leftarrow ReLU(a_i + b)$ 
12:   end if
13:   采用 Max Pooling 对数据降维
14:    $z^{\max} \leftarrow \max f_i, i=1, 2, \dots, m-h+1$ 
15:   将池化层得到的向量紧进行连接
16:   Dropout 层用于缓解过拟合
17:   将连接后的向量变换为有关各类别  $y_{z^{\max}}$  输出
18:   网络优化: 采用梯度下降优化中医论治模型, 选择交叉熵损失函数
   作为优化函数, 并在损失函数上添加 L2 正则化
19: end while

```

中医论治的目标是对于一个患者，医生通过对患者的观察和了解来对患者做出治法诊断，为此，需要根据获得的中医证候数据集训练本文提出的基于文本卷积神经网络的中医论治模型，并利用该算法对网络中的权重、偏置量进行训练，从而得到最优的网络。最终，当有未出现过的患者的数据出现时，只需要运用该算法，输入已经训练好的网络，就可以得到最终的治法诊断。

为了评估算法的性能指标，除了采用有效的实验方法外，还需要制定模型的性能指标。针对分类模型，算法的性能指标通用准确率、精准率、召回率和 F1 分数。在现实生活中，性能指标的评价也会有所差异，比如，在预测股票时，他们更在意的是精确率，也就是预测的股票中到底有多大的涨跌幅，因为这些预测的股票都是经过投资的。而在预测患者的中医治法诊断场景下，更关注召回率，即真的患者那些人里预测错了情况应该越少越好。

但是针对本研究的中医论治模型的分类问题，则使用精确率、召回率、精度、F1 分数来评价该方法。具体公式如下所示。其中 TP、TN、FP、和 FN。TP 表示一个正例被正确地预言为正例，TN 表示一个反例被正确地预言为反例，FP 表示一个反例被错误地预言为正，FN 表示一个正例被错误地预言为反例。

准确率是指在全部的中医数据集中，对中医治法分类的预测是正确的。计算公式如下：

$$Accuracy = \frac{N_{correct}}{N} \quad (3-8)$$

精度是针对预测结果而言的，在正确的中医治法分类中，有多少把握可以是预测正确的。计算公式如下：

$$Precision = \frac{TP}{FP + TP} \quad (3-9)$$

召回率是指在一个实际为正地样本中，所预测地实际为正的可能性。计算公式如下：

$$Recall = \frac{TP}{TP + FN} \quad (3-10)$$

准确率和召回率是中医论治模型效能地体现，仅凭单一指标无法对其进行较为全面地评估。因此引入 F1 分数，更加全面地分析中医论治模型的性能指标。计算公式如下：

$$F1 = \frac{2PR}{P + R} \quad (3-11)$$

3.4 实验

3.4.1 实验数据

为了验证该模型的有效性,本研究使用中医数据集的证候和治法数据集进行训练,并对比实验结果。本文选取了全国中医药行业高等教育“十三五计划”中的中医古典书籍,以疾病为单位,整理出证候和治法数据集作为训练案例。该治法主要包含23大类和400余种治则治法小类,其中23种治法大类主要包括解表治法、清热治法、安神治法、理气治法、理血治法、润燥治法、温里治法等。

将10065个中医证候治法文本按比例划分为治疗数据集,并将其分为训练集和测试集,两者的比例分别为80%和20%。通过对训练集的训练,可以计算出中医论治模型的权重、偏置量等,测试集主要通过中医论治模型进行验证,通过验证可以分析出模型的准确率,以此判断是否达到本研究的预期目标。通过测试集训练集的结果的对比,可以避免中医论治模型算法出现欠拟合和提高本研究算法的泛化能力。

在开始训练之前,本研究对中医数据集进行预处理并作为中医论治模型的输入。首先,本研究使用 jieba 分词来对句子进行分词,从中医文本中删除停用词和符号。接下来,本研究预训练词嵌入,预训练词嵌入在中医数据集上训练时可以捕获单词的语义和句法意义,可以提高自然语言处理模型的性能,它通过将文本表示成数字,用于计算机读取文本内容,所以适用于神经网络的输入。传统的词嵌入表示方法比如独热编码,会使得输入数据非常稀疏且维度特别大。预训练的词嵌入模型主要包括 Word2Vec 模型和 GloVe 模型,Word2Vec 使用从谷歌新闻中训练的 1000 亿个单词的本地上下文窗口模型,GloVe 是一个基于全局单词共现统计的模型,本研究使用的 GloVe 模型是在 8400 亿标记的网络数据语料上训练的。目前因此在中医论治模型中使用的是 GloVe 模型,它对其他的模型如 Word2Vec、CBOW、Skip-Gram、SVD 等进行了优化,做到了充分兼顾全局信息和局部信息,使得效果有了一定的提升,算法复杂度也得到一定的优化。

3.4.2 实验设置及分析

本研究的中医论治模型实验环境配置主要包括操作系统,专业显存,内存, GPU, CPU 以及硬盘大小,并且也介绍了实验所用到的编程语言和机器学习框架。如下表 3-1 所示。

表 3-1 实验环境配置

Tab. 3-1 Experimental environment Configuration

参数名称	参数值	参数名称	参数值
操作系统	Window10	硬盘大小	1TB
专用显存	4GB	CPU	4 核
内存	16GB	GPU 核个数	2 核
机器学习框架	Pytorch	编程语言	Python

中医论治模型中的参数是本研究训练神经网络的最终要学习的目标，通过训练，最终就是要找到一套好的模型参数，用于预测未知的结果。中医论治模型中的参数是通过训练生成的，调节的方式也因不同的场景而有所不同。网络中超参数的选择一般需要根据数据集的不同而进行适当地调节，是通过在实际情况中反复实验的。在本文中医论治模型中的参数，主要包括：序列长度、卷积核尺寸、词向量维度、Dropout 保留比率、学习速率、L2 正则化参数以及优化方式等。具体参数如表 3-2 所示。

本研究为了避免模型过度拟合，中医论治模型将 Dropout 应用在的词向量输入层和隐藏层中，并将其设置为 0.5，通过防止特征检测器的联合作用从而提高了神经网络的处理能力。当我们将学习率设置为 0.001 时，中医论治模型是最优的。模型训练的损失函数采用最小化交叉熵函数。为了保证实验的准确性，对所有模型的层数设置相同。

表 3-2 实验参数表

Tab. 3-2 Experimental parameter table

实验参数	实验设置
序列长度	256
卷积核尺寸	[2, 3, 4]
词向量维度	100
Dropout 保留比率	0.5
学习速率	0.001
L2 正则化参数	0.01
优化方式	交叉熵

本研究将提出的中医治法诊断模型与 Bi-LSTM 和 Bi-LSTM Attention 在中医

证候数据集上进行比较,并且使用 *Precision* (精确率), *Recall* (召回率), *F1* (*F1* 分数) 作为评价指标对实验结果进行测评比较。准确率和召回率反映了中医论治模型的性能两个方面,单独依靠某个指标并不能较为全面地评价一个模型的性能,因此引入 *F1* 作为综合指标,更加全面地分析模型。在不同的应用场景下,各个性能的评价指标不同,比如在预测股票的时候,更关心精确率,也就是预测升的那些股票里,真的升了有多少,因为那些预测升的股票都是投钱的。而在对患者治法诊断的场景下,更关注召回率,也就是真的患者里预测错的情况越少越好。本研究对三种神经网络算法的实验结果进行对比分析,下表 3-3 为实验结果分析表。

表 3-3 实验结果分析

Tab. 3-3 analysis of experimental results

Model	precision	Recall	F1-score	Accuracy
TextCNN	0.83	0.83	0.83	0.71
Bi_LSTM	0.05	0.25	0.08	0.24
Bi_LSTM Attention	0.81	0.80	0.08	0.68

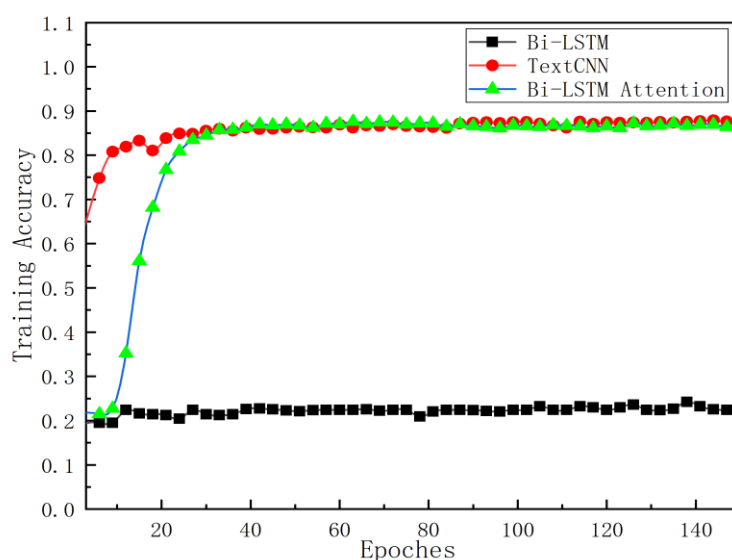


图 3-2 训练精度对比图

Fig. 3-2 Training accuracy comparison

表 3-3 所示的是不同治法推荐模型的 *Precision* 值, *Recall* 值和 *F1* 值的比较, 本文的中医治法模型相比其他两种对比模型具有较高的精确率。中医论治模

型准确率为 71%，比 Bi_LSTM 高出 47%，比 Bi_LSTM Attention 高出 3%。这也充分体现 TextCNN 的优势，它可以自动地将中医证候的特点进行组合、筛选，从而获取各种抽象层面地语义信息。因此，本研究将 TextCNN 模型应用在中医治法中是可行的。

