

基于正则化 GRU 模型的洪水预测^①



段生月, 王长坤, 张柳艳

(南昌航空大学 信息工程学院, 南昌 330063)

通讯作者: 王长坤, E-mail: 2076885@qq.com

摘要: 针对传统神经网络模型在洪水预测过程中存在准确性低、过拟合等问题, 本文以赣江流域外洲水文站每月平均水位为研究对象, 提出基于正则化 GRU 神经网络的洪水预测模型来提高洪水预报精度. 选用 relu 函数作为整个神经网络的输出层激活函数, 将弹性网正则化引入到 GRU 模型中, 对网络中输入权重 w 实施正则化处理, 以提升 GRU 模型的泛化性能, 并将该模型应用于外洲水文站每月平均水位的拟合及预测. 实验对比表明, 弹性网正则化优化后的模型预测拟合程度较高, 合格率提高了 9.3%, 计算出的均方根误差较小.

关键词: 时间序列; 门结构循环单元; 弹性网正则化; 洪水预报; 水位

引用格式: 段生月, 王长坤, 张柳艳. 基于正则化 GRU 模型的洪水预测. 计算机系统应用, 2019, 28(5): 196-201. <http://www.c-s-a.org.cn/1003-3254/6883.html>

Flood Forecast Based on Regularized GRU Model

DUAN Sheng-Yue, WANG Chang-Kun, ZHANG Liu-Yan

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: Aiming at the problems of low accuracy and over-fitting of traditional neural network model in flood forecasting process, this study takes the monthly average water level of Waizhou Hydrological Station in Ganjiang River Basin as the research object, and proposes a flood forecasting model based on regularized GRU neural network to improve the accuracy of flood forecasting. Relu function is selected as the output layer activation function of the whole neural network. To improve the generalization performance of GRU model, regularization of elastic network is introduced into GRU model, and regularizes the input weights in the network. The model is applied to the fitting and prediction of the monthly average water level at Waizhou Hydrological Station, and the experimental comparison shows that the model optimized by regularization of elastic network has a higher fitting degree, the qualified rate is increased by 9.3%, and the calculated root mean square error is small.

Key words: time series; GRU; ElasticNet normalization; flood forecast; water level

我国是世界上洪水灾害频繁且严重的国家之一, 洪水灾害不仅范围大、发生频繁、突发性强, 而且损失大. 特别是近年来, 受全球气候变化和人类活动影响, 极端天气事件和人为突发水事件明显增加. 为了防治洪涝灾害, 我国投入了大量的人力、财力与物力, 建设

了一批又批防洪工程, 大大加强了我国抗灾御洪的能力. 与此同时, 我国还加强了非工程防洪措施. 其中, 洪水预报就是预防洪水灾害的重要非工程措施之一, 也一直是专家学者关心的课题. 因此提高洪水预测的准确性, 减轻洪水带来的一系列灾害, 对整个流域的防洪

① 基金项目: 国家自然科学基金 (61866028, 61741312)

Foundation item: National Natural Science Foundation of China (61866028, 61741312)

收稿时间: 2018-11-08; 修改时间: 2018-12-03; 采用时间: 2018-12-07; csa 在线出版时间: 2019-05-01

安全有着非常重要的意义. 一般来说洪水预测有 3 种不同的方法, 经验法、统计法以及模型方法, 但水文系统中多变量之间的非线性关系^[1]远非传统方法能解决. 随着科学技术的快速发展, 如何把最新的科学技术引入到水文预报中, 提高水文预报精度, 成为大家的关注点. 随着神经网络^[2]等非线性科学技术在各个行业的广泛应用, 为解决上述问题提供了机遇.

目前国内已有很多学者尝试将神经网络引入到洪水预测中. 兰州理工大学的李晓丽等^[3]提出了不确定, 确定支持向量机在洪水预测模型中的应用, 该算法利用数据间的关系去掉冗余的信息, 简化算法运算, 从而提高了洪水预测的精度; 晋中市水文水资源勘测局的梁存峰等^[4]提出了基于混沌自适应模型在洪水预测中的研究, 该模型亦可在数据较少的情况下, 取得精度较好的预测结果; 福建省南平市水利局的金保明^[5]在闽江十里庵流量预测中应用 BP 神经网络, 将反向传播 BP 神经网络模型应用于洪水预测, 并且预测精度符合要求; 何勇、李妍琰^[6]提出改进粒子优化 BP 神经网络的洪水智能预测模型研究, 以 BP 神经网络为基础, 提取水位站往年平均径流量作为洪水属性, 采用 POS 算法对 BP 神经网络的各个参数进行优化, 得出最优的 BP 神经网络预测适应度值, 提高了预测的效率. 以上算法都是比较传统的神经网络算法, 虽然在一定程度上提高了洪水预测精确度, 但是精度依旧不是特别理想.

