

基于信息扩散近似推理的年降水量预测模型

黄 强¹, 原文林¹, 陈晓楠², 张洪波¹, 王义民¹

(1 西安理工大学 水利水电学院, 陕西 西安 710048; 2 南水北调中线干线建设管理局, 北京 100053)

【摘 要】 **【目的】** 研究信息扩散近似推理方法在年降水量预测中的应用, 并对比其他方法分析其推广前景。**【方法】** 通过分析年降水量时间序列的特性, 提出了基于当前趋势以及相邻年份降水量的年降水预测规则。利用信息扩散近似推理描述年降水量间的复杂非线性关系, 并以某灌区长系列降水资料为样本进行实例计算。**【结果】** 信息扩散近似推理方法预测效果较好, 该方法误差绝对值和为 1.673, 小于人工神经网络和线性自回归方法的统计结果。**【结论】** 信息扩散近似推理可将样本点转换成模糊集, 部分弥补了由于数据的不完备性所造成的信息空白, 并可将其矛盾模式转换成兼容模式。通过与传统预测方法相比较发现, 该模型能够很好地光滑样本数据以及较好地发掘知识, 有较高的预测精度和推广应用价值。

【关键词】 信息扩散近似推理; 年降水量; 预测模型

【中图分类号】 S274.1; P333.1

【文献标识码】 A

【文章编号】 1671-9387(2009)02-0213-07

Prediction model of annual precipitation based on information-diffusion approximate reasoning

HUANG Qiang¹, YUAN Wen-lin¹, CHEN Xiao-nan²,
ZHANG Hong-bo¹, WANG Yi-min¹

(1 Institute of Water Resources and Hydroelectric Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China;

2 Construction and Administration Bureau of South-to-North Water Diversion Middle Route Project, Beijing 100053, China)

Abstract: **【Objective】** The method of applying information-diffusion approximate reasoning to predict annual precipitation was studied and compared with other methods in order to analyze the application prospects. **【Method】** According to features of precipitation time-series, prediction rules were suggested based on current tendency and the neighbor year precipitation. This may help the information-diffusion approximate reasoning describe the complex nonlinear relation in precipitation data. Applying the rules a precipitation time-series in an irrigation area as an example, the forecasting result was obtained. **【Result】** The information-diffusion approximate reasoning method had fewer errors and a better effect on precipitation prediction than artificial neural network and linear autoregressive method. **【Conclusion】** The method can transfer sample points to fuzzy sets and take advantage of more information, and even may switch conflict mode to compatible mode. Results indicate that annual precipitation prediction with information-diffusion approximate reasoning model is good at the mining of uncertain knowledge, and can find out more information and make the data series more smoothly than traditional methods. In fact, this is a new data mining model for time-series. Any similar problem can be dissolved better by the method. It is significant to spread this method.

Key words: information-diffusion approximate reasoning; annual precipitation; prediction model

降水规律分析与预报是农业干旱风险管理中一项十分重要的内容。如果能够较准确地预测出未来

* [收稿日期] 2008-04-16

[基金项目] 国家自然科学基金项目(50679070, 50709027); 水利部公益性行业专项(2007SHZ1-19)

[作者简介] 黄 强(1958—), 男, 四川梓潼人, 教授, 博士, 博士生导师, 主要从事水资源系统工程研究。E-mail: sy-sj@xaut.edu.cn

年的年降水量,就可以以此为基础对作物产量进行预估,并可以评价未来年份农业的干旱程度,从而提前做好抗旱准备。因此,预测年降水量是一项重要而有意义的研究。目前,由于科学技术水平的限制,很难从物理成因上完成中长期降水的预测,因而多采用时间序列、神经网络或灰色预测等方法,通过挖掘历史数据的规律来实现年降水量的预测^[1-6]。但是,当用于训练的模式互相矛盾时,神经网络就会由于不知道如何适应矛盾模式,而产生较大的系统误差,导致神经网络不收敛。当实际提供的数据样本不完备,例如小样本数据,则会令传统方法的预测效果不理想。为了能够更多地从已有数据中获得信息,且克服训练模式矛盾的问题,黄崇福^[7-9]在模糊推理的基础上,提出了信息扩散近似推理技术,对预测进行了初步研究,对预测效果有了很大的改善。但目前该方法的进一步研究相对较少^[10-12]。本研究重点探讨了信息扩散近似推理方法在年降水量预测中的应用,并对比其他方法分析了其推广前景。