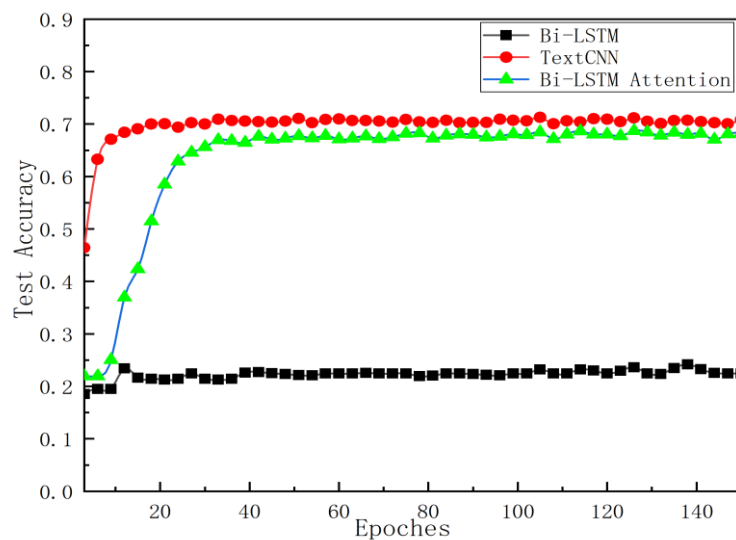


图 3-3 测试精度对比图

Fig. 3-3 Test accuracy comparison

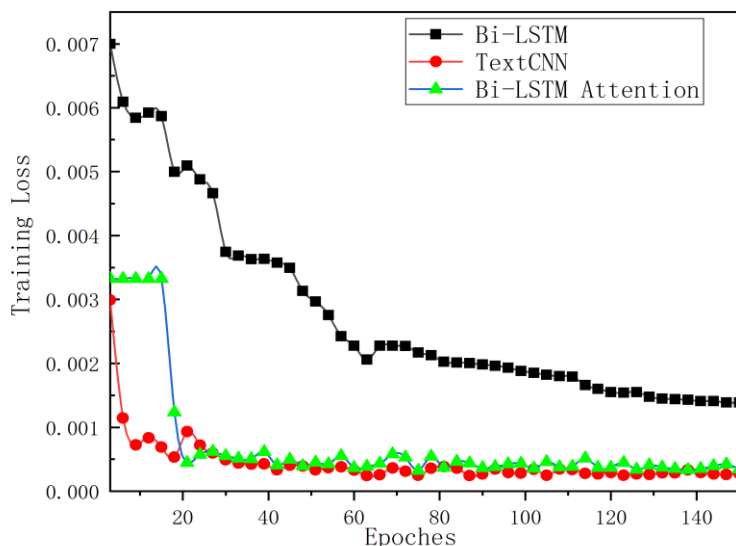


图 3-4 训练损失对比图

Fig. 3-4 Training loss comparison

图 3-2、图 3-3 是训练精度和测试精度的曲线图，可以看出，本研究的中医

论治模型效果最好。而且我们使用的模型相比另外两种模型计算量少，训练速度快。文本卷积神经网络共享参数使得网络只需要学习一个参数的集合。而在中医证候单词中，由于证候文本短并且上下文联系不敏感，使得本文的中医论治模型的整体效果优于其他两种对比算法。

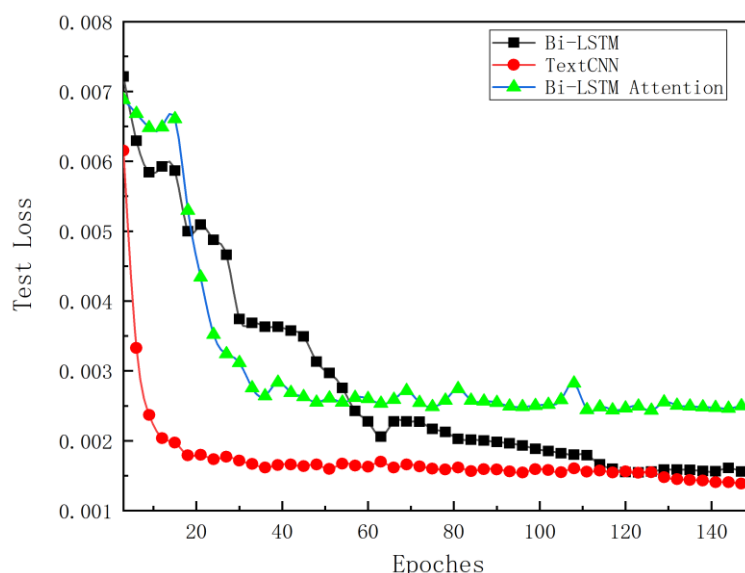


图 3-5 测试损失对比图

Fig. 3-5 Test loss comparison

图 3-4、图 3-5 是训练和测试损失值的曲线图，显示了三种模型在实验中损失值的大小，图中可以看出，本研究的中医论治模型的损失值最小，因为我们的中医证候论治模型充分利用 TextCNN 网络结构的优点，避免了 LSTM 由于序列上的前后依赖关系，导致获取特征的能力更弱。而且，由于每次卷积中采用了共享权重的机制，因此 TextCNN 的训练速度相对较快。在实际的中医证候到治法推荐中取得了不错的效果。随着迭代次数增加，Bi-LSTM 和 Bi-LSTM Attention 模型的损失值均大于本研究的中医论治模型，并且损耗值波动较大。当迭代次数较大时，本研究的中医证候论治模型的损失值相对稳定。

3.5 本章小结

本研究将 TextCNN 模型应用于中医论治方法中。其效果优于 Bi-LSTM 和 Bi-LSTM Attention，实现了中医证候特征的提取，并通过模型给出了治疗方法。本研究主要针对中医证候治法文本的分类，选择一维卷积神经网络对中医文本进行预测治法，文本选用 GloVe 作为中医论治模型的输入，以提高治法分类的精

度，并在预测的结果中选择最大类别最为输出，从而大大提高分类的准确性。为中医辅助诊疗的诊断提供了帮助。

本研究模型可以帮助中医医生解决治疗方法过程中遇到的问题：如何更好地解决中医从证候预测治法的问题，实现证候的治则治法适配策略，确保证候以及治法的合理性。本研究模型通过卷积层获得文本的关键特征，并使用最大池化层提取主导特征，并在实验部分展示了三种神经网络训练结果的不同比较。最后，通过对中药证候治法数据集的实验表明，本文提出的模型在准确性方面优于 Bi-LSTM 和 Bi-LSTM Attention 模型。

第4章 基于注意力分解机的中医处方推荐算法研究

4.1 引言

中医是古代建立的医学实践体系，与现代医学在本质、方法论和哲学上都有所不同。几千年来，中医在我国的治疗防病中发挥着不可或缺的作用，并且在西方国家也越来越频繁地被使用^[42,43]。中医治病对证不对病，不同的病产生同样的证候会采取同样的中医处方。处方是指多种中药材的集和，在我国悠久的历史中，人们发明了许多治疗证候的处方，并且已经记录在案的药方超过 10 万余种。中医传承发扬着中华民族几千年的文化和历史，它具有独特的诊疗方法和显著的临床疗效。处方中的中草药组成以及相应的治则治法对临床疗效和新药开发具有重要意义^[44,45]。

13140 麻黄汤 (Ephedra Decoction)

【方源】《伤寒论》。

(Source: "Treatise on Cold Damage Diseases".)

【组成】麻黄 9g 桂枝 6g 杏仁 6g 甘草 3g

(Composition herbs: Ephedra 9g, Cassia Twig 6g, Apricot Seed 6g, Licorice Root 3g.)

【用法】水煎服。

(Usage: decocted in water for oral dose.)

【主治】外感风寒，恶寒发热，头身疼痛，无汗而喘，口不渴，舌苔薄白，脉浮紧。

(Indication symptoms: exogenous wind cold, aversion to cold with fever, headache and body pain, asthma without sweat, no thirst, white and thin coating of the tongue, floating and tense pulse.)

图 4-1 中医处方图

Fig. 4-1 TCM prescription

中医药在漫长的发展历史中积累了大量的文献和治疗记录，由中草药组成的处方是中医治疗的最重要形式。中医的诊断过程可以称为“理法方药”，这在临床实践中至关重要^[46,47]。理法方药分别是原理、方法、处方和中药材，它指出了诊断和治疗的三个基本步骤：根据证候确定病因以及治法，根据治法最终选择处方和合适的药材。处方就是对草药进行配伍，这意味着根据临床情况和草药的性质，缓和两种或多种中草药，结合中药可以改善疗效，防止副作用。草药对是两

种相对固定的药材的独特组合，是草药配位的最基本和最简单的方式^[48,49]。在形成处方中，草药对始终被用作基本单位。例如在这本书 *Dictionary of Traditional Chinese Medicine Prescriptions* 中的处方如图 4-1 中，麻黄和桂枝可以结合可以诱导出汗，如果我们只使用其中一种，则诱导出汗的效果会弱很多。同样，制附子和干姜在许多驱寒处方中总是在一起使用。

在中医药悠久的历史中，由杰出的中医研究者和实践者撰写的中医古典中有许多治疗记录和方法^[50]，在现实生活中，中医师为患者开处方时经常参考这些经典记录，因此，这就启发我们设计一个模型，通过学习这些古典书籍中的经典记录自动生成处方。当前，随着我国人民对中医药的需求越来越大，人工智能与医疗健康的融合不断加深，借助人工智能领域，语音交互，计算机视觉合认知计算等技术来辅助中医师进行诊断已经成为当前研究的热点，这也使得古老中医再次收到青睐和关注^[51,52]。

目前，针对给定证候推荐处方在中医师之间较为普遍。WEI 等^[53]从中医证候推荐处方，提出一种基于文本证候描述的中药处方自动生成方法，使用序列到序列模型探索中医药处方生成任务的潜在端到端的解决方法。章亚东等^[54]提出一种基于复杂网络的中医治疗肺癌的处方推荐算法，旨在确定治疗肺癌的核心药物，帮助患者减轻病情。但是他们都是从证候直接到处方，违背了中医理法方药的诊断过程，从而导致推荐的中医处方的传承越来越少。此外处方中的药材排列很不整齐，所谓“弱序”，就是药材的效果不受处方中的顺序排列的影响。然而，药材的顺序反映了构造处方时的思维模式。因此，药材之间是相互联系的，最重要的通常列在第一位。由于缺乏数字化和规范化，中医药在人工智能界尚未引起足够的重视。

将中医处方推荐应用到因子分解机（FM）模型中，提高中医信息化智能化发展。为了促进中医药处方自动生成的研究，本研究从中医古典数据中收集了大量的处方和相应治法的描述，使用注意力因子分解机（AFM）^[55]来描述中药方剂的生成过程。该处方推荐模型以中医古典书籍中的处方为例，反应了中医的处方模式，它可以帮助中医师开处方和制药公司决定使用哪种中医药材进行组合。中医处方推荐智能化技术对于中医的发展和继承具有重要意义，利用人工智能技术研究治法和处方间的关系，准确提取其特征，进行多维定量分析，使其更加科学化、客观化和条理化。利用深度学习技术提高中医诊断的正确率已经成为一种新的中药推荐方法，并日益显示其优越性^[56]。此外还需要注意，由于在真实的环境中，本研究的工作目标是生成候选处方，以促进中医处方智能化，而不是完全代替中医师。

4.2 系统模型

4.2.1 模型应用场景

中医传承着中华民族几千年的文化和历史，其治疗手段，临床效果明显。在人工智能飞速发展的今天，将现有的计算机技术与中医理论知识相结合，让计算机担当起“名医”的角色，通过选法定方为病人开出个性化处方，将会在很大程度上提高中医证候开处方的诊断效率^[16,17]。中医辅助诊疗系统提高了中医生诊断效率，利于名老中医诊疗经验的数字化、智能化传承，对打造健康中国、构建人类卫生健康共同体，具有战略性的意义。

4.2.2 注意力分解机

