

编号：____

硕士学位论文

中长期径流智能预报技术 研究与应用

申请学位级别：____工学硕士____

专业名称：____信号与信息处理____

研究方向：____中长期径流智能预报____

论文提交日期：____2018年3月30日____

摘要

准确的中长期径流预报对水资源的综合开发利用、科学管理和优化调度具有重要意义。但是由于水文过程复杂，且受客观认知能力和科学技术水平的限制，形成中长期径流的物理机制还没有完全弄清楚。因此，目前基于物理成因分析的预报方法实施起来困难，且不能适用于所有流域，而常用的基于统计学的方法对于预报模型研究较多，对预报因子的筛选方法研究相对较少。基于此，本文在对国内外中长期径流预报方法研究的基础上，提出了互信息和核主成分分析驱动的神经网络径流预报模型，并开发了相应的系统，为中长期径流预报提供一定的参考依据。主要的研究内容及成果如下：

（1）从神经网络的拓扑结构、神经网络的训练过程及神经网络的适用范围三个方面分别介绍了标准神经网络和循环神经网络，为循环神经网络作为中长期径流预报模型提供理论依据。

（2）提出了标准互信息和核主成分分析驱动的神经网络径流预报模型。首先对预报因子的初选技术、初选因子的特征提取技术以及中长期径流智能预报模型的构建技术进行了介绍，然后对不同的智能预报技术体系进行了分析，最终确定以“基于标准互信息的预报因子的初选→基于核主成分分析的初选因子的特征提取→基于单个循环神经网络的径流量预报→基于多个循环神经网络的径流量集合预报”为中长期径流智能预报技术体系，该体系能够有效地挖掘预报因子之间复杂的相互关系，并能够对筛选的因子进行降噪和去冗余，以这些因子作为神经网络的输入，能有效地缓解神经网络的过拟合现象，同时，多个神经网络的集合预报在一定程度上减少了神经网络模型参数的不确定性对中长期径流预报造成的不利影响。

（3）开发了中长期径流智能预报系统。首先建立了中长期径流智能预报系统的数据库。采用 SQL Server 2008 为数据库管理系统，对数据库的各种表结构进行了设计，实现了对中长期径流预报相关的各种信息的存储和管理，为中长期径流预报提供数据支持。接着设计了中长期径流智能预报系统的界面。主要包括预报方案制作界面、预报方案管理界面、中长期径流预报界面以及中长期径流预报管理界面。其中预报方案制作界面是对方案基本信息、预报对象、预报模型、预

报因子以及模型参数进行设置，完成模型的参数率定，并对训练好的模型进行试预报，检验模型的预报效果。预报方案管理界面可以对制作的预报方案进行统一管理。中长期径流预报界面可以采用制作好的方案对径流量进行预报。中长期径流预报管理界面可以对预报的结果进行统一管理。

(4)利用中长期径流预报系统对锦屏一级水电站的年平均径流量进行预报。首先利用“预报因子初选→基于互信息的预报因子的二次筛选→基于核主成分分析的因子特征提取”的方法对预报因子进行遴选，然后构建基于循环神经网络的集合预报模型，实现对锦屏一级水电站年平均径流量的预报。结果表明，本文提出的基于标准互信息和核主成分分析驱动的神经网络径流预报模型的预报稳定性和精确度较高，能够满足实际生产的需求，具有一定的参考价值。

关键词：中长期径流智能预报；技术体系；预报系统；集合预报；锦屏一级水电站

Abstract

Accurate medium-and-long term runoff forecasting is of great significance to the comprehensive development, scientific management and optimal dispatching of water resources. But the hydrological process is complex. Because of the limitation of objective cognitive ability and technical level, the physical mechanism to form medium-and-long term runoff has not been fully understood. Therefore, the present prediction method based on physical cause analysis is difficult to implement and can not be applied to all basins. In this paper, the medium-and-long term runoff forecasting methods are studied, and then based on the preliminary selection of factors, the feature extraction of factors, the construction of prediction model based on recurrent neural network, and the ensemble forecasting model based on recurrent neural network, the system of medium-and-long term runoff intelligent forecasting is established and the corresponding systems are developed. This provides some reference for the medium-and-long term runoff forecast. The main research contents and results are as follows:

(1) The standard neural network and the recurrent neural network are introduced from the network topology structure, the training process of the network and the applicable range of the network, which provides the theory support for the long-term-runoff forecast model.

(2) The technology system of medium-and-long term runoff intelligent forecasting is established. Firstly, this paper introduces the primary technique of the forecast factor, the feature extraction technique of primary factor and the construction technology of the medium-and-long term runoff intelligent forecast model, and then analyzes the different intelligent forecasting technology system, and finally determines the " factor screening based on standard mutual information → feature extraction of factors based on kernel principal component analysis (KPCA) → prediction model based on recurrent neural network → ensemble prediction model based on multiple recurrent neural networks " as an intelligent forecasting system for medium-and-long term runoff. The system can effectively excavate the complex relationship between the forecast factor

and the runoff, improve the stability and accuracy of the forecast, and provide the reference for the medium and long term runoff forecast.

(3) The medium-and-long term runoff intelligent forecasting system is developed. First, a database of medium-and-long term runoff intelligent forecasting system is established. Using SQL Server 2008 as database management system, various table structures of database are designed to realize the storage and management of various information related to medium-and-long term runoff forecast, and provide data support for medium-and-long term runoff forecast. Secondly, the interface of the medium-and-long term runoff intelligent forecasting system is designed. It mainly includes the interface of making forecast scheme, the interface of management forecast scheme, the interface of runoff forecast and the interface of managing runoff forecast. The interface of making forecast scheme is to set up the basic information, the forecast object, the forecast model, the forecast factor and the model parameter, and then to forecast the runoff with the model and to check the prediction effect of the model. The forecast scheme management interface can manage the forecast scheme. The runoff prediction interface can be used to forecast the runoff by using the scheme already produced. The runoff forecast management interface can manage the results of the forecast.

(4) Using the medium-and-long term runoff forecast system to forecast the annual average runoff of Jinping I hydropower station. Firstly, the factors are selected, then the factors are further screened by using the mutual information method, then the factor features is extracted by KPCA, and the annual average runoff of Jinping I hydropower station is forecasted by using multiple cyclic neural networks. The results show that the forecast stability and accuracy of the medium-and-long term runoff intelligent forecasting system is high, which can meet the demand of actual production and has some reference value.

Key Words: mid-and-long term runoff forecasting; technology system; forecasting system ;ensemble forecast; Jinping I Hydropower Station

目录

摘 要.....	I
Abstract	III
第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	1
1.2 研究现状及存在的问题.....	2
1.2.1 研究现状.....	2
1.2.2 存在的问题.....	5
1.3 研究内容与技术路线.....	6
1.3.1 研究内容.....	6
1.3.2 技术路线.....	7
1.4 本章小结.....	8
第二章 人工神经网络理论.....	9
2.1 引言.....	9
2.2 标准神经网络.....	9
2.2.1 标准神经网络的拓扑结构.....	9
2.2.2 标准神经网络的前向传播过程.....	10
2.2.3 计算标准神经网络的成本函数.....	12
2.2.4 标准神经网络的反向传播过程.....	13
2.2.5 标准神经网络的参数更新算法.....	14
2.2.6 标准神经网络的适用范围.....	17
2.3 循环神经网络.....	17
2.3.1 循环神经网络的拓扑结构.....	17
2.3.2 循环神经网络的前向传播过程.....	18
2.3.3 计算循环神经网络的成本函数.....	19
2.3.4 循环神经网络的反向传播过程.....	20
2.3.5 循环神经网络的参数更新算法.....	20
2.3.6 循环神经网络的适用范围.....	20
2.4 本章小结.....	21
第三章 中长期径流智能预报技术体系研究.....	22
3.1 引言.....	22

3.2 预报因子的初选技术.....	22
3.2.1 物理机制分析.....	22
3.2.2 Pearson 相关系数法.....	24
3.2.3 Spearman 等级相关系数法.....	24
3.2.4 互信息.....	25
3.3 初选因子的特征提取技术.....	27
3.3.1 基于主成分分析的特征提取.....	27
3.3.2 基于核主成分分析的特征提取.....	28
3.4 中长期径流智能预报模型的构建技术.....	30
3.4.1 循环神经网络结构的确定.....	30
3.4.2 循环神经网络成本函数及正则化方法的确定.....	31
3.4.3 循环神经网络参数更新算法的确定.....	32
3.4.4 基于循环神经网络的集合预报模型构建.....	33
3.5 智能预报技术体系的分析及选取.....	33
3.6 预报结果的精度评价方法.....	35
3.7 本章小结.....	36
第四章 中长期径流智能预报系统设计与实现.....	37
4.1 系统总体架构设计思路.....	37
4.2 数据库的设计.....	37
4.2.1 数据库的设计原则.....	38
4.2.2 数据库的总体架构.....	38
4.2.3 数据库的表结构设计.....	39
4.3 预报方案库的设计.....	45
4.4 预报模型库的设计.....	46
4.5 人机交互平台设计.....	47
4.5.1 人机交互平台设计原则.....	48
4.5.2 中长期径流智能预报系统界面设计.....	48
4.6 本章小结.....	51
第五章 中长期径流智能预报系统应用实例.....	52
5.1 研究区域概况.....	52
5.2 预报对象与数据来源.....	52
5.3 实验数据的预处理.....	52
5.4 预报因子的遴选.....	53
5.4.1 基于物理机制的因子初选.....	53

5.4.2 基于互信息的因子二次筛选	53
5.4.3 基于 KPCA 的因子特征提取	54
5.5 预报模型的构建	55
5.5.1 成本函数、正则化方法、参数更新算法的确定	55
5.5.2 网络结构的确定	56
5.6 预报结果分析	59
5.7 本章小结	60
第六章 总结与展望	61
6.1 主要结论	61
6.2 展望	61
参考文献	63

第一章 绪论

1.1 研究背景及意义

1.1.1 研究背景

水是人类的生命之源，人类生产实践活动离不开水。但是我国的水资源储量并不丰富，并且呈现出东多西少、南多北少的时空分布不均性。为此流域内各部门希望水文部门能提供比较准确水文预报，为人类生产实践活动提供指导。

水文预报是指根据水文要素或其影响因素的过去或现在状态，对其未来状态做出回答^[1]。水文预报有不同的分类方法。如果按预见期可分为短期水文预报和中长期水文预报。一般以流域汇流时间为界，凡预报的预见期小于或等于流域汇流时间的称为短期预报，否则称为中长期预报^[2]。具体来说，一般认为预见期为数小时至数天的为短期预报；预见期在 2~5 天以上，10~15 天以内的为中期预报；预见期在 15 天以上，1 年之内的为长期预报；预见期在一年以上的称超长期预报^[3]。如果按照预报的结果形式可分为定量预报或者定性预报。定量预报是指预报的结果需要提供具体的数值，定性预报是指预报结果需要提供定性的评价，如偏枯、偏丰，或提供等级预报，如 I ~ V 级。

鉴于中长期水文预报的预见期较长，所以较高精度的定量预报能够使人们尽早采取措施进行统筹安排，争取获得水资源开发利用的最大综合效益^[4]。为此，研究者们提出了很多的预报方法，包括传统的物理成因分析法、数量统计法以及人工智能的方法。但是，由于水文系统是一个开放的、动态的非线性系统，它不仅与气候、地形等自然因素相互作用、相互依赖，而且会受到不同程度的当地人类生产活动的影响，从而形成了其复杂的演化规律，为中长期径流的准确预报带来了很大的困难^[5]。这也导致目前各种中长期径流预报模型的预报精度还有待进一步提高，同一预报模型对不同地区的径流量预报的通用性还差强人意。但是，随着国民经济的发展以及生产实践的需求，中长期水文预报的研究也会越来越受到学者和有关部门的重视。

1.1.2 研究意义

准确的中长期径流预报是水资源综合利用、科学管理和优化调度的重要

前提,对防洪、航运、水环境以及国民经济的可持续发展等方面也有重要的积极影响^[6]。但是,目前的中长期径流预报模型与方法不能适用于所有流域,常用的物理成因分析法、统计学方法的预报精度也需要进一步提高。本文针对这些问题,提出了基于标准互信息和核主成分分析驱动的神经网络径流预报模型,为中长期径流预报提供一定的参考依据。

由于中长期径流预报方法复杂,实施起来比较困,对预报人员的专业性要求较高。开发中长期径流预报系统,能够使预报人员在不掌握预报模型底层细节的情况下,通过友好的人机交互界面,完成中长期径流预报工作,提高效率。

1.2 研究现状及存在的问题

1.2.1 研究现状

随着国民经济的发展和生产实践的需要,中长期水文预报也越来越受到世界各国研究者和有关部门的重视,形成了一系列的预报方法。这些方法大概可以分为三大类。第一类是物理成因分析法,第二类是基于统计学的方法,包括传统统计方法和现代统计方法,第三类是统计分析与水文模型耦合的预报方法^[7]。下面从这三类方法来介绍国内外的研究现状。

(1) 物理成因分析法

对于中长期径流预报而言,物理成因分析法是指对影响中长期径流形成过程的物理因子进行分析的一种方法。分析形成中长期径流的物理因素是水文学研究的一个重要方向^[8]。在 20 世纪初,国外提出了“世界天气法”的中长期水文预报模型,并将其应用于尼罗河下游的春汛洪水预报,随后在世界其他一些国家也有广泛应用^[9]。Hamlet、Whitaker^{[10][11]}运用物理成因分析方法研究了尼尔尼诺和太平洋 10 年涛动现象与北美哥伦比亚河流的旱涝关系、南方涛动与印度北部地区恒河的流量关系。在 50 年代,长江科学研究所东亚大气环流的前后期演变规律为依据,对长江做了中长期径流预报^[12]。章淹^[13]从水文气象学角度论述了近年国内在暴雨中期预报研究方面的若干新进展,探索了中期预报的成因分析方法等。李永康等^[14]分析了长江中下游梅雨期和夏季旱涝的一般特征,在此基础上着重研究了大涝(旱)和特大涝(旱)年前期大气环流的各种因子的特征,进而确定了可供预测旱涝趋势的若干环流因子。刘清仁^[15]以太阳活动为中心,

以长期和超长期水文预报为目标,用数理统计分析方法,分析了太阳黑子、厄尔尼诺事件对松花江区域水文影响特征及其水、旱灾害发生的基本规律。黄忠恕等^[16]分析了北太平洋和青藏高原下垫面热状况变化与长江流域汛期旱涝之间的关系,初步揭示出一些相关性。

总之,影响中长期径流的因子归纳起来主要包括大气环流、太阳活动、人类活动、下垫面情况(如海洋表层水温、陆地表面状况的异常等)、其他天文以及地球物理因素(如地球自转速度的变化、火山爆发、行星相对位置等)。这些因素之间相互作用、相互依赖,使得中长期径流的物理形成过程错综复杂。因此,物理成因分析法虽然有一定的物理意义,但是由于目前的技术水平的限制,这些水文过程还没有完全弄清楚,影响了物理成因分析法的预报精度,并且该方法在实际运用中,实施起来比较困难,限制了其应用与推广。

(2) 统计学方法

统计学方法是一种寻找预报因子与预报对象之间的关系或寻找水文要素自身历史变化的统计规律,进而实现预报的一种方法。一般可以分为传统的统计学方法和现代的统计学方法。

传统的统计学方法是应用比较广泛的一种中长期水文预报方法,它从大量的历史资料中应用数理统计的理论和方法,寻找预报对象和预报因子之间的统计规律或水文要素自身历史变化的统计规律,建立预报模型进行预报^[17]。通过寻找预报对象和预报因子之间的统计规律进行预报的方法典型代表是回归分析法,如多元回归、逐步回归、门限回归等。通过分析水文要素自身历史变化的统计规律进行预报的方法的典型代表是时间序列法,如自回归滑动平均法、马尔科夫转移法、周期分析法等。除了回归分析法和时间序列法,传统的统计学方法还包括各种聚类(如系统聚类法)和判别方法(如贝叶斯判别)。葛朝霞等^[17]利用长江宜昌站年平均流量资料,运用多因子逐步回归周期分析法,建立了中长期水文预报模型,实例应用结果表明,该方法具有较高的可信度。Box 和 Jenkins^[18]于 1970 年提出了时间序列 ARMA 模型,此后该方法逐步被应用于中长期径流预报^[19]。张铭^[20]等将贝叶斯判别应用在中长期径流预报中,指出以预报概率分布期望作为预报值,比相应的确定性预报具有更高的精确度。

近年来,随着计算机计算能力的提高、各种新的算法的涌现以及可用的数据量的增长,研究者们将一系列新的现代统计学方法应用于中长期径流预报。主要包括模糊分析、灰色系统理论、小波分析、混沌分析、支持向量机以及当下发展迅猛的人工神经网络。陈守煜^{[21][22][23]}教授于 20 世纪 80 年代将模糊数学应用到水文水资源领域,逐渐形成了模糊水文学体系^[24]。但是因为模糊分析方法包含的信息带有明显的主观性,所以在生产实践中的应用受到了一定的限制。邓聚龙教授于 1982 年创立了灰色系统理论^[25]。蓝永超等^[26]利用河西地区典型河流的径流观测资料,基于灰色系统的思想理论,建立了残差序列周期修正的 GM(1,1)径流预测模型,并对河流的来水量进行了长期预报,预报结果较好。灰色系统理论能够很好的解决呈指数趋势增长的问题,但是水文径流序列本身具有非线性、随机的特点,将灰色系统理论模型运用于中长期径流预报具有一定的局限性。Morlet 在 1980 年将小波分析用于水文时间序列分析中,该方法能够很好的分析序列周期和趋势。刘晓安等^[27]将一种基于小波分析的自回归滑动平均求和模型用于月径流的预测,预报效果较好。但是,小波分析存在的尺度水平选择等难题,其分析理论和方法还不太完善,需要进一步的研究^[28]。混沌理论分析法把对水文时间序列的确定性分析和随机性分析结合起来,为水文预测工作开辟了新的途径。Sivakumar 基于确定性的混沌理论建立了二阶局域预测模型,并应用于巴西 Araguari 流域的月径流预测,结果验证了模型的可行性,也得出该流域的月径流动力系统具有低维混沌特性的结论。但是就水文预报而言,由于混沌吸引子对初始条件的敏感性,使得水文预测从短期到长期,相轨道的发散越来越大,信息损失越来越多,因此,混沌分析法对短期预测效果较好,对中长期径流的预测效果有待提高^[6]。Vapnik 于 20 世纪 90 年代中期基于统计学习理论,以结构化风险理论最小为原则,提出了支持向量机^[29]。Liong 和 Sivapragasm^[30]将 SVM 应用于以日时段、遇见期为 7 天洪水水位预报,取得了较好的预报精度。20 世纪 40 年代,人工神经网络理论被提出。在水文预报领域,基于神经网络的智能预报方法也到了广泛的应用。Kang 等^[31]将 ANN 和 ARMA 模型应用于韩国 Pyung Chang 河流域的以日和小时为时段的径流预测,并测试了不同的三层 ANN 网络结构,证实了 ANN 是一种有效的径流预报方法。屈亚玲、周建中等^[32]提出了改进的采用局