本文提出一种基于正则化 GRU 神经网络的洪水预测模型来提高洪水预报精度, 选用 relu 函数作为整个神经网络的输出层激活函数^[7-9], 将弹性网正则化引入到 GRU 模型中, 对网络中输入权重 w 实施正则化处理, 以提升 GRU 模型的泛化性能. 并将该模型应用于外洲水文站每月平均水位的拟合及预测, 实验对比表明, 弹性网正则化改进后模型的拟合效果和预测效果明显提高, 该方法计算出的均方根误差较小, 预测拟合程度较高.

1 GRU 模型简介

传统的人工神经网络又被称为前馈神经网络 FNN, 是一种最简单的神经网络, 包含输入层、隐藏层. 这种模型在处理序列数据时只能利用当前时刻的信息, 无法利用历史信息. Elman 在 1990 年提出了循环神经网络 (Recurrent Neural Network, RNN), 其隐藏层之间的节点也是有连接的, 因此隐藏层的输入不仅包括当前

时刻输入层的输入还包括上一时刻隐藏层的输出. 理论上, RNN 能够对任何长度的序列数据进行处理, 但是在实践中, 当相关信息和当前预测位置之间的间隔不断增大时, RNN 就会丧失学习远距离信息的能力. 这是因为 RNN 通常使用的是 BPTT 反向传播算法对网络进行优化, 但是 BPTT 无法解决长时依赖问题, 因此该算法会带来神经网络梯度消失问题和梯度爆炸问题.

为了解决 RNN 网络的梯度消失和梯度爆炸问题, 研究人员在优化学习算法和配置网络的技巧方面提出了较多的改进方法. 文献[10,11]中介绍 Schimidhuber H 于 1997 年提出了长短期记忆单元 (Long Short-Term Memory, LSTM), 用于改进传统的 RNN 模型. LSTM 神经单元是由一个或多个存储器外加三个自适应乘法门组成. 它通过输入门、输出门和遗忘门来控制信息的流动和传递, 通过对输入的信息进行剔除或增强到神经细胞单元中实现对细胞状态的控制, 使得 LSTM 可以记忆、更新长距离的信息, 从而实现对长距离信息的处理. 受 LSTM 模型门机制启发, Stanford 在 2014 年提出了 GRU 模型, GRU 网络将 LSTM 结构中的输入门和遗忘门结合成一个单独的更新门, 合并了记忆细胞和隐含状态, 同时也做了一些调整. 已有研究表明, GRU 模型的性能与 LSTM 相当, 但计算效率更高, 参数更少. 整体的 GRU 神经网络结构如图 1 所示.

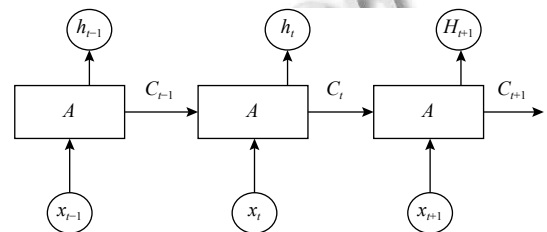


图 1 GRU 神经网络结构示意图^[5]

GRU 中各个门的表达式如下:

GRU 更新门表达式:

$$Z_t = s(W_z \cdot [h_{t-1}, x_t]) \quad (1)$$

GRU 重置门表达式:

$$r_t = s(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

GRU 输出部分表达式:

$$\tilde{h}_t = \tanh(W_k \cdot [r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

$$y_t = s(W_0 \cdot h_t) \quad (5)$$

式(1)中 z_t 表示更新门, W_z 是更新门的权重矩阵, h_{t-1} 表示上一个神经元的输出, x_t 表示本次神经元的输入, s 是sigmoid函数用来激活控制门. 更新门的作用是帮助模型决定要将多少过去的信息传送到未来, 或者前一时间步和当前时间步的信息有多少是需要继续传递的, 模型通过复制过去信息以减少梯度消失的风险. 式(2)中 r_t 表示重置门, W_r 表示重置门的权重矩阵, 从本质上说, 重置门主要决定有多少过去的信息需要被遗忘.