传统机器学习模型假设预测变量之间相互独立，忽略他们之间的作用，他将无法准确预测并进行推荐，与线性回归相似，机器学习将各属性的权值联系在一起，并将其作为各属性的权值之和。但是，不是所有的特征对结果的贡献程度都相同。因此，对具有不太有用的特征的交互应该被赋予较低的权重，因为他们的预测贡献较低。

注意力分解机^[55]将特征的权重参数转化为组成特征嵌入向量的内积，通过学习每个特征的嵌入向量，并且可以通过矩阵分解估计任何交叉特征的权重，从而解决传统机器学习中特征相互独立以及只显式地增加特征向量泛化问题，比如观察到少数交叉特征的稀疏问题，无法估计未观察到的交叉特征参数问题。同时，利用注意力网络，对目标有用的特征赋予更高的权重，从而区分处方中不同组合特征的重要程度。因此注意力分解机可以观察到哪些交叉特征对预测的贡献更大，这使得注意力分解机的可解释性和透明度更强，能够使研究人员对其模型进行更深入地分析。

注意力分解机是针对电影的个性化标签推荐背景下讨论的，在推荐中每个标签（用户 ID、电影 ID 和标签）都经独热编码转为一个特征向量。对于标签推荐问题中的原始特征经过独热编码表示后会变得很稀疏且维度增大，如果直接将其作为神经网络的输入，会导致网络的计算量就会非常大，最后导致模型对各个参数的学习效果不理想。一种解决办法是通过嵌入层对输入特征稀疏表示，并将每个非零特征嵌入到一个稠密向量中。将高维稀疏矩阵通过矩阵分解压缩成低维稠密矩阵，通过将得到的稠密向量输入到注意力分解机中进行训练，能够得到一个适用于多场景的推荐模型，但是对所有的特征没有进行特征域区分。

4.2.3 中医处方推荐模型

基于注意力分解机的中医处方推荐模型通过嵌入层和两两特征交叉层，解决中医处方特征和治法诊断信息特征在编码后过于稀疏以及特征自动交叉组合的问题，并根据处方推荐任务将处方特征与患者个人信息特征进行自动交叉组合，得到患者特征和处方交叉特征的权重，通过注意力网络，对中医处方和患者个人信息的有用交叉特征赋予更高的权重，从而使本研究的模型提取对目标预测起到更高作用的交叉特征，并除去噪音等信息。由于中药处方药物间的复杂性、患者的个体差异等特点，因此将 AFM 模型应用到中医智能处方推荐中，从而使中医诊断更加智能化。

在中药处方推荐中，患者的信息以及治法和处方数据形成了一个数据集，并进一步用于模型的训练。由于中医处方数据是多特征类别无序特征（例如治法）、二元特征（例如性别）或连续特征（例如年龄），经过独热编码后包含高度稀疏并且维度也很大，因此很难学习关键特征之间的相互作用，并且也很难训练模型。因此，通过引入特征域的结构，把不同的特征分为不同的域，例如治法分为一个域，年龄分为一个域，可以有效学习各个特征之间的加权交互，而无需进行特征工程。

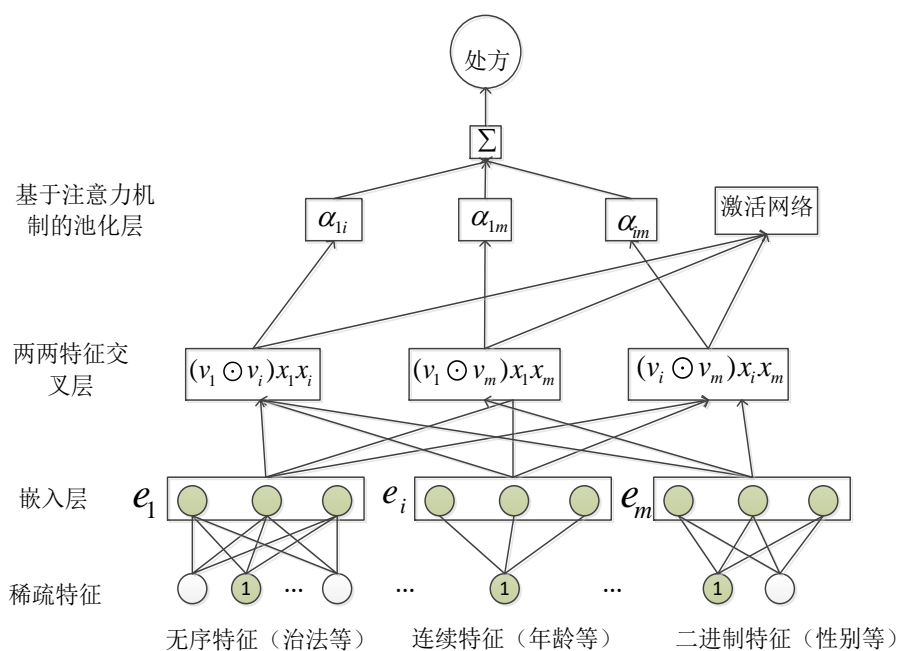


图 4-2 中医处方推荐模型结构图

Fig. 4-2 The Structure of TCM prescription recommendation model

如图 4-2 所示，将一个包含 x_1, \dots, x_m 的处方推荐文本数据集输入到处方推荐

模型中, 其中 x_i 表示处方推荐的特征类别, 经过嵌入层后, 得到每个特征域的嵌入特征向量, 且向量维度保持一致, 每个二阶特征向量经过两两特征交互层, 进行特征向量两两交叉的操作, 获得二阶特征项。同时, 将每个二阶特征向量加入注意力网络的池化层, 输出该特征向量的权重值, 再用池化层加权求得综合所有特征信息的向量, 再由全连接层输出处方信息。

在嵌入层, 对数据进行向量转换。训练集中包含 n 条数据, 原始特征向量 x 包含无序特征 (例如治法)、二元特征 (例如性别) 或连续特征 (例如年龄) 等 m 个类别, 即 $x=(x_1, \cdots, x_m)$, 所以多字段分类形式广泛应用于中药处方推荐中数据, 每个原始处方特征都被视为一个字段, 针对原始输入处方数据集进行独热编码: 对于二分类特征字段, 例如性别, 只包含一个值, 直接通过独热编码作为输入特征, 对于无序特征字段, 通过独热编码加和后整体作为输入特征, 避免了同域的特征后续断断续续做特征交叉, 增加无用交叉, 对于连续特征, 每个连续特征值表示为值本身。例如输入一个例子: [性别=女, 年龄=28, 治法=理气和中, 处方=香苏散, 药材一: 苏叶 120g, 药材二: 香附 120g, 药材三: 陈皮 60g, 药材四: 炙甘草 30g, ……], 本文通过使用独热编码将中医处方信息转换成高维稀疏特征。

由于中药处方中的数据是具有分类、二进制等多特征类别, 经过独热编码处理后, 比如针对上千个药材名称, 对药材 id 进行独热编码后, 经过嵌入层处理后的特征, 简单拼接携带的信息太少, 因此 FM 通过线性模型的基础上添加了交叉项, 并根据处方推荐任务将处方特征与患者个人信息特征进行自动交叉组合, 并得到任何交叉特征的权重, 经过 FM 离散化的表达如下:

$$y_{FM}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} x_i x_j \quad (4-1)$$

其中, n 是 one-hot 编码的样本特征数量, x_i 是第 i 个特征的值, x_j 是第 j 个特征的值。本研究使用矩阵分解的方式对中药处方二阶参数 w_{ij} 进行因式分解, 对于每个用户治法信息和处方, 都能使用一个隐向量将其表示。因此 x_i 对应的隐向量与 x_j 对应的隐向量 v_i , v_j 的内积等于 x_i 和 x_j 的交叉项系数, k 表示隐向量的长度, 通过矩阵分解二阶交叉项稀疏 w_{ij} 为:

$$w_{ij} = \langle v_i, v_j \rangle = \sum_{f=1}^k v_{if} v_{jf} \quad (4-2)$$

因此, 本研究通过训练模型可以得到中医治法信息和处方信息的交互特征项系数, 可以将 FM 的表达式改写为:

$$y_{FM}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \quad (4-3)$$

本研究利用嵌入层将其转化为稠密向量，然后将其输入到神经网络中进行训练。利用嵌入层的优势在于，可以根据查表的形式，获得非零元素对应的权重，并将其相加，这样做的效果会更好。嵌入层得到的权重就是对应独热向量中的一个位置的权重。输入的特征向量 x 中的非零特征集合为 χ ，中医处方推荐模型嵌入层的输出 a 为：

$$a = [e_1, \dots, e_m] \quad e = \{v_i x_i\}_{i \in \chi} \quad (4-4)$$

本模型将嵌入层的输出结果 a 输入到 AFM 的两两特征交叉层，本研究使用两两特征交互层将 n 个特征向量压缩到 $n(n-1)/2$ 交叉向量，任意两个交叉向量都是患者信息和处方两个向量之间的元素乘积，用于编码中药处方特征的交互。

注意力分解机模型添加注意力网络用于学习处方中用户信息和药材信息的加权特征，学习出加权特征的权重得分从而进行推荐。因为在实际的推荐中，不是所有的特征交互对推荐的贡献值相同。因此，引入注意力机制，以捕获数据的显著结构，从而提取对推荐模型很重要的特征交叉。注意力模型也可以被视为输出的加权平均。中医处方推荐模型将两两交互层的的患者信息和处方信息输出向量如图 4-2 所示 $\langle v_i, v_m \rangle x_i x_m$ 作为注意力单元输入向量，返回的是一个基于注意力得分的向量，该向量是输入向量加权算术平均数，并根据向量中每个元素的重要性选择权重。 α_{ij} 是经过注意力网络处理后的特征交互权重，表示不同的特征组合对最终预测的贡献程度。因此本模型只保留特征交互的重要部分。对于中医处方推荐模型中训练的时候没有出现过的交叉特征，本模型无法估计交叉特征的注意力值。为了解决这个问题，本模型使用感知器对注意力进行参数化，从而形成一个注意力网络。权重是通过注意力网络学习的，使用以下公式计算处方推荐的注意力权重：

$$\begin{cases} a'_{ij} = h^T \sigma(W(v_i \odot v_j) x_i x_j + b) \\ a_{ij} = \frac{\exp(a'_{ij})}{\sum_{i=1}^m \sum_{j=i+1}^m \exp(a'_{ij})} \end{cases} \quad (4-5)$$

其中， σ 表示激活函数 ReLU， $W \in R^{t \times k}$, $b \in R^t$, $h \in R^t$ 是模型的参数， t 表示注意力隐层神经元的个数，是通过网络优化得到。