部回归的 Elman 神经网络,并将该方法应用于凤滩水库的径流预报中,通过与回归分析和 BP 神经网络比较,表明了所提方法的有效性与优越性。近年来,随着人工神经网络的基本理论的完善,计算机计算能力的加强以及各种新的算法的提出,使得人工神经网络在模式识别、智能机器人、自动控制、预测估计、生物、医学、经济等领域得到了广泛的应用,成功地解决了许多现代计算机难以解决的实际问题,并表现出了良好的智能特性。人工神经网络具有很强的鲁棒性、记忆能力、非线性映射能力以及强大的自学习能力,能够很好的挖掘预报因子和径流量之间复杂的关系。因此将人工神经网络应用于中长期径流预报,是一种不错的选择。

(3) 组合预测方法

Bates 和 Granger 于 1969 年提出最优组合预测的方法。他们认为,人们常在选择一个较优预测值的同时舍弃了另外的预测值是很不明智的,因为舍弃的预测值一般都蕴含某些有用的信息;另一方面,任何时间序列模型的参数都不可能得到准确的识别,不同模型组合往往能得到较好的预测值^[33]。曹启辉等^[34]基于小波分析理论和人工神经网络各自的优点和适用条件,提出了一种新型的小波网络组合模型,以长江三峡水电站年平均流量为例,运用此组合模型进行了研究并得到了比较满意的结果。因此,对于中长期径流预报,除对方法本身进行改进外,采取两种及以上的方法藕合,在各种方法之间取长补短,能够有效提高预报的精度。

1.2.2 存在的问题

根据预报方法的类别,分别从物理成因分析法存在的问题、统计学预报方法存在的问题两方面来介绍目前中长期径流预报方法存在的问题。

(1) 物理成因分析法存在的问题

基于物理成因分析的中长期径流预报方法,需要弄清楚中长期径流的物理形成过程,但是影响中长期径流形成过程的因素众多,基于目前的技术水平,这些水文过程还不能完全弄清楚,影响了物理成因分析法预报的精度,并且不同地域的中长期径流形成过程可能会有差异,这导致物理成因分析法在实际应用过程中实施起来比较困难,限制了其发展与推广。

(2) 统计学预报方法存在的问题

现行的基于统计学的中长期径流预报方法存在以下三方面的问题：

第一，常用的用于因子初选的相关分析方法不能反应径流量与预报因子之间的非线性关系。水文过程复杂，预报因子与预报对象之间除了线性关系外还存在非线性关系。目前常用的预报因子初选的方法是线性相关分析法（如 Pearson 相关分析、Spearman 相关分析等），即通过计算气象水文资料（包括遥相关因子、本地相关因子等）与历史径流序列之间的相关系数，选出相关系数高的因子的方法。但是这些方法大多是线性相关分析法，该方法只能描述变量间的线性关系，不能反映变量间的非线性关系。

第二，常用的用于初选因子特征提取的主成分分析法（PCA，Principal Component Analysis）忽略了预报因子之间高于 2 阶的相互关系^[35]，所以抽取的主成分并不是最优的。由于用相关性分析方法筛选因子时，筛选出的高相关性的因子往往较多，且不同的因子时间序列之间存在较高的复共线性^[36]，因子时间序列本身也存在一定的噪声。因此，需要对初选出来的因子进行降噪和去冗余。目前常用的方法是主成分分析法。主成分分析法能用几个较少的综合指标来代替原来较多的指标，而这些较少的综合指标不但能尽多地反映原来较多指标的有用信息，且相互之间正交。但是，用于初选因子降噪和去冗余的 PCA，实质是一种线性映射方法，得到的主成分是由线性映射生成的，这种方法忽略了数据之间高于 2 阶的相互关系^[35]，所以抽取的主成分并不是最优的。

第三，采用的预报模型是线性模型，或者使用的是人工神经网络，但没有考虑模型参数不确定对预报结果造成的影响。对于预报对象与预报因子之间的最优数学关系的建立，目前常用的模型有多元回归、随机森林、人工神经网络、支持向量机等。常用的多元回归实际上也是一种线性拟合，不能反映预报对象与预报因子之间的非线性关系。与其他模型相比，人工神经网络因鲁棒性好、非线性映射和自学习能力强，在中长期径流预报中得到了较为广泛的应用。但是单个神经网络模型的参数存在一定的不确定性，这会造成每次预报的结果之间存在一定的差异，影响预报的精度。

1.3 研究内容与技术路线

1.3.1 研究内容

为了解决目前中长期径流预报方法存在的问题,本研究探讨的内容主要包括以下几个方面:

(1) 分析中长期径流智能预报技术体系,提出一种预报对象与预报因子之间的非线性系统,提高预报的精度。具体包括预报因子的遴选技术、特征提取技术的分析与选取,人工神经网络模型的分析与构建。其中,人工神经网络模型的构建包括人工神经网络种类的确定、网络训练算法的分析选取、网络最优结构的确定以及多模型集合预报。

(2) 设计中长期径流智能预报系统,降低对预报人员的要求。该系统主要包括四个部分的设计,即数据库模块的设计、预报方案库模块的设计、预报模型库的设计以及人机交互界面的设计。其中人机交互界面由四个子界面组成,具体包括预报方案制作界面、预报方案管理界面、径流预报参数设置界面以及预报结果查询界面。

1.3.2 技术路线

根据以上的研究内容,本研究首先在第一章介绍了研究的背景与意义,以及目前的研究现状与存在的问题。接着,在第二章介绍了人工神经网络的基本理论,为第三章的中长期径流智能预报技术体系的构建提供理论支持。第三章详细探讨了中长期径流预报因子的筛选、对筛选因子的特征提取以及神经网络模型的构建。第四章详细探讨了中长期径流智能预报系统的设计技术。第五章以锦屏一级水电站水库年入库流量的预报为例,验证了模型的有效性 & 预报系统的可行性。最后在第六章对本研究工作进行了总结,并阐述了研究的不足及展望。具体的技术路线图如图 1-1 所示。

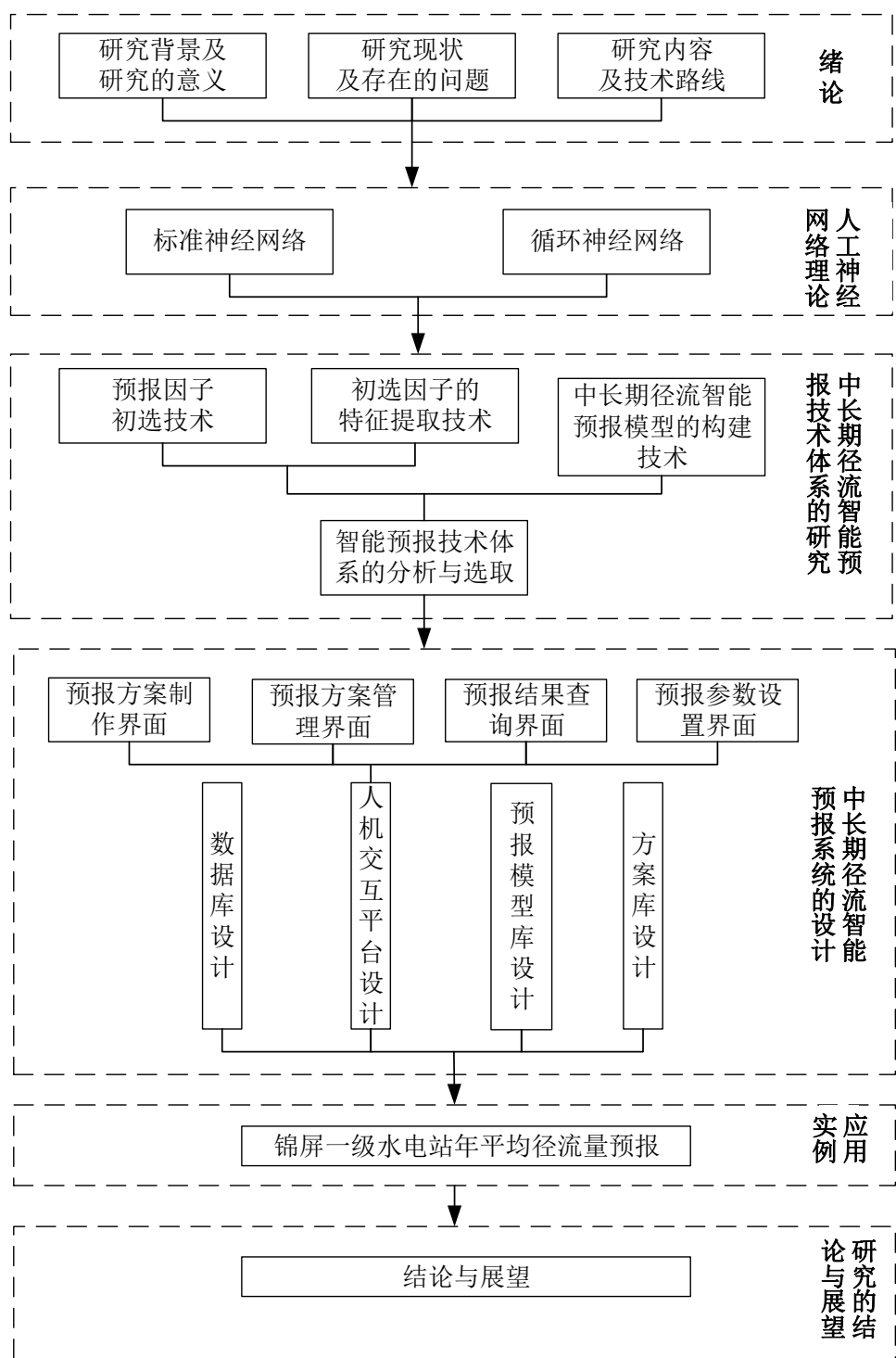


图 1-1 技术路线

1.4 本章小结

本章首先介绍了研究背景及意义，然后根据研究现状指出了目前常用的中长期径流预报方法存在的一些问题。最后介绍了本研究的主要内容及技术路线。

第二章 人工神经网络理论

2.1 引言

近年来,随着计算机计算能力的增强、数据量的增加以及各种新算法的提出,以人工神经网络为基础的人工智能快速崛起,人工神经网络得到了极大地发展。与其他模型相比,人工神经网络因鲁棒性好、非线性映射和自学习能力强,在中长期径流预报中也得到了广泛的应用。人工神经网络总类繁多,一般来说可以分为 3 大类。第一类是标准神经网络 (Standard Neural Networks),第二类是循环神经网络 (RNN , Recurrent Neural Networks),第三类是卷积神经网络 (CNN , Convolutional Neural Networks)。其中,循环神经网络特别适合时间序列数据 (如径流时间序列) 的处理,而卷积神经网络主要用于计算机视觉领域,不适合用来处理时间序列数据。因次,本章将从网络的拓扑结构、网络的训练过程及网络的适用范围三方面来着重介绍标准神经网络和循环神经网络。其中,神经网络的一次训练过程包含前向传播 (Forward Propagation)、计算成本函数 (Cost Function)、反向传播 (Backward Propagation) 及参数更新四个阶段。下面分别进行介绍。

2.2 标准神经网络

2.2.1 标准神经网络的拓扑结构

标准神经网络是人工神经网络中最基本的一类神经网络,其典型代表是 BP 神经网络^[37]。图 2-2 是标准神经网络的拓扑结构,包含输入层 (Input Layer)、隐含层 (Hidden Layer) 以及输出层 (Output Layer)。

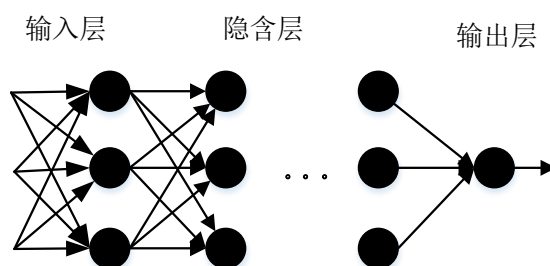


图 2-2 标准神经网络的拓扑结构

图 2-2 是标准神经网络拓扑结构的示意图,这种神经网络的输入层、输出层、隐含层的层数及各隐含层的节点数需要根据实际的应用来确定。对于中长期径流

预报来说，这种神经网络的输入层节点数等于预报因子的个数；输出层节点数等于预报对象的个数；隐含层的层数根据预报因子时间序列的长度确定，一般选择1~3层。因为如果神经网络的隐含层数目过大，则神经网络需要训练的参数会增加很多，而用于训练神经网络的因子时间序列长度有限，这会很容易导致神经网络出现过拟合现象；对于各隐含层的节点数，除了一些经验公式外，目前还没有一个系统和标准的方法来确定^[38]，但是其对神经网络的泛化性能有重要影响。本研究将采用试错法，即通过采用不同输入层节点数和隐含层节点数的组合，观察神经网络的预报效果，从而确定输入层节点和隐含层节点的数目。

2.2.2 标准神经网络的前向传播过程

(1) 前向传播过程的数学描述

标准神经网络的前向传播过程是数据从输入层输入，经过隐含层，最后从输出层输出的过程。前向传播计算过程的数学描述如下：

$$\begin{aligned}
 A^{[0]} &= X \\
 Z^{[1]} &= W^{[1]} A^{[0]} + b^{[1]} \\
 A^{[1]} &= g^{[1]}(Z^{[1]}) \\
 Z^{[2]} &= W^{[2]} A^{[1]} + b^{[2]} \\
 A^{[2]} &= g^{[2]}(Z^{[2]}) \\
 &\vdots \\
 Z^{[L]} &= W^{[L]} A^{[L-1]} + b^{[L]} \\
 A^{[L]} &= g^{[L]}(Z^{[L]}) = \hat{Y}
 \end{aligned} \tag{2-1}$$

其中， X 表示神经网络的输入数据，对于中长期径流预报来说， X 表示输入的预报因子； $W^{[1]}$ 、 $W^{[2]}$ 及 $W^{[L]}$ 分别表示神经网络第一、第二及第 L 层的权值； $b^{[1]}$ 、 $b^{[2]}$ 及 $b^{[L]}$ 分别表示神经网络第一、第二及第 L 层的偏置； $Z^{[1]}$ 、 $Z^{[2]}$ 及 $Z^{[L]}$ 分别表示神经网络第一、第二及第 L 层的中间计算结果； $g^{[1]}$ 、 $g^{[2]}$ 及 $g^{[L]}$ 分别表示神经网络第一、第二及第 L 层的激活函数； $A^{[1]}$ 、 $A^{[2]}$ 及 $A^{[L]}$ 分别表示神经网络第一、第二及第 L 层的激活值。 \hat{Y} 表示神经网络的输出，对于中长期径流预报来说， \hat{Y} 表示径流量的预报值。

(2) 激活函数

对于神经网络的激活函数，常用的有四种，即 Sigmoid 函数、tanh 函数、ReLU 函数及 Leaky ReLU 函数。

第一种：Sigmoid 函数。Sigmoid 函数的表达式为 $Sigmoid(z) = \frac{1}{1+e^{-z}}$ 。该函数的输出值在 0 到 1 之间。对于二元分类问题，神经网络的输出层激活函数一般选用该函数。因为该函数的输出值为 0 或者 1，而 0 和 1 正好可以代表是或者不是。但是对于神经网络其他层的激活函数，现在一般不使用 Sigmoid 函数了。

第二种：tanh 函数。该函数的表达式为 $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ 。tanh 函数又叫双曲正切函数。对于隐含层的激活函数，选 tanh 函数总比 Sigmoid 函数效果要好。因为 tanh 函数的输出值在 -1 到 1 之间，激活函数输出值的平均值就更接近 0，这让 tanh 函数有类似数据中心化的效果，使得网络的训练算法会快很多。

第三种：ReLU 函数。该函数的表达式为 $ReLU(z) = \max(0, z)$ ，ReLU 函数又叫修正线性单元。与 ReLU 函数相比，Sigmoid 函数和 tanh 函数都有一个缺点，就是当 z 非常大或非常小时，Sigmoid 函数和 tanh 函数的导数就可能很小，甚至接近于 0，这会使得在反向传播过程中更新参数时，梯度下降算法收敛的很慢。然而对于 ReLU 函数，只要 z 为正，其导数就是 1，当 z 为负时，其倒数为 0。即对于很多 z 空间，ReLU 函数的导数和 0 差很远，这使得神经网络的学习速度通常会比使用 tanh 函数或 Sigmoid 函数快得多。

第四种：Leaky ReLU 函数。该函数的表达式为 $LeakyReLU(z) = \max(0.01z, z)$ 。Leaky ReLU 函数是 ReLU 函数的修正版本。ReLU 的一个缺点，即当 z 为负时，导数等于零。而对于 Leaky ReLU 函数，当 z 为负时，导数不再为 0，其有一个很平缓的斜率，使得导数为 0.01。这使得在神经网络的实际使用过程中，Leaky ReLU 函数通常比 ReLU 函数表现的更好。

总之，对于激活函数的选择，如果是二元分类问题，则神经网络输出层的激活函数选 Sigmoid 函数，隐含层的激活函数都选用 ReLU 函数或者 Leaky ReLU 函数。而对于非二元分类问题，ReLU 函数或者 Leaky ReLU 函数通常是神经网络

络隐含层激活函数最好的选择。

2.2.3 计算标准神经网络的成本函数

(1) 成本函数

标准神经网络训练的过程就是通过不断的减小成本函数 (Cost Function)，寻找最优参数的过程。成本函数有时又称损失函数 (Loss Function)，两者没有明显的区分。一般来说，损失函数是针对单个样本的， m 个样本的损失函数的平均称为成本函数。对于损失函数 L 或成本函数 J ，根据实际应用可以有不同的选择，如均方误差函数 (Mean Squared Error)、对数损失函数 (Binary Crossentropy) 等。下面介绍一种常用的损失函数 L 和成本函数 J ，其表达式分别如下：

成本函数 J 的表达式：

$$J = \frac{1}{m} \sum_{i=1}^m L^{(i)} \quad (2-2)$$

损失函数 L 的表达式：

$$L \in [L^{(1)}, L^{(2)}, \dots, L^{(i)}, \dots, L^{(m)}]$$
$$L^{(i)} = -\left[y^{(i)} \log(a^{[L](i)}) + (1 - y^{(i)}) \log(1 - a^{[L](i)}) \right] \quad (2-3)$$