2 改进 GRU 模型

2.1 正则化技术

在深度神经网络中, 随着层数以及各层神经元个数的增加, 模型中参数的个数会以极快的速度增长. 当网络参数个数过多或训练数据集很小时, 神经网络模型往往会出现过拟合问题. 所谓过拟合, 就是模型在训练数据上拟合效果好, 但是它在验证集上效果很差, 即模型泛化能力差. 目前, 防止过拟合的一般方法是增大训练数据集, 但当数据集有限时, 防止过拟合的方法则是在训练数据集中加入正则化技术. 现在我们主要讨论一些旨在改善过拟合的正则化方法.

正则化的方法就是在模型 $cost$ 函数中加入某种正则项或几种正则项的组合, 表达式如下:

$$\min \left\{ \sum_{i=1}^T l(y_i, f(x_i, w)) + \sum_{i=1}^m \lambda_i \rho_i \right\} \quad (6)$$

式中, $l(\cdot, \cdot)$ 为模型中损失函数, 用来评价模型的泛化性能; $\lambda \rho(w)$ 为正则项, λ 为正则化可调参数, 用来控制正则项与损失函数之间的平衡关系, w 为模型中待估参数. 在正则项中, $\rho(w)$ 具有多种不同的表达形式, 常用的是 L_1 范数和 L_2 范数. 当选取不同的范数罚正则项, 整个模型会有不同的泛化方向. L_2 正则化会对待估参数 w 进行一定程度压缩使之正态分布, 但并不能将其压缩为零, 因此不能产生稀疏解. 而 L_1 正则化将待估参数 w 的先验分布约束为拉普拉斯分布, 从而使模型具有稀疏性^[12], 进而控制模型的过拟合问题, 但是由于参数值大小和模型复杂度成正比; 若 L_1 范数较大, 最终可能影响模型的预测性能. 为了克服不同范数正则项的缺点, 这里引入弹性网正则化, 也就是将 L_1 范数和 L_2 范数线性组合作为一个正则项.

本文将弹性网引入到 GRU 模型中, 对网络中输入权重 w 实施正则化处理, 以提升 GRU 模型的泛化性

能, 模型如下:

$$\min_w \left\{ \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^m (y_{ij} - \bar{y}_{ij})^2 + \lambda_1 \|w_1\| + \lambda_2 \|w_2\|^2 \right\} \quad (7)$$

其中, λ 为正则化可调参数用于平衡原有损失函数与正则项之间的关系. 通过式(7)可知, 修改正则可调参数 λ_1 和 λ_2 可得到不同组合的正则项, 当 $\lambda_1=0$ 和 $\lambda_2=0$ 时, 为普通的 GRU 模型; 当 $\lambda_1 \neq 0$ 和 $\lambda_2=0$ 时, 为 L_1 正则化网络; 当 $\lambda_1=0$ 和 $\lambda_2 \neq 0$ 时, 为 L_2 正则化网络; 当 $\lambda_1 \neq 0$ 和 $\lambda_2 \neq 0$ 时, 为弹性网正则化网络.

2.2 正则化 GRU 模型的洪水预测步骤

由于 GRU 在训练过程中, 容易出现过拟合现象, 而正则化方法通过限制网络中权重的大小, 可以对一些因子施加惩罚, 以够弥补网络本身的不足. 所以这里利用正则化方法来优化 GRU 模型, 从而实现对外洲站水位的预测. GRU 改进模型构建步骤如下:

步骤 1: 将时间序列的水文数据进行预处理, 采用极差法使数据标准化, 首先计算指标值的最小值和最大值, 然后计算极差, 通过极差法将指标值映射到 [0-1] 之间, 公式为:

$$x^* = \frac{x - \min}{\max - \min} \quad (8)$$

步骤 2: 将时间序列水文数据集 $F_0 = \{f_1, f_2, f_3, \dots, f_n\}$, 划分为训练集和测试集, 分别表示为 $F_{tr} = \{f_1, f_2, \dots, f_m\}$ 和 $F_{te} = \{f_{m+1}, f_{m+2}, \dots, f_n\}$, 满足约束条件 $m < n$ 和 $m, n \in N$. 设置分割窗口长度 ($batch-size$) 取值为 L , 分割后的模型输入为:

$$X = \{X_1, X_2, \dots, X_L\} \quad (9)$$

$$X_p = \{f_p, f_{p+1}, \dots, f_{m-L+p-1}\} \quad (10)$$

其中, $1 \leq p \leq L; p, L \in N$.

