日径流预报贝叶斯回声状态网络方法

李保健, 程春田*, 武新宇, 王森

大连理工大学水电与水信息研究所, 大连 116024

* E-mail: ctcheng@dlut.edu.cn

收稿日期: 2013-12-25; 接受日期: 2014-05-19

国家高技术研究发展计划专项经费(编号: 2012AA050205)和国家自然科学基金(批准号: 51109024)资助项目

摘要 回声状态网络(ESN)相比传统递归神经网络, 具有模型简单、参数训练速度快的特点. 针对标准 ESN 因常采用线性回归率定模型参数容易出现过拟合问题, 提出了基于贝叶斯回声状态网络(BESN)的日径流预报模型. 该模型将贝叶斯理论与 ESN 模型相结合, 通过权重后验概率密度最大化而获得最优输出权重, 提高了模型的泛化能力. 通过安砂和新丰江两座水库日径流预测实例表明, BESN 模型是一种有效、可行的预测方法, 与传统 BP 神经网络和 ESN 模型对比, 进一步表明 BESN 模型具有更好的预测精度.

关键词

递归神经网络
回声状态网络
贝叶斯理论
径流预报

1 引言

较高精度的径流预报是合理利用水资源、实现水库安全运行及优化调度的有效保证^[1,2], 但其变化情势受各种因素(如降水、人类活动等)的影响. 由于径流过程形成机理极为复杂^[3], 且具有极强的非线性特征, 如何获得更高预报精度的实用化模型一直是国内外研究的热点和难点. 近年来, 多种方法被应用于水文预报, 其中过程驱动模型^[4]和数据驱动模型^[5,6]为两大基本方法. 作为数据驱动模型之一, 人工神经网络具有较强的非线性映射能力, 能够从历史径流数据中学习隐含的系统规律且无须显式表述, 并能提供较好的预报精度, 因而在水文预报领域得到了广泛应用. 在这些应用中, 常使用基于梯度下降法的静态神经网络^[7,8], 但由于径流形成过程具有高度复杂的动态特征, 因此, 直接应用静态神经网络在预报精度方面具有一定的影响. 尽管在如何提高静态神经网络的泛化能力方面已获得较多研究成果^[9-12], 但

为更好地满足实际生产需求, 仍然需要进一步深入研究. 递归神经网络^[13]作为动态神经网络, 能够较好地模拟径流形成过程, 但由于其学习算法耗时长、计算量大, 直接应用还有一定的限制.

回声状态网络^[14,15](ESN)是一种新的递归神经网络, 因具备模型简单、参数训练速度快等特点而相对传统递归神经网络模型拥有一定的优越性, 虽然已被应用于水文预报领域, 但仅有少量成果得到报道, 其中: Sacchi 等人^[16]以水电站月入库流量预报为实例, 表明 ESN 模型比径向基神经网络、自适应模糊推理模型具有更好的预测精度; Coulibaly^[17]以美国 4 个湖泊的月平均水位预报为例, 验证了 ESN 模型具有较强的预测能力, 适用于水文时间序列模拟; de Vos^[18]以 12 个流域的日降雨-径流数据为实例, 证实了 ESN 模型在多数情况下能够取得比 BP 神经网络和 Elman 递归神经网络更好的预测精度. 由于标准 ESN 常依据线性回归获得输出权重, 容易出现过拟合问题, 致使模型泛化能力不足, 但以上文献并未对此进行深入

引用格式: 李保健, 程春田, 武新宇, 等. 日径流预报贝叶斯回声状态网络方法. 中国科学: 技术科学, 2014, 44: 1004-1012

Li B J, Cheng C T, Wu X Y, et al. Daily rainfall-runoff forecasting using Bayesian echo state network (in Chinese). Sci Sin Tech, 2014, 44: 1004-1012, doi: 10.1360/N092013-00036

入研究. 针对该问题, 岭回归技术^[19]被用于优化 ESN 的输出权重, 但岭回归参数选择比较困难; Shi 和 Han^[20]提出基于支持向量机的正则化方法, 但参数正则化比较困难, 同时相应交叉验证过程比较费时.

依据贝叶斯理论^[21,22]考虑权重空间的概率密度, 通过权重后验概率分布最大化获得最优输出权重, 可以避免模型参数率定的过拟合现象^[23~25]. 为此, 本文通过对传统参数率定目标函数引入正则项, 提出了基于贝叶斯回声状态网络(BESN)的径流预报模型. 以安砂和新丰江两座水库的径流预报为例, 通过与传统 BP 神经网络和标准 ESN 模型对比, 验证了 BESN 模型在径流预报中的有效性.