其中， i 表示第 i 个训练样本， m 表示样本个数， $y^{(i)}$ 表示第 i 个样本的实测值；

$a^{[L](i)}$ 表示第 L 层第 i 个样本对应的激活值。

(2) 正则化方法

正则化方法 (Regularization Methods) 有利于避免神经网络的过拟合。常用的正则化方法有 $L1$ 正则化、 $L2$ 正则化、Dropout 正则化、数据扩增 (Data Augmentation) 以及提前停止训练 (Early Stopping)。下面分别进行介绍。

① $L1$ 正则化： $L1$ 正则化是指在损失函数的后面加上 $L1$ 正则项，即

$$J = \frac{1}{m} \sum_{i=1}^m L^{(i)} + \frac{\lambda}{2m} \sum_{l=1}^L \sum_{i=1}^{n_x} |w_i^{[l]}| = \frac{1}{m} \sum_{i=1}^m L^{(i)} + \frac{\lambda}{2m} \sum_{l=1}^L \|W^{[l]}\|_1 \quad (2-4)$$

其中， n_x 为权值 W 所在层的神经元的个数； m 为样本数； λ 为正则化参数； $\|W^{[l]}\|_1$ 表示第 l 层的 W 的 $L1$ 范数，即矩阵 W 中所有元素绝对值求和。在实际应用中，

$L1$ 正则化用的比较少，人们更倾向于 $L2$ 正则化。

② $L2$ 正则化: $L2$ 正则化是指在损失函数的后面加上 $L2$ 正则项，即

$$J = \frac{1}{m} \sum_{i=1}^m L^{(i)} + \frac{\lambda}{2m} \sum_{l=1}^L \sum_{i=1}^{n_x} (w_i^{[l]})^2 = \frac{1}{m} \sum_{i=1}^m L^{(i)} + \frac{\lambda}{2m} \sum_{l=1}^L \|W^{[l]}\|_F^2 \quad (2-5)$$

其中 $\|W^{[l]}\|_F^2$ 表示第 l 层的 W 的弗罗贝尼乌斯范数(Frobenius Norm),即矩阵 W 中所有元素的平方和;其他参数与 $L1$ 正则化中的参数含义一致。在实际训练网络的过程中, $L2$ 正则化更为常用。

③ Dropout 正则化: Dropout 正则化是指用每一个训练样本训练神经网络时,随机设置某些层神经元保留或者消除的概率,这样就可以训练一个更小的网络,达到正则化的效果。Dropout 正则化主要应用在计算机视觉领域,因为该领域通常没有足够的数据,神经网络一直存在过拟合现象,需要 Dropout 正则化。在其它领域应用得比较少。

④ 数据扩充: 更多的训练数据,可以减少神经网络的过拟合,达到正则化的效果。但在实际问题中,有时无法获取更多的数据,这时候可以通过一些手段,扩充数据集。例如对机器视觉领域一些应用,我们需要大量的图片集,在无法获得更多的图片情况下,可以通过水平翻转已有的图片或随意裁剪已有的图片,然后将这些图片添加到数据集中,达到扩充数据集的目的。

⑤ 提前停止训练: 提前停止训练是提前结束神经网络的训练过程,从而达到防止神经网络过拟合的目的。这种方法有一个缺点,就是一旦停止了训练过程,也就停止了成本函数的优化,虽然防止了过拟合的发生,但是这也可能导致成本函数还不够小。

总之,在实际应用过程中, $L2$ 正则化用的比较多,也可以同时使用几种正则化方法,达到减少过拟合的目的。

2.2.4 标准神经网络的反向传播过程

前向传播过程执行完后,需要用反向传播算法(Backpropagation Algorithm)从网络的最后一层往前计算成本函数(Cost Function)对各层参数的导数。反向传播过程中计算各层导数的公式如下:

$$\begin{aligned}
dZ^{[L]} &= \frac{\partial L}{\partial Z^{[L]}} = dA^{[L]} * g^{[L]}(Z^{[L]}) \\
dW^{[L]} &= \frac{\partial J}{\partial W^{[L]}} = \frac{1}{m} dZ^{[L]} A^{[L-1]T} \\
db^{[L]} &= \frac{\partial J}{\partial b^{[L]}} = \frac{1}{m} \sum_{i=1}^m dZ^{[L](i)} \\
dA^{[L-1]} &= \frac{\partial L}{\partial A^{[L-1]}} = W^{[L]T} dZ^{[L]} \\
db^{[L]} &= \frac{\partial J}{\partial b^{[L]}} = \frac{1}{m} \sum_{i=1}^m dZ^{[L](i)} \\
&\vdots \\
dZ^{[1]} &= dA^{[1]} * g^{[1]}(Z^{[1]}) \\
dW^{[1]} &= \frac{\partial J}{\partial W^{[1]}} = \frac{1}{m} dZ^{[1]} A^{[0]T} \\
db^{[1]} &= \frac{\partial J}{\partial b^{[1]}} = \frac{1}{m} \sum_{i=1}^m dZ^{[1](i)}
\end{aligned} \tag{2-6}$$

其中， L 为损失函数（Loss Function）； J 为成本函数（Cost Function）； $dZ^{[L]}$ 、 $dZ^{[1]}$ 分别为损失函数 L 对 $Z^{[L]}$ 、 $Z^{[1]}$ 的导数； $dW^{[L]}$ 、 $dW^{[1]}$ 分别为成本函数 J 对 $W^{[L]}$ 、 $W^{[1]}$ 的导数； $db^{[L]}$ 、 $db^{[1]}$ 分别为为成本函数 J 对 $b^{[L]}$ 、 $b^{[1]}$ 的导数。

2.2.5 标准神经网络的参数更新算法

（1）梯度下降算法

通过反向传播过程计算各层参数的导数后，就可以采用梯度下降算法更新各层的参数。各层权值 W 和偏置 b 的更新公式如下：

$$W = W - \alpha dW \tag{2-7}$$

$$b = b - \alpha db \tag{2-8}$$

其中 α 表示学习率（Learning Rate），默认值为0.01。

（2）动量梯度下降算法

动量梯度下降算法（Gradient Descent with Momentum）与梯度下降算法不同的是，动量梯度下降算法先计算参数导数的指数加权滑动平均（Exponentially Weighted Moving Averages），然后使用这个加权平均值来更新权重和偏置。

首先计算 W 导数和 b 导数的加权平均值:

$$v_{dw} = \beta v_{dw} + (1 - \beta) dW \quad (2-9)$$

$$v_{db} = \beta v_{db} + (1 - \beta) db \quad (2-10)$$

然后更新参数, 新的参数为:

$$W = W - \alpha dW \quad (2-11)$$

$$b = b - \alpha db \quad (2-12)$$

其中 v_{dw} 、 v_{db} 分别为 dW 、 db 的指数加权滑动平均值; β 为超参数 (Hyperparameters)。

与梯度下降算法相比, 动量梯度下降算法能有效减少算法迭代过程中的震荡, 加快算法的收敛速度。

(3) RMSprop 算法

RMSprop 算法, 也称作为均方根传递(Root Mean Square prop)。RMSprop 算法首先计算参数导数平方的指数加权滑动平均, 然后使用这个加权平均值来更新权重和偏置。

首先计算 W 导数的平方和 b 导数的平方的指数加权滑动平均值:

$$S_{dw} = \beta S_{dw} + (1 - \beta) dW^2 \quad (2-13)$$

$$S_{db} = \beta S_{db} + (1 - \beta) db^2 \quad (2-14)$$

然后更新参数, 新的参数为:

$$W = W - \alpha \frac{dw}{\sqrt{S_{dw}}} \quad (2-15)$$

$$b = b - \alpha \frac{db}{\sqrt{S_{db}}} \quad (2-16)$$

其中 dw^2 、 db^2 分别表示 dW 、 db 的平方; S_{dw} 、 S_{db} 分别表示 dw^2 、 db^2 指数加权滑动平均值。

RMSprop 算法与动量梯度下降算法一样, 能够降低梯度下降过程中的振荡, 并且能使用更大的学习率 α , 从而提高算法的学习速度。

(4) Adam 优化算法

Adam 优化算法是一种将动量梯度下降算法和 RMSprop 算法结合起来的算法。具体如下介绍。

首先计算 W 导数和 b 导数的加权平均值：

$$v_{dw} = \beta_1 v_{dw} + (1 - \beta_1) dW \quad (2-17)$$

$$v_{db} = \beta_1 v_{db} + (1 - \beta_1) db \quad (2-18)$$

然后计算 W 导数的平方和 b 导数的平方的指数加权滑动平均值：

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) dW^2 \quad (2-19)$$

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2 \quad (2-20)$$

接着进行偏差修正：

$$v_{dw}^{\text{corrected}} = \frac{v_{dw}}{1 - \beta_1^t} \quad (2-21)$$

$$v_{db}^{\text{corrected}} = \frac{v_{db}}{1 - \beta_1^t} \quad (2-22)$$

$$S_{dw}^{\text{corrected}} = \frac{S_{dw}}{1 - \beta_2^t} \quad (2-23)$$

$$S_{db}^{\text{corrected}} = \frac{S_{db}}{1 - \beta_2^t} \quad (2-24)$$

最后进行参数更新：

$$W = W - \alpha \frac{v_{dw}^{\text{corrected}}}{\sqrt{S_{dw}^{\text{corrected}} + \epsilon}} \quad (2-25)$$

$$b = b - \alpha \frac{v_{db}^{\text{corrected}}}{\sqrt{S_{db}^{\text{corrected}} + \epsilon}} \quad (2-26)$$

其中 $v_{dw}^{\text{corrected}}$ 、 $v_{db}^{\text{corrected}}$ 、 $S_{dw}^{\text{corrected}}$ 、 $S_{db}^{\text{corrected}}$ 分别表示 v_{dw} 、 v_{db} 、 S_{dw} 、 S_{db} 修正后的值；

ϵ 是一个数值很小的参数，添加这个参数为了避免算法除以零； β_1 、 β_2 以及 α

都是算法需要设置的超参数，通常 β_1 设置为 0.9， β_2 设置为 0.999， ϵ 设置为 10^{-8} 。

Adam 算法结合了动量梯度下降算法和 RMSprop 算法的各种优势，被广泛使用，而且已经被证明在很多不同种类的神经网络构架中都十分有效。本研究也将采用这种算法对网络的参数进行更新。

2.2.6 标准神经网络的适用范围

标准神经网络是最基本的神经网络，大多数其他类型的神经网络都是在标准神经网络的基础上改进或者衍生出来的。标准神经网络具有鲁棒性好、非线性映射能力和自学习能力强等优点，可以用来预测房价的走势、在线广告的推荐等等。不过在目前的很多应用中，都不是直接应用标准神经网络，而是其各种改进的版本。对于中长期径流预报，标准神经网络并不是最合适的。因为对于中长期径流时间序列， t 时刻的径流值不仅与当前神经网络的输入有关，而且与 t 时刻之前的径流量值也有一定的关系，比如， $t-1$ 时刻的径流量可以反映该时刻的大气湿度、降雨等情况，而这些因素对预测 t 时刻的径流值也会有一定的参考意义。但是标准神经网络是一种前向网络，网络在 t 时刻的输出仅仅与 t 时刻的输入有关，而与 t 时刻之前的输入无关，这样就不能反映径流时间序列数据在时间维度上前后之间的联系，因此，标准神经网络更适合非时间序列数据的处理，对于时间序列数据（如中长期径流时间序列），循环神经网络更合适。

2.3 循环神经网络

2.3.1 循环神经网络的拓扑结构

循环神经网络(RNN, Recurrent Neural Networks)的结构类型较多，根据输入、输出的不同，可以分为多对多 (Many to Many)、多对一 (Many to One)、一对多 (One to Many)、一对一 (One to One) 四种类型。多对多的 RNN 最典型的应用是机器翻译 (Speech Recognition)，每一个输入序列有很多的输入，对应的一个输出序列也有很多的输出；多对一的 RNN 最典型的应用是情感分类 (Sentiment Classification)，比如电影情感分类，输入一句话，RNN 输出一个 1 到 5 的数字，代表不同的评分等级；一对多的 RNN 最典型的应用是音乐生成 (Music generation)，输入任意一个字符，RNN 会输出一段合成的音乐；一对一的 RNN 也就是一个小型的标准神经网络，输入 x 然后得到输出 y ，与 2.2 节介绍的标准

神经网络差异不大。如果将 RNN 应用于中长期径流预报，输入是因子的时间序列数据，输出则是径流量的时间序列数据，因此对应的 RNN 为多对多类型，下面将着重介绍这种结构，对于其他结构，可以参考相关文献。

图 2-3 是多对多的 RNN 在时间维度展开后的拓扑结构图。图中的方框表示一个模型单元，该单元可以是标准神经网络,也可以是门控循环单元 (GRU, Gated Recurrent Unit),或者长短期记忆单元 (LSTMUnit, Long Short Term Memory Unit)。

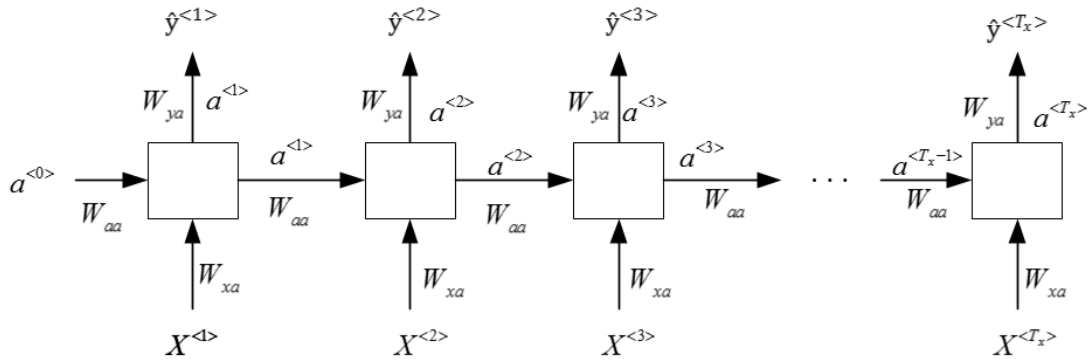


图 2-3 循环神经网络在时间维度展开后的拓扑结构

图中， $X^{<1>}$ 、 $X^{<2>}$ 、 $X^{<3>}$ 、 $X^{<T_x>}$ 分别表示时间序列的第一个样本、第二个样本、第三个样本、第 T_x 个样本， $\hat{y}^{<1>}$ 、 $\hat{y}^{<2>}$ 、 $\hat{y}^{<3>}$ 及 $\hat{y}^{<T_x>}$ 分别为对应的预测值， $a^{<1>}$ 、 $a^{<2>}$ 、 $a^{<3>}$ 、 $a^{<T_x>}$ 分别为对应的激活值； W_{aa} 、 W_{xa} 、 W_{ya} 为神经网络的权值参数。

2.3.2 循环神经网络的前向传播过程

RNN 的前向传播过程与标准神经网络的前向传播过程基本一致，也是数据从输入层输入，经过隐含层，最后从输出层输出的过程。唯一不同的是， t 时刻隐含层的激活值会在 $t+1$ 时刻再次作为网络隐含层的输入。下面介绍前向传播过程的数学推导。

如图 2-4，先给网络输入 $a^{<0>}$ 与 $x^{<1>}$ ， $a^{<0>}$ 一般初始化为零向量。接着开始前向传播过程，计算激活值 $a^{<1>}$ 与 $\hat{y}^{<1>}$ ：

$$a^{<1>} = g_1(W_{aa}a^{<0>} + W_{ax}x^{<1>} + b_a) \quad (2-27)$$

$$\hat{y}^{<1>} = g_2(W_{ya}a^{<1>} + b_y) \quad (2-28)$$

然后把 $a^{<1>}$ 与 $x^{<2>}$ 输入网络，计算激活值 $a^{<2>}$ 与 $\hat{y}^{<2>}$ ：

$$a^{<2>} = g_1(W_{aa}a^{<1>} + W_{ax}x^{<2>} + b_a) \quad (2-29)$$

$$\hat{y}^{<2>} = g_2(W_{ya}a^{<2>} + b_y) \quad (2-30)$$

更一般的情况下，在 t 时刻，计算激活值 $a^{<t>}$ 与 $\hat{y}^{<t>}$ ：

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \quad (2-31)$$

$$\hat{y}^{<t>} = g_2(W_{ya}a^{<t>} + b_y) \quad (2-32)$$

上述公式中 $x^{<1>}$ 、 $x^{<2>}$ 及 $x^{<t>}$ 分别表示时间序列的第一个样本、第二个样本、第 t 个样本， $\hat{y}^{<1>}$ 、 $\hat{y}^{<2>}$ 及 $\hat{y}^{<t>}$ 分别为对应的预测值； $a^{<1>}$ 、 $a^{<2>}$ 、 $a^{<t>}$ 分别为对应的激活值； W_{aa} 、 W_{xa} 、 W_{ya} 为神经网络的权值参数， b_a 与 b_y 分别为神经网络的偏置参数； g_1 与 g_2 是激活函数。

2.3.3 计算循环神经网络的成本函数

循环神经网络成本函数也需要根据实际应用的不同选择不同的函数，如均方误差函数（Mean Squared Error）、对数损失函数（Binary Crossentropy）等。循环神经网络成本函数与标准神经网络成本函数基本是一样的，唯一不同的地方是，标准神经网络的成本函数是所有单个样本或者一个批量（batch）的样本的损失函数求和后的平均，而循环神经网络成本函数是把每一个时间步的损失函数求和后再取平均。下面以常用的标准逻辑回归损失函数（Standard Logistic Regression Loss）为例，对循环神经网络的成本函数进行介绍。

对于标准循环神经网络，在 t 时刻的单样本损失函数为：

$$L^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -y^{<t>} \log \hat{y}^{<t>} - (1 - \hat{y}^{<t>}) \log(1 - \hat{y}^{<t>}) \quad (2-33)$$

整个序列数据的成本函数则是对所有时间步的单样本损失函数求和再取平均：

$$J(\hat{y}, y) = \frac{1}{T_x} \sum_{t=1}^{T_x} L^{<t>}(\hat{y}^{<t>}, y^{<t>}) \quad (2-34)$$

其中 $\hat{y}^{<t>}$ 表示 t 时刻的样本对应的预测值； $y^{<t>}$ 为 t 时刻样本的实测值； T_x 为序列数据的长度。

为了减少循环神经网络在训练过程中的过拟合，可以采用一些正则化方法。比如对循环神经网络的损失函数加上 $L1$ 正则项或 $L2$ 正则项等。标准神经网络中介绍的 $L1$ 正则化、 $L2$ 正则化、Dropout 正则化、数据扩增以及提前停止训练等正则化方法完全适用于循环神经网络，这里不再赘述。