1 信息扩散

1.1 信息扩散的定义

信息分配法是用模糊集方法研究函数逼近的一种新途径。该方法的优点是其估计精度较传统直方图高,特别是当不能保证给定样本的总体假设正确且样本容量较小时,为了避免数据结构的人为破坏,并充分利用每一个知识样本点携带的信息,利用该方法则更具有优势。但是在工程应用中,如何选取合适的控制点空间 $U=\{u_1,u_2,\cdots,u_m\}$ 和分配函数 μ ,一直是困扰人们的实际问题。为此还需对小样本进行较深入地分析。

令 $X=\{x_1,x_2,\cdots,x_n\}$ 是一个随机样本,令 U 代表随机变量 X 所有可能值的集合,称为 X 的论域,即 U 是随机变量的定义域。 X 和 U 分别简称为样本和论域。

定义 1 令 X 是一个样本, V 是论域 U 的一个子集。从 $X\times V$ 到 $[0,1]$ 的一个映射:

$$\begin{aligned} \mu: X\times V &\rightarrow [0,1], \\ (x,v) &\mapsto \mu(x,v), \\ \forall (x,v) &\in X\times V, \end{aligned}$$

称为 X 在 V 上的一个信息扩散,如果它是递减的,即: $\forall x\in X, \forall v',v''\in V, \|v'-x\|\leq \|v''-x\|$, 则 $\mu(x,v')\geq \mu(x,v'')$ 。 μ 称为一个扩散函数, V 称为一个监控空间。可将 μ 简称为扩散。

推论 1 一个分配函数是一个扩散函数。

定义 2 $\mu(x,v)$ 是充分的,当且仅当 $V=U$,用 $\mu(x,\mu)$ 表示。

推论 2 一个分配函数是不充分的。

由推论 1 可知,信息分配是信息扩散的特例。在形式化表达方面,它们的主要区别是:在信息分配模型中, U 是 X 的一个离散论域,称为控制点空间;在信息扩散模型中, U 是一个随机变量的定义域,它的某个子集可能被拿来担当监控空间。当监控空间是随机变量定义域本身时,信息扩散是充分的。信息分配模型中的控制点空间是随机变量定义域的真子集,因此,一个分配函数是不充分的。

定义 3 $\mu(x,v)$ 是守恒的,当且仅当 $\forall x\in X$, 其在论域 U 上的积分值是 1。即:

$$\int_U \mu(x,u) du = 1. \tag{1}$$

如果随机变量的定义域 U 是离散的,假设 $U=\{u_1,u_2,\cdots,u_m\}$,则守恒条件是:

$$\sum_{j=1}^m \mu(x,u_j) = 1, \forall x\in X. \tag{2}$$

例如,一维线性信息分配函数是一个守恒扩散函数。

定义 4 令 X 是一个给定样本, $I_j=[u_0+(j-1)h,u_0+jh], j=1,2,\cdots,m$, 是其论域 U 上的 m 个区间,一个区间扩散定义为:

$$\mu(x,u) = \begin{cases} \frac{1}{h}, & \text{如果 } x \text{ 在 } u \text{ 所在的区间;} \\ 0, & \text{其他.} \end{cases} \tag{3}$$

式中:“ u 所在的区间”即区间 $I_j, j=1,2,\cdots,m$ 。

容易验证,区间扩散是守恒的。由此可以用一个充分的扩散区间 $\mu(x,u)$ 得到概率密度函数的一个直方图估计,即:

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \mu(x_i,x), x\in R. \tag{4}$$

式中: n 是给定样本的容量。注意区间宽度 h 在该式中没有出现,这是因为区间扩散函数 $\mu(x,u)$ 中已包含这一平均化因子。

定义 5 令 X 是一个给定样本, U 是其论域。 $\mu(x,u)$ 是 X 在 U 上的一个信息扩散。对于一个 x , 由 $\mu(x,u)$ 在 U 上定义了一个模糊集,对于所有的 x , 由 $\mu(x,u)$ 产生的模糊集集合:

$$D(X) = \{\mu(x,u) | x\in X, u\in U\}, \tag{5}$$

称其为 X 在 U 上由 $\mu(x,u)$ 产生的模糊集样本。

1.2 信息扩散的原理

当 X 是一个不完备样本时,将 X 变成模糊样本,有助于弥补进行函数逼近时由于数据不足造成的缺陷。可用信息扩散原理描述这个优点。

1.2.1 相伴特征函数 设给定样本 $X = \{x_i | i=1, 2, \dots, n\}$, 其定义域为 U 。样本点 x_i 是分明观测值或分明矢量。