2 回声状态网络

经典的 ESN 网络结构由输入层、储备池(即隐含层)和输出层 3 个部分组成, 如图 1 所示, 其中储备池是由大规模稀疏且随机连接的神经元组成(神经元个数通常取值范围为 50~500). 假定 ESN 网络有 K 个输入节点, N 个储备池神经元和 L 个输出层节点, 则 ESN 储备池状态更新及输出按下式进行:

$$\mathbf{x}(k) = f(\mathbf{W}^{\text{in}}\mathbf{u}(k) + \mathbf{W}\mathbf{x}(k-1) + \mathbf{W}^{\text{back}}\mathbf{y}(k-1)), \quad (1)$$

$$\mathbf{y}(k) = f^{\text{out}}(\mathbf{W}^{\text{out}}(\mathbf{u}(k), \mathbf{x}(k))), \quad (2)$$

其中, $\mathbf{u}(k) = [u_1(k), u_2(k), \dots, u_K(k)]^T$ 表示输入向量; $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_N(k)]^T$ 表示储备池状态; $\mathbf{y}(k) = [y_1(k), y_2(k), \dots, y_L(k)]^T$ 表示输出向量; $\mathbf{W}^{\text{in}}$ 表示输入层与储备池神经元的连接权重矩阵; $\mathbf{W}$ 表示储备池内部的连接权重, 同时为提供有效的记忆能力, 大约保持 1%~5% 的稀疏连接且通常谱半径小于 1; $\mathbf{W}^{\text{back}}$ 表示输出层与储备池单元的反馈权重矩阵; $\mathbf{W}^{\text{out}}$ 表示输出权重矩阵; $f(\cdot)$ 表示储备池单元的激活函数(一般取双曲正切函数); $f^{\text{out}}(\cdot)$ 为输出层激活函数. 在建立 ESN 网络时, $\mathbf{W}^{\text{in}}$, $\mathbf{W}$ 和 $\mathbf{W}^{\text{back}}$ 一旦被初始化, 在网络训练过程中即保持不变, 只需对 $\mathbf{W}^{\text{out}}$ 进行训练即可.

对于 $\mathbf{W}^{\text{out}}$, 常采用线性回归方法求解, 如下式所示:

$$\mathbf{W}^{\text{out}} = \mathbf{M}^{-1}\mathbf{H}, \quad (3)$$

其中, $\mathbf{M} = [\mathbf{x}(H_0)^T, \mathbf{x}(H_0+1)^T, \dots, \mathbf{x}(H)^T]^T$,

$$\mathbf{H} = [\mathbf{y}(H_0), \mathbf{y}(H_0+1), \dots, \mathbf{y}(H)]^T.$$

图 1 标准 ESN 网络拓扑图

实际上, 应用线性回归进行参数率定容易导致病态解问题, 使得输出权重能够达到 10^8 以上, 而较大的输出权重则表示模型的泛化能力较差^[26], 因此, 需要对 ESN 输出权重的确定方式进行改进.

3 贝叶斯回声状态网络

3.1 贝叶斯回声状态网络

对于 ESN 模型给定的输入和储备池状态 $\mathbf{z}$ 及期望输出 t , 通常选取误差平方和为性能评价函数, 即

$$E_D = \frac{1}{2} \sum_{i=1}^n (f(\mathbf{z}_i; \mathbf{W}^{\text{out}}) - t_i)^2. \quad \text{本文依据正则化技术对}$$

性能评价函数引入正则项 $E_W = \frac{1}{2} \sum_{i=1}^L (\mathbf{W}_i^{\text{out}})^2$ 而构建总误差函数, 定义为

$$F(\mathbf{W}^{\text{out}}) = \beta E_D + \alpha E_W, \quad (4)$$

其中, α 和 β 为超参数.

贝叶斯理论注重于权重在权空间的概率分布^[27]. 无样本数据时, 若 $p(\mathbf{W}^{\text{out}})$ 为输出权重的先验概率分布, 根据贝叶斯理论, 对于给定样本数据集 D , 可得输出权重的后验概率分布为

$$p(\mathbf{W}^{\text{out}} | D) = \frac{p(D | \mathbf{W}^{\text{out}}) p(\mathbf{W}^{\text{out}})}{p(D)}, \quad (5)$$

其中, $p(D | \mathbf{W}^{\text{out}})$ 为似然函数; $p(D)$ 为归一化因子.