隐藏神经元的输出值定义公式为:

$$y = W \cdot x + b \quad (11)$$

X 经过隐藏层后的输出可表示为:

$$Y = \{Y_1, Y_2, \dots, Y_L\} \quad (12)$$

$$Y_p = \{f_{p+1}, f_{p+2}, \dots, f_{m-L+p}\} \quad (13)$$

其中, $1 \leq p \leq L; p, L \in N$.

选用均方误差作为误差计算公式, 训练中的损失函数定义公式为:

$$loss = \frac{1}{n} \sum_{i=1}^n (y_i - y_{real})^2 \quad (14)$$

损失函数是用来评估神经网络目标输出和实际输出差距的函数,函数值越小说明实际输出与目标输出的差值越小,也就说明权值越合适.将损失函数与正则项之和设置为目标函数,给定网络初始化的学习率以及训练步数,如式(7)再应用 Adam 优化算法实现目标函数的最小化并不断更新网络权重,从而得到最终的隐藏层网络.

步骤 3: 预测过程采用迭代的方法.理论输出 P 的最后一行数据:

$$P_f = \{p_{m-L+1}, p_{m-L+2}, \dots, p_m\} \quad (15)$$

将 P_f 输入隐藏层后,输出结果表示为:

$$Y_f = \{y_{m-L+2}, y_{m-L+3}, \dots, y_{m+1}\} \quad (16)$$

则 $m+1$ 时刻的预测值为 y_{m+1} ,然后将 P_f 后的数据与 y_{m+1} 合并成新的一行数据:

$$P_{f+1} = \{p_{m-L+2}, p_{m-L+3}, \dots, y_{m+1}\} \quad (17)$$

将 P_{f+1} 输入隐藏层,则 $m+2$ 时刻的预测值为 y_{m+2} ,以此类推,得到的预测序列为:

$$Y_0 = \{y_{m+1}, y_{m+2}, \dots, y_n\} \quad (18)$$

步骤 4: 对 Y_0 作 z -score 反标准化,得到最终的与测试集 F_{te} 对应的预测序列为 Y_{te} ,通过计算 Y_{te} 和 F_{te} 的偏差定量的给出模型的拟合与预测精度.

3 实验与分析

3.1 实验设置

本实验环境为:硬件配置 CPU i7-7700K, GPU Gtx1060, 运行内存 16 GB; 软件配置 CUDA9.0、cudnn7.1; 操作系统 Ubuntu16.04; 语言 python3.6; 框架 tensorflow.

本实验的代码参数设置如下: 输入层神经元个数设为 3, 输出层神经元个数设为 1, 其中的隐藏单元的数量选取为 10, 每批数据量的大小 $batch_size$ 设为 10, 通过调节 λ 值得到三种正则化网络模型: L_1 正则化网络、 L_2 正则化网络和弹性网正则化网络. 对比每组的 RMSE 来确定最佳参数配置, 最终确定三种模型中正则化可调参数 λ 分别为 $\lambda_1=0.02$ 、 $\lambda_2=0.004$ 和 $\lambda_1=0.0043$ 、 $\lambda_2=0.38$. 本文针对江西省赣江外洲水文站 1954 年到 2002 年每月平均流量、平均水位、月份数

据进行分析,将次月平均水位作为输出变量,其他 3 个指标作为模型的输入变量^[13-16].以 1954 年到 1997 年各月平均水位作为训练数据,对 1998 年到 2002 年共 60 个月的平均水位数据进行预测.