2.3.4 循环神经网络的反向传播过程

计算完循环神经网络的成本函数后,然后就可以使用反向传播算法计算成本函数对各层参数的导数,计算过程与标准神经网络的计算过程是一致的，此处不再赘述。

2.3.5 循环神经网络的参数更新算法

运用反向传播算法计算各参数的导数后,就可使用参数的更新算法对参数进行更新。标准神经网络部分介绍的梯度下降算法、动量梯度下降算法、RMSprop 算法以及 Adam 优化算法都可以直接用于循环神经网络的参数更新。此处不再赘述。

2.3.6 循环神经网络的适用范围

与标准神经网络不同的是，RNN 在 t 时刻的隐含层的激活值会作为 $t+1$ 时刻的隐含层的输入。如图 2-3 中， $a^{<1>}$ 为网络在输入样本 $X^{<1>}$ 时产生的激活值；在输入下一个样本 $X^{<2>}$ 时， $a^{<1>}$ 也同时作为网络的输入，参与输出值 $\hat{y}^{<2>}$ 的计算，而 $a^{<1>}$ 是由输入样本 $X^{<1>}$ 计算得到的，因此预测值 $\hat{y}^{<2>}$ 不仅与当前的输入样本 $X^{<2>}$ 有关，而且与前一时刻的输入样本 $X^{<1>}$ 有关；在输入下一个样本 $X^{<3>}$ 时， $a^{<2>}$ 也同时作为网络的输入，参与输出值 $\hat{y}^{<3>}$ 的计算，而 $a^{<2>}$ 是由输入样本 $X^{<2>}$ 以及 $a^{<1>}$ 计算得到的， $a^{<1>}$ 是由输入样本 $X^{<1>}$ 计算得到的，因此，预测值 $\hat{y}^{<3>}$ 不仅与当前的输入样本 $X^{<3>}$ 有关，而且还与在这之前的时刻输入的样本 $X^{<2>}$ 及 $X^{<1>}$ 有关。同理，对于输入样本预测值 $\hat{y}^{<T_x>}$ ，不仅与当前的输入样本 $X^{<T_x>}$ 有关，

而且与之间所有时刻的输入样本有关。这种特殊的结构使得 RNN 能够共享从时间序列数据不同位置上学习到的特征，因此，与标准神经网络相比，RNN 能够更有效的处理时间序列数据。

语音识别、机器翻译、情感分类等都是 RNN 用于处理时间序列数据的典型代表。对于中长期径流预报，径流时间序列与预报因子序列也都是时间序列数据，用 RNN 进行中长期径流预报， t 时刻的预测值不仅仅是用 t 时刻的因子来预报，还会用上 t 时刻之前的所有的预报因子的信息，即预报的结果不仅与当前的输入有关，还与这之前的所有的输入有关。因此，与标准神经网络相比，RNN 更适合中长期径流的预报。

2.4 本章小结

本章介绍了人工神经网络的基本理论。分别从网络的拓扑结构及其训练过程两方面详细介绍了标准神经网络和循环神经网络，并结合网络各自的特点，分析了两种网络各自的应用范围。为第三章中长期径流智能预报技术体系的分析与选取提供了理论基础。

第三章 中长期径流智能预报技术体系研究

3.1 引言

中长期径流预报就是据已知信息对未来一定时期内的径流状态做出定性或定量的预测^[39]。建立中长期径流智能预报技术体系,首先,要进行物理机制分析,从太阳活动、地球及其他天文物理因素、大气环流、下垫面情况、人类活动及前期气象水文要数等因素里面初步选出一些因子作为预报因子。然后,运用统计学方法对这些与径流有物理联系的因子进行分析,选出相关性高的因子。在此基础上,运用特征提取技术,对初选的因子进行特征提取,降噪和去冗余。接着,从确定模型结构与训练算法两方面来建立神经网络模型。最后,对前面各部分提到的方法进行对比分析,确定中长期径流智能预报技术体系。

3.2 预报因子的初选技术

对预报因子进行初选,首先要对形成径流的物理机制进行分析,选出与径流量有物理联系的预报因子。但是,这些预报因子,只是与径流量有物理联系,还需进一步的运用统计学的方法对这些因子作进一步筛选,选出高相关的因子。常用的统计学方法包括 Pearson 相关系数法和 Spearman 等级相关系数法,此外,还有互信息法等。

3.2.1 物理机制分析

中长期径流过程具有随机性、模糊性的特点,受诸多因素的影响,具体包括太阳活动、地球及其他天文物理因素、大气环流、下垫面情况、人类活动及前期气象水文要数等^[40]。图 3-1^[41]是径流过程的影响因素示意图。对预报因子的筛选,首先要从物理机制出发,从众多影响因素中选出物理联系紧密的因子。

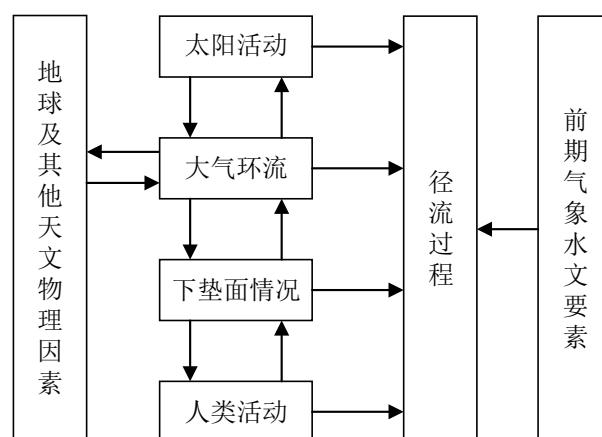


图 3-1 径流过程的影响因素

(1) 太阳活动：太阳辐射是大气运动和水文循环的能量来源，太阳活动的增强和减弱会引起大气运动状态的改变从而影响水文过程，使其发生相应的变化。如不少地区的旱涝灾害与太阳黑子相对数的变化存在一定的对应关系。

(2) 大气环流：大气环流决定着一个流域（或地区）的降水和蒸发的变化，影响水文循环的各个环节。异常的旱涝现象总与大气环流的异常联系在一起。

(3) 下垫面情况：海洋表层水温异常会引起大气运动的异常,陆地表面状况的异常,也会影响大气环流的形势，从而影响水文过程的长期变化。

(4) 地球及其他天文物理因素：地球自转速度的变化、火山爆发、行星相对位置等对大气环流和水文形势变化也有一定影响。

(5) 人类活动：人类活动不仅直接影响当地水文情况的变化，而且对水文过程产生间接影响。这些影响因素与水文过程的联系见示意图。

(6) 前期气象水文资料：前期气象水文资料与中长期径流紧密相关。考虑到气象因子对径流的影响存在滞后性，因此采用前期的气象水文资料对中长期径流进行预报。可以采用的前期气象水文资料包括大气环流特征（如历年逐月的环流指数）、高空气压场（如北半球 100hpa 、500hpa 逐月平均高度场）、海表温度（如北太平洋逐月海温场，SST）、以及一些地表观测数据（如当地降雨、径流、蒸发、日照等）。

中长期径流预报因子的初选，可以从以上介绍的影响因子中选取。国家气候中心在 74 项环流指数的基础上提供了新的 130 项指数，包括 88 项历年逐月的大气环流指数数据、26 项历年逐月温度指数数据以及 16 项其他指数的历年逐月数据。这 130 项指数基本涵盖了前期主要的气象水文要素，为中长期径流预报提供

了方便。本研究将直接从这新的 130 项指数中筛选预报因子。

3.2.2 Pearson 相关系数法

Pearson 相关系数适用于研究两个变量之间的相关程度。例如，径流量和预报因子之间的线性相关关系。计算径流量和预报因子之间的 Pearson 相关系数的计算公式为：

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (3-1)$$

公式中： r 为 Pearson 相关系数； x_i 、 y_i 分别为 X 和 Y 的第 i 的样本， X 和 Y 分别为预报因子时间序列和径流时间序列； \bar{x} 、 \bar{y} 分别为样本 X 和 Y 的样本均值； n 为时间序列的长度。利用相关系数 r 可以判断变量间的相关程度，具体见下表：

表 3-1 相关程度等级

相关系数的值	线性相关的程度
$ r = 0$	完全不相关
$0 < r \leq 0.3$	微弱相关
$0.3 < r \leq 0.5$	低度相关
$0.5 < r \leq 0.8$	显著相关
$0.8 < r < 1$	高度相关
$ r = 1$	完全相关

需要注意的是，Pearson 相关系数适用于线性相关的情形，对于非线性相关的情形，系数的大小并不能代表其相关性的强弱。中长期径流的预报因子与径流时间序列之间除了线性关系外，还存在着非线性的关系，因此，在中长期径流预报中用 Pearson 相关系数来筛选因子的做法是不够准确的。

3.2.3 Spearman 等级相关系数法

Spearman 等级相关系数适合用来计算序列变量之间的线性相关关系。这种

方法把变量的数值序列由小到大的顺序进行重新排序后,记录下变量每个数值对应的序号,即秩,然后用秩取代数值并计算相关系数。计算径流量和预报因子之间的 Spearman 等级相关系数的计算过程为:

令因子时间序列 X 和径流时间序列 Y 分别为

$$X = (x_1, x_2, x_3, \dots, x_n)$$

$$Y = (y_1, y_2, y_3, \dots, y_n)$$

X 、 Y 的值经过秩替换后的顺序列分别为

$$C = (c_1, c_2, c_3, \dots, c_n)$$

$$D = (d_1, d_2, d_3, \dots, d_n)$$

则变量 X 、 Y 的 Spearman 等级相关系数计算公式为:

$$r_R = 1 - \frac{6 \sum_{i=1}^n (c_i - d_i)^2}{n(n^2 - 1)} \quad (3-2)$$

其中, r_R 表示 Spearman 等级相关系数; c_i 表示 X 的第 i 个样本对应的秩; d_i 表示 Y 的第 i 个样本对应的秩; n 为样本容量, 即时间序列的长度。

Spearman 等级相关系数法的突出优点是对数据的总体分布、样本大小都不做要求。但缺点是计算精度不高。此外, Spearman 等级相关系数也只适合用来计算顺序变量之间的线性相关关系。径流时间序列与其对应的因子序列满足顺序序列的要求, 可以用 Spearman 等级相关系数法来计算彼此之间的线性相关性, 但是对于非线性相关, 该方法也不适用。

3.2.4 互信息

在概率论与信息论中, 两个变量之间的互信息 (MI, Mutual Information) 是两个变量之间相互依存的一种度量^[42]。以计算径流时间序列与因子时间序列之间的互信息为例, 计算公式如下:

$$MI(X, Y) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) \quad i, j = 1, 2, \dots, n \quad (3-3)$$

其中, X 为径流时间序列, $X = (x_1, x_2, x_3, \dots, x_n)^T$, Y 为因子时间序列, $Y = (y_1, y_2, y_3, \dots, y_n)^T$, 分子 $p(x_i, y_i)$ 为 X 和 Y 的联合分布律, $p(x_i)$ 和 $p(y_i)$ 分别为 X 与 Y 的边缘分布律, n 为时间序列的长度。

为了方便比较, 需要将互信息标准化。标准化后互信息的值在 0 与 1 之间。标准互信息 (NMI, Normalized Mutual Information) 的计算公式如下:

$$NMI(X, Y) = 2 \frac{MI(X, Y)}{H(X) + H(Y)} \quad (3-4)$$

其中, $H(X)$ 和 $H(Y)$ 分别为 X 和 Y 的熵, $H(X)$ 和 $H(Y)$ 的计算公式如下:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (3-5)$$

$$H(Y) = - \sum_{j=1}^n p(y_j) \log_2 p(y_j) \quad (3-6)$$

关于标准互信息的显著性检验, Sharma 提出采用自举法^[43], 具体包括以下步骤^[44]:

步骤 1: 计算原径流时间序列与指标时间序列的标准互信息 NMI 值;

步骤 2: 随机同时打乱两对具有对应关系的时间序列的顺序 K 次(一般取 100 次), 计算乱序后 NMI 值并按由大到小顺序排列;

步骤 3: 取顺序排列 NMI 的概率分位数作为对应该概率显著性水平的 NMI 阈值;

步骤 4: 若原时间序列 NMI 值大于某概率阈值(一般取 95%)对应的 NMI 值, 则认为这两组数据显著相关。

步骤 5: 选出通过显著性检验且标准互信息大于某一阈值(一般取 0.9, 但根据时间序列长短的不同会有差异, 可自行调整)的指标作为初选的预报因子。

基于互信息的预报因子的选择方法, 不仅能够反映因子与径流之间的线性关系, 还能反应他们之间的非线性关系。而传统的基于线性相关分析(如 Pearson 相关分析, Spearman 等级相关等)筛选因子的方法, 只能反应因子与径流之间的线性关系。因此, 基于互信息的预报因子选择方法选出来的因子更具有代表性。

3.3 初选因子的特征提取技术

由于用统计学方法初选出来的因子，彼此之间高相关性的因子往往较多，不同的因子的时间序列之间存在较高的复共线性，且因子时间序列本身也存在一定的噪声。因此，需要对初选出来的因子进行降噪和去冗余。本研究将介绍两种特征提取方法：主成分分析法和核主成分分析法。

3.3.1 基于主成分分析的特征提取

对于初选出来的因子的降噪和去冗余，目前常用的方法是主成分分析法（PCA, Principal Component Analysis）。主成分分析法主要包含以下步骤：

步骤 1：将初选的预报因子数据 z-score 标准化，计算公式如下：

$$y^* = \frac{y - \mu}{\sigma} \quad (3-7)$$

公式中， y^* 为 z-score 标准化后的因子数据， y 为初选的单个预报因子数据， μ 为 y 所处的预报因子时间序列的平均值， σ 为 y 所处的因子时间序列的标准差。

步骤 2：计算初选的因子的协方差矩阵 COV 。如果初步选取了 m 个因子，则 COV 为 $m \times m$ 的矩阵，第 i 行第 j 列的元素 $COV_{i,j}$ 的计算公式如下：

$$COV_{i,j} = \frac{\sum_{i,j=1}^n (Y_i^* - \bar{Y}_i^*)(Y_j^* - \bar{Y}_j^*)}{n-1} \quad i, j = 1, 2, \dots, n \quad (3-8)$$

公式中： Y_i^* 、 Y_j^* 代表不同的初选的预报因子 z-score 标准化后的时间序列； \bar{Y}_i^* 、 \bar{Y}_j^* 分别为 Y_i^* 、 Y_j^* 时间序列的平均值； n 为所选因子时间序列的长度。

步骤 3：求协方差矩阵 COV 的特征值与特征向量，并把特征值按照由大到小的顺序排列，特征向量的顺序按照特征值做对应的调整。排序后得到的特征值矩阵为 Λ ，特征向量矩阵为 U ，表示如下：

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix}, \quad U = (u_1, u_2, \dots, u_m) = \begin{pmatrix} u_{11} & & u_{1m} \\ & u_{ij} & \\ u_{n1} & & u_{nm} \end{pmatrix} \quad (3-9)$$

其中 $u_i = (u_{i1}, u_{i2}, \dots, u_{im})^T$ ； $i, j = 1, 2, \dots, m$ ， m 为初选因子的数目。

步骤 4：提取主成分，一般提取前 2 到 3 个主成分作为预报因子。第 i 个主成分的计算公式如下：

$$PCA_i = (Y_1, Y_2, \dots, Y_i, \dots, Y_m)u_i \quad i, j = 1, 2, \dots, m \quad (3-10)$$

其中 Y_i 为初步选取的第 i 个因子的时间序列； u_i 为步骤 3 求得特征向量。

PCA 能用几个较少的综合指标来代替原来较多的指标，而这些较少的综合指标不但能尽量多地反映原来较多指标的有用信息，且相互之间正交。这样得到的因子作为神经网络模型的输入，能有效缓解人工神经网络在训练阶段的过拟合，提高网络的泛化能力。但是 PCA 实质也是一种线性映射方法，得到的主成分是由线性映射生成的。这种方法忽略了数据之间高于 2 阶的相互关系^[35]，所以抽取的主成分并不是最优的。

3.3.2 基于核主成分分析的特征提取

核主成分分析法（KPCA）是主成分分析法（PCA）的非线性扩展，即通过映射函数 Φ 把原始向量映射到高维特征空间 F ，在 F 上进行 PCA 分析。核主成分分析法主要包含以下步骤：

步骤 1：首先，与 PCA 一样，将初选的预报因子数据 z-score 标准化，计算公式如下：

$$y^* = \frac{y - \mu}{\sigma} \quad (3-11)$$

公式中， y^* 为 yz-score 标准化后的因子数据， y 为初选的单个预报因子数据， μ 为 y 所处的预报因子时间序列的平均值， σ 为 y 所处的因子时间序列的标准差。

步骤 2：计算初选的预报因子的核矩阵 K 。 K 为 $n \times n$ 的矩阵，第 i 行第 j 列的元素 $K_{i,j}$ 的计算公式如下：

$$K_{i,j} = k(Y_i^*, Y_j^*) \quad (3-12)$$

公式中， Y_i^*, Y_j^* 代表不同的初选的预报因子 z-score 标准化后的时间序列； n 为所选因子时间序列的长度， $i, j = 1, 2, \dots, n$ ； k 为核函数，常用的核函数有以下几种：

(1) 线性核（Linear Kernel）：

$$k(Y_i^*, Y_j^*) = Y_i^{*T} Y_j^* \quad (3-13)$$

(2) 多项式核 (Polynomial Kernel):

$$k(Y_i^*, Y_j^*) = (b Y_i^{*T} Y_j^* + c)^p \quad (3-14)$$

(3) 径向基核函数 (Radial Basis Function):

$$k(Y_i^*, Y_j^*) = \exp\left(-\frac{\|Y_i^* - Y_j^*\|^2}{2\delta^2}\right) \quad (3-15)$$

(4) Sigmoid 核 (Sigmoid Kernel):

$$k(Y_i^*, Y_j^*) = \tanh(\nu Y_i^{*T} Y_j^* + \xi) \quad (3-16)$$

公式 (3-13)、(3-14)、(3-15) 与 (3-16) 中的 b 、 c 、 p 、 δ 、 ν 、 ξ 均为常数，是各种核函数的参数。

步骤 3: 计算中心化的核矩阵。经过中心化后的核矩阵用 K_C 表示， K_C 计算公式如下:

$$K_C = K - J \square K - K \square J + J \square K \square J \quad (3-17)$$

公式 (3-17) 中 J 为 $n \times n$ 的矩阵， n 为所选因子时间序列的长度， J 的形式如下所示:

$$J = \begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix} \quad (3-18)$$

步骤 4: 计算中心化后的核矩阵 K_C 的特征值与特征向量, 并把特征值按照由大到小的顺序排列, 特征向量的顺序按照特征值做对应的调整。排序后得到的特征值矩阵为 Λ , 特征向量矩阵为 U , 表示如下:

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, U = (u_1, u_2, \dots, u_n) = \begin{pmatrix} u_{11} & \dots & u_{1n} \\ \dots & u_{ij} & \vdots \\ u_{n1} & \dots & u_{nn} \end{pmatrix} \quad (3-19)$$

公式中 λ_1 、 λ_n 表示特征值， u_1 、 u_2 、 u_n 表示特征向量， n 为所选因子时间序列的长度。

步骤 5: 计算归一化的特征向量矩阵 A , A 的形式如下:

$$A = (a_1, a_2, \dots, a_i, \dots, a_n) = \begin{pmatrix} a_{11} & \dots & a_{n1} \\ \dots & a_{ij} & \vdots \\ a_{1n} & \dots & a_{nn} \end{pmatrix} \quad i, j = 1, 2, \dots, n \quad (3-20)$$

其中 $a_i = \frac{u_i}{\sqrt{\lambda_i}}$, n 为所选因子时间序列的长度。

步骤 6: 提取主成分, 一般提取前 2 到 3 个主成分作为预报因子。第 i 个主成分的计算公式如下:

$$KPC_i = a_i^T K_C \quad i = 1, 2, \dots, n \quad (3-21)$$

K_C 为步骤 3 计算得到的中心化的核矩阵, n 为所选因子时间序列的长度。

KPCA 通过将原始空间中的因子时间序列数据映射到高维空间, 使得在低维空间中线性不可分的因子时间序列数据在高维空间中几乎都能线性可分, 此时在高维空间做 PCA, 提取的主成分更具有代表性。因此基于 KPCA 的特征提取方法大大提高了非线性数据的处理能力, 与传统的基于 PCA 的特征提取方法相比, 更有优势。此外, 经 KPAC 提取的主成分之间也相互正交, 并且数据经过了降噪去冗余, 能够很好的防止神经网络的过拟合, 提高网络的泛化能力。

3.4 中长期径流智能预报模型的构建技术

近年来, 随着计算机计算能力的增强、数据量的增加以及各种新算法的提出, 以人工神经网络为基础的人工智能快速崛起, 因此, 人工神经网络也得到了极大地发展。与其他模型相比, 人工神经网络因鲁棒性好、非线性映射和自学习能力强, 在中长期径流预报中也得到了广泛的应用^[45]。循环神经网络因为其特有的结构, 特别适合径流时间序列的处理, 所以本研究将用循环神经网络来构建中长期径流智能预报模型。本小节将从循环神经网络的结构、成本函数及正则化方法、网络的参数更新算法三方面来介绍单个循环神经网络预报模型的构建, 最后用多个循环神经网络来构建集合预报模型。

3.4.1 循环神经网络结构的确定

构建基于循环神经网络的单一预报模型, 需要确定其结构, 即神经网络的层数以及每一层的节点数。将循环神经网络应用于中长期径流预报, 网络包含一个

输入层、一个输出层，因此，还需要确定隐含层层数、输入层节点数与隐含层节点数、输出层节点数。下面分别进行介绍。

(1) 输出层节点数的确定：输出层节点数等于预报对象的个数。如果是对径流量进行单值预报，则输出层节点数为 1。

(2) 隐含层层数的确定：由于在时间维度上展开后，循环神经网络会变得很大，并且对深层的 RNN 训练需要很多计算资源及很长的时间，因此 RNN 的层数一般不超过 3 层，与标准神经网络可以达到 100 多层相比，RNN 的层数会少很多。对中长期径流预报而言，隐含层的层数选择 1 层即可。一方面，含有一个隐含层的 RNN 足以学习任何复杂的函数关系；另一方面，因为中长期径流预报因子的时间序列长度一般不超过 60 年，年分久远的数据没有记录，可用来训练 RNN 的数据量并不多。而随着 RNN 层数的增加，需要训练的网络参数会成倍的增长，如果数据量过少，很容易使 RNN 发生过拟合，降低 RNN 的泛化能力，影响预报的精度。

(3) 输入层节点数和隐含层节点数的确定：循环神经网络输入层的节点数等于输入神经网络的特征个数。对于隐含层的节点数，除了一些经验公式外，目前还没有一个系统和标准的方法来确定^[34]，但是其对网络的泛化性能有重要影响。本研究将采用试错法，即通过采用不同的输入特征个数与隐含层节点数的组合，观察网络的预报效果，从而确定输入层节点数和隐含层节点数。

3.4.2 循环神经网络成本函数及正则化方法的确定

在循环神经网络的训练过程中，成本函数的选取重要，通过最小化成本函数，达到训练模型参数的目的。对于中长期径流预报，本研究做的是定量的预报，每一次预报的结果都是一个确定的值，因此，可以用预测值与实测值之间的差值平方和作为成本函数，即均方误差函数（MSE, Mean Squared Error）。通过不断减小均方误差函数的值，优化模型参数。

均方误差函数的数学表达式如下：

$$J = \sum_{t=1}^{T_x} (\hat{y}^{<t>} - y^{<t>})^2 \quad (3-22)$$

其中， $\hat{y}^{<t>}$ 表示径流量的预测值； $y^{<t>}$ 表示径流量的实际值； T_x 为一次梯度更新

所用的样本数， J 代表成本函数。

对于正则化方法，有 $L1$ 正则化、 $L2$ 正则化、Dropout 正则化、数据扩增以及提前停止训练等。对于 Dropout 正则化，该方法主要应用在计算机视觉领域的，该领域通常没有足够的数据，神经网络一直存在过拟合现象，可以采用 Dropout 正则化。处理计算机视觉相关问题的神经网络主要是用卷积神经网络，卷积神经网络通常有很多的层数，因此可以使用 Dropout 正则化方法让一些层的神经元失活，在不影响网络性能的情况下，减小网络的结构，从而减轻或者避免过拟合。但是对于用于中长期径流预报的循环神经网络，只有一个隐含层，网络结构很小，因此不适合使用这种随机失活的方法；数据扩充法可以用来对图片等数据的处理，对于用来做中长期径流预报的数据，数据没有办法扩充，因此数据扩充法也不合适；对于提前停止训练法，如果用该方法来避免中长期径流预报过程中神经网络的过拟合，虽然可以通过提前停止对网路的训练达到避免过拟合的目的，但是这也停止了对模型参数的寻优，因此这种方法有一定的缺陷，同时需要一定经验； $L1$ 正则化、 $L2$ 正则化是常用的正则化方法。相比于 $L1$ 正则化， $L2$ 正则化方法应用的更多，且能够很好的减少神经网络的过拟合，因此，本研究将选用 $L2$ 正则化。通过给成本函数加上 $L2$ 正则项，即可避免或者减轻循环神经网络的过拟合。加上 $L2$ 正则项的成本函数如下：

$$J = \sum_{t=1}^{T_x} (\hat{y}^{<t>} - y^{<t>})^2 + \frac{\lambda}{2T_x} \sum_{l=1}^L \|W^{[l]}\|_F^2 \quad (3-23)$$

其中 λ 为正则项参数。

3.4.3 循环神经网络参数更新算法的确定

用反向传播算法计算成本函数对神经网络各层参数的导数后，就可以用参数更新算法更新参数。在 2.2.5 节介绍的梯度下降算法、动量梯度下降算法、RMSprop 算法以及 Adam 优化算法都可以用来对循环神经网络的参数进行更新。

目前，常用的参数更新算法是梯度下降算法，但是梯度下降算法在寻找最优参数的过程中容易发生震荡，并且容易陷入局部最优值。与梯度下降算法相比，动量梯度下降算法能有效减少算法迭代过程中的震荡，加快算法的收敛速度，而 RMSprop 算法不仅能够降低梯度下降过程中的振荡，并且能使用更大的学习率，

提高算法的学习速度。因此动量梯度下降算法和 RMSprop 算法比梯度下降算法性能更好。进一步的, Adam 优化算法将动量梯度下降算法和 RMSprop 算法结合起来, 集合了两种算法的优势, 且该算法已经被证明在很多不同种类的神经网络构架中都十分有效。因此, 本研究采用 Adam 优化算法对 RNN 的参数进行更新。

3.4.4 基于循环神经网络的集合预报模型构建

反向传播算法和参数更新算法实现了对 RNN 的参数空间的搜索, 通过不断的训练, 减小历史径流数据的实际值和 RNN 的预测值间的误差。但是, 误差曲面可能含有多个不同的局部极小值, 在对 RNN 参数空间的搜索过程中, 可能会停留在局部极小值点, 而不一定是全局最小值点。因此, 即使各个经过训练后的 RNN 的结构相同, 但是模型的参数也是不同的, 这导致各个单一 RNN 模型预测结果之间存在差异。为了减少这种由于模型参数不确定性而导致的预测结果的偏差, 本研究多次进行径流量的单模型预报, 将多次预报结果的平均值作为最终的预报结果, 实现基于循环神经网络的中长期径流集合预报。

3.5 智能预报技术体系的分析及选取

预报因子的初选技术、初选因子的特征提取技术、中长期径流智能预报模型的构建技术构成中长期径流智能预报技术体系。这三种技术不同的组合, 可以构成不同的技术体系。表 3-2 为不同的中长期径流智能预报技术体系及各自所属的类型。

表 3-2 智能预报技术体系

方案编号	技术体系			类型		
	预报因子	初选因子	预报	预报因子	初选因子	预报
	初选	特征提取	模型	初选	特征提取	模型
1	相关分析	→ PCA	→ 回归分析	线性	→ 线性	→ 线性
2	相关分析	→ PCA	→ RNN	线性	→ 线性	→ 非线性
3	互信息	→ PCA	→ RNN	非线性	→ 线性	→ 非线性
4	互信息	→ KPCA	→ RNN	非线性	→ 非线性	→ 非线性
5	互信息	→ KPCA	→ 多个 RNN	非线性	→ 非线性	→ 非线性

本研究介绍的预报因子的初选方法，有相关分析法和互信息法。相关分析法常用的有 Pearson 相关系数法及 Spearman 等级相关系数法。这两种方法描述的都是变量之间的线性关系，因此属于线性的分析方法。互信息是两个变量之间相互依存的一种度量，不仅能够反映变量之间的线性关系，还能反映变量之间的非线性关系，因此是一种非线性的分析方法；对于初选因子的特征提取方法，有 PCA 和 KPCA 两种方法。PCA 忽略了变量之间高于 2 阶的相互关系，得到的主成分是由线性映射生成的，因此 PCA 实质也是一种线性映射方法。KPCA 将原始空间中变量的数据通过核函数映射到高维空间，使得在低维空间中线性不可分的数据在高维空间中能线性可分，然后再在高维空间做 PCA。因此 KPCA 是一种非线性的方法；对于预报模型的选择，常用的回归分析也是一种线性分析方法，而 RNN 是一种非线性映射模型，还有鲁棒性好、自学习能力强的特点。

方案 1、2、3 与 4 之间通过控制变量法，每次改变一个技术要数，比较不同预报技术体系之间的优劣。方案 1 用相关分析法进行预报因子初选，用 PCA 进行初选因子的特征提取，用回归分析作为预报模型。然而这三种方法都是线性的分析方法，只能反映径流量与预报因子之间的线性关系，对于线性关系外的非线性关系则无能为力；方案 2 与方案 1 不同的是，将回归分析预报模型替换成了 RNN。虽然 RNN 具有非线性映射能力，但是因子的初选和初选因子的特征提取方法都是线性的分析方法，这样提取出来的特征也只能反映与径流量之间的线性关系，与径流量具有非线性关系的那些特征都缺失了，如果将这些提取出来的特征作为 RNN 的输入，即使 RNN 具有非线性映射能力，也无法挖掘出径流量与预报因子之间的非线性关系，因此方案 2 本质也相当于是一种线性预报体系；方案 3 在方案 2 的基础上，把预报因子的初选技术由相关分析换成了互信息，其他的不变。互信息是一种非线性的因子初选方法，能够充分挖掘出预报因子之间的相互关系。但是为了对初选的因子进行降噪、去冗余，方案 3 采用的特征提取方法是 PCA，PCA 实质是一种线性的映射方法，基于这种方法提取的特征不是最优的，会丢失部分有用的信息。因此，将这些提取的特征输入 RNN，最终的预报效果也会受到不好的影响。相比于方案 2，方案 3 的因子提取方法和预报模型都有非线性映射的能力，能够反映一部分径流量与预报因子之间的非线性关系，

因此方案 3 在方案 2 的基础上有所改进；在方案 3 的基础上，方案 4 把初选因子的特征提取技术由 PCA 换成了 KPCA。KPCA 具有非线性映射能力，能够充分挖掘因子的特征，并且提取的特征相互正交，没有复共线性。把这样的特征作为 RNN 的输入，在尽量保留原有的因子与径流量之将的关系同时，还能够大大提高 RNN 的泛化能力，提高预报的精度。因此，方案 4 是一个真正具有非线性映射能力的技术体系；为了消除 RNN 模型参数不确定性带来的不利影响，在方案 4 的基础上，方案 5 用多个 RNN 集合预报的方式代替单一 RNN 预报，提高预报的精度。

综合以上分析，与方案 1、2、3 及 4 相比，方案 5 是一种最好的技术体系。因此，本研究将采用方案 5 来构建中长期径流智能预报技术体系。

3.6 预报结果的精度评价方法

本研究以平均绝对百分误差((MAPE , Mean Absolute Percentage Error)、相对误差 (RE , Relative Error)、最大相对误差(MRE , Maximum Relative Error)、合格率 (QR , Qualified Rate) 为预报的评价指标。各指标的计算公式如下：

$$MAPE = \frac{1}{T_x} \sum_{t=1}^{T_x} \frac{|\hat{y}^{<t>} - y^{<t>}|}{y^{<t>}} \quad (3-25)$$

$$RE = \frac{\hat{y}^{<t>} - y^{<t>}}{y^{<t>}} \times 100\% \quad (3-26)$$

$$MRE = \max_{1 \leq t \leq T_x} \frac{|\hat{y}^{<t>} - y^{<t>}|}{y^{<t>}} \times 100\% \quad (3-27)$$

$$QR = \frac{T_{\text{qual}}}{T_{\text{total}}} \times 100\% \quad (3-28)$$

其中 $\hat{y}^{<t>}$ 为径流量的预测值, $y^{<t>}$ 为对应的实测径流值, T_x 为检验样本的个数；

T_{qual} 为合格的预报次数, T_{total} 为总的预报次数。

根据《水文情报预报规范》(GI3/T22482-2008)^[59]中评定中长期径流预报精度的方案，本研究以预报的最大相对误差小于 20%的预报为合格预报，模型的表现采用合格率来评估，评估标准如下：

表 3-3 预报精度等级

合格率	精度等级
$QR \geq 85.0\%$	甲
$85.0\% > QR \geq 70.0\%$	乙
$70.0\% > QR \geq 60.0\%$	丙

3.7 本章小结

本章首先从预报因子的初选、初选因子的特征提取以及中长期径流智能预报模型构建三个方面介绍了中长期径流智能预报技术体系的构建。接着分析了 5 种不同的技术体系的优劣，通过分析比较，最终选取方案 5 作为中长期径流智能预报技术体系。最后介绍了中长期径流预报结果的精度评价方法。

第四章 中长期径流智能预报系统设计与实现

4.1 系统总体架构设计思路

中长期径流智能预报系统遵循科学性、实用性、规范性和可扩充性原则^{[46][47][48][49]}进行设计，基于 B/S 模式进行开发。系统总体架构如图 3-2 所示，按照运行逻辑分三层：数据层、服务层和应用层^[50]。数据层执行数据存储和检索任务，为服务层提供透明的数据访问。具体包括设计数据库结构，存储中长期径流数据、因子数据、制作的方案及预报的结果，执行数据的增删改查等；服务层实现系统的数据处理和业务规则，提供的服务可以为所有的应用程序使用。具体包括预报因子的计算分析、预报模型的参数率定等；应用层负责应用调度并向客户端发送结果，实现数据输入和显示功能，并在一定程度上允许对数据进行编辑^[51]。具体包括制作预报方案时相关数据的输入、制作好的各种预报方案的展示以及具体预报事件结果的可视化等。

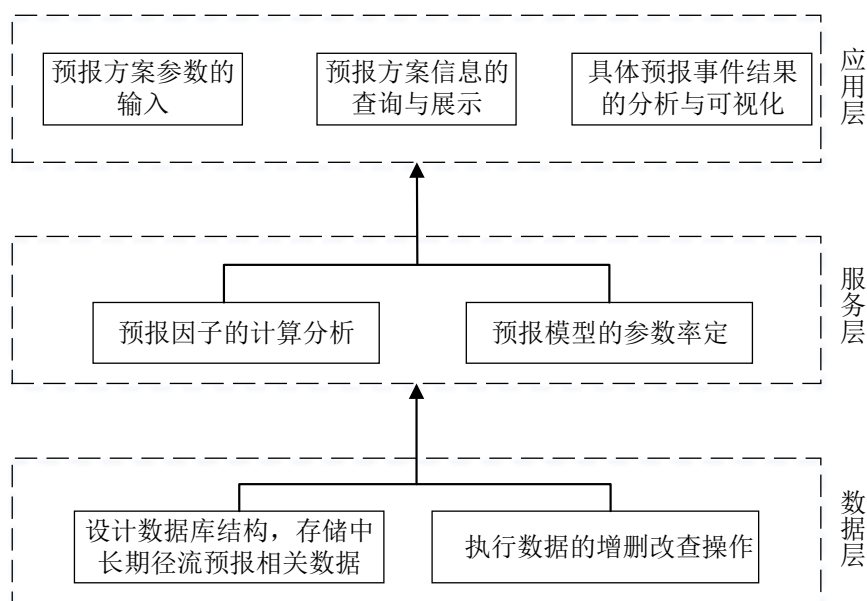


图 4-1 系统总体架构

根据数据库总体架构，本研究将从数据库设计、预报方案库设计、预报模型库设计、人机交互界面设计四个部分来构建中长期径流智能预报系统。

4.2 数据库的设计

数据库设计是中长期径流智能预报系统设计的重要环节，本研究在商业数据

库软件系统平台基础上，选择 SQL Server2008 作为数据库管理系统。下面从数据库的设计原则、数据库的总体架构及数据库的表结构设计三方面介绍数据库的设计。

4.2.1 数据库的设计原则

为提高软件的开发质量和效率，针对中长期径流智能预报系统的实际情况，在数据库建设中遵循以下的原则^{[52][53]}：

第一，规范命名：所有的库名、表名必须遵循统一的命名规则，并进行必要说明，以方便设计、维护、查询。

第二，数据类别界定清晰，定义明确（无二义性），字段说明翔实：术语的采用、定义尽可能符合有关国家和行业规范，对没有国家或行业标准的自行设计。

第三，保证数据结构化、规范化、编码标准化。

第四，数据具有独立性和可扩展性：尽量做到数据库的数据具有独立性，独立于应用系统，使数据库的设计以及其结构的变化不影响程序，反之亦然。应用系统在不断变化，数据库设计时应考虑到其扩展接口，使得系统增加新的应用或新的需求时，不至于引起整个数据库体系的重新改写。