无样本数据时, 由于对输出权重分布具有较少的知识, 因此, 先验分布可认为是一个较宽的分布, 一旦有了数据, 才可转化为后验分布. 为得到后验分布, 需先获得先验分布 $p(\mathbf{W}^{\text{out}})$ 和似然函数 $p(D | \mathbf{W}^{\text{out}})$. 假定 $p(\mathbf{W}^{\text{out}})$ 服从常见的高斯分布, 则

有

$$p(\mathbf{W}^{\text{out}}) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W), \quad (6)$$

其中, $Z_W(\alpha)$ 为归一化因子, 可用下式表示:

$$Z_W(\alpha) = \int e^{(-\alpha E_W)} d\mathbf{W}^{\text{out}} = \left(\frac{2\pi}{\alpha} \right)^{\frac{L}{2}}. \quad (7)$$

若样本数据是独立选择的, 则似然函数 $p(D | \mathbf{W}^{\text{out}})$ 为

$$p(D | \mathbf{W}^{\text{out}}) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D), \quad (8)$$

其中, $Z_D(\beta)$ 为归一化因子; 假定期望输出 t 由高斯噪声均值为 0 的平滑函数生成, 则期望输出 t 的概率为

$$p(t | \mathbf{z}, \mathbf{W}^{\text{out}}) \propto \exp \left[-\frac{\beta}{2} (y(\mathbf{z}; \mathbf{W}^{\text{out}}) - t)^2 \right]. \quad (9)$$

则似然函数式(8)可写为

$$p(D | \mathbf{W}^{\text{out}}) = \frac{1}{Z_D(\beta)} \times \exp \left(-\frac{\beta}{2} \sum_{i=1}^n (f(\mathbf{z}_i; \mathbf{W}^{\text{out}}) - t_i)^2 \right), \quad (10)$$

其中 $Z_D(\beta)$ 可用下式表示:

$$Z_D(\beta) = \int e^{-\beta E_D} dD = \left(\frac{2\pi}{\beta} \right)^{\frac{n}{2}}. \quad (11)$$

结合(6)和(9)式, 利用似然函数不依赖于 α , 先验概率不依赖于 β , 可得后验概率分布为

$$\begin{aligned} p(\mathbf{W}^{\text{out}} | D) &= \frac{1}{Z_M(\alpha, \beta)} \exp(-\beta E_D - \alpha E_W) \\ &= \frac{1}{Z_M} \exp[-F(\mathbf{W}^{\text{out}})], \end{aligned} \quad (12)$$

其中,

$$Z_M(\alpha, \beta) = \int \exp(-\beta E_D - \alpha E_W) d\mathbf{W}^{\text{out}}. \quad (13)$$

由于 $Z_M(\alpha, \beta)$ 与 $\mathbf{W}^{\text{out}}$ 无关, 而(12)式作为高微积分, 无法直接解析得到, 因此求解网络最优输出权重, 即通过最小化误差函数 $F(\mathbf{W}^{\text{out}})$ 等价于最大化后验概率分布实现. 将 $F(\mathbf{W}^{\text{out}})$ 在最小值 $\mathbf{W}_{\text{MP}}^{\text{out}}$ 附近以二阶泰勒级数展开式, 可得

$$\begin{aligned} F(\mathbf{W}^{\text{out}}) &= F(\mathbf{W}_{\text{MP}}^{\text{out}}) \\ &+ \frac{1}{2} (\mathbf{W}^{\text{out}} - \mathbf{W}_{\text{MP}}^{\text{out}}) \mathbf{A} (\mathbf{W}^{\text{out}} - \mathbf{W}_{\text{MP}}^{\text{out}}), \end{aligned} \quad (14)$$

其中, $\mathbf{A}$ 为 $F(\mathbf{W}^{\text{out}})$ 的海森矩阵.