本模型得到每个二阶特征向量的权重后，AFM 通过对所有特征向量进行加权池化层求和的方式，得到综合所有特征信息的向量，然后将其输入到整个全连接层，通过 Softmax 计算出最后的预测值。最后将 α_{ij} 的值添加到 AFM 模型中。

$P \in R^k$ 表示预测层的权重。公式如下：

$$y_{AFM}(x) = w_0 + \sum_{i=1}^m w_i x_i + P^T \left(\sum_{i=1}^m \sum_{j=i+1}^m \alpha_{ij} \langle v_i \odot v_j, x_i x_j \rangle \right) \quad (4-6)$$

通过以上算法完成了中药处方推荐模型的训练，最终模型输出最佳推荐处方概率集合，由医生进行选择最适合病人的处方。

基于注意力分解机的中医处方推荐模型，在嵌入层通过独热编码对输入数据进行预处理，通过引入场景域的概念，把相同性质的特征归于同一个场景，避免了后续得特征连续的特征交叉，可以避免无用特征的交叉，这样对于无序特征的推荐问题，使得中医处方推荐模型的时间复杂度大大减少，并且提高处方推荐的准确率。

4.3 算法设计与分析

基于注意力分解机的中医处方推荐算法的具体步骤流程如表算法 2 所示。

算法 2 中医处方推荐算法

输入：中医治法名称 x
 输出：对应治法名称 x 的治疗处方

- 1: while 输入治法名称 do
- 2: 利用 one-hot 编码处理输入的中医治法文本
- 3: 处方推荐模型嵌入层的输入特征为 $e \leftarrow \{v_i x_i\}$
- 4: if e not null
- 5: 使用 FM 算法对特征集进一步特征处理
- 6: $y_{FM}(x) \leftarrow w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} x_i x_j$
- 7: end if
- 8: 使用矩阵分解将治法到处方特征转化为低维稠密特征
- 9: $w_{ij} \leftarrow \langle v_i, v_j \rangle$
- 10: 注意力机制提取对推荐模型的交叉特征，计算注意力权重 a_{ij}
- 11: for $i < \text{交互向量数}$ do
- 12: $a'_{ij} \leftarrow h^T \sigma(W(v_i \odot v_j) x_i x_j + b)$
- 13: end for
- 14: $a_{ij} \leftarrow \frac{\exp(a'_{ij})}{\sum_{i=1}^m \sum_{j=i+1}^m \exp(a'_{ij})}$
- 15: 将处方推荐结果转换成概率值输出，得出推荐的处方信息
- 16: 网络优化：优化函数为交叉熵损失函数，并在损失函数上添加 L2 正则化项
- 17: end while

在该算法步骤流程中，本研究将中医治法名称集分为训练集合测试集，主要包含患者个人信息数据、治法名称。根据中医师的诊断结果，将中药治法对应的标签处方信息分为训练标签集和测试标签集。

首先，从中药处方推荐的数据集合中读取一个代表向量形式的样本。采用因子分解机的二阶交叉项将患者和处方的多类别属性进行交叉组合，在注意力网络将处方中有效特征赋予更高的权重，从而区分处方中不同组合特征的重要程度，采用梯度下降优化卷积神经网络，通过训练得到最终的中医处方推荐网络结构。同样，本研究使用中医处方测试集和对应的中医处方诊断结果测试中医处方推荐算法的准确度。

为了评估算法的性能指标，除了采用有效的实验方法外，还需要制定模型的性能指标。针对该研究的推荐算法，使用准确率、召回率来评价该方法，具体计算公式如下其中 TP、TN、FP、和 FN。TP 表示一个正例被正确地预言为正例，TN 表示一个反例被正确地预言为反例，FP 表示一个反例被错误地预言为正，FN 表示一个正例被错误地预言为反例。

准确率是指在全部的中医数据集中，对中医治法分类的预测是正确的。计算公式如下：

$$Accuracy = \frac{N_{correct}}{N} \quad (4-7)$$

召回率是指在一个实际为正地样本中，所预测地实际为正的可能性。计算公式如下：

$$Recall = \frac{TP}{TP + FN} \quad (4-8)$$

4.4 实验

4.4.1 实验数据

本研究对中医处方数据集进行实验，通过对古典书籍中的处方数据集进行训练来检验算法的正确性，并通过实验结果进行对比。对以上古典书籍进行整理，整理的治法大类有 23 种，治法小类 400 余种，处方中包含 4174 种常用中医药材名称、药材剂量、来源等信息，经典名方包含 900 余种。在训练和测试数据集中，中医处方信息特征被分为 21 类，包含患者年龄、性别、是否怀孕、治法、处方名称以及药材等类别，最终的目标是为患者推荐处方。此外，我们还按照比例将 11990 正负样本中医治法到处方数据集进行了划分，将其划分为训练集和测

试集，两者的比例分别是 80%、20%。

在处方推荐模型中，把中医治法名称、处方信息以及患者的个人信息数据的原始特征映射成嵌入向量，从而将中医治法的高维度稀疏特征转化为稠密特征，嵌入层由单个神经网络层组成。中医处方数据是多特征类别的无序特征（例如治法）、二元特征（例如性别）以及连续特征（例如年龄），本研究通过将各个类别的中医数据特征进行区分，将中医类别信息进行独热编码处理，例如，患者年龄等连续数据，除了进行归一化处理外可以直接使用，性别等二元特征将使用 [0,1] 表示女性，[1,0] 表示男性，以此类推。将中医处方实验数据通过进行简单的预处理，主要包括数据清洗、数据转换以及类别编码。另外，还要构建数值特征和类别特征列表，将数据集中所有特征对应的特征进行标记，构建输入层字典，并以稠密和稀疏两类字典的形式返回。将所有的稀疏特征进行嵌入拼接，将稀疏特征筛选出来，获取输入层和对应的嵌入层，如果嵌入列表最终是直接输入到全连接层中，需要进行展平，否则不需要。最后，将输入数据转化成字典的形式，等待输入模型。

4.4.2 实验设置及分析

本研究的中医处方推荐模型实验环境配置主要包括操作系统，专业显存，内存，GPU，CPU 以及硬盘大小，并且也介绍了实验所用到的编程语言和机器学习框架。如下表 4-1 所示。

表 4-1 实验环境配置

Tab. 4-1 Experimental environment Configuration

参数名称	参数值	参数名称	参数值
操作系统	Window10	硬盘大小	1TB
专用显存	4GB	CPU	4 核
内存	16GB	GPU 核个数	2 核
机器学习框架	Tensorflow	编程语言	Python

网络中超参数的选择一般需要根据数据的不同而有所调节。为了实现每个模型的最佳性能，对参数进行了仔细研究，中医处方推荐模型进行监督训练使用最小化预测误差，由于本研究的中医处方推荐模型是一个二元模型，最常用的目标函数是最小化交叉熵函数。

表 4-2 实验参数表

Tab. 4-2 Experimental parameter table

实验参数	实验设置
批大小 batch-size	64
Dropout 保留比率	0.5
学习速率	0.001
L2 正则化参数	0.01
优化器	Adam

本研究使用的最小化交叉熵函数作为中医处方推荐目标函数。本研究处方推荐模型和对比模型使用相同超参数，使用的非线性函数 ReLU 作为中医处方推荐模型的激活函数，Dropout 设置为 0.5，采用 L2 正则化和 Dropout 来防止中医处方推荐网络过拟合，将 Adam 作为优化器，自适应调整学习率。如表 4-2 所示。

本文在中医处方推荐数据集上将本研究提出的中医处方推荐模型与（Factorization-Machine based Neural Network，DeepFM）和（Neural Factorization Machine，NFM）进行了对比。在中医处方推荐的场景下，更关注的是召回率，通过公式(4-7)和(4-8)可以得知，本实验的精度和召回率分别为 0.67 和 0.65，本研究模型的召回率和精度相对于对比试验效果更好。中医处方推荐模型对比的实验结果如图 4-2 至 4-5 所示。

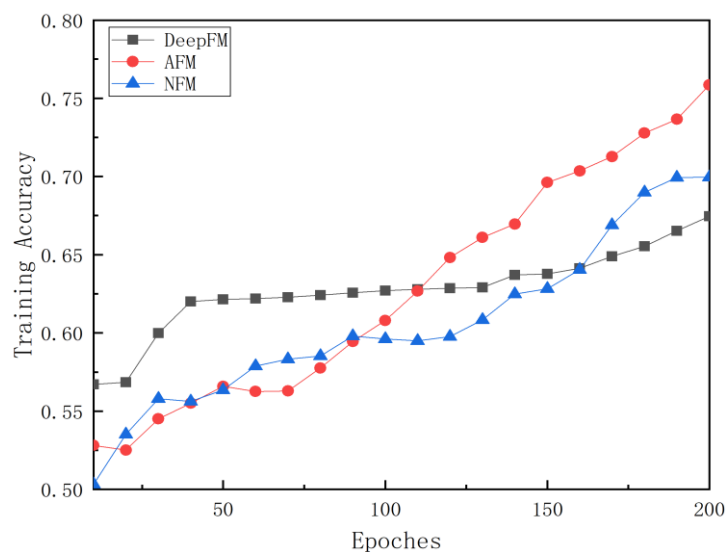


图 4-2 训练精度对比图

Fig. 4-2 Training accuracy comparison

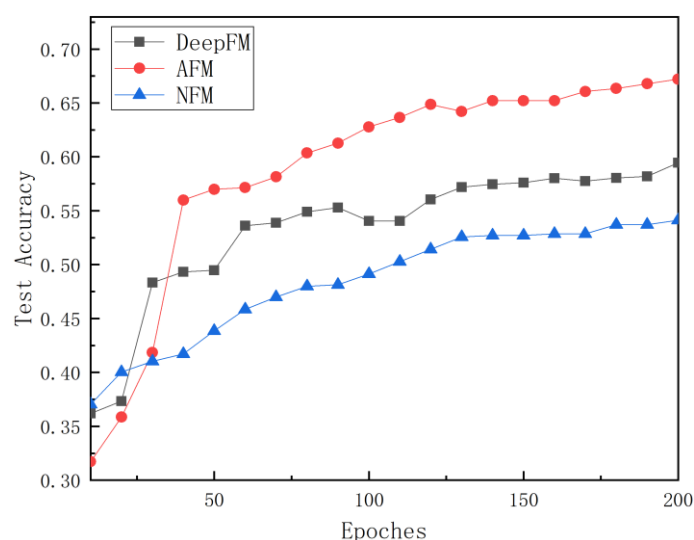