定义6 X 在 U 上的相伴特征函数定义为:

$$\chi_{X,U}(x_i, u) = \begin{cases} 1, u = x_i; \\ 0, u \neq x_i. \end{cases} \quad (6)$$

式中: $x_i \in X, u \in U$ 。

1.2.2 扩散估计 若给定一个扩散函数, 则可以利用其将一个给定样本 X 变成一个模糊样本 $D(X)$ 。用某个算子(模型)处理模糊样本 $D(X)$ 得到的估计, 称作一个扩散估计。形式定义如下。

定义7 令 X 是一个可以用算子(模型) γ 估计关系 R 的给定样本。如果估计是用模糊样本 $D(X)$ 得到, 则此估计称为 R 的扩散估计, 可表示为:

$$\hat{R}[\gamma, D(X)] = \{\gamma(\mu(x_i, u)) | x_i \in X, u \in U\}. \quad (7)$$

式中: $\mu(x_i, u)$ 是 X 在 U 上的扩散函数。

相应地, 如果估计是用 X 得到的, 则此估计称为 R 的非扩散估计, 表示为

$$\hat{R}(\gamma, X) = \{\gamma(\chi_{X,U}(x_i, u)) | x_i \in X, u \in U\}. \quad (8)$$

1.2.3 信息扩散原理 令 $X = \{x_1, x_2, \dots, x_n\}$ 是用来估计论域 U 上关系 R 的一个给定样本。假设 γ 是一个合理的算子, $\chi_{X,U}(x_i, u)$ 是相伴特征函数, 则所得非扩散估计为:

$$\hat{R}(\gamma, X) = \{\gamma(\chi_{X,U}(x_i, u)) | x_i \in X, u \in U\}. \quad (9)$$

当且仅当 X 不完备时, 一定存在一个合理的扩散函数 $\mu(x_i, u)$ 和一个相应算子 γ' , 用 $\mu(x_i, u)$ 取代 $\chi_{X,U}(x_i, u)$, γ' 调整 γ , 所得扩散估计为:

$$\hat{R}[\gamma', D(X)] = \{\gamma'(\mu(x_i, u)) | x_i \in X, u \in U\}. \quad (10)$$

使得:

$$\|R - \tilde{R}\| < \|R - \hat{R}\|. \quad (11)$$

式中: $\|\cdot\|$ 表示估计关系和真实关系间的误差。

信息扩散原理保证, 在给定样本不完备时, 存在可以改进非扩散估计的合理扩散函数。换言之, 当 X 不完备时, 一定存在某种途径能够拣起 X 的模糊信息, 以便更精确地估计作为函数逼近的一个关系 R 。

1.3 信息扩散的近似推理

定义8 令 U 是变量 x 的论域, V 是变量 y 的论域, 从 $U \times V$ 到 $[0, 1]$ 的一个映射:

$$\mu_R: U \times V \rightarrow [0, 1],$$

$$(x, y) \mapsto \mu_R(x, y),$$

$$\forall (x, y) \in U \times V,$$

可称作 x 和 y 之间的一个模糊关系。

若其表达了 x 和 y 之间的某种因果关系, 则称 μ_R 是 x 和 y 之间的一个因果型模糊关系。

在许多现有方法中, 用于构造模糊规则和模糊关系的模糊集是由研究者给定的^[13-15]。选择不同的模糊集将会得出不同的结果。虽然模糊集可以根据推理结果进行反馈调整, 但是许多时候没有足够的信息来支持选用较好的最初模糊集, 调整不能解决问题。如果有大量的数据, 样本数据的聚类轨迹曲线被认作输入、输出模糊关系函数。聚类分析技术可以用来选择合理的模糊集。当样本很小或者分散时, 不存在任何聚类描述曲线。聚类途径将会被信息扩散方法所代替。

假设 $\mu(x_i, u)$ 是 x_i 在 U 上的信息扩散函数。取最大值将它归一化, 记为 $\mu_{x_i}(u)$, 可以得到一个模糊分类函数, 它指明 u 属于模糊集“ x_i 周围”的程度, 利用正态扩散函数:

$$\mu(x_i, u) = \frac{1}{h_x \sqrt{2\pi}} \exp\left[-\frac{(u-x_i)^2}{2h_x^2}\right], \text{ 和}$$

$$\mu(y_i, v) = \frac{1}{h_y \sqrt{2\pi}} \exp\left[-\frac{(v-y_i)^2}{2h_y^2}\right], \quad (12)$$