由于以概率形式考虑了参数的不确定性, 贝叶斯理论能够克服神经网络模型参数率定的过拟合现象, Nabney^[22]对此进行了详细阐述. 假定有两个不同的模型 M_1 和 M_2 , 且 M_2 比 M_1 拥有更多的模型参数, 则依据贝叶斯理论每个模型的后验概率可写为

$$p(M_i | D) = \frac{p(D | M_i) p(M_i)}{p(D)}. \quad (15)$$

由于 $p(D)$ 与模型相互独立, 因此模型比较时可忽略. 对于 M_1 和 M_2 一般认为先验分布 $p(M_i)$ 是相同的, 可以似然函数 $p(D | M_i)$ 进行比较, 则

$$p(D | M_i) = \int p(D | \mathbf{W}, M_i) p(\mathbf{W} | M_i) d\mathbf{W}. \quad (16)$$

$p(D | \mathbf{W}, M_i)$ 为似然函数, $p(\mathbf{W} | M_i)$ 为权重后验分布. 由于模型 M_2 较为复杂, 最优的参数 $\mathbf{W}_{\text{MP}}$ 可能更适于数据拟合, 因此, 后验分布 $p(D | \mathbf{W}_{\text{MP}}, M_2)$ 的峰值大于 M_1 的峰值. 但是, 模型越复杂, 对参数调整越敏感, $p(D | \mathbf{W}_{\text{MP}}, M_1)$ 使得 M_2 的后验分布 $\Delta \mathbf{W}_{\text{posterior}}$ 比 M_1 的后验分布 $\Delta \mathbf{W}_{\text{posterior}}$ 较为狭窄. 可直观认为后验分布峰值明显, 则(16)式的积分近似为

$$p(D | M_i) \approx p(D | \mathbf{W}_{\text{MP}}, M_i) p(\mathbf{W}_{\text{MP}} | M_i) \Delta \mathbf{W}_{\text{posterior}}, \quad (17)$$

其中, 先验概率 $p(\mathbf{W}_{\text{MP}} | M_i)$ 为某个较大区域 $\Delta \mathbf{W}_{\text{prior}}$ 的均匀分布, (17)式可简化为

$$p(D | M_i) \approx p(D | \mathbf{W}_{\text{MP}}, M_i) \left(\frac{\Delta \mathbf{W}_{\text{posterior}}}{\Delta \mathbf{W}_{\text{prior}}} \right). \quad (18)$$

(18)式中, $p(D | \mathbf{W}_{\text{MP}}, M_i)$ 是训练数据在最优权重值处的似然函数, 而 $\Delta \mathbf{W}_{\text{posterior}} / \Delta \mathbf{W}_{\text{prior}}$ (小于 1) 表示一个使模型具有特定权重后验概率分布的惩罚因子, 且与模型参数数目大约呈指数关系, 并随着模型逐渐复杂而减小, 也就是, 该因子惩罚 M_2 超过 M_1 , 可防止网络复杂时出现过拟合问题. 因此, 无论神经网络模型多么复杂, 基于贝叶斯理论框架能够自动惩罚高度复杂的模型, 可避免模型出现过拟合现象^[23].

3.2 超参数选择

根据 Laplace 逼近^[22], 则网络输出权重的后验概率分布近似为: $p(w|D) \approx p(w|\alpha^*, \beta^*, D)$, 其中 α^* 和 β^* 均表示最优值. 因此, 为使得输出权重的后验概率最大化, 需要获得 α^* 和 β^* . 依据文献[22,24], 可得 α 和 β 的后验分布为

$$p(\alpha, \beta|D) = \frac{p(D|\alpha, \beta)p(\alpha, \beta)}{p(D)}, \quad (19)$$

其中, $p(\alpha, \beta)$ 为 α 和 β 的先验分布. 由于归一化因子 $p(D)$ 与 α 和 β 是相互独立, 则最大化 $p(\alpha, \beta|D)$ 等价于最大化似然函数 $p(D|\alpha, \beta)$. 考虑到似然函数与 β 相互独立以及先验分布与 α 相互独立, 同时结合(7)和(11)式可得

$$\begin{aligned} p(D|\alpha, \beta) &= \int p(D|W^{\text{out}}, \beta)p(W^{\text{out}}|\alpha)dW^{\text{out}} \\ &= \frac{1}{Z_D(\beta)Z_W(\alpha)} \int e^{-M(W^{\text{out}})} dW^{\text{out}} \quad (20) \\ &= \frac{Z_M(\alpha, \beta)}{Z_D(\beta)Z_W(\alpha)}. \end{aligned}$$

将 $Z_W(\alpha)$, $Z_D(\beta)$, $Z_M(\alpha, \beta)$ 代入(20)式, 并取对数, 然后分别对 α 和 β 求一阶偏导, 可得最优值 α^* 和 β^* :

$$\alpha^* = \gamma/2E_W; \beta^* = (n-\gamma)/2E_D. \quad (21)$$

其中, $\gamma = \sum_{i=1}^L \frac{\lambda_i}{\lambda_i + \alpha}$, λ_i 为误差函数 E_D 海森矩阵的特征值.