图 4-3 测试精度对比图

Fig. 4-3 Test accuracy comparison

图 4-2、图 4-3 分别是训练精度和测试精度的曲线图，由图中可以看出，本研究的处方推荐模型效果最好，相比其他两种模型具有较高的精度，在测试集中的对比模型中，所有的特征彼此都是平等的，不能反映类别和层次信息。因为并非所有的特征交互都能实现正增益，添加噪声特征甚至会损害模型的精度。而 AFM 模型为每个特征分配一个注意力权重后可以起到筛选重要交叉特征。因此本研究的模型效果优于其他两种对比算法。

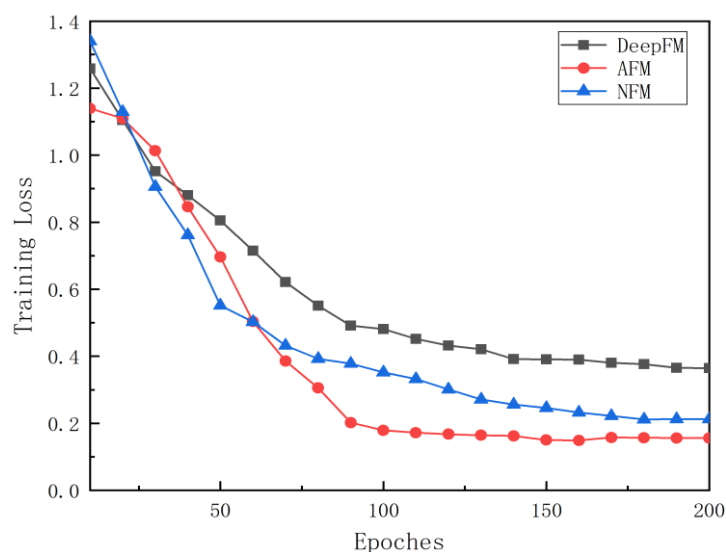


图 4-4 训练损失对比

Fig. 4-4 Training loss comparison chart

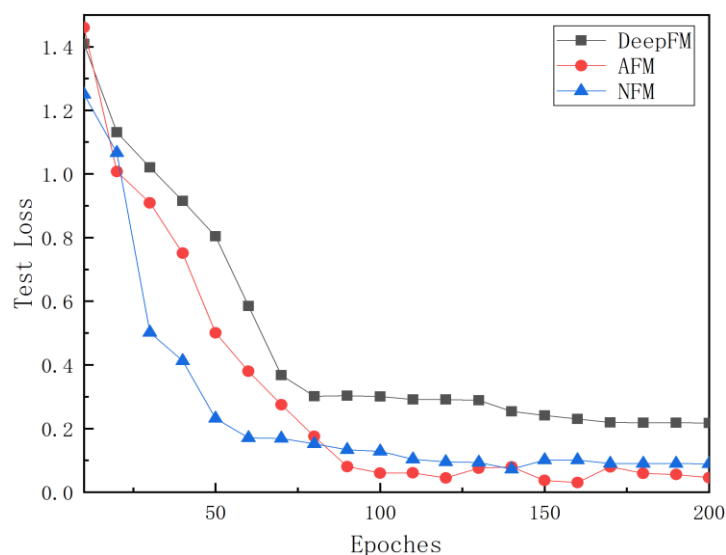


图 4-5 测试损失对比图

Fig. 4-5 Test loss comparison chart

图4-4、图4-5分别是中医处方推荐的训练损失和测试损失的曲线图，当本研究将Dropout设置为0.5时，本研究的模型以及两个对比的算法的泛化能力都有显著提高，但注意力分解机模型相比其他两个对比算法效果更显著，即使没有使用Dropout，本研究所提出的推荐算法模型也高于因子分解机，这也更加证明了注意力网络对特征交叉的重要性，也表明注意力分解机对于未见的数据具有更广泛的分解能力。

4.5 本章小结

目前，由于数据采集技术、算法和集成技术的突破，我国的人工智能技术发展突飞猛进。对于具有连续、分类和二进制等多维的数据，很难进行特征工程和训练推荐模型。在本研究中，我们提出一种基于注意力分解机的处方推荐模型，将 AFM 运用到处方推荐中，并取得了不错的效果，实现了从中医治法到处方的推荐。作为一种复杂的神经网络模型，AFM 容易过度拟合，对给定的训练数据反应过度，导致性能不佳。实验结果表明，尽管使用了 L2 正则化和 Dropout，但是该模型依然存在过拟合问题，本文的下一个目标是如何解决过拟合问题。

第5章 系统设计与实现

5.1 系统需求分析

本节对中医辅助诊疗系统在实际中所面临的需求进行了分析，主要包括系统目标、功能需求分析和性能需求分析。

5.1.1 系统目标

本文旨在对中医辅助诊疗系统进行设计与实现，该系统拟从基本的中医治法理论入手，通过收集中医古典书籍中真实数据，采用数据挖掘技术，建立了“证候”、“治法”、“处方”数据仓库，并且通过深度学习建立中医论治和中医药处方推荐学习模型，完成相应的诊断策略，最终实现中医辅助诊疗系统的学习过程，为患者进行个性化处方推荐。该系统是辅助医生诊断，医生根据患者描述的证候和患者自身没有感觉出的一些具体证候，录入到该系统，医生结合该系统和自己的经验、临床知识选择相应的中医治法、处方、以及一些注意事项等信息。以此规范在职中医的诊疗过程，检查出是否存在误治、失治等现象，从而提高临床疗效，为中医的信息化发展奠定基础。

5.1.2 功能需求分析

中医辅助诊疗系统是针对医生用户设计的集中医论治、处方推荐和经典方库等为一体的中医辅助诊断系统。以名老中医选法定方经验的思维模式，将中医处方推荐与数据挖掘、深度学习技术相结合，辅助医生决策，帮助中医师学习中医理论知识，给患者更加安心的治疗，切实提升中医的诊断水平。本系统各个功能需求分析如下。

中医论治需求分析：论治是根据症状、舌苔、脉搏等证候信息确定中医治法的过程。首先医生将患者所描述的证候，录入到辅助诊断系统，通过该系统得出最佳的治疗方法。并且医生可以对患者的信息进行添加、删除等修改操作。该治法主要包括解表法、表里双解法、涌吐法和清热法等 23 大类和中医常用 400 余种治法小类。

中医辅助诊疗系统需求分析：处方推荐是针对相应的治法给出的诊疗方案，处方推荐是诊断过程中至关重要的一步，如果诊断失误，将会影响患者病情治愈程度，更甚者导致病情加重。医生结合患者的年龄、性别以及是否怀孕等信息，来对患者实现个性化推荐。推荐出的处方有多条，因此医生可以通过系统给出的多个处方结合自身经验选择，针对特定的人群，筛选掉有副作用的药材和添加一

些附加药材。中医辅助处方推荐主要包含处方药材名字、药材的剂量、来源以及药品的解释信息。

中医经典方库需求分析：它是包含众多古今名方的中药经典方库。中医方库是将中医古典书籍的名方进行整理，为医生决策提供了帮助，医生可以查看每个经典方库的具体信息，强化自己的中医临床知识，对自身未来发展提供支撑。针对每个药方做出详细解释包括方解、来源、用法等信息。它是在处方推荐的基础上将所有的处方结果数据集进行收集，呈现出来供医生查看巩固专业知识。

5.1.3 性能需求分析

性能需求分析是针对系统运行过程中系统的各项指标是否达到要求，对于一个完整的系统除了功能需求分析外，系统的性能分析也至关重要。

1. 易用性：中医辅助诊疗系统设计与开发是为了增加医生对病人的诊断效率，为不同的医护人员提供便捷的使用功能，因此该系统的使用流程一定要简洁明了，尽可能地与医护人员平时的操作习惯一致，确保提升医护人员的效率。同时，针对年龄稍大的医生的计算机水平不高，设计时应该重视系统的易用性，使大部分的医护人员都能参与并使用该系统。

2. 稳定性：系统的稳定性是指用户在对该系统进行操作过程中，系统的运行不会出现错误，稳定性是系统设计最为基础的保障。例如，当医生的操作频繁或多名医生同时访问时，系统不会出现崩溃、界面混乱以及数据报错等现象。同时，当医生由于失误造成点击错误时，系统应该能相应的处理，系统不会因部分模块失误导致整体模块瘫痪。

3. 可维护和可扩展性：系统的可维护性是指在智慧中医辅助诊疗系统出现故障的时候，开发人员能以较低成本进行维护，因此设计方案和架构必须符合国际主流标准，开放性好。由于患者的需求会发生变化，医生的诊断操作也发生相应的变化，因此我们必须考虑系统的可扩展性，当增加新模块时，新增模块不会对其他原有模块造成影响并且能很好地与原有模块相融合。

4. 安全性：该系统中存储着患者大量的个人信息数据，而该信息一般都是相对隐私的，因此要保证操作用户的合法性，同时应建立完整的网络防毒机制和使用规范，以防被黑客盗取数据，此外也要防止数据被非法篡改，当被篡改时系统也能及时地修复。

5.1.4 可行性分析

本系统以 Java 为开发语言，具有很高的可移植性。只要稍作修改，就可以在不同的平台上运行，该系统使用 MySQL 数据库技术，所以该系统的准确性与

安全性得到了很大的提升，并且也一定程度改善了用户的并行操作与用户管理方面，而且该系统易于修改和扩展，有利于研发人员后续进行开发。本系统采用 Asynchronous Javascript+Xml (AJAX) 技术，在后台和服务器之间进行信息的交互，从而达到网页的非同步传送，所以，从技术上讲，这是一种可行的方法。探索智慧医院建设，开发辅助医生诊断的中医辅助诊疗系统是教育、医疗、养老等民生的迫切需求。传统的中医诊断方法，多依赖于个人的经验积累，主观意识和环境的变化，而且无法满足中医的传承与发展。因此，将现代化技术与中医选法定方相结合，设计智能中医辅助诊疗系统，对于提高医生决策效率、培养锻炼年轻中医生、提升中医现代化发展速度具有重要意义，而且该系统具有普遍性、且覆盖率较高，因此该智能中医辅助诊疗系统具有可行性。

5.2 系统概要设计

上节对智能中医辅助诊疗系统的需求分析做了详细介绍，本章主要介绍了系统的设计。

5.2.1 总体设计

智能中医辅助诊疗系统首先从中医内科学、外科学等中医古典名籍中获取患者证候、治法和处方等数据信息，利用数据挖掘技术对采集的信息进行处理，主要包括空值、异常值的清洗和不同格式数值之间的转换，作为神经网络的输入值，继而构建神经网络模型，调整参数配置并训练模型。该系统主要分为三大模块：中医论治诊断模块，智能处方推荐诊断模块，经典方库诊断模块。使用过程中，主要受众的是中医师。医生根据患者的描述，对患者的证候等信息有了初步了解，然后结合本研究的中医论治算法和处方推荐模型，输出患者最终所需要的处方以及处方的详细解释。

5.2.2 架构设计

本系统选用了数据层、逻辑层和表现层三层处理结构。数据管理层是对本系统所收集的数据进行存储、访问和管理，利用各层逻辑之间的复用，为该系统提供了数据层面的支撑。逻辑层处于架构设计核心位置，是最为关键的部分，主要负责中医论治算法和处方推荐算法的实现以及模型参数的配置。它通过调用数据层的数据进行操作，并将其处理后的数据传送至表现层，表现层是使用者与该体系的互动，也就是展示给中医师的系统界面，主要接收中医师的请求，以及处方数据的返回。如图 5-1 为智能中医辅助诊疗系统整体架构图。