3.2 实验性能评价指标

为了验证本文所验证的预测模型的可行性以及有效性,本文所采用的精度评定标准来自于中华人民共和国水利部《水文情报预报规范》(SL250-2000)^[17],一次预报的误差小于许可误差时,为合格预报.合格预报次数与预报总次数的百分比为合格率.本文结合《水文情报预报规范》的有关规定设置水位许可误差为 0.2 m. 合格率 QR 的表达公式如下:

$$QR = \frac{n}{m} \times 100\% \quad (19)$$

其中, m 是预报的总次数、 n 是合格预报(预报误差小于许可误差)的次数;当 $QR \geq 85.0\%$ 时,为甲级精度; $70.0\% \leq QR < 85.0\%$ 时,为乙级精度; $60\% \leq QR < 70.0\%$ 时,为丙级精度.

采用均方根误差 RMSE(Root Mean Square Error) 来定量的评价模型的预测性能, RMSE 通过计算预测值对观察值的平均偏差程度来反映模型的预测性能,值越小预测效果越好.公式定义如下:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (\bar{y}_t^c - y_t^c)^2}{T}} \quad (20)$$

3.3 实验结果分析

通过对上文中每种模型结构及参数的确定,共得到四种不同的 GRU 模型.为了比较这四种网络模型预测性能,将四种模型分别对外洲站水文数据做训练及预测,每组的误差统计见表 1.

表 1 几种预测模型 RMSE 值和合格率(QR)统计表

	RMSE	合格率(QR)%	预测精度
GRU 模型	0.962	82.315	乙
L1 正则化网络	0.737	88.410	甲
L2 正则化网络	0.666	90.409	甲
弹性网正则化网络	0.631	91.644	甲

由表 1 可以看出 3 种正则化 GRU 模型的 RMSE 的值均比普通 GRU 模型的 RMSE 值小, L_2 正则化网络的 RMSE 比 L_1 正则化网络稍小一些, 弹性网正则化 GRU 模型的 RMSE 是实验中最低的, 仅为 0.631. 三种正则化 GRU 模型的合格率均比普通 GRU 模型大,

L_2 正则化网络的合格率比 L_1 正则化网络的合格率高一些, 实验中弹性网正则化 GRU 模型的合格率 QR 最高为 91.644%, 相比普通的 GRU 模型合格率提高了 9.3%。以上的验证表明改进的 GRU 神经网络对水文时间序列的验证是有效的, 弹性网正则化 GRU 模型预测结果更加准确, 误差更小, 所以在洪水预测预测方面, 弹性网正则化 GRU 模型预测值更接近洪水实测值, 是一种有效可靠的预测方法。

这里选取各自最小 RMSE 的实验预测结果图, 如图 2-图 5 所示。

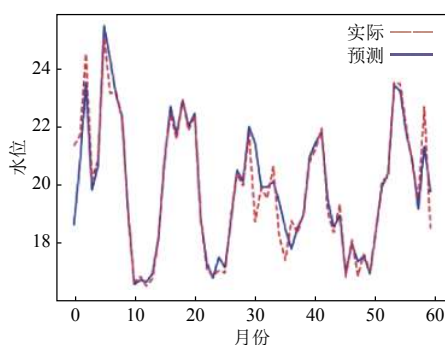


图 2 GRU 模型预测结果图 ($\lambda_1=0, \lambda_2=0$)

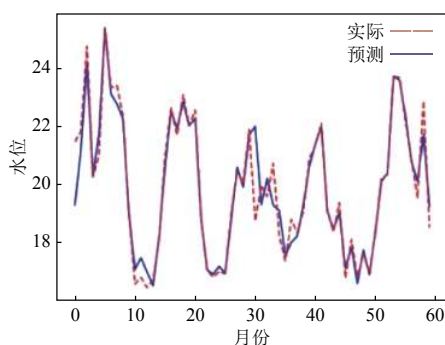


图 3 L_1 正则化网络预测结果图 ($\lambda_1=0.02, \lambda_2=0$)

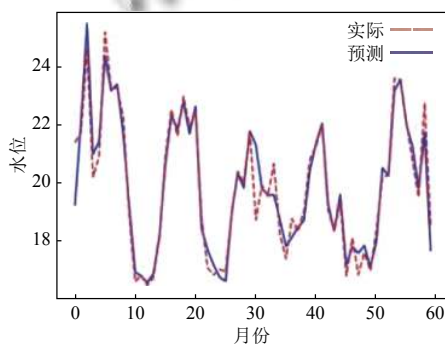


图 4 L_2 正则化网络预测结果图 ($\lambda_1=0, \lambda_2=0.004$)