3.3 模型参数率定

BESN 参数率定步骤如下:

- 1) 设置 BESN 相关参数, 并初始化权重 W^{in} , W 和 W^{back} ;
- 2) 依据样本和公式(1), 求出储备池状态 $x(k)$;
- 3) 初始化超参数 α 和 β 及输出权重 W^{out} ;
- 4) 依据 $\alpha, \beta, W^{\text{out}}$ 和公式(14), 计算 $F(W^{\text{out}})$;
- 5) 使得 $F(W^{\text{out}})$ 最小化, 利用 Levenberg-Marquardt 算法求解最优输出权重;
- 6) 依据当前输出 W^{out} , 计算 E_W, E_D , 并根据(21)式更新 α 和 β ;
- 7) 检验是否满足计算精度要求, 若满足, 输出最优权重, 停止计算; 否则, 转向 4 继续迭代计算.

4 实例应用

为验证本文所提模型的性能, 选择安砂和新丰江两座水库 1999/1/1~2010/12/31 的降雨径流资料进行预见期为 1 日的日径流模拟. 因资料所限, 本文对于两座水库不考虑雨量站点的分布状况, 均采用流域日平均降雨径流数据开展模拟工作. 分别选择两座水库 1999/1/1~2007/12/31 共 9 年的降雨径流资料作为训练数据, 其余 3 年的资料作为检验数据. 为更好地评测 ESN 及 BESN 的模型性能, 采用常用的三层 BP 神经网络模型作为对比, BP 神经网络模型输入输出与 ESN 和 BESN 模型相同, 而隐层节点数通过试错法确定. 采用均方根误差(RMSE)、确定性系数(CE)、相关系数(R)和相对误差(MAPE)4 个常用评价指标, 作为模型性能评价标准. 由于本文重点研究标准 ESN 模型的过拟合问题, 因此仅主要分析检验段各模型的预测效果.

4.1 实例一: 安砂水库日径流预测

安砂水库位于闽江支流沙溪上游, 是沙溪流域的龙头水库; 水库控制流域面积 5184 km², 库容 6.4×10⁸ m³, 装机容量 11.5×10⁴ kW; 水库以防洪、发电为主, 兼有灌溉、养殖和改善航运等功能. 由于受亚热带季风气候影响, 流域内汛枯期降雨量变化范围较大, 径流情势比较复杂.

对于神经网络输入确定问题, 常用方法主要三种, 即经验法, 互信息法和相关分析法. 相关分析法(包括自相关、偏自相关和交叉相关)依据相关系数大小确定影响滞时, 以此确定输入因子. 本文主要采用相关分析法. 根据相关分析, 确定将 2 日降雨和 3 日径流数据作为各模型的输入因子. 对于 BESN 模型, 采用与 ESN 模型相同的模型结构, 依据试验选择储备池维数为 100, 稀疏连接设置为 5%, 连接权重谱半径选为 0.85, 储备池激活函数为双曲正切函数; 对于 BP 神经网络模型, 确定模型结构为 5-10-1.

图 2 为安砂水库各模型检验段拟合过程. 从图 2 可看出, 除个别峰值点预测略有明显偏差外, 三种模型大致上均能提供较为满意的预测结果. 为了进一步评价模型性能, 图 3 给出了检验段各模型预测值与实际值的散点图. 从图 3 中可看出, 当实测流量大约低于 800 m³/s 时, 三种模型预测值的离散性非常接近; 当实测流量大于等于 800 m³/s 时, BP 神经网络和 ESN

模型相比 BESN 模型具有较大的离散性, 表明 BESN 模型具有更好的预测效果。

为直观理解各模型的预测精度, 图4给出了各模型在检验段各时段的预测绝对误差。结合图 2, 从图 4 整体上可看出, BESN 的预测误差分布相对比较均

匀, 尽管一些较大流量值的预测误差值较大, 但与 BP 神经网络和 ESN 模型的预测误差相比, BESN 模型的预测误差仍然可以接受。为进一步量化评价各模型性能, 表 1 给出了各模型在率定和检验段的指标统计。在训练段, ESN 模型各统计指标均优于 BP 神经