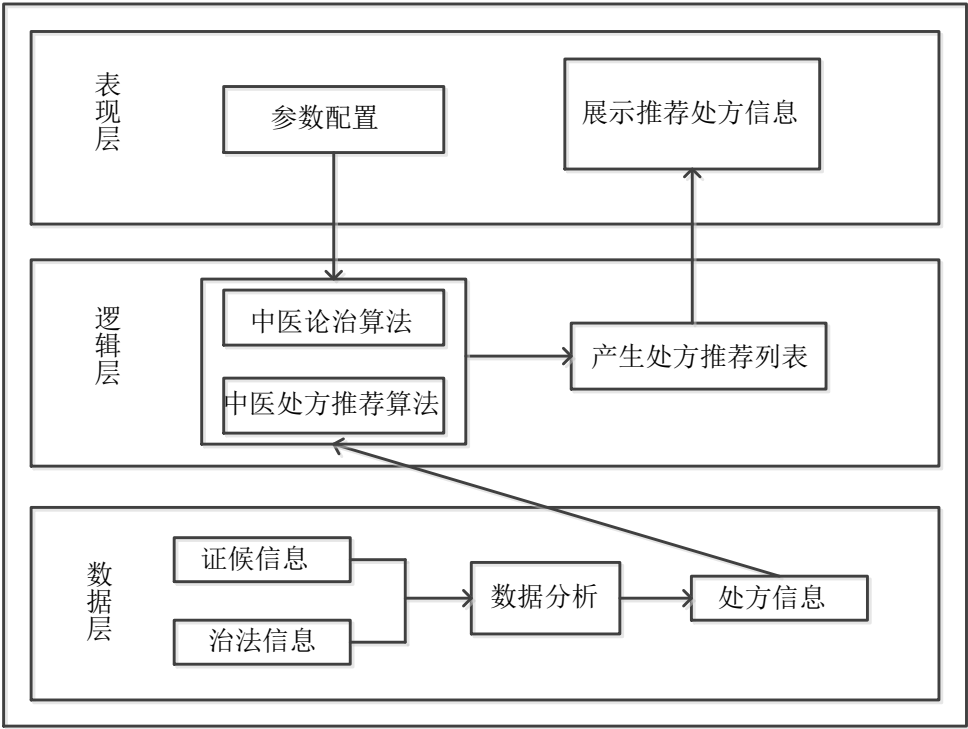


图 5-1 系统架构图

Fig. 5-1 System architecture diagram

5.2.3 功能模块设计

本系统的功能主要包括中医论治模块、处方推荐模块和经典方库三个模块。下图 5-2 为本系统的功能模块图。

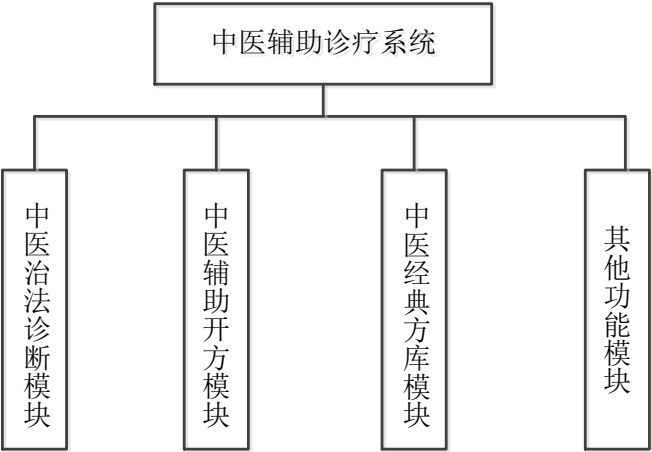


图 5-2 系统模块图

Fig. 5-2 System module diagram

中医论治模块：中医论治模块是从证候到治法的过程。首先进入系统的基本信息界面，医生输入患者的年龄、性别、是否怀孕和证候等数据信息，提交之后系统将数据读入，通过建立好的中医论治模型对读取的数据进行预测分类，系统预测结束后返回最佳的治法方案。在本模块，我们需要将训练好的中医论治神经网络部署到辅助诊疗系统中，通过将中医论治模型转换成序列化的模型，并且以 Torch Script 存储起来，选择合适的最终在 Windows 上以本地的形式集成模型数据，将代码进行移植并在本系统中添加中医论治模型，即可实现最终的部署，下图为中医论治流程。

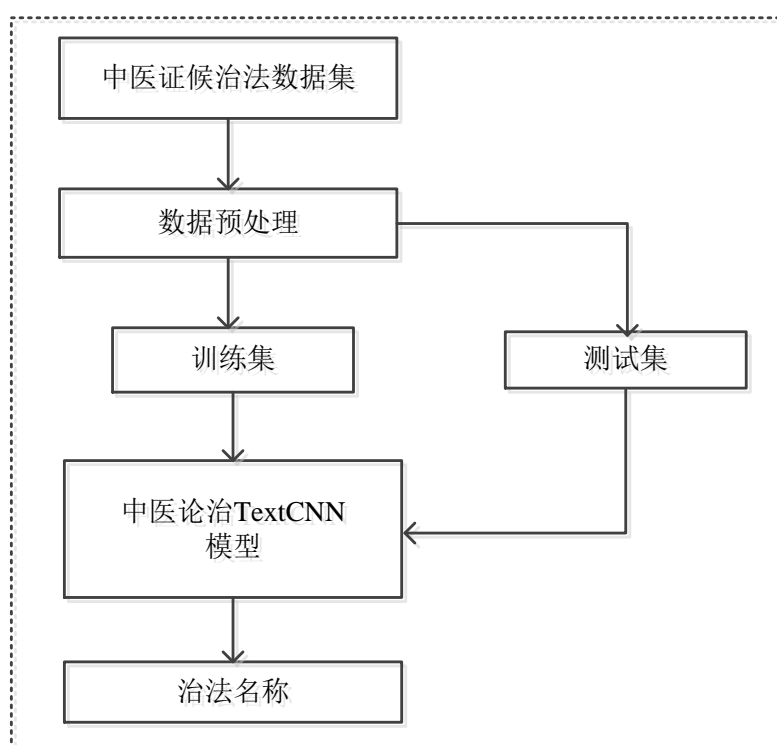


图 5-3 中医论治模块图

Fig. 5-3 TCM therapeutic methods module diagram

中医处方推荐模块：处方推荐是针对相应的治法给出的处方诊疗方案。该部分主要包含中药处方的名称、剂量等详细解释。医生点击上一步所出现的治法，由处方推荐模型结合患者的个人具体信息给出个性化处方。系统会展示出患者所需要的处方详情。在此模块，对处方推荐模型进行打包，选用 pytorch 离线训练框架和在线 inference 的框架，对处方推荐模型进行部署，能在一定程度上缩短模型的部署周期。下图为中医处方推荐模块图。

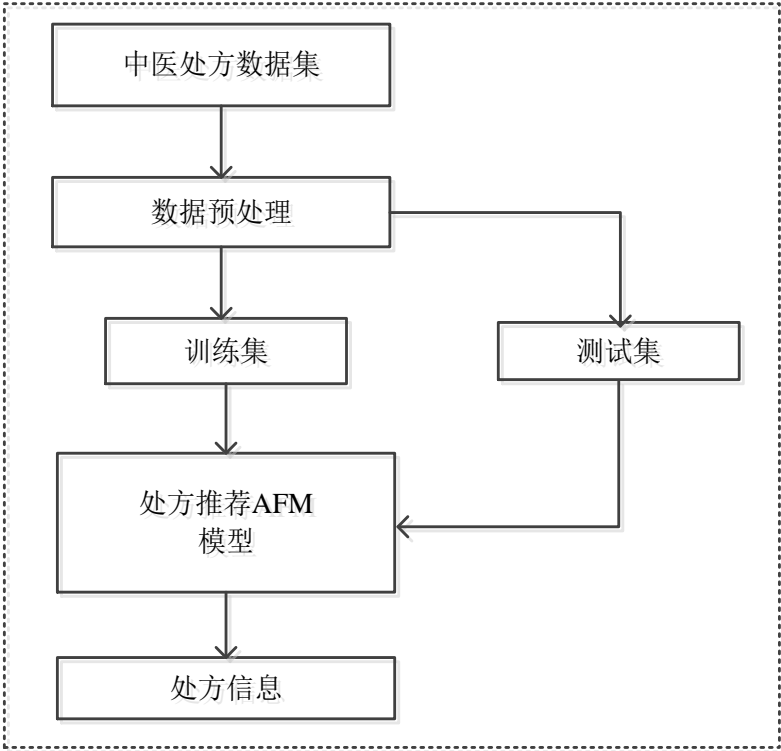


图 5-4 中医处方推荐模块图

Fig. 5-4 TCM prescription recommendation module diagram

经典方库模块：经典方库为医生提供了中医经典处方查看机制，医生对处方推荐的模块有疑惑时，可以使用该模块对处方进行查看并解惑，本模块对收集的经典名方进行存储，包括处方名称、药材、来源、治法、用法等信息，将该信息存储在数据库中，以备医生查看和巩固自己的中医临床理论。

5.2.4 数据库设计

数据库设计是处方推荐系统的基础，它是针对中医辅助诊疗的应用环境，设计的存储结构和存储方法，并由此运用到中医辅助诊疗系统中，从而满足不同医生的使用需求。一个好的数据库设计会使系统操作更简单、使用效率更高以及更易于实现等特点。本系统以中医临床诊疗国家标准典籍进行汇总，整理出患者的证候、年龄、是否怀孕等信息，以及对应的治法、推荐的中药处方以及多种经典名方。

在中医论治模块中，我们从中医学典籍共整理了 10065 条治法数据。其中，按照治法分为 23 大类，下表为治法名称分类数据表。

表 5-1 治法分类数据表

Tab. 5-1 Therapeutic classification data table

治法大类	治法数据量
解表法/发表发	195
攻下法/泻下法	105
和解法	439
清热法	903
理气法	825
祛湿法	877
.....
温里法	350
祛痰法	563
安神法	280
消导法	270
治痢疡法	440
治五官法	535
总计：23 大类	总计：10065 条

在中医辅助诊疗系统的智能处方推荐模块中，本研究从中医学古典书籍中按病证的类型对治法的处方进行整理，共整理出 327 种病证类。每种病证对应多种治法，其中，针对病证中的不同的治法对应的处方类型共整理出 11990 种。下表为病证分类数据。

表 5-2 病证分类数据表

Tab. 5-2 Disease classification data table

书籍名称	病证分类数量
中医外科学	88
中医内科学	51
中医眼科学	52
中医儿科学	48
中医妇科学	43
中医急诊学	45
总计	327

数据库表的设计对于智能处方推荐系统非常重要，前期考虑不全会导致重构起来非常麻烦，养成良好的设计习惯，可以使系统开发事半功倍。通过对中医论治模型、处方推荐模型以及经典方库进行需求分析，设计了以下数据库表。

患者信息表（Patient_information）主要包含患者 ID、患者姓名、年龄、性别、是否怀孕、证候等信息。下表为患者信息表。

表 5-3 患者信息表

Tab. 5-3 Patient_information

字段名	数据类型	约束类型
Patient_ID	int	主键
Patient_Name	char	非空
Patient_Age	int	非空
Patient_Sex	int	非空
Patient_pregnant	int	非空
Patient_Symptom	varchar	非空

治法名称表（Therapy_name_ID）主要包含中医治法 ID、治法大类以及小类，下表为治法名称表。

表 5-4 治法名称表

Tab. 5-4 Therapy_name_ID

字段名	数据类型	约束类型
Therapy_name_ID	int	主键
Therapy_name_da	varchar	非空
Therapy_name_xiao	varchar	非空

处方名称表（TCM_Prescription_Name）主要包括处方名称 ID、处方名称大类和小类信息。下表为处方名称表。

表 5-5 处方名称表

Tab. 5-5 TCM_Prescription_Name

字段名	数据类型	约束类型
Prescription_Name_ID	int	主键

续表 5-5

Continued 5-5

字段名	数据类型	约束类型
Prescription_Name	vachar	非空
Therapy_Name_xiao	varchar	非空

药材配方表（TCM_Pres_details）主要包括配方详情 ID、处方推荐表 ID、药材 ID、药材剂量。下表为药材配方表。

表 5-6 药材配方表

Tab. 5-6 TCM_Pres_details

字段名	数据类型	约束类型
TCM_Pres_details_ID	int	主键
Herbs_ID	vachar	外键
Herbs_Dosage	varchar	非空
Prescription_Name_ID	int	外键

经典方库表（Prescription_Intro）中主要包括方库 ID、处方来源、治法名称 ID、处方用法、药材名称 ID、药材剂量、主治等信息，下表为经典方库表。

表 5-7 经典方库表

Tab. 5-7 Prescription_Intro

字段名	数据类型	约束类型
Prescription_Intro_ID	int	主键
Prescription_Intro	text	外键
Prescription_Source	varchar	非空
Therapy_name_ID	int	外键
Prescribing_Instructions	varchar	
Herbs_ID	int	外键
Herbs_Dosage	varchar	非空
Prescription_ZH	varchar	

ER 图是用来描述智能处方推荐中数据对象与数据对象之间的关系，它主要

有实体、属性和关系三个基本要素，下图为本系统的智能处方推荐 ER 图。

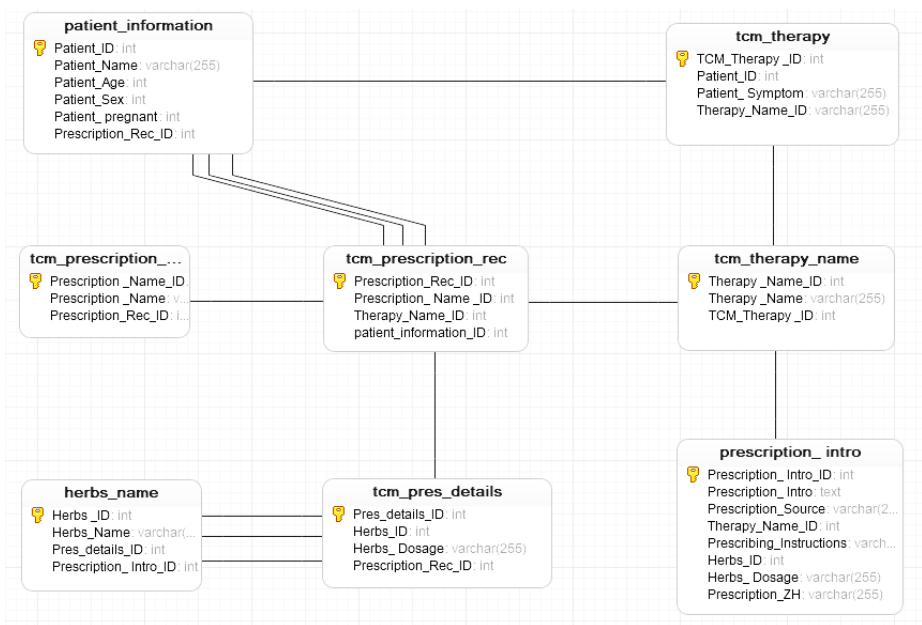


图 5-5 系统 ER 图

Fig. 5-5 System ER diagram

5.3 系统实现与测试

本模块主要对中医辅助诊疗系统的开发环境与技术、功能模块实现以及系统设计进行主要介绍。

5.3.1 运行环境与开发平台

该体系以Eclipse为基础，使用了当前使用最为普遍的Java程序设计语言，Java基于JVM的虚拟机的跨平台语言，使得该程序的可移植性高，只要稍作修改，就可以在不同的平台上运行，并且Eclipse非常适合Java语言开发编写，它的插件功能强大，是一个免费开源的软件，用Linux云服务器对我们的项目进行部署。其中中医论治模型和智能处方推荐模型是在Python应用开发框架下开发实现，使用NVIDIA (R) Cuda compiler driver进行训练，他们拥有强大的第三方库，可以直接进行调用，能降低系统开发周期。

5.3.2 功能模块实现

1. 中医论治模块

在中医论治模块，医生根据患者描述的证候信息以及对患者舌苔、脉搏等