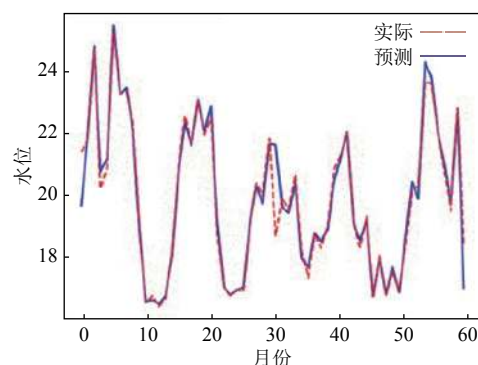


图 5 弹性网正则化网络预测结果图 ($\lambda_1=0.0043, \lambda_2=0.38$)

4 结论

本文使用正则化 GRU 模型对水文站往年每月平均水位进行预测, 通过比较三种正则化方法对 GRU 模型的改进, 发现用弹性网正则方法优化 GRU 模型后合格率提高了 9.3%, 预测精度从乙等升到了甲等。由于洪水预测受到多方因素以及突发因素的影响, 本文所使用的神经网络训练数据资料有限导致模型预测精度不是非常高, 有待结合更多相关影响因子来进行研究处理。

参考文献

- 1 范睿. 基于遗传算法的神经网络洪水预报研究与应用[硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2005.
- 2 Lippmann RP. An introduction to computing with neural nets. ACM SIGARCH Computer Architecture News, 1988, 16(1): 7-25. [doi: 10.1145/44571]
- 3 李晓丽, 周小健, 沈钢纲, 等. 不确定支持向量机在洪水预测模型中的应用. 兰州理工大学学报, 2012, 38(3): 107-110. [doi: 10.3969/j.issn.1673-5196.2012.03.025]
- 4 梁存峰. 基于混沌 Volterra 自适应模型的洪水预测研究. 水资源与水工程学报, 2011, 22(1): 146-150.
- 5 金保明. BP 神经网络在闽江十里庵流量预测中的应用. 水电能源科学, 2010, 28(9): 12-14. [doi: 10.3969/j.issn.1000-7709.2010.09.004]
- 6 何勇, 李妍琰. 改进粒子群优化 BP 神经网络的洪水智能预测模型研究. 西南师范大学学报(自然科学版), 2014, 39(5): 75-80.
- 7 叶小舟, 陶飞飞, 戚荣志, 等. 循环神经网络结构中激活函数的改进. 计算机与现代化, 2016, (12): 29-33. [doi: 10.3969/j.issn.1006-2475.2016.12.006]
- 8 Semeniuta S, Severyn A, Barth E. Recurrent dropout without memory loss. arXiv preprint arXiv:1603.05118, 2016.

- 9 刘洋. 基于 GRU 神经网络的时间序列预测研究[硕士学位论文]. 成都: 成都理工大学, 2017.
- 10 张玉环, 钱江. 基于两种 LSTM 结构的文本情感分析. 软件, 2018, 39(1): 116–120. [doi: [10.3969/j.issn.1003-6970.2018.01.023](https://doi.org/10.3969/j.issn.1003-6970.2018.01.023)]
- 11 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测. 北京航空航天大学学报, 2018, 44(4): 772–784.
- 12 刘建伟, 崔立鹏, 刘泽宇, 等. 正则化稀疏模型. 计算机学报, 2015, 38(7): 1307–1325.
- 13 崔东文. 多隐层 BP 神经网络模型在径流预测中的应用. 水文, 2013, 33(1): 68–73. [doi: [10.3969/j.issn.1000-0852.2013.01.013](https://doi.org/10.3969/j.issn.1000-0852.2013.01.013)]
- 14 丁海蛟. 基于 LS-SVM 的河道洪水预报研究[硕士学位论文]. 昆明: 昆明理工大学, 2016.
- 15 孙卫刚, 王正勇. 克尔古提水文站站月水量预报方案的编制. 大科技·科技天地, 2010.
- 16 朱星明, 卢长娜, 王如云, 等. 基于人工神经网络的洪水水位预报模型. 水利学报, 2005, 36(7): 806–811. [doi: [10.3321/j.issn:0559-9350.2005.07.007](https://doi.org/10.3321/j.issn:0559-9350.2005.07.007)]
- 17 中华人民共和国水利部. SL 250–2000 水文情报预报规范. 北京: 中国水利水电出版社, 2001.