图 2 安砂水库检验段各模型日径流预测结果对比

图 3 安砂水库检验段各模型日径流预测散点图

图 4 安砂水库检验段各模型日径流预测误差

表 1 安砂水库日径流预测各模型参数率定及检验性能统计

模型	训练段				检验段			
	RMSE	CE	R	MAPE(%)	RMSE	CE	R	MAPE(%)
BP	53.51	0.948	0.962	26.17	74.92	0.919	0.929	32.24
ESN	50.17	0.955	0.977	18.13	78.98	0.911	0.955	21.13
BESN	55.57	0.942	0.971	19.05	63.08	0.944	0.972	20.53

网络. 尽管 BESN 模型的 RMSE 值比 BP 神经网络模型的 RMSE 值略差, 但其 CE, R 和 MAPE 值则非常接近或优于 BP 神经网络的 CE, R 和 MAPE 值; 与 ESN 模型相比, BESN 模型除指标 R 非常相近外, 其余指标均稍劣于 ESN 模型. 在检验段, ESN 模型与 BP 神经网络相比, 并不具有优势; 由于可能受到流域降雨-径流形成要素、输入因子选择、模型结构及模型泛化能力等复杂因素影响, BP 神经网络模型在 RMSE 和 CE 两指标方面略优于 ESN 模型, 而 ESN 模型在 R 和 MAPE 两指标方面优于 BP 神经网络. BESN 模型各统计指标相对于 BP 神经网络和 ESN 模型优势均比较明显, 其中, 与 BP 神经网络模型相比, BESN 模型在 RMSE, CE, R 和 MAPE 指标方面分别提高了 15.8%, 2.7%, 4.7% 和 36.3%; 与 ESN 模型相比, BESN 模型在 RMSE, CE, R 和 MAPE 指标方面分别提高了 20.1%, 3.5%, 1.8% 和 2.9%; 此外, 在峰值预测方面, BP 神经网络、ESN 和 BESN 的预测误差分别为 21.4%, 21.6% 和 15.5%. 从统计指标看, 安砂水库实例应用中, BESN 模型优于 ESN 模型和 BP 神经网络模型, 而 BP 神经网络模型和 ESN 模型相比预测性能无明显差别.

4.2 实例二: 新丰江水库日径流预测

新丰江水库位于东江支流新丰江上; 水库控制流域面积 5813 km², 库容 139×10⁸ m³, 装机容量 33.5×10⁴ kW; 水库以发电为主, 兼有防洪、灌溉和改善航运等功能. 与安砂水库类似, 新丰江水库同样受亚热带季风气候影响, 流域内汛枯期降雨量变化范围较大, 径流情势比较复杂.

根据相关分析法, 确定将 5 日降雨和 4 日径流数据作为各模型的输入因子. ESN 和 BESN 采用相同的模型结构, 依据试验选择储备池维数为 80, 稀疏连接设置为 5%, 连接权重谱半径选为 0.8, 储备池激活函数为双曲正切函数; 对于 BP 神经网络模型, 确定模型结构为 5-9-1.

图 5 为新丰江水库各模型检验段拟合过程. 从图 5 可看出, 除个别峰值点预测略有明显偏差外, 三种

模型大致上均能提供较为满意的预测结果. 图 6 给出了检验段各模型预测值与实际值的散点图. 从图 6 中可看出, 当实测流量大约低于 500 m³/s 时, BP 神经网络预测值的离散性略大于 ESN 和 BESN 预测值的离散性, 而 ESN 和 BESN 相比模型预测性能差别不大; 当实测流量大于等于 500 m³/s 时, BP 神经网络与 ESN 模型预测的离散性相近, 但两者的离散性略大于 BESN 模型的离散性, 总体表明 BESN 模型预测效果较好.