第五，保证数据的安全可靠：数据库是整个系统的核心，它的设计要保证其可靠性和安全性，不能因数据库的临时故障而导致整个系统的瘫痪。

4.2.2 数据库的总体架构

数据库存储着与中长期径流预报相关的各种数据，为了弄清楚这些数据的关系，设计了中长期径流预报 E-R 图(Entity Relationship Diagram)，如图 4-2 所示。

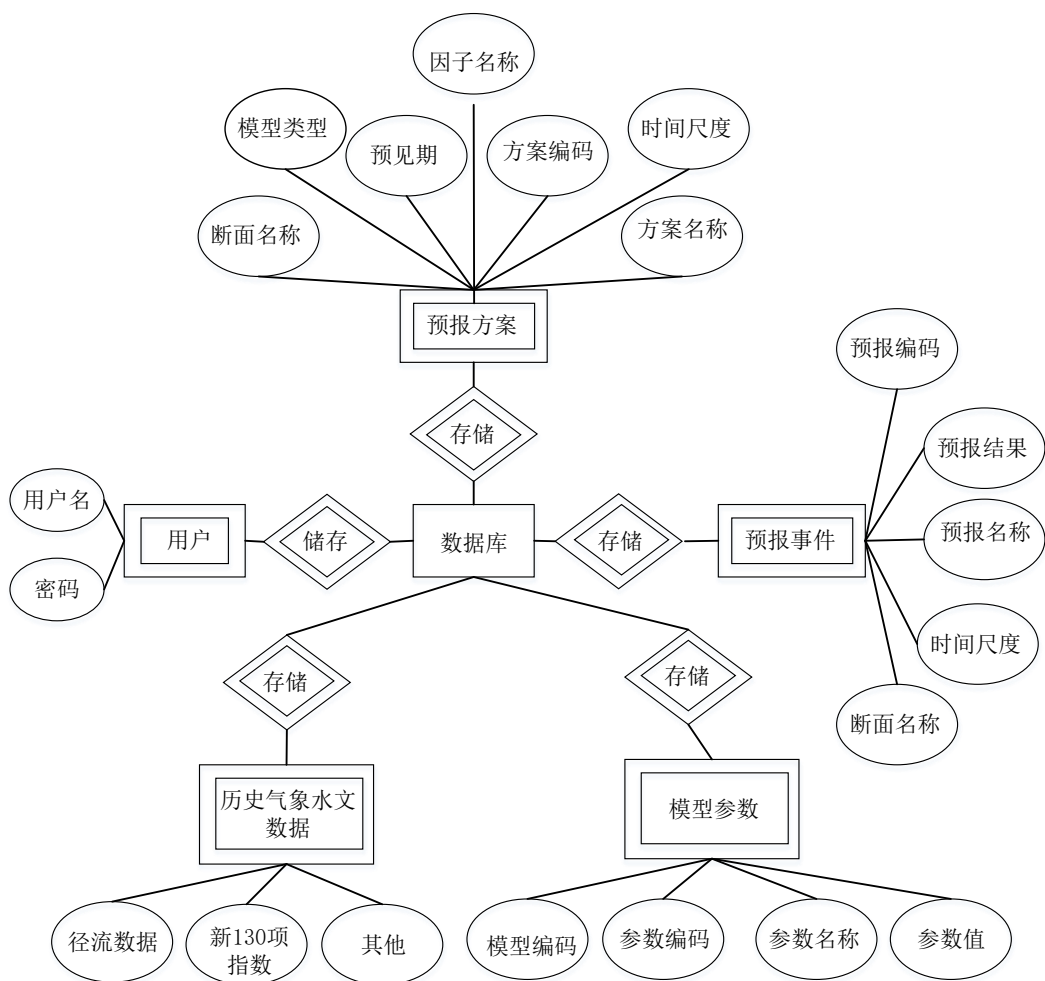


图 4-2 中长期径流预报 E-R 图

4.2.3 数据库的表结构设计

根据 E-R 图的实体、属性和关系，就可以设计数据库的表结构。设计的表可以分为 7 大类:用户信息表、预报断面基本信息表、历史气象水文数据表、模型信息表、预报方案相关信息表、预报事件信息表、预报因子信息表。下面分别介绍。

(1) 用户信息表：用户信息表存储用户的用户名和密码。

表 4-1 用户信息表

序号	列名	数据类型	是否主键	备注
1	Id	numeric	自动编码	
2	User	nvarchar(50)	主键	用户名
3	password	nvarchar(50)		用户密码

(2) 预报断面基本信息表：该表记录了哪些断面有预报任务，不同的研究

区之间不共享预报断面。一个预报断面只有一个预报输出，使用 `dataTypeCode` 对预报输出的数据类型进行规定。

表 4-2 预报断面基本信息表

序号	列名	数据类型	是否主键	备注
1	Id	numeric	自动编码	
2	sectionCode	nvarchar(50)	主键	预报断面的编码
3	sectionName	nvarchar(50)		预报断面名称
4	RDField	nvarchar(50)		研究区标识
5	stationCode	nvarchar(50)		站点编码
7	dataTypeCode	nvarchar(50)		输出的数据类型
8	remark	nvarchar(MAX)		备注

(3) 历史气象水文数据表：历史气象水文数据表是一大类，具体包含历年逐月的径流量数据表、新 130 项指标数据表。

表 4-3 历年逐月的径流量数据表

序号	列名	数据类型	是否主键	备注
1	id	numeric		自增长
2	stationCode	nvarchar(50)	联合主键	站点编码
3	DT	datetime	联合主键	时间
4	Q	float		流量观测值

表 4-4 新 130 项指标数据表

序号	列名	数据类型	是否主键	备注
1	id	numeric		自增长
2	indexCode	nvarchar(50)	联合主键	因子编码
3	indexName	nvarchar(50)		因子名称
4	DT	datetime	联合主键	时间
5	Q	float		因子观测值

(4) 模型信息表：模型信息表是一大类，具体包含模型的基本分类信息表、模型的注册信息表、模型信息说明表。模型的基本分类信息表中记录中长期预报模型的基本分类信息，如中长期预报模型分为智能预报模型、遥相关预报模型、数值预报模型等；模型的注册信息表中记录中长期预报模型的注册信息；模型信息说明表是指当用户编写了一个模型，并且针对一个区域建完模之后，应该在这

张中将模型参数的基本信息进行说明，主要包括参数编码，模型中用于设置该参数的函数名称，模型中用于获取该参数的函数名称，该模型参数取值的最小值，最大值，默认值，以及是否是率定参数等。

表 4-5 模型的基本分类信息表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	modelTypeCode	模型类型编码	nvarchar(50)	主键
3	modelTypeName	模型类型名称	nvarchar(50)	
4	remark	备注	nvarchar(MAX)	

表 4-6 模型的注册信息表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	modelCode	模型编码	nvarchar(50)	主键
3	modelName	模型名称	nvarchar(50)	
4	modelTypeCode	模型类型编码	nvarchar(50)	外键
5	RDField	研究区编码	nvarchar(50)	
6	modelFormat	模型类型	nvarchar(50)	外键
7	registerDT	注册时间	datetime	
8	author	作者（注册人）	nvarchar(50)	作者或者注册人
9	DBUsageModeCode	数据库使用模式	nvarchar(50)	外键
10	stateCode	模型状态编码	nvarchar(50)	外键
11	remark	备注	nvarchar(MAX)	备注

表 4-7 模型信息说明表

序号	列名	数据类型	是否主键	备注
1	id	numeric	自动编码	
2	modelCode	nvarchar(50)	联合主键	模型编码
3	paramCode	nvarchar(50)	联合主键	参数编码
4	paramName	nvarchar(50)		参数名称
5	setClassName	nvarchar(50)		参数设置函数名
7	getClassName	nvarchar(50)		参数获取函数名
8	paramMin	float		参数最小值
9	paramMax	float		参数最大值
10	paramDefault	float		参数默认值
11	calibratable	nchar(10)		是否是率定参数
12	remark	nvarchar(MAX)		备注

(5) 预报方案相关信息表：预报方案相关信息表是一大类，具体包含预报方案的相关信息表、预报方案的相关信息缓存表、预报方案因子选择表、预报方案对应的模型参数表、预报方案检验结果表。预报方案的相关信息表存储预报方案的编码、名称、预报断面编码、预报时间尺度等；预报方案的相关信息缓存表存储模型训练集相关参数及率定算法的相关参数；预报方案因子选择表存储中长期径流预报中，某一个预报方案所选择的预报因子；预报方案对应的模型参数表存储某一个预报方案下的模型取得的具体参数值；预报方案检验结果表存储该方案对径流量试预报的检验结果。

表 4-8 预报方案的相关信息表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	forecastSchemeCode	预报方案编码	nvarchar(50)	
3	forecastSchemeName	预报方案名称		
4	sectionCode	预报断面编码		外键
5	timeScale	预报时间尺度		
6	modelCode	模型编码	nvarchar(50)	外键
7	indexSetCode	指数集合编码	nvarchar(50)	外键
8	paramSetCode	参数集合编码	nvarchar(50)	外键
9	remark	备注	nvarchar(MAX)	备注

表 4-9 预报方案的相关信息缓存表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	forecastSchemeCode	预报方案编码	nvarchar(50)	主键
3	status	方案状态	Int	0、1、2、3、4
4	beginSampleDt	径流样本起始年	Int	例如：1954
5	endSampleDt	径流样本结束年	Int	例如：2017
6	beginTrainDt	径流训练集开始年份	nvarchar(50)	例如：2015
7	EndTrainDt	径流训练集结束年份	nvarchar(50)	例如：2015
8	beginVerifyDt	径流检验集开始年份	nvarchar(50)	例如：2015
9	endVerifyDt	径流检验集结束年份	nvarchar(50)	例如：2015
10	foresightPeriod	预见期	Int	单位：月份
11	posPCoount	Pso 种群数	Int	
12	psoMaxIterations	Pso 最大迭代次数	Int	
13	psoMaxObjectValue	Pos 目标函数最大值	Float	
14	createDt	该方案的创建时间	Datetime	

表 4-10 预报方案因子选择表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	longTermForecastSchemeCode	预报方案编码	nvarchar(50)	联合主键
3	climateMonthIndexCode	气象月因子编码	nvarchar(50)	联合主键
4	headTime	提前时间量	nchar(10)	
9	remark	备注	nvarchar(MAX)	

表 4-11 预报方案对应的模型参数表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	longTermForecastSchemeCode	预报方案编码	nvarchar(50)	联合主键
3	paramCode	参数编码	nvarchar(50)	联合主键
4	paramValue	参数值	nchar(10)	

表 4-12 预报方案检验结果表

序号	列名（代码）	名称	数据类型	备注
1	longTermForecastSchemeCode	预报方案编码	nvarchar(50)	联合主键
2	Year	年份	nvarchar(50)	联合主键
3	realValue	实测值	nchar(10)	
4	ForecastValue	预报值	nvarchar(MAX)	
5	relativeError	相对误差	nvarchar(MAX)	绝对值

（6）预报事件信息表：预报事件信息表是一大类，具体包含预报事件信息表、预报事件对应方案信息表、预报事件结果表。预报事件信息表存储中长期径流预报事件的信息；预报事件对应的方案信息存储表某一具体预报事件对应的方案信息；预报事件结果表存储某一预报事件的预报值。

表 4-13 预报事件信息表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	longTermForecastCode	预报编码	nvarchar(50)	联合主键
3	longTermForecastName	预报名称	nvarchar(50)	
4	sectionCode	预报断面编码	nvarchar(50)	外键
5	RDField	研究区编码	nvarchar(50)	
6	timeListCode	预报时间列表编码	nvarchar(50)	外键

表 4-14 预报事件对应方案信息表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	timeListCode	时间列表编码	nvarchar(50)	主键
3	longTermForecastCode	预报编码	nvarchar(50)	
4	DT	预报时间	datetime	
5	timeScale	时间尺度	nvarchar(50)	
6	longTermForecastSchemeCode	预报方案编码	nvarchar(50)	

表 4-15 预报事件结果表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	longTermForecastCode	预报编码	nvarchar(50)	联合主键
3	DT	预报时间	datetime	
4	timeScale	时间尺度	nvarchar(50)	
5	value	预报值	float	

（7）预报因子信息表：预报因子信息表是一大类，具体包含预报因子集合表、模型默认选择的因子表。预报因子集合表存储某预报方案可选的因子样本集合；模型默认选择的因子表存储中长期径流预报中，某一个模型所选择的默认预报因子。

表 4-16 预报因子集合表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	longTermForecastSchemeCode	预报方案编码	nvarchar(50)	联合主键
3	climateMonthIndexCode	气象月因子编码	nvarchar(50)	联合主键
4	headTime	提前时间量	nchar(10)	
5	isRecoment	是否选定	nvarchar(MAX)	1 选 0 不选

表 4-17 模型默认选择的因子表

序号	列名（代码）	名称	数据类型	备注
1	id	自动编号	numeric(18, 0)	自增长
2	modelCode	模型编码	nvarchar(50)	联合主键
3	climateMonthIndexCode	气象月因子编码	nvarchar(50)	联合主键
4	headTime	提前时间量	nchar(10)	
5	isRecoment	备注	nvarchar(MAX)	是否默认选中
6	remark	备注	nvarchar(MAX)	备注

4.3 预报方案库的设计

预报方案设计包含两部分,一部分是预报方案的设计,另一部分是方案管理的设计。预报方案是指在系统中进行一次径流预报的流程。该流程包含基本信息的输入、预报对象创建、模型选择、因子选择及模型训练与检验五个部分;预报方案管理是指对用户制作的方案进行管理,包括查看或者删除等。下面分别进行介绍。

(1) 方案基本信息的输入

创建预报方案,首先需要用户输入方案编码及方案名称。其中方案编码是指径流预报方案的唯一标识性信息。方案编码由 14 为数字和一位字母“T”组成,格式为:YYYYMMDDTHHFFMM。前四位 YYYY 代表当下进行预报方案制定的年份,MM 为当前月份,DD 为当前日期,HH 为当前小时,FF 为当前分钟,MM 为当前秒钟。方案名称用户可以自行设计,但最好做到见名知意,说明方案的一些基本信息。

(2) 预报对象创建

创建预报对象,首先需要确定预报的断面。预报的断面是指具体针对什么地区或者流域进行预报,例如漓江、汉江等。确定预报的断面后,需要根据预报的要求,选择年尺度的预报或者月尺度的预报,确定预报的时间尺度。

(3) 模型选择

创建预报方案,模型的选择是很重要的一部分。中长期径流智能预报系统采用循环神经网络模型作为预报模型。但是考虑到系统的可扩展性,系统也为其他模型的接入提供了接口。

(4) 因子选择

选择预报因子,首先需要确定从哪些水文气象资料中来选择。中长期径流智能预报系统数据库中存储了历年逐月的 130 项指数时间序列数据,用户可以输入采用的指数时间序列的起止年份,然后选择预见期,系统会自动分析所选的时间序列与对应的径流时间序列的相关性,对因子进行初选并进行特征提取。最后选定因子会在人机交互界面进行展示,用户可以对因子进行进一步修正。如果用户没有进行因子的筛选,系统也会提供一套默认的因子。

（5）模型训练与检验

预报因子确定后，就可以对模型进行参数率定及预报效果检验。对于循环神经网络模型，就是对模型进行训练及检验。系统会采用默认模型参数或者用户输入的模型参数对模型进行训练，并对训练的模型进行检验，然后分析检验结果，筛选出最优的网络结构。

至此，针对特定的预报对象，循环神经网络模型的结构和参数都确定了，一个方案创建完成，保存方案到预报方案库即可。

（6）预报方案的管理

用户创建了很多方案后，可以通过人机交互界面，到方案库对创建的方案进行管理，具体包括各方案的试预报结果查看，或者对不满意的方案进行删除等。

4.4 预报模型库的设计

模型库是计算机中按照一定组织结构形式存储的多个模型的集合体，是中长期径流智能预报系统的核心^{[54][55]}。下面从模型库的设计要求及模型的参数设置对模型库的设计进行介绍。

（1）模型库的设计要求

模型库应遵循系统开放性、可扩充性及较长的生命周期的开发原则，可对预报模型以“黑匣子”方式进行增加、删除及修改管理^[56]。中长期径流智能预报系统中所用的模型和方法都按照统一接口进行开发，经整理规范化后接入模型库，使系统能够适应环境的变化要求（新增和更新预报模型、实时校正模型等），并实现对多种模型的综合管理、模型选择、模型的分步运行和多种方案的比较。

（2）预报因子遴选模型

预报因子的遴选模型采用 Java 语言进行开发，包含因子初选和初选因子特征提两个部分。对于因子初选部分，模型采用基于互信息的因子选择方法；对于初选因子特征提取的部分，模型采用核主成分分析法。用户只需设置用于筛选因子的指数时间序列的起止年份以及预见期，预报因子遴选模型会自动提取出初选的预报因子的特征，然后采用试错法，通过依次选取 1 到 n 个特征作为神经网络的输入，然后比较预报效果，确定神经网络输入特征的个数。用户可以设置 n 的数值，系统默认为 5 个。

(3) 基于循环神经网络的单一预报模型

循环神经网络单一预报模型采用 **Matab** 语言, 根据中长期径流智能预报系统预先定义的接口进行开发, 模型开发完成后, 编译成二进制文件, 接入中长期径流智能预报系统。

对循环神经网络模型进行训练前, 用户可以通过系统设置的模型参数包括成本函数、正则化方法、参数更新的算法、隐含层节点的个数, 如果用户没有设置这些参数, 系统会采用默认的参数。系统提供的成本函数包括均方误差函数、对数损失函数, 系统默认的成本函数为均方误差函数; 系统提供的正则化方法包括 $L1$ 正则化、 $L2$ 正则化, 系统默认的正则化方法为 $L2$ 正则化方法; 系统提供的参数更新的算法包括梯度下降算法、动量梯度下降算法、RMSprop 算法以及 Adam 优化算法, 系统默认的为 Adam 优化算法, 算法里的超参数都是用默认值; 对于输入层节点数和隐含层节点数, 系统采用试错法确定其数目。首先, 用户需要设置最多可以提取的主成分的个数 n 以及隐含层最大可以采取的节点数目 m 。 n 的系统默认值为 5, m 的系统默认值为 30。如果最多可以提取的主成分的个数被设置成 5 个, 隐含层最大可以采取的节点数目被设置为 15 个, 则系统会循环遍历主成分数分别为 1、2、3、4、5 和隐含层节点数分别为 2、3、...、15 的两两组合方案, 通过比较各方案检验期的预报结果, 确定输入层节点数和隐含层节点数。