并分别用:

$$\frac{1}{h_x \sqrt{2\pi}} \text{ 和 } \frac{1}{h_y \sqrt{2\pi}},$$

除之而实现归一化, 从而可以将样本点 (x_i, y_i) 转化为 2 个模糊集:

$$A_i = \int_U \mu_{x_i}(u) / u \text{ 和 } B_i = \int_V \mu_{y_i}(v) / v. \quad (13)$$

也就是:

$$A_i: \mu_{x_i}(u) = \exp\left[-\frac{(u-x_i)^2}{2h_x^2}\right], u \in U; \quad (14)$$

$$B_i: \mu_{y_i}(v) = \exp\left[-\frac{(v-y_i)^2}{2h_y^2}\right], v \in V. \quad (15)$$

显然, (x_i, y_i) 支持 $A_i \rightarrow B_i$ 的模糊规则成立。

为了能保留较多的信息, 可用关联乘积编码产生基于 $A_i \rightarrow B_i$ 的模糊关系, 即:

$$\mu_{R_i}(u, v) = \mu_{A_i}(u) \mu_{B_i}(v), u \in U, v \in V. \quad (16)$$

所以, 从 n 个样本点可以得到 n 个模糊关系。在此模型中, 所有模糊“若一”规则“若 u 是 A_i , 则 v 是 B_i ”的权重是相同的, 可以看作 1。

若分明输入值 x_0 是已知的, 则需推断从 x_0 和关系 R_i 得到的结果 y_0 。

在实际计算中, U 通常不连续, 所以 x_0 不等于 U 中的任何值。在此条件下可以用信息分配公式(17)得到一个模糊集, 作为输入:

$$\mu_{x_0}(u_j) = \begin{cases} 1 - |x_0 - u_j| / \Delta, & |x_0 - u_j| \leq \Delta; \\ 0, & |x_0 - u_j| > \Delta. \end{cases} \quad (17)$$

这里 $\Delta = u_{j+1} - u_j$ 。

从模糊集 $\tilde{x}_0$ 和 R_i 得到模糊推论 $\tilde{y}_0$, 其隶属函数为:

$$\mu_{y_0} = \sum_u \mu_{x_0}(u) \mu_{R_i}(u, v)。(18)$$

利用所谓的“取大”原理, 可以将 $\tilde{y}_0$ 非模糊化为一个分明值 $\hat{y}_i$, 并取得权值 w_i 。假定:

$$\mu_{y_0}(v') = \max_{v \in V} \{ \mu_{y_0}(v) \}。(19)$$

则最大非模糊化值 $\hat{y}_i = v'$, 权值为 $w_i = \mu_{y_0}(v')$ 。因而, 组合所有由 $R_1, R_2, \dots, R_n$ 得到的结果, 相应的输出值为:

$$\bar{y}_0 = (\sum_{i=1}^n w_i \hat{y}_i) / (\sum_{i=1}^n w_i)。(20)$$

由式(13)~(20)构成的模型称作信息扩散近似推理。

2 基于信息扩散近似推理的年降水量预测步骤

使用信息扩散近似推理来描述相邻年份降水量的非线性关系。具体步骤如下:

步骤 1 根据 n 年降水序列 $E = (e_1, e_2, \dots, e_n)$, 得出输入、输出向量 X, Y , 其中, 输入向量的分量 $x_i = e_i, i = 1, 2, \dots, n-1$; 输出向量的分量 $y_i = e_{i+1}, i = 1, 2, \dots, n-1$ 。

步骤 2 利用式(13)~(20)建立信息扩散近似推理模型。

步骤 3 利用信息扩散近似推理模型得到相邻年份降水量的关系。同时, 以当时连续 t 年的降水量平均值作为此时期的趋势, 利用上述方法建立当前趋势与未来年份的降水量关系。

步骤 4 利用上述 2 种非线性关系进行预测, 取平均值为最终结果。

3 算 例

以某灌区 1961~1992 年 32 年的年降水量数据作为模型训练的样本, 用 1993~2002 年 10 年的数据作为验证模型精度的样本。在进行信息扩散近似推理时, 仍然考虑相邻年份对未来年份的影响以及趋势对未来的影响。因此, 分别使用信息扩散近似推理模型, 对相邻年份的降水量关系以及趋势与未来相邻年份之间的关系进行推理, 并将这 2 种情况下所得的预测结果进行平均, 作为最终的输出。