的观察，对此进行概括总结，录入辅助诊疗系统中，点击提交，系统会智能地给出针对该患者的治法信息。本系统主要包含23种治法大类以及400余种治法小类。下图为中医治法诊断界面图。

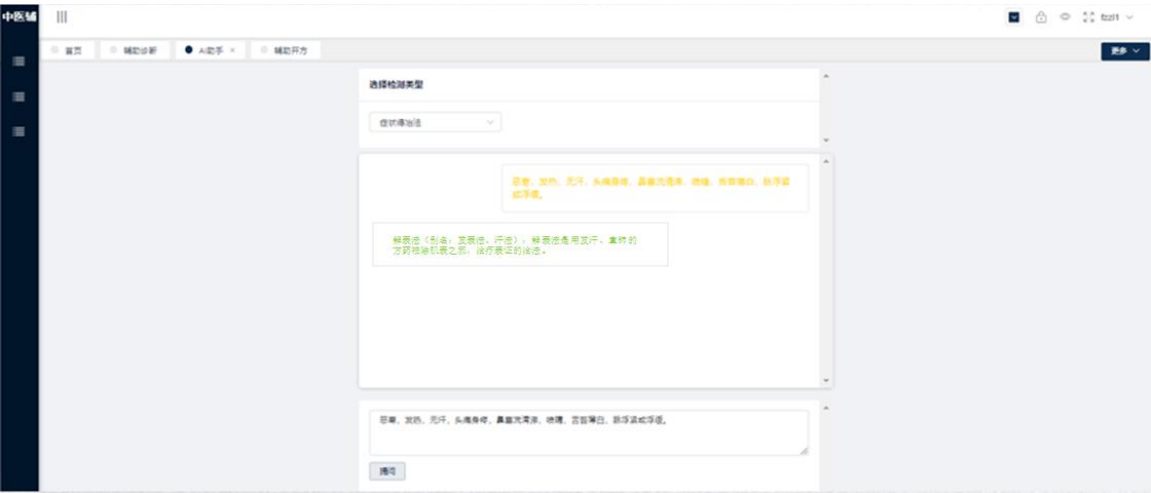


图 5-6 中医论治界面图

Fig. 5-6 TCM therapeutic methods interface

2. 处方推荐模块

在本处方推荐模块，医生根据系统提示的治法信息，选择合适的治法种类，点击下一步，则系统会根据选择的治法推荐对应的多个处方，再由医生根据用户实际情况选择首选处方和次选处方。该处方界面主要包含药材的名称、剂量以及查看方解等信息。界面如图5-7、5-8、5-9所示。



图 5-7 中医治法界面图

Fig. 5-7 TCM therapeutic interface



图 5-8 中医处方界面图

Fig. 5-8 TCM prescription recommendation interface



图 5-9 中医处方详情界面图

Fig. 5-9 TCM prescription recommendation details interface

3. 经典方库

祖国中医在历史的长河里，通过选法定方，积累了许多安全、方便和效用显著的经典名方。中医名方以其简单的配方以及出众的疗效让人惊叹不觉。经典方库模块是中医论治模块和处方推荐模块之外的模块，经典方库为医生提供了中医经典处方查看机制，医生对处方推荐的模块有疑惑时，可以使用经典方库模块对处方进行查看并解惑，经典方库模块通过点击处方名可以查看每个处方的详细解释，主要包括中医方解、药材的来源、中医师的主要治法、药材如何使用等信息，以供医生查看和学习。通过学习经典名方，将中医药的文化进行传承与发扬。中医经典名方界面如图5-10、5-11所示。

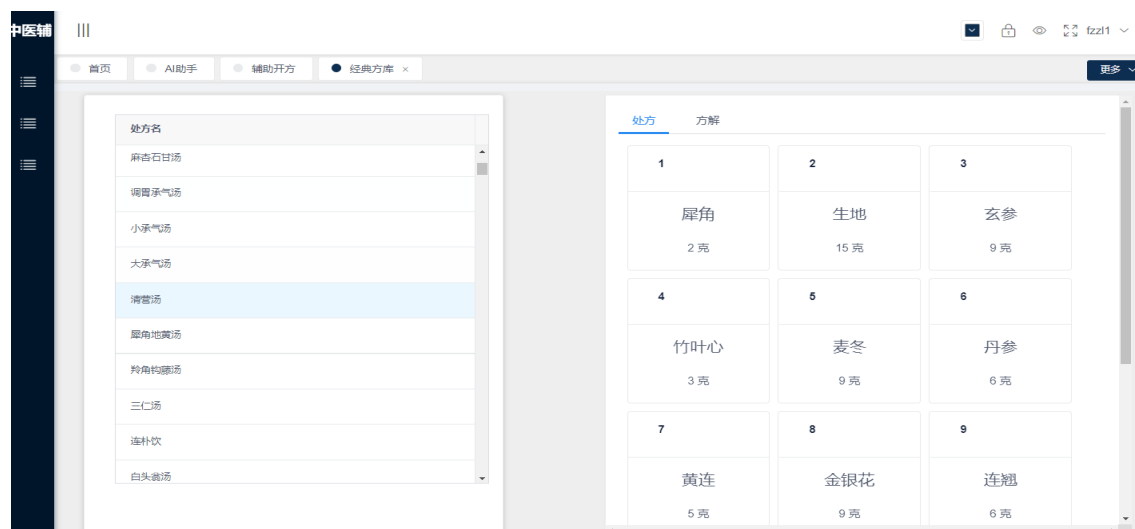


图 5-10 经典方库界面图

Fig. 5-10 TCM Classic library interface

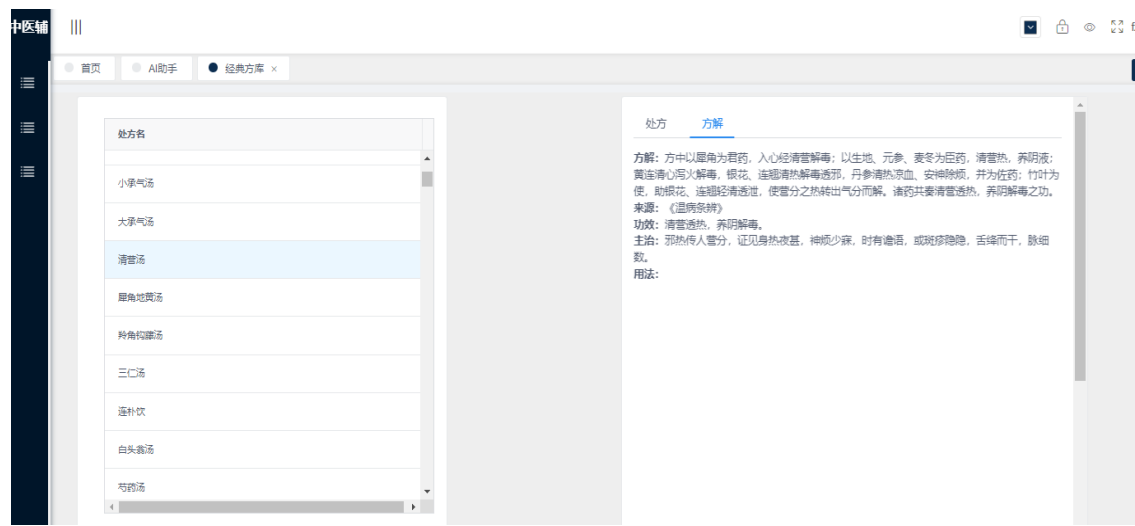


图 5-11 经典方库详情界面图

Fig. 5-11 TCM Classic library details interface

5.3.3 功能测试

智能中医辅助诊疗系统可以辅助医生为患者提供健康管理服务。中医处方辅助系统可以辅助医生为患者提供健康管理服务。本系统主要对中医论治模块和智能处方推荐模块进行测试。该系统根据患者的个人信息、证候推荐相应的中医，再由中医治法和患者的不同信息推荐处方效果较好，能够辅助中医生进行决策。下表 5-8 是本系统的用例测试表。

表 5-8 测试表

Tab. 5-8 Test table

用例编号	用例描述	预期结果
1	输入本系统网址	展示登录页面
2	输入医生账号密码，点击登录	登录成功，展示首页
3	点击 AI 问诊	展示治法模块功能
4	医生输入证候信息	显示该证候的对应治法
5	点击治法	显示治法对应的处方名称
6	点击处方名称	显示药材名称、剂量信息
7	点击方解	显示处方的来源、治法等信息

以上用例是对本系统的具体模块进行了验证，测试效果与预期结果一致，达到了本系统设计的预期要求。

5.4 本章小结

近年来，中医诊断在人们生活中占据举足轻重的作用，因此将中医诊断和信息化技术进行结合是必然趋势。而中医的传统诊断，则是依靠主观意识和经验的积累，不仅缺乏客观的参考标准，而且无法满足研究上的可重复性要求，这些问题严重制约着中医处方推荐的进一步传承与发展。因此，本章将传统中医和人工智能结合，对智能处方推荐系统进行了设计和实现。该系统主要围绕中医论治和处方推荐算法设计相关功能。首先，根据医生临床诊断真实过程对本系统的需求分析进行了简要概述，针对本系统需求对系统的架构和模块进行具体设计，以便更好地辅助医生诊疗，从而提升医生的看病效率。最后对系统开发中使用的开发技术和环境、功能实现以及测试进行了说明。

第6章 结论

6.1 总结

中医传承着中华民族几千年的文化和历史，其治疗手段，临床效果明显。近年来，随着社会生活的需求及国家政策的支持，“人工智能+医疗”的发展迅猛，借助图像处理、模式识别、深度学习等人工智能技术来辅助医疗诊断已成为研究的热点，这是让古老中医在科技时代再次焕发光彩的必要途径^[51,52]。面向中医证候的智能处方推荐系统利用现代计算机技术以及大数据的分析能力，对患者的证候信息进行输入，经过中医论治模型得出治法信息，再经智能处方推荐模型得出中药配方信息。中医辅助诊疗系统提高了中医生诊断效率，利于名老中医诊疗经验的数字化、智能化传承，对打造健康中国、构建人类卫生健康共同体，具有战略性的意义。

本文根据中医师选法定方的过程设计并实现了智能中医辅助诊疗系统。本系统不仅可以使患者快速了解自己的病情并且为医生临床诊断提供了思路，在一定程度上节省了医疗人力。本系统主要分为中医论治、智能处方推荐和经典方库三大模块，主要工作包括以下几点。

1. 中医药数据的质量对我们的模型起着至关重要的作用，高质量的数据是建模的基础。本论文的中医药数据主要从中医古典书籍中获取，由于获取的中医药数据是半结构化的，本文根据中药“半结构化”的特征，将中药资料进行了规范化，为以后的工作提供了必要条件。

2. 设计了一种基于文本卷积神经网络的中医论治算法。首先通过 GloVe 嵌入方式将中医数据自然语言数值化处理，将 CNN-static 模型的输出的词向量作为文本卷积神经网络模型的输入，使用隐藏层提取中医药数据的关键信息，最终通过输出层输出预测分类。试验结果表明基于文本卷积神经网络不仅训练速度快、稳定性好，而且预测分类的结果准确率也较高。能解决中医证候与治法间非线性的复杂关系问题。

3. 构建了基于注意力分解机的处方推荐模型，AFM 是在 FM 基础上添加了一个具有全连接层和 Softmax 输出层的注意力机制，通过注意力网络对本模型的每一个交叉特征提供权重，使其对每个患者的不同信息更有针对性，可以更加智能地进行个性化处方推荐。结果表明，加入注意力网络的模型具有更好的非线性表达能力并且泛化能力更强。

4. 本文对面向中医证候的智能处方推荐系统的需求分析、系统概要设计、

系统的设计与实现和测试等进行了详细地阐述。该系统对中医生诊断效率，老中医诊疗经验的数字化、智能化传承，具有战略性的意义。

6.2 展望

近年来，随着社会生活的需要及国家政策的支持，“人工智能+医疗”的发展迅速，借助数据处理、深度学习等人工智能技术来辅助医疗诊断已成为研究的热点。本研究针对当前医生对患者选法定方的诊断过程，设计与实现了智慧中医辅助诊疗系统。本文的研究以中医药数据的挖掘为基础，中医药数据集的质量在一定程度上对方剂生成的准确率有一定的影响。在中医临床决策中，中医生开处方针对不同的情况随机应变，只在经典方库中选择并不是特别准确。

本研究以中草药资料为依据，中药资料的质量直接影响到方剂的产生。同时，由于本研究是从经典方药中选取合适的处方药材，但是中医师在实践中的选择具有一定的弹性，仅从经典的方药中选取不太精确。针对以上问题，后期的研究将会围绕以下几个方面展开。

1. 扩展和规范中医药方剂数据的收集渠道。通过人工整理的数据存在林林总总的缺陷，比如属性缺失、数据格式不统一等问题，此时必须对数据进行清洗，这就导致前期的工作量有所加大并且可能由于专业性不够等原因，中医药数据预处理之后的准确性也许还会存在问题。因此通过和中医馆合作、购买等其他方式，收集更规范的中医药数据，这会对提升模型的准确率有很大帮助并且也会降低工作量。

2. 利用中药配伍的方式以药物组合加减进行开方。方剂的功效，既有方药的成分决定，也由证候所决定。中医师会以经典名方作为参考基础，针对患者的不同情况，与其他中药材相结合创造出新的方剂。中药材配伍是有一定的结构框架的。例如石膏加小青龙汤是在小青龙汤的基础上增加石膏来解决口苦、苔黄等热象证候。通过对药物配方进行加减进行开方会更加有针对性和准确性。

参考文献

- [1] 金榜, 梁保丽, 孙树椿. 新型冠状病毒(COVID-19)感染性肺炎现代中医诊疗建议方案与探讨[J]. 世界中医药, 2020, 15(1): 35-46.
- [2] 蔡菲菲, 马立新, 胡元佳, 等. “AI+中医药”: 迈向智慧中医药时代的新途径[J]. 世界科学技术-中医药现代化, 2019.
- [3] 陈建清, 马祯一, 邵建强, 等. 基于信息化的智慧中医服务模式研究与实践[J]. 中国乡村医药, 2020, v. 27(09): 68-69.