(4) 基于循环神经网络集合预报模型

针对某一特定的预报对象, 单一循环神经网络模型训练好后, 用户可以重复该步骤, 运用筛选的因子训练多个循环神经网络模型, 然后将多个模型的预报结果进行平均, 作为最终的预报结果。

4.5 人机交互平台设计

人机交互平台是用户和计算机的接口, 在操作者、模型库、数据库和方案库之间起着传送、转换命令和数据的重要作用, 其设计应当遵循一定的原则, 其核心是人机界面^[57]。在中长期径流智能预报系统中, 使用 **Adobe Flex** 开发了四个子界面, 分别是预报方案制作界面、预报方案管理界面、径流预报管理界面及径流预报界面。下面分别进行介绍。

4.5.1 人机交互平台设计原则

中长期径流预报系统的人机交互平台以友好性、一致性与可靠性为原则进行设计。友好性是指用户在操作系统时主观操作复杂性低，系统容易使用。要做到这一点，需要对界面各窗口或者按钮进行合理的布局，显示方式简洁明了；一致性是指界面中体现的中长期径流预报流程应该尽量和人的思维方式一致，其次，不同的界面之间，应该有相似的交互显示风格和相似的人机操作方式。一致性好的交互系统，可以帮助用户把他们当前知识经验推广应用到新界面、新操作中去，从而减轻用户重新学习和记忆的负担；可靠性是指系统正常工作的能力。交互系统应该能使用户正确的使用系统，保证有关程序和数据的安全，同时，系统应该具有一定的容错能力^[58]。

4.5.2 中长期径流智能预报系统界面设计

（1） 预报方案制作界面

图 4-1 为预报方案制作界面。预报方案制作界面是根据 4.3 节中介绍的预报方案库的设计来编写开发的。在该界面，通过对方案基本信息、预报对象、预报模型、预报因子以及模型参数的设置，完成模型的参数率定，并能够对训练好的模型进行试预报，检验模型的预报效果，预报的结果以图像化的方式显示在界面的正下方。如果预报效果良好，用户可以点击创建方案按钮，即可把预报方案保存到方案库。



图 4-1 预报方案制作界面

(2) 预报方案管理界面

图 4-2 为预报方案管理界面。预报方案管理界面是根据 4.3 节中介绍的预报方案管理来编写开发的。在该界面,用户可以通过输入与预报方案相关的信息(如预报名称、开始时间、结束时间、断面等)来查找已经保存的预报方案。预报方案列表显示了所有满足条件的预报方案,方案结果对应的框图则会显示该方案对应模型相关的信息(如模型编码、参数编码、参数名称、参数的大小等等)以及模型的试预报结果。



图 4-2 预报方案管理界面

（3）径流预报界面

图 4-3 为径流预报界面。在径流预报界面选择需要预报的断面、时间尺度、及预报的时间后，下拉列表中会自动显示预报方案库中满足条件的预报方案，用户选择预报方案后，点击开始预报按钮，即可进行预报。预报的结果会分两大块在界面的右边显示。右上部分是预报结果的图形化显示，右下部分则是对应的预报结果表格，详细的记录了预报的年份、预报方案的编码、预报的名称、预报的径流值及与其他时间径流值的比较分析等。



图 4-3 径流预报界面

（4）径流预报管理界面

图 4-4 为径流预报管理界面。在该界面通过输入与每一次具体预报事件相关的信息（如方案名称、径流预报开始时间、径流预报结束时间、断面等），即可查看对应的预报结果。预报列表选项展示了每一次预报事件的详细信息，点击预报结果按钮，即可查看对应预报事件的预报结果。

第五章 中长期径流智能预报系统应用实例

5.1 研究区域概况

本章采用中长期径流预报系统,对锦屏一级水电站年平均径流量进行预报。锦屏一级水电站水库位于四川省雅砻江流域内,是雅砻江干流下游河段的龙头水库。水库库容约 77.6 亿 m^3 。水库所属流域地形复杂,高山河谷交错。流域内的气候主要受到西南季风和高空西风环流的影响。所建的水库电站主要用于发电,兼有蓄水、拦沙和防洪的作用。对水电站水库年平均径流量的准确预报,有利于指导该地区水资源综合开发利用、科学管理和优化调度^[60]。

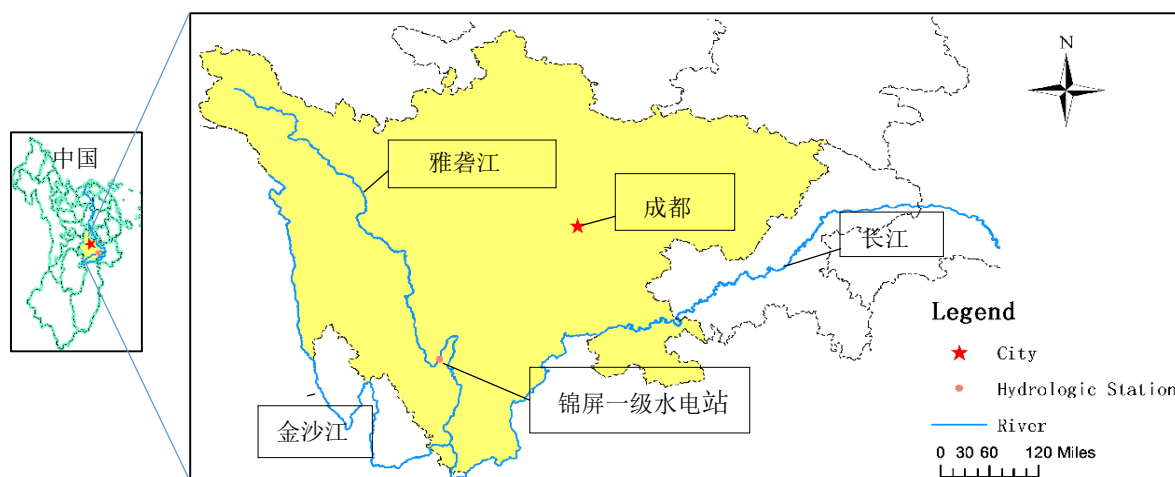


图 5-1 雅砻江流域主要河流概况图

5.2 预报对象与数据来源

本应用实例的预报对象是锦屏一级水电站年平均径流量。采用的数据资料包括锦屏一级水电站水库断面 1960~2011 年的逐年年平均径流量数据(中国水利水电科学研究院提供)和 1959~2010 年的逐月 130 项指数数据(国家气候中心提供)。系统将采用 1960~1999 年的数据对网络模型进行训练,2000~2011 年的数据对网络模型进行检验。

5.3 实验数据的预处理

由于是对年平均径流量预报,因此因子不能从同期当年的时间内选择,同时,考虑到气象因子对径流的影响存在滞后性,所以建立锦屏一级水电站逐年(1960~2011 年)年平均径流量与前一年(1959~2010 年)逐月的 130 项指数

的一一对应关系。其中某一项指数时间序列与径流时间序列的对应关系如表 5-1 所示，其他指标类似。

表 5-1 某指数时间序列与径流时间序列的对应关系

年径流时间序列		某一项指数时间序列		
1960 年径流量	1959 年 1 月数据	1959 年 2 月数据	...	1959 年 12 月数据
1961 年径流量	1960 年 1 月数据	1960 年 2 月数据	...	1960 年 12 月数据
1962 年径流量	1961 年 1 月数据	1961 年 2 月数据	...	1961 年 12 月数据
...
2009 年径流量	2008 年 1 月数据	2008 年 2 月数据	...	2008 年 12 月数据
2010 年径流量	2009 年 1 月数据	2009 年 2 月数据	...	2009 年 12 月数据
2011 年径流量	2010 年 1 月数据	2010 年 2 月数据	...	2010 年 12 月数据

在中长期径流预报系统的预报方案制作界面，用户只需设置用于训练的径流时间序列的起止时间、时间尺度以及预见期，预报因子遴选模型则会根据这些参数自动完成上述数据的预处理工作。

5.4 预报因子的遴选

5.4.1 基于物理机制的因子初选

锦屏一级水电站年平均径流量受到的影响因数众多,如太阳活动、大气环流、下垫面情况、地球及其他天文物理因素、人类活动、前期气象水文要素等。国家气候中心提供的新的 130 项指数（包括 88 项历年逐月的大气环流指数、26 项历年逐月温度指数数据以及 16 项其他指数的历年逐月数据）基本涵盖了前期主要的气象水文要素，为中长期径流预报提供了方便。本研究将直接从这新的 130 项指数中筛选预报因子。中长期径流智能预报系统会自动从国家气候中心下载这些数据，保存到数据库并定时更新。

5.4.2 基于互信息的因子二次筛选

对于初选的因子，选用互信息（MI, Mutual Information）的方法对因子二次筛。选出通过显著性检验且标准互信息（NMI, Normalized Mutual Information）大于 0.9 的指标作为二次筛选的因子。本应用案例中，标准互信息大于 0.9 的指标有 205 个，前 20 个指标的信息如下：

表 5-2 前 20 个二次筛选的预报因子

初选的因子	NMI	MI
8 月太阳黑子	0.988375	5.426929
4 月太阳黑子	0.988375	5.426929
7 月太阳黑子	0.988375	5.426929
10 月太阳黑子	0.988375	5.426929
12 月太阳黑子	0.988375	5.426929
2 月太阳黑子	0.98444	5.384376
9 月太阳黑子	0.98444	5.384376
11 月太阳黑子	0.98444	5.384376
1 月太阳黑子	0.98444	5.384376
3 月太阳黑子	0.98444	5.384376
5 月太阳黑子	0.98444	5.384376
8 月北半球副高强度指数(5E-360)	0.980474	5.341823
3 月北半球极涡面积指数(5 区,0-360)	0.980474	5.341823
6 月北非大西洋北美副高强度指数(110W-60E)	0.976477	5.299270
6 月北半球副高强度指数(5E-360)	0.976291	5.256717
4 月北半球副高强度指数(5E-360)	0.972448	5.256717
7 月北非大西洋北美副高强度指数(110W-60E)	0.972448	5.256717
9 月北非大西洋北美副高强度指数(110W-60E)	0.972448	5.256717
6 月太阳黑子	0.972448	5.256717
6 月太平洋副高强度指数(110E-115W)	0.970919	5.240655

5.4.3 基于 KPCA 的因子特征提取

初选出的 205 个因子不宜直接作为神经网络的输入。一方面,因子数目过多,这些因子如果直接作为神经网络的输入,会导致网络的权重矩阵过大,不易训练;另一方面,这 205 个因子之间往往存在多重共线性,重复的信息和噪声会直接影响神经网络的泛化能力。由于这两方面的原因,因此需要对这 205 个因子进行特征提取、降噪去冗余。本实例选用核主成分分析法进行特征提取,核函数选用径向基核函数。调用系统的预报因子遴选模型,计算得到的主成分按照方差贡献率的值由大到小的顺序排列,其中前 5 个主成分的方差贡献率如表 5-3,对应的前 5 个主层分的数据如表 5-4。

表 5-3 前 5 个主成分的方差贡献率

主成分	主成分_1	主成分_2	主成分_3	主成分_4	主成分_5
方差贡献率	25.7%	6.9%	5.6%	5.1%	3.9%

表 5-4 KPCA 提取的前 5 个主成分

样本序号	主成分_1	主成分_2	主成分_3	主成分_4	主成分_5
1	-0.00605	-0.025852	0.007556	0.013533	0.017703
2	-0.02725	-0.006741	0.001084	0.017781	0.009853
3	0.007912	0.0031653	-0.00755	0.007487	-0.02185
4	0.002089	0.0131535	0.006936	-0.00139	-0.00235
⋮	⋮	⋮	⋮	⋮	⋮
48	-0.05354	0.0081831	0.013478	-0.03407	0.018846
49	-0.04108	0.0216432	0.00613	-0.01603	-0.02593
50	-0.01539	0.0190786	0.005982	-0.02812	0.013842
51	-0.03675	0.0243891	0.00997	-0.00845	0.013842
52	-0.07989	0.0283544	-0.04816	0.007058	-0.00822

从表 5-3 中可以看出，第一个主成分的方差贡献率已经达到了 25.7%，包含了所选因子的大部分信息。其他主成分的方差贡献率越来越小，包含所选的因子的信息也越来越少。值得注意的是，为了训练样本和检验样本提取主成分时所用的标准是一致的，需要将训练样本序列和检验样本序列组合在一起进行 KPCA。本实施例中，训练样本序列的长度为 47，检验样本序列的长度为 5，序列样本和检验样本序列组合的长度为 52，因此，表表 5-3 中提取的主成分的序列长度为 52。

5.5 预报模型的构建

构建基于循环神经网络的单一预报模型，需要确定成本函数、正则化方法、参数更新算法及网络结构。对于网络的结构，神经网络的输入层、输出层都是一层，隐含层也采用一层，因此还需要确定输出层节点数、输入层节点和隐含层节点数。单个神经网络结构确定好后，就可以运用多个单一预报模型，实现对中长期径流的集合预报。

5.5.1 成本函数、正则化方法、参数更新算法的确定

在第 3.5 节分析并比较了不同预报方案的优劣，最终确定方案 5 为中长期径流预报方案，即基于“互信息-KPCA-多个 RNN”的预报方案，该方案也是中长期径流预报系统所采用的系统方案，因此成本函数、正则化方法、参数更新算法均采用系统默认的设置，即成本函数采用均方误差函数，正则化方法采用 L2 正

则化，参数更新算法采用 Adam 算法。

5.5.2 网络结构的确定

（1）输出层节点的确定

因为是对锦屏一级水电站年平均径流量的单值预报，所以输出节点数为 1。

（2）输入层节点数和隐含层节点数的确定

隐含层节点数对神经网络的泛化系能有很重要的影响。而隐含层节点数又会受到输入节点数的影响，输入节点数改变，隐含层节点数也会相应的改变。因此，输入层节点数和隐含层节点数需要一起确定。本应用案例将采用试错法确定输入层节点数和隐含层节点数。实验结果及分析介绍如下。

将主成分数据序列和对应的径流时间序列的前 47 年的数据作为训练样本，后 5 年的数据作为检验样本。以前 3 个主成分与 15 个隐含层节点的不同组合为网络结构的模型的预报表现如表 5-5 所示。

表 5-5 以第一主成分作为网络模型输入时不同隐含层节点对应的网络的表现

编号	结构	训练	检验		
		MAPE	MRE	MAPE	QR
1	1-2-1	0.160	0.272	0.133	60%
2	1-3-1	0.161	0.257	0.129	60%
3	1-4-1	0.160	0.255	0.128	60%
4	1-5-1	0.160	0.260	0.129	60%
5	1-6-1	0.156	0.348	0.146	60%
6	1-7-1	0.155	0.371	0.146	60%
7	1-8-1	0.154	0.375	0.149	60%
8	1-9-1	0.153	0.382	0.152	60%
9	1-10-1	0.154	0.362	0.145	60%
10	1-11-1	0.154	0.363	0.147	60%
11	1-12-1	0.156	0.330	0.143	60%
12	1-13-1	0.154	0.357	0.148	60%
13	1-14-1	0.154	0.369	0.149	60%
14	1-15-1	0.162	0.286	0.139	60%

表 5-5 为以主成分_1 作为神经网络输入，隐含层节点数分别为 2 到 15 时，网络模型的表现。从表格 5-5 可以看出，当以主成分_1 作为神经网络输入时，随着隐含层节点数的增加，网络验证期的预报最大相对误差均大于 20%，合格率均只有 60%，其中预报效果最好的模型是 Model_3，检验期的预报最大相对误差为 25.5%，合格率为 60%，预报效果欠佳。

表 5-6 以第一、二主成分作为网络模型输入时不同隐含层节点对应的网络的表现

编号	结构	训练	检验		
		MAPE	MRE	MAPE	QR
1	2-2-1	0.155	0.295	0.138	60%
2	2-3-1	0.148	0.216	0.115	80%
3	2-4-1	0.131	0.289	0.142	80%
4	2-5-1	0.138	0.158	0.078	100%
5	2-6-1	0.132	0.170	0.086	100%
6	2-7-1	0.132	0.178	0.090	100%
7	2-8-1	0.136	0.165	0.080	100%
8	2-9-1	0.135	0.127	0.069	100%
9	2-10-1	0.129	0.155	0.097	100%
10	2-11-1	0.131	0.152	0.086	100%
11	2-12-1	0.140	0.159	0.092	100%
12	2-13-1	0.134	0.156	0.089	100%
13	2-14-1	0.134	0.166	0.097	100%
14	2-15-1	0.134	0.155	0.084	100%

表 5-6 以主成分_1 和主成分_2 作为神经网络输入，隐含层节点数分别为 2 到 15 时，网络模型的表现。从表 5-6 可以看出，当以主成分_1 和主成分_2 作为神经网络输入时，随着隐含层节点数从 2 增加到 9 时，验证期的预报最大相对误差呈现下降趋势，当隐含层节点数从 9 增加到 15 时，验证期的预报最大相对误差呈现先上升后基本保持稳定的趋势。当隐含成节点数大于 4 后，验证期预报的合格率都达到了 100%。其中预报效果过好的是模型 8，验证期的预报最大相对误差为 6.9%，合格率为 100%，预报效果良好。

表 5-7 以第一、二、三主成分作为网络模型输入时不同隐含层节点对应的网络的表现

编号	结构	训练	检验		
		MAPE	MRE	MAPE	QR
1	3-2-1	0.147	0.274	0.119	80%
2	3-3-1	0.147	0.202	0.106	80%
3	3-4-1	0.122	0.438	0.149	80%
4	3-5-1	0.121	1.106	0.319	60%
5	3-6-1	0.125	0.197	0.115	100%
6	3-7-1	0.119	0.471	0.197	60%
7	3-8-1	0.113	0.239	0.136	60%
8	3-9-1	0.136	0.191	0.110	100%
9	3-10-1	0.117	0.280	0.143	60%
10	3-11-1	0.129	0.169	0.114	100%
11	3-12-1	0.126	0.207	0.106	80%
12	3-13-1	0.115	0.379	0.135	60%
13	3-14-1	0.114	1.170	0.294	60%
14	3-15-1	0.128	0.179	0.113	100%

表 5-7 为以主成分_1、主成分_2 和主成分_3 作为神经网络输入，隐含层节点数分别为 2 到 15 时，网络模型的表现。从表 4-7 可以看出，当以主成分_1、主成分_2 和主成分_3 作为神经网络输入时，随着隐含层节点的增加，网络预报效果的评价指标值波动性较大，预报的效果欠佳。其中预报效果过好的模型是 Model 10，验证期的预报最大相对误差为 16.9%，合格率为 100%。