3.1 相邻年份间的关系推理

令 1961~2002 年的归一化降水量数据为 $E = \{e_1, e_2, \dots, e_{42}\}$ 。将 1961~1991 年的数据作为输入样本 X , 1962~1992 年的数据作为输出样本 Y 。即有:

$$X = \{x_1, x_2, \dots, x_{31}\} = \{e_1, e_2, \dots, e_{31}\} = \{0.703, 0.570, \dots, 0.469\},$$

$$Y = \{y_1, y_2, \dots, y_{31}\} = \{e_2, e_3, \dots, e_{32}\} = \{0.570, 1.000, \dots, 0.405\}。$$

可利用 (x_i, y_i) 推导相邻年份之间的模糊关系。

令输入数据的离散论域为:

$$U = (u_1, u_2, \dots, u_{30}) = \{0.220, 0.248, \dots, 1.028\}。$$

此处步长 $\Delta = 0.028, u_1 = a_m - \Delta, u_{30} = b_m + \Delta$, 其中

$$a_m = \min_{1 \leq i \leq 31} \{x_i\}, b_m = \max_{1 \leq i \leq 31} \{x_i\}。$$

输出数据的离散论域与输入的相同。即:

$$V = \{v_1, v_2, \dots, v_{30}\} = \{0.220, 0.248, \dots, 1.028\}。$$

然后, 对 X 和 Y 取正态扩散系数, 有:

$$h_x = h_y = 0.051。$$

用正态信息扩散公式(14)和(15), 可将样本点 (x_i, y_i) 转化成 2 个模糊子集, 并按式(16)将其转化为模糊关系矩阵。例如, 令 $i = 1$, 则 $(x_1, y_1) = (0.703, 0.570)$, 模糊关系矩阵为:

	v_1	$\cdots$	v_{12}	v_{13}	v_{14}	v_{15}	v_{16}	$\cdots$	v_{30}
	0.220	$\cdots$	0.526	0.554	0.582	0.610	0.638	$\cdots$	1.028
$u_1(0.220)$	0.000	$\cdots$	0.000	0.000	0.000	0.000	0.000	$\cdots$	0.000
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$u_{12}(0.526)$	0.000	$\cdots$	0.026	0.031	0.032	0.027	0.019	$\cdots$	0.000
$u_{13}(0.554)$	0.000	$\cdots$	0.070	0.085	0.086	0.073	0.053	$\cdots$	0.000
$u_{14}(0.582)$	0.000	$\cdots$	0.161	0.194	0.196	0.168	0.120	$\cdots$	0.000
$u_{15}(0.610)$	0.000	$\cdots$	0.311	0.374	0.379	0.323	0.233	$\cdots$	0.000
$R_1 = u_{16}(0.638)$	0.000	$\cdots$	0.506	0.608	0.616	0.526	0.378	$\cdots$	0.000
$u_{17}(0.666)$	0.000	$\cdots$	0.694	0.834	0.844	0.721	0.518	$\cdots$	0.000
$u_{18}(0.694)$	0.000	$\cdots$	0.801	0.963	0.975	0.832	0.598	$\cdots$	0.000
$u_{19}(0.721)$	0.000	$\cdots$	0.779	0.937	0.949	0.809	0.582	$\cdots$	0.000
$u_{20}(0.749)$	0.000	$\cdots$	0.639	0.768	0.778	0.664	0.477	$\cdots$	0.000
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$u_{30}(1.028)$	0.000	$\cdots$	0.000	0.000	0.000	0.000	0.000	$\cdots$	0.000

当给定前一年的数据,假定为 $x=e_{32}=0.405$, 则可以利用式(17)中的信息分配公式将其转化成 U 上的一个模糊子集:

$$\tilde{x}=\sum_{j=1}^{30}\mu_{x_{32}}(u_j)/u_j=0/0.220+\cdots+0.357/0.387+0.643/0.415+\cdots+0/1.028。$$

利用式(18)的合成运算,可得到基于 R_1 的模糊推断,并根据“取大”原理(式 19),可得出估计值

表 1 基于相邻年份推理的 1993~2002 年降水量年型预测

Table 1 Annual forecast values from 1993 to 2002 based on near year reasoning

年份 Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
预测值 Forecast value	0.538	0.572	0.576	0.534	0.542	0.519	0.568	0.539	0.481	0.507