图 7 给出了各模型在检验段各时段的预测绝对误差. 结合图 5, 从图 7 整体上可看出, BESN 的预测误差分布相对比较均匀, 尽管一些较大流量值的预测误差较大, 但与 BP 神经网络和 ESN 模型的预测误差相比, BESN 模型的预测误差仍然可以接受. 表 2 给出了各模型在率定和检验段的指标统计. 在训练段, ESN 模型除指标 R 与 BP 神经网络模型接近外, 其余指标均明显优于 BP 神经网络模型. BESN 模型在 RMSE, CE, R 和 MAPE 4 个统计指标方面均优于 BP 神经网络模型; 与 ESN 模型相比, BESN 模型除指标 CE 和 R 非常相近外, 其余两指标均稍劣于 ESN 模型. 在检验段, 除指标 R 外, ESN 模型在指标 RMSE, CE 和 MAPE 方面均明显优于 BP 神经网络模型, 分别提高了 15.9%, 2.4% 和 30.3%; BESN 模型在统计指标 RMSE, CE 和 MAPE 方面相对于 BP 神经网络和 ESN 模型优势均比较明显, 其中, 与 BP 神经网络模型相比, BESN 模型在指标 RMSE, CE 和 MAPE 方面分别提高了 26.3%, 3.8% 和 33.4%; 与 ESN 模型相比, BESN 模型在指标 RMSE, CE 和 MAPE 方面分别提高了 12.4%, 1.4% 和 4.4%; 在峰值预测方面, BP 神经网络、ESN 和 BESN 的预测误差分别为 26.0%, 13.2% 和 11.5%. 从统计指标看, 新丰江水库实例应用中, BESN 模型优于 ESN 模型和 BP 神经网络模型, ESN 模型优于 BP 神经网络模型.

综合以上两个实例, BESN 模型预报精度相比 BP 神经网络和 ESN 模型具有一定程度的提高. 为表明本文重点关心的过拟合问题, 可对 ESN 和 BESN 两个

图5 新丰江水库检验段各模型日径流预测结果对比

图6 新丰江水库检验段各模型日径流预测散点图

图7 新丰江水库检验段各模型日径流预测误差

表2 新丰江水库日径流预报各模型参数率定及检验性能统计

模型	训练段				检验段			
	RMSE	CE	R	MAPE(%)	RMSE	CE	R	MAPE(%)
BP	92.43	0.917	0.976	40.56	86.42	0.925	0.970	33.04
ESN	66.44	0.958	0.979	23.84	72.71	0.947	0.972	23.02
BESN	68.68	0.955	0.977	24.73	63.70	0.960	0.980	22.01

模型再次进行分析, 从表 1 和 2 中可看出, 虽然 ESN 模型在训练段结果较优, 但在检验段各指标明显劣于 BESN 模型, 这个结果可以从图 3, 4, 6 和 7 中进一步得到证实. 主要原因可归结于: ESN 模型采用了传统性能评价函数, 将模型参数率定假定为确定性问题, 较少考虑模型泛化能力; 而 BESN 模型在整个权重空间内考虑了模型参数的不确定性, 通过对传统目标函数引入正则项而构建新的目标函数, 并依据贝叶斯理论将其转化为求解权重后验概率分布最大化问题. 考虑权重在权空间的概率密度分布使得 BESN 模型具有更好的泛化能力, 能够模拟复杂的日降雨-径流过程, 并能提供较好的预测精度, 因此, BESN 模型是一种有效、可行的预报方法.

5 结论

回声状态网络(ESN)作为一种新的递归神经网络, 具有模型简单、参数训练速度快的优点, 在水文预报领域已有少量应用, 但由于标准 ESN 输出权重直接由线性回归获得, 容易出现过拟合问题. 将贝叶斯理论和 ESN 相结合, 提出了基于贝叶斯回声状态网络(BESN)的日径流预报模型. 通过对传统性能评价函数引入正则项, 依据权重后验概率密度最大化而获得最优输出权重, 提高了 BESN 模型的泛化能力. 实例结果表明, BESN 模型比传统 BP 神经网络和标准 ESN 模型具有更好的预报精度, 是一种有效、可行的日径流预报方法.