- [4] 张治霞, 牛婷立, 王京平, 等. 中医四诊合参辅助诊疗仪的应用[C]. 中华中医药学会中医诊断学分会第十六次中医诊断学术年会论文集. 2015: 60-60.
- [5] 陈智慧, 张会永, 裴宇鹏, 等. 基于中医证候量表探讨建立智能化中医药临床决策与疗效评价模式的设想[J]. 北京中医药大学学报, 2021, 44(3): 221-225.
- [6] W Li, Z Yang. Exploration on Generating Traditional Chinese Medicine Prescriptions from Symptoms with an End-to-End Approach[M]. 2019.
- [7] L Yao, Y Zhang, B Wei, et al. A Topic Modeling Approach for Traditional Chinese Medicine Prescriptions[J]. IEEE Transactions on Knowledge & Data Engineering, 2018:1-1.
- [8] 王宏武, 王峰, 王晓洒, 等. 人工智能在中医诊察中的应用综述[J]. 电脑知识与技术, 2019 (19): 87.
- [9] 周昌乐, 张志枫. 智能中医诊断信息处理技术研究进展与展望[J]. Journal of Integrative Medicine, 2018, 4(6): 560-566.
- [10] 陈敏. “互联网+医疗健康”: 打造智慧医疗服务新模式[J]. 中国党政干部论坛, 2018.
- [11] 洗向阳, 张志强, 李纪麟. 基于推荐算法和主成分分析的智能处方推荐系统[J]. 中国数字医学, 2018, 13(12): 23-24, 67.
- [12] 秦芳. 浅论中西医的区别及“中医西化”的危害性[J]. 中国中医药现代远程教育, 2012, 10(6): 94-96.
- [13] 赵文, 李书楠, 徐佳君, 等. 智能化中医辅助诊疗系统模型构建[J]. 中华中医药杂志, 2020, 35(5): 2421-2424.
- [14] 张腾宇, 张静莎, 徐功铖, 王峥, 张学敏, 李增勇. 融合脑功能和运动评估的脑卒中康复训练处方推荐模型构建[J]. 中国生物医学工程学报, 2021, 40(04): 394-400.
- [15] 闫恩亮. 基于医案和机器学习的中西医结合糖尿病专家系统设计[D]. 燕山大学.
- [16] 张少聪. 中医医疗辅助诊断系统研究与实现[D]. 电子科技大学, 2018.
- [17] 张艺品, 关贝, 吕荫润, 等. 深度学习基础上的中医实体抽取方法研究[J]. 医学信息学杂志, 2019, 40(02): 62-67.

- [18] 谢天宇, 曹继忠, 赵姝婷, 等. 基于大数据深度学习的中医“证候”到“方剂”的新型算法研究[J]. 亚太传统医药, 2018, 014(001):51-53.
- [19] 罗瑞静, 何建成. 中医智能化问诊系统开发及应用前景[J]. 时珍国医国药, 2014(7).
- [20] 张晓航, 石清磊, 王斌, 等. 机器学习算法在中医诊疗中的研究综述[J]. 计算机科学, 2018, 45(S2): 32-36.
- [21] Z. Li, W. Shang, M. Yan. News Text Classification Model Based on Topic Model[C]. IEEE/ACIS International Conference on Computer & Information Science. IEEE, 2016.
- [22] Y. Zhang, Z. Rao. n-BiLSTM: BiLSTM with n-gram Features for Text Classification[C]. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2020.
- [23] M. Gao and J. Li, Chinese Short Text Classification Method Based on Word Embedding and Long Short-Term Memory Neural Network[C]. 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), 2021, pp. 91-95.
- [24] 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[J]. 计算机应用, 2017(z1).
- [25] W. Li, Z. Yang, X. SUN, et al. Exploration on Generating Traditional Chinese Medicine Prescription from Symptoms with an End-to-End method[J]. Eprint Arxiv, 2018.
- [26] 李鹏飞, 鲁法明, 包云霞, 等. 基于医疗过程挖掘与患者体征的药物推荐方法[J]. 计算机集成制造系统, 2020, 26(6): 1668-1678.
- [27] 徐旭程. 基于注意力机制的文本分类研究[D]. 电子科技大学, 2019.
- [28] 覃忠台, 张明军. 基于协同过滤算法的学习资源推荐模型研究[J]. 计算机技术与发展, 2021.
- [29] Rendle S. Factorization Machines[C]. ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010. IEEE, 2010.
- [30] 吴平, 燕海霞, 吕仪, 等. 以“人工智能+中医诊断”辅助新型冠状病毒肺炎防控策略分析[J]. 辽宁中医杂志, 2021, 48(9): 54-56.1719.2021.09.014.
- [31] 王斌, 刘涛, 王广志, 等. 支持新型冠状病毒肺炎的中医智能处方推荐和知识库系统[J]. 中国数字医学, 2020, 15(5): 25-27.
- [32] E. Luo, D. Zhang, H. Luo, et al. Treatment efficacy analysis of traditional Chinese medicine for novel coronavirus pneumonia (COVID-19): an empirical study from Wuhan, Hubei Province, China[J]. Chinese Medicine, 2020, 15(1).
- [33] D. Guo, C. Liu. Rudolf Bauer: China anti-COVID-19 success is closely related with the role TCM played[J]. 2020(4): 2.

- [34] 孙鹏. 慢性胃炎中医治法探析[J]. 江苏中医药, 2020, 52(1): 11-13.
- [35] V. Kaldo, J. Ramner, S. JernelV. Involving Clients in Treatment Methods: A Neglected Interaction in the Therapeutic Relationship[J]. Journal of consulting and clinical psychology, 2015, 83(6): 1136-1141.
- [36] Y. Jin, W. Zhang, X. He, et al. Syndrome-aware Herb Recommendation with Multi-Graph Convolution Network[C]. 2020 IEEE 36th International Conference on Data Engineering (ICDE), 2020, pp. 145-156.
- [37] S. Vangala, R. Vadlamani. ATM Cash demand forecasting in an Indian Bank with chaos and deep learning[J]. 2020.
- [38] C. Ruan, Y. Wu, Y. Yang, et al. Semanticaware Graph Convolutional Networks for Clinical Auxiliary Diagnosis and Treatment of Traditional Chinese Medicine[C]. IEEE Access, vol. PP, no. 99, pp. 1-1, 2021.
- [39] G Zhao, X Zhuang, X Wang, et al. Data-Driven Traditional Chinese Medicine Clinical Herb Modeling and Herb Pair Recommendation[C]. 2018 7th International Conference on Digital Home (ICDH). 2018.
- [40] 杨玉凤, 亢小雨, 李远. 从数据处理中挖掘乳腺癌术后的中医证治规律[J]. 医学研究与教育, 2017, 34(5): 16-23.
- [41] L. Yao, Y. Zhang, B. Wei, et al. Discovering treatment pattern in Traditional Chinese Medicine clinical cases by exploiting supervised topic model and domain knowledge[J]. Journal of Biomedical Informatics, 2015, 58:260-267.
- [42] Y. Kim, Kim Y. Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.
- [43] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [44] Y. Chen, J. Elenee Argentinis, G. Weber. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research[J]. Clinical Therapeutics, 2016, 38(4):688-701.
- [45] 黄璐琦. 中医药防治新型冠状病毒肺炎的优势与挑战[J]. 科学通报, 2021, 66(26): 3370-3371.
- [46] 李红. 基于机器学习的中医药配方评估研究[D]. 南京大学, 2016.
- [47] 管飞诗. 基于深度学习的中医药研究应用[J]. 电子技术与软件工程, 2019(21): 236-237.
- [48] 于兴文. 基于机器学习与数据挖掘的名老中医用药规律研究[D]. 南京中医药大学, 2019.
- [49] 田同良, 位梦霞. 基于数据挖掘方法对周仲瑛治疗糖尿病医案处方用药配伍规律及其病机分析[J]. 中国中医基础医学杂志, 2019, 25(3): 337-339, 352.

- [50] 徐静雯, 夏菁, 邸若虹, 等. 数据挖掘技术在中医药研究中的应用进展[J]. 医学综述, 2019, 25(18): 3672-3676, 3681.
- [51] 李丹溪, 关静, 李峰. 舌诊仪的发展及其在舌诊客观化研究中的应用现状[J]. 世界中医药, 2017 (2): 456-460.
- [52] 王永宏, 刘娟. 中医面部色诊客观化研究概况[J]. 湖南中医学院学报, 2005, 25(1): 59-61.
- [53] L. Deng, D. Yu. (2014). Deep Learning: Methods and Applications[J]. Foundations & Trends in Signal Processing, 2014, 7(3):197-387.
- [54] Eli, Schwartz, Raja, et al. DeepISP: Towards Learning an End-to-End Image Processing Pipeline.[J]. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 2018.
- [55] W. Li, Z. Yang, X. Sun. Exploration on Generating Traditional Chinese Medicine Prescription from Symptoms with an End-to-End method[J]. 2018.
- [56] 章亚东, 胡孔法, 杨涛, 等. 基于复杂网络的中医治疗肺癌的处方推荐算法[J]. 时珍国医国药, 2019, 30(5): 1257-1260.
- [57] 叶豪. 注意力分解机[D]. 浙江: 浙江大学, 2018.
- [58] 罗文奇. 计算机科学在中药信息化建设中的应用探讨[J]. 数字化用户, 2017, 23(24): 98.

致 谢

转眼间，研究生三年的生活即将接近尾声，回首三年来的点点滴滴，老师的鼓励，同门之间的帮助，同学之间的友情，仿佛历历在目。心中有太多的不舍，更心存感激。

首先，向我敬爱的导师张明川老师表达最衷心的感谢，感谢他三年的精心培养和淳淳教导，以及生活中给予的关怀。他平时的言传身教，让我知道对科研地态度有了转变，必须用严谨的态度面对科研，不能存有一丝侥幸心理；老师更教会了我在逆境中成长，面对科研中地困难或者失败，老师总是用生动地比喻简洁明了地告诉我如何去突破，并且总是在适当时刻对我进行鼓励，让我更有勇气面对科研路上的绊脚石。也感谢实验室所有的在我学习、论文完成期间帮助过我的老师，他们为我提供了一个专心学术的平台，带领我们进行项目的实战，为我们今后的工作奠定基础。

感谢我实验室的同门师兄姐妹，在这三年中实验室几乎所有事情都有他们的参与，几乎所有的麻烦他们也都会热情帮助，他们更像我的家人，我们一起学习、一起聚餐畅谈未来，分享彼此之间快乐。研究生的生活有你们而变得更加充实，这段旅程虽然结束，但是也代表着一个新的开始，希望在未来的生活中，大家能够生活顺利、友谊长存。

感谢我的室友，相遇就是缘分在我最美好的年华与你们相遇。从素未谋面的陌生人到无话不谈的好朋友，从客气梳理到亲密无间的学习、生活，我们一起度过了很多美好的时光，感谢这一路的相伴让我的校园生活不在孤单和落寞，我们一起吃饭上课回宿舍，一起吃喝玩乐、分享最新的电影，通宵聊天……。我们在同一个屋檐下，一起憧憬未来的诗篇。名虽各性，情同手足。我会珍惜我们这来之不易的室友情，希望未来的生活依旧有你们。

最后特别感谢我的父母，他们无私的爱就像一盏永不熄灭的灯，永远指引着我不断前进和奋斗。每次遇到困难或者心情不好时，跟他们倾诉一下，听听他们的声音，烦恼就会消失很多，听完他们的疏导，心中就会再次充满斗志。是他们的支持和鼓励，才让我没有后顾之忧地专注学习。羊有跪乳之情，鸦有反哺之意，今后就让我来为你们撑起一片天。

攻读学位期间的研究成果

一、发表论文

- [1] D. Liu, Y. Zhang, L. Wang, et al. "A TCM Therapeutic Methods Based on Text Convolutional Neural Network", 2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2021, pp. 751-756.

二、攻读硕士学位期间参加的科研项目

- [1] 基于图像处理的智能中医望诊系统的关键技术，省部级，参与人员。

三、攻读硕士学位期间所获奖励

- [1] 河南科技大学二等奖学金，2019 年 11 月。
[2] 河南科技大学二等奖学金，2020 年 11 月。
[3] 河南科技大学一等奖学金，2021 年 11 月。