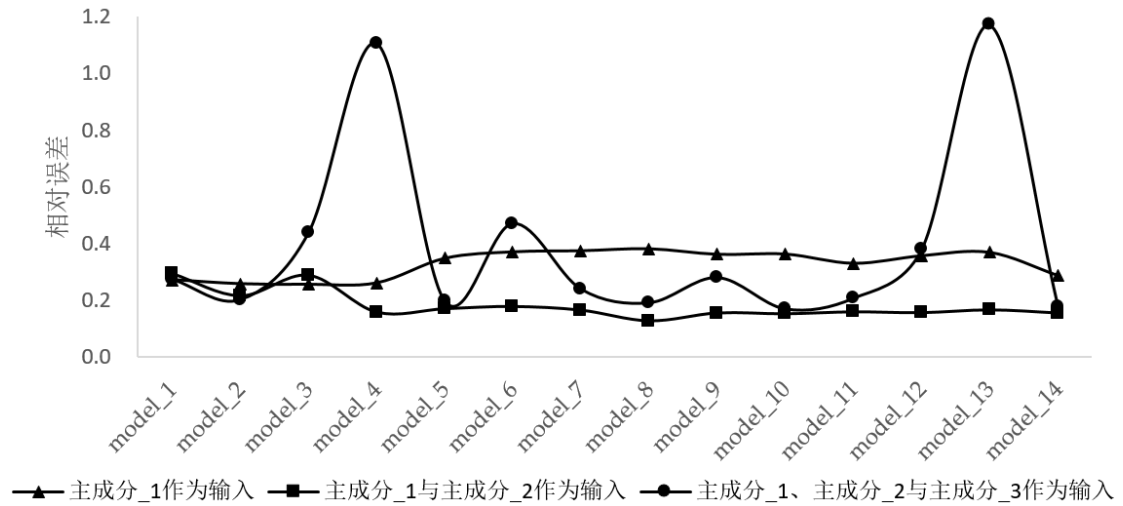


图 5-2 各模型在检验期预报的最大相对误差

了更加直观的比较以不同主成分组合作为输入时，隐含层节点数不同的各网络模型预报效果，绘制了图 5-2。图 5-2 为表 5-5、表 5-6 与表 5-7 中各模型在检验期预报的最大相对误差。从图 5-2 中可以可以看出，当以主成分_1 和主成分_2

同时作为网络输入时，各模型在验证期的表现明显好于以主成分_1 或者同时以主成分_1、主成分_2 和主成分_3 作为网络输入时的表现。综合以上分析，本研究案例选择主成分_1 和主成分_2 作为网络输入，因此输入层节点确定为 2。此外，从表 5-6 中可以看出，当隐含层节点数为 9 时，模型在验证期各评价指标均好于其他模型，因此，本应用案例确定隐含层节点数为 9。

至此，单一循环神经网络模型构建完成。

为了减少由于模型参数不确定性而导致的预测结果的偏差，本应用实例对中长期径流量进行了 5 次单模型预报，并将这 5 次预报结果的平均值作为最终的预报结果，实现基于循环神经网络的中长期径流集合预报。

5.6 预报结果分析

根据以上确定的模型结构，采用 1960~1999 年的锦屏一级水电站年平均径流量数据以及提取的特征数据对网络模型进行训练，2000~2011 年的锦屏一级水电站年平均径流量数据对网络模型进行检验。

表 5-8 是 5 次单模型预报检验期预报误差统计。预报结果以平均绝对百分误差(MAPE, Mean Absolute Percentage Error)、最大相对误差(MRE, Maximum Relative Error)、合格率（QR, Qualified Rate）为评价指标。

表 5-8 各模型检验期预报误差统计

评价指标	年份	第一次	第二次	第三次	第四次	第五次	集合预报
RE	2007	14.10%	13.15%	11.41%	12.86%	15.24%	13.35%
	2008	-14.85%	-14.20%	-12.93%	-14.76%	-11.26%	-13.60%
	2009	1.63%	2.26%	3.70%	1.91%	6.57%	3.21%
	2010	4.63%	6.80%	7.97%	4.96%	6.78%	6.23%
	2011	-4.51%	-11.77%	-13.70%	-2.06%	-12.95%	-9.00%
QR	—	100%	100%	100%	100%	100%	100%
MRE	—	14.85%	14.20%	13.70%	14.76%	15.24%	13.60%
MAPE	—	7.94%	9.64%	9.94%	7.31%	10.56%	9.09%

从表 5-8 中可以看出，单预报模型检验期内预报的最大相对误差都在 16% 以内，根据文献 38 中对中长期径流预报的精度评价方案，合格率都达到了 100%，精度等级为甲。这说明了基于循环神经网络的单一预报模型具有良好的泛化能力和预报稳定性。此外，从表 4-5 中还可以看出，单个网络模型对于不同年份的预报误差和 5 个网络模型对于同一年的预报误差都不相同。各模型的预报误差有一

部分原因是因为训练数据的不确定性造成的，而预报误差之间的差异则主要是由模型网络参数的不确定性造成的。模型的不确定性是由于模型参数的随机性造成的。反向传播算法和参数更新算法实现了对 RNN 的参数空间的搜索，通过不断的训练，减小了历史径流数据的实际值和 RNN 的预测值间的误差。但是，误差曲面可能含有多个不同的局部极小值，在对 RNN 参数空间的搜索过程中，可能会停留在局部极小值点，而不一定是全局最小值点^[61]。因此，即使各个 RNN 模型经过训练后的网络结构相同，但是模型的参数也是不同的，这导致各个单一 RNN 模型预测结果之间存在差异。

当预报因子数据和历史径流数据确定以后，模型的训练数据也就确定了，不管这些数据是否存在误差，模型预测结果的不确定主要取决于模型参数的不确定性。为了减少这种由于模型参数不确定性而导致的预测结果的偏差，本应用实例进行了 5 次径流量的单模型预报，并将这 5 次预报结果的平均值作为最终的预报结果。从表 5-8 可以看出，集合预报模型最大的预报相对误差小于 13.60%，较 5 个单一预报模型的预报精度有进一步提高。

5.7 本章小结

本章采用中长期径流智能预报系统对锦屏一级水电站年平均径流量进行预报。详细介绍了预报因子筛选、预报模型构建的过程，并对预报的结果进行了分析。结果表明，单一预报模型对 5 年预报的最大相对误差控制在 13.70%到 15.24%之间，平均相对误差控制在 7.31%到 10.56%之间，预报的稳定性较高，预报的精度也能满足要求。集合预报模型对 5 年预报的最大相对误差为 13.6%，与单一预报模型相比，预报的精度进一步提高。

第六章 总结与展望

6.1 主要结论

综合课题的研究背景及意义、研究现状及存在的问题，本文主要研究内容和成果如下：

（1）根据中长期径流预报的 E-R 图，设计了中长期径流预报系统的数据库。采用 SQL Server 2008 为数据库管理系统，通过设计各种数据库表结构，实现对中长期径流预报相关的数据以及信息的管理，为中长期径流预报提供数据支持。

（2）设计了中长期径流预报方案库，形成了“基本信息的输入、预报对象创建、预报模型选择、预报因子选择与特征提取、预报模型训练以及预报效果检验”的预报流程；设计了中长期径流预报的模型库，并提出了基于标准互信息和核主成分分析驱动的循环神经网络径流预报模型。通过分析比较，该模型能够有效地挖掘预报因子之间复杂的相互关系，并能够对筛选的因子进行降噪和去冗余，以这些因子作为循环神经网络的输入，预报的稳定性较高，预报的精度能够满足实际生产的需要；采用多个神经网络集合预报的方式，解决模型参数不确定性的问题，提高预报的精度。

（3）采用 Adobe Flex 设计了中长期径流智能预报系统，主要包括预报方案制作界面、预报方案管理界面、径流预报管理界面及径流预报界面。其中，预报方案制作界面是对方案基本信息、预报对象、预报模型、预报因子以及模型参数进行设置，完成模型的参数率定，并对训练好的模型进行试预报，检验模型的预报效果。预报方案管理界面可以对制作的预报方案进行统一管理。中长期径流预报界面可以采用制作好的方案对径流量进行预报。中长期径流预报管理界面可以对预报的结果进行统一管理。

6.2 展望

对于本文提出的基于标准互信息和核主成分分析驱动的循环神经网络径流预报模型以及开发的中长期径流智能预报系统，还存在不足之处，需要在以下方面进一步开展研究：

(1) 神经网络输入层节点数和隐含层节点数确定方法需要改进：本文中对于神经网络输入层节点数和隐含层节点数是采用试错法确定的，这种方法需要系统把所有的组合情况遍历一边，比较各种组合对应的模型的预报效果，才能确定输入层节点数和隐含层节点数。然而计算机遍历一遍，需要花费大量的时间。因此，需要进一步研究确定神经网络隐含层节点数的方法，提高网络的泛化能力，同时节省计算成本。

(2) 防止神经网络过拟合技术还需要进一步研究：循环神经网络模型参数率定，需要很多的历史数据。而对于中长期径流预报而言，用于训练模型的数据比较少，例如对于锦屏一级水电站，有观测记录的径流数据只有 1959 年以后的数据，对于之前的数据则没有记录。如果训练神经网络模型的数据量比较少，神经网络就很容易发生过拟合，影响网络的泛化性能。因此，在无法获得更多数据量的情况下，如何防止神经网络过拟合，是以后需要研究的问题。

(3) 研究多模型的耦合预报技术：本文运用多个神经网络的集合预报技术来缓解由于神经网络模型参数不确定性造成的预报误差，提高预报的精度。在以后的研究中，可以将神经网络与其他模型进行耦合，结合各个模型的优点，进一步提高预报的精度及预报的稳定性。

参考文献

- [1] 王燕生.工程水文学[M].北京:水利电力出版社, 1991.
- [2] 中华人民共和国水利部.水文情报预报规范(SL250-2000) . 2000.
- [3] 汤成友, 官学文, 张世明.现代中长期水文预报方法及其应用[M].北京:中国水利水电出版社, 2008.
- [4] 李崇浩.中长期水文预报理论方法的研究及应用[D].武汉:武汉大学, 2003.
- [5] 包为民, 水文预报(第三版)[M].北京:中国水利水电出版社, 2007.
- [6] 刘佳.中长期径流预报技术及应用系统研究[D].东华大学, 2011.
- [7] 王文, 马骏.若干水文预报方法综述[J].水利水电科技进展, 2005, 25(1):6-60.
- [8] 李保健.水电站群中长期径流预报及发电优化调度的智能方法应用研究[D].大连理工大学, 2005.
- [9] 狄源硕. 中长期径流预报模型与方法综合应用研究[D].大连理工大学, 2015.
- [10] Hamlet A F, Lettenmaier D P. Columbia river streamflow forecasting based on ENSO and PDO climate signals[J]. Journal of Water Resources Planning and Management, 1999, 125(6) : 333-341.
- [11] Whitaker DW, Wasimi S A, Islam S. The El Nio2SouthernOscillationand long2range forecasting of flows in the Ganges [J]. International Journal of Climatology, 2001,21(1):77-87.
- [12] 范钟秀.中长期水文预报[M]. 南京:河海大学出版社, 1999.
- [13] 章淹. 致洪暴雨中期预报进展. 水科学进展, 1995, 6 (2) :162-168.
- [14] 李永康, 陈方维, 马开玉, 陆菊中.长江中下游夏季特大旱涝预测研究.水科学进展, 2000, 11(3): 265-271.
- [15] 刘清仁.松花江流域水旱灾害发生规律及长期预报研究.水科学进展, 1994, 5(4):319-327.
- [16] 黄忠恕, 王钦梁, 匡奇. 北太平洋和青藏高原下垫面热状况与长江流域汛期旱涝关系初步探讨. 全国水文预报学术讨论会. 北京:水利电力出版社, 1985.
- [17] 葛朝霞, 薛梅, 宋颖玲. 多因子逐步回归周期分析在中长期水文预报中的应用[J].河海大学学报(自然科学版), 2009, 37(3): 255-257.
- [18] Box G. E. P, Jenkins G. M. Time Series Analysis: Forecasting and Control[M]. San Francisco Holden Day, 1970.
- [19] 丁晶, 刘权授. 随机水文学[M].北京:中国水利水电出版社, 1997.
- [20] 张铭, 李承军, 张勇传.贝叶斯概率水文预报系统在中长期径流预报中的应用[J].水科学进展, 2009,20(1):40-44.
- [21] 陈守煜.模糊水文学[J].大连理工大学学报, 1988, (3): 93-97.
- [22] 陈守煜.模糊水文学与水资源系统模糊优化原理[M].大连:大连理工大学出版社, 1990.
- [23] 陈守煜, 周惠成.水文点值预报与分组预报精度评价的一种数学方法[J].水能技术经济. 1986, (1):23-27.
- [24] 陈守煜, 工程模糊集理论与应用[J].北京:国防工业出版社, 1998.
- [25] 邓聚龙. 灰色控制系统明.华中工学院学报[J], 1982, 10(3): 9-18.
- [26] 蓝永超, 杨志怀, 权建民等. 灰色预测模型在径流长期预报中的应用. 中国沙漠[J], 1997, 17(1):49-52.
- [27] 晓安, 王金文, 王海伟.基于小波分析的月径流 ARIMA 预测方法[J]. 水电自动化与大坝监测, 2006, 30 (4):77-79.
- [28] 王文圣, 丁晶, 向红莲.小波分析在水文学中的应用研究及展望. 水科学进展, 2002, 13(4):515 -520.
- [29] Vapnik V. The Nature of Statistical Learning Theory [M]. New York, USA: Springer Verlag, 1995.
- [30] Liong S. Y, Sivapragasm C. Flood stage forecasting with support vector machines[J].Journal of the American Water Resources Association, 2002, 38(1): 173-186.
- [31] Kang K. W., Kim J. H., Park C. Y., et al. Evaluation of hydrological forecasting system based on neural network model[C]. Proceedings of 25th Congress of International Association for Hydrological Research, Delft, The Netherlands, 1993, 257-264.
- [32] 屈亚玲, 周建中, 刘芳等. 基于改进的 Elman 神经网络的中长期径流预报闭.水文, 2006,

- (1): 45-50.
- [33] 王富强. 中长期水文预报及其在平原洪水资源利用中的应用研究[D].大连: 大连理工大学, 2008.
- [34] 曹启辉, 王文圣, 汤成友. 一种新的小波网络组合预测模型[J]. 人民长江, 2006, 37 (11): 65-67.
- [35] 赵丽红, 张西礼, 徐心和. 基于二维对称主成分分析的人脸识别[J]. 仪器仪表学报, 2008, 29(6).
- [36] 王元, 文兰, 陈木法. 数学大辞典[M]. 北京: 科学出版社, 2010.
- [37] 蔡煌东, 姚林声. 径流长期预报的人工神经网络方法[J]. 水科学进展, 1995, 6(1): 62-65.
- [38] Chiang, Y.M.; Chang, L.C.; Chang, F.J. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling [J]. Journal of Hydrology, 2004, 290: 297-311.
- [39] 李崇浩. 中长期水文预报理论方法的研究及应用[D]. 武汉: 武汉大学, 2003.
- [40] 冯小冲, 王银堂, 刘勇, 胡庆芳. 基于物理统计方法的丹江口水库月入库径流预报[J]. 河海大学学报(自然科学版), 2011, 39(3): 242-247.
- [41] 汤成友, 官学文, 张世明. 现代中长期水文预报方法及其应用[M]. 北京: 中国水利水电出版社, 2008.
- [42] Yaseen, Z.M.; El-Shafie, A.; Jaafar, O.; Afan, H.A.; Sayl, M.N. Artificial intelligence based models for stream-flow forecasting: 2000-2015[J]. Journal of Hydrology, 2015, 530: 829-844.
- [43] Duan, Q.Y.; Sorooshian, S.; Gupta, V. Effective and efficient global optimization for conceptual Rainfall-runoff models [J]. Water Resour Res., 1992, 28: 1015-1031.
- [44] 赵铜铁刚, 杨大文. 神经网络径流预报模型中基于互信息的预报因子选择方法[J]. 水力发电学报, 2011, 30(1): 25-30.
- [45] 陈守煜, 王大刚. 基于遗传算法的模糊优选 BP 网络模型及其应用[J]. 水利学报, 2003, 5(1): 116-120.
- [46] 雷晓辉, 廖卫红, 蒋云钟等. 分布式水文模型 EasyDHM 模型 (I) 理论方法[J]. 水利学报, 2010, 41(7): 893-899.
- [47] 唐勇, 胡和平, 田富强等. 鳌江流域洪水预报调度决策支持系统[J]. 水科学进展, 2003, 14(3): 295-299.
- [48] 周之英. 现代软件工程[M]. 北京: 科学出版社, 2000.
- [49] Gopalaswamy Ramesh. 全球化软件项目管理[M]. 机械工业出版社, 2003.
- [50] 何文社, 戴会超, 曹叔尤, 袁杰. 三峡水库水文泥沙信息分析管理系统设计[J]. 水利发电学报, 2005, 24(6): 95-99.
- [51] 蒋云钟, 鲁帆, 雷晓辉等. 水资源综合调配模型技术与实践[M]. 北京: 中国水利水电出版社, 2009.9.
- [52] 李春葆, 曾平. 数据库原理与应用 - 基于 SQL SERVER 2000(第二版)[M]. 清华大学出版社, 2007.11.
- [53] Yuanyuan Shen, Xiaohui Lei, Yu Tian, Yunzhong Jiang. Development of a groundwater flow simulation system base on RCP mode. The Fourth Conference of Asia Pacific Association of Hydrology and Water Resources (APHW), 2008.
- [54] 贾仰文, 王浩, 倪广恒, 杨大文, 王建华, 秦大庸. 分布式流域水文模型原理与实践[M]. 北京: 中国水利水电出版社, 2005.1.
- [55] 旦木仁加甫. 常用中长期水文预报 VisualBasic6.0 应用程序及实例[M]. 郑州: 黄河水利出版社, 2004.
- [56] 张利平, 王德智, 牛存稳, 夏军. 水库中长期径流预报系统的开发研究[J]. 灾害学, 2003, 18(3): 24-27.
- [57] 黄梯云, 李一军. 管理信息系统[M]. 北京: 高等教育出版社, 2002.
- [58] 单美贤. 人机交互设计[M]. 北京: 电子工业出版社, 2016.
- [59] GB/T 22482-2008, 水文情报预报规范[S]. 北京: 中国标准出版社, 2008: 4-10.
- [60] 贺志尧, 杨明祥等. 基于 Elman 网络的锦屏一级水电站年平均径流量集合预报研究[J]. 水电能源科学, 2017.10(10): 25-28.
- [61] Mitchell T M. Machine Learning [D]. USA: McGraw-Hill Education, 1997: 88 -89.