参考文献

- 1 刁艳芳, 王本德. 基于不同风险源组合的水库防洪预报调度方式风险分析. 中国科学: 技术科学, 2010, 40: 1140–1147
- 2 张艳平, 王国利, 彭勇, 等. 考虑洪水预报误差的水库汛限水位动态控制风险分析. 中国科学: 技术科学, 2011, 41: 1256–1261
- 3 蒋晓辉, 刘昌明, 王煜, 等. 用混合回归模型预测黄河三门峡站年径流变化的初步研究. 中国科学 E 辑: 技术科学, 2004, 34: 95–102
- 4 张小琴, 包为民, 王涛, 等. 新安江模型与水动力学模型结合修正研究. 水文, 2008, 29: 18–21
- 5 李彦彬, 尤凤, 黄强, 等. 多元变量径流预测的最小二乘支持向量机模型. 水力发电学报, 2010, 29: 28–33
- 6 Wang W C, Chau K W, Cheng C T, et al. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol, 2009, 374: 294–306
- 7 胡铁松, 袁鹏, 丁晶. 人工神经网络在水文水资源中的应用. 水科学进展, 1995, 6: 76–82
- 8 Hsu K L, Gupta H V, Sorooshian S. Artificial neural network modeling of the rainfall-runoff process. Water Resour Res, 1995, 31: 2517–2530
- 9 Coulibaly P, Anctil F, Bobee B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J Hydrol, 2000, 230: 244–257
- 10 Hu T S, Lam K C, NG S T. River flow time series prediction with range-dependent neural network. Hydrolog Sci J, 2001, 46: 729–745
- 11 Hu T S, Lam K C, NG S T. A modified neural network for improving river flow prediction. Hydrolog Sci J, 2005, 50: 299–318
- 12 Wu C L, Chau K W. Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol, 2011, 399: 394–409
- 13 Coulibaly P, Anctil F, Rasmussen P, et al. A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. Hydrol Process, 2000, 14: 2755–2777
- 14 Jaeger H. The echo state approach to analyzing and training recurrent neural networks. German National Research Center for Information Technology, Bremen. 2001
- 15 Jaeger H, Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 2004, 304: 78–80
- 16 Sacchi R, Ozturk M C, Principe J C, et al. Water inflow forecasting using the echo state network: A brazilian case study. In: International Joint Conference on Neural Networks. Orlando, Florida, 2007
- 17 Coulibaly P. Reservoir computing approach to Great Lakes water level forecasting. J Hydrol, 2010, 381: 76–88
- 18 de Vos N. Echo state networks as an alternative to traditional artificial neural networks in rainfall-runoff modelling. Hydrol Earth Syst Sci, 2013, 17: 253–267
- 19 Wyffels F, Schrauwen B, Stroobandt D. Stable output feedback in reservoir computing using ridge regression. In: Artificial Neural Networks-ICANN 2008. Berlin: Springer, 2008. 808–817
- 20 Shi Z, Han M. Support vector echo-state machine for chaotic time-series prediction. IEEE T Neural Networ, 2007, 18: 359–372
- 21 李向阳, 程春田, 林剑艺. 基于 BP 神经网络的贝叶斯概率水文预报模型. 水利学报, 2006, 37: 354–359

- 22 Nabney I T. Netlab Algorithms for Pattern Recognition. New York: Springer, 2004
- 23 Khan M S, Coulibaly P. Bayesian neural network for rainfall-runoff modeling. *Water Resour Res*, 2006, 42: W07409
- 24 Maiti S, Tiwari R K. Neural network modeling and an uncertainty analysis in Bayesian framework: A case study from the KTB borehole site. *J Geophysical Res*, 2010, 15: B10208
- 25 Liu Y, Liu Q, Wang W, et al. Data-driven based model for flow prediction of steam system in steel industry. *Inform Sci*, 2012, 193: 104–114
- 26 Jaeger H. Reservoir riddles: Suggestions for echo state network research. In: *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, 2005
- 27 Hippert H S, Taylor J W. An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting. *Neural Networks*, 2010, 23: 386–395

Daily rainfall-runoff forecasting using Bayesian echo state network

LI BaoJian, CHENG ChunTian, WU XinYu & WANG Sen

Institute of Hydropower System and Hydroinformatics, Dalian University of Technology, Dalian 116024, China

The echo state network (ESN) is simpler and costs less training time than traditional recurrent neural networks. Due to linear regression algorithm usually adopted by standard ESN to calibrate model parameters, the over-fitting phenomenon easily occurs. To overcome this shortcoming, a Bayesian echo state network (BESN) model is proposed for daily rainfall-runoff forecasting. The BESN model combined Bayesian theory and ESN obtains the optimal output weights via maximizing posterior probabilistic density and improves its generalization ability. Two Case studies on daily inflow forecasting for Ansha Reservoir and Xinfengjiang Reservoir show that the BESN model is effective and feasible and can provide better forecast accuracy than the traditional BP neural network and ESN models.

recurrent neural network, echo state network, Bayesian theory, rainfall-runoff forecasting

doi: 10.1360/N092013-00036