3.2 降水趋势对未来年份的影响规律

以连续 5 年降水量数据的平均值作为近来的趋势值,通过原始 32 年的训练样本可以得到 27 组新的数据,即:

$$x_i=\frac{1}{5}\cdot\sum_{j=i}^{i+4}e_i,i=1,2,\cdots,27,$$
$$y_i=e_{i+5},i=1,2,\cdots,27。$$

表 2 基于趋势推理的 1993~2002 年降水量年型预测

Table 2 Annual forecast values from 1993 to 2002 based on trend reasoning

年份 Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
预测值 Forecast value	0.515	0.526	0.560	0.515	0.518	0.524	0.522	0.583	0.505	0.517

表 3 1993~2002 年降水量预测值和实际值的比较

Table 3 Forecast and real value from 1993 to 2002

年份 Year	预测值 Forecast value	实际值 Real value	年份 Year	预测值 Forecast value	实际值 Real value
1993	0.526	0.682	1998	0.522	0.703
1994	0.549	0.661	1999	0.545	0.411
1995	0.568	0.385	2000	0.561	0.824
1996	0.525	0.435	2001	0.493	0.307
1997	0.530	0.340	2002	0.512	0.333

3.3 信息扩散近似推理方法与其他方法的对比

将信息扩散近似推理得出的结果与人工神经网络和传统的线性自回归方法计算结果进行比较,并使用实际数据与预测数据的误差绝对值和作为评判标准。将 1961~1991 年的降水量数据作为输入样本 X ,以 1962~1992 年的数据作为输出样本。人工神经网络的输入层神经元个数为 1,输出层神经元个数为 1,隐含层神经元个数为 2,取学习效率 $\eta=0.5$,训练次数取为 1 000。网络训练成功后,其输入、输出层权矩阵 W_m 和 W_{out} 分别为:

$$W_m=[0.910,-0.833];$$

$$W_{out}=[0.861,-1.932]^T。$$

利用相同的训练样本,使用传统的一阶线性自

$\hat{e}_{33}=0.582$ 与权重(隶属度) 7.427×10^{-5} 。对于 $x=0.405$,分别用 $R_1,R_2,\cdots,R_{31}$ 进行模糊推理,可得估计向量和权重向量。

然后,利用式(20)对 $R_1,R_2,\cdots,R_{31}$ 得到的结果进行组合,得输出值为 0.538,此数据为 1993 年的预测值。用同样方法可以得出 1994~2002 年的预测值,其结果如表 1 所示。

式中: e_i 为 1961~1992 年归一化后的降水量数据。

利用相同的方法得出 1993~2002 年的预测值,其结果如表 2 所示。

将上述 2 种结果进行平均,得出最终的预测值。表 3 为最终的预测值以及各年份实际值的比较。

回归方法进行模型的建立,所得出的线性自回归方程为:

$$y_t=0.482+0.123y_{t-1}。$$

相关系数仅为 0.123,可见线性关系很不明显。利用上述 2 种方法进行预测,其结果如表 4 所示。

以实际数据与预测数据的误差绝对值和作为评价算法优劣的标准。即:

$$Err=\sum_{i=33}^{42}|\hat{e}_i-e_i|。$$

式中: Err 为误差绝对值和, $\hat{e}_i$ 为 1993~2002 年的预测值, e_i 为 1993~2002 年的实际值。

利用信息扩散近似推理、人工神经网络和线性

自回归模型预测的误差绝对值和分别为 1. 673, 1. 809, 1. 804。可见信息扩散近似推理的精度最高。

由于训练样本的非线性特性,使得线性相关系数很小,线性回归预测的结果并不理想。虽然神经网络可以进行非线性预测,但这种信息处理模型很大程度上假定训练网络的学习模式是兼容的,因而,

表 4 基于人工神经网络和线性自回归的 1993~2002 年降水量预测值

Table 4 Forecast value from 1993 to 2002 based on artificial neural network and linear autoregressive					
年份 Year	人工神经网络 Artificial neural network	线性自回归 Linear autoregressive method	年份 Year	人工神经网络 Artificial neural network	线性自回归 Linear autoregressive method
1993	0. 530	0. 532	1998	0. 524	0. 524
1994	0. 556	0. 566	1999	0. 557	0. 569
1995	0. 554	0. 563	2000	0. 531	0. 533
1996	0. 528	0. 529	2001	0. 598	0. 583
1997	0. 533	0. 536	2002	0. 521	0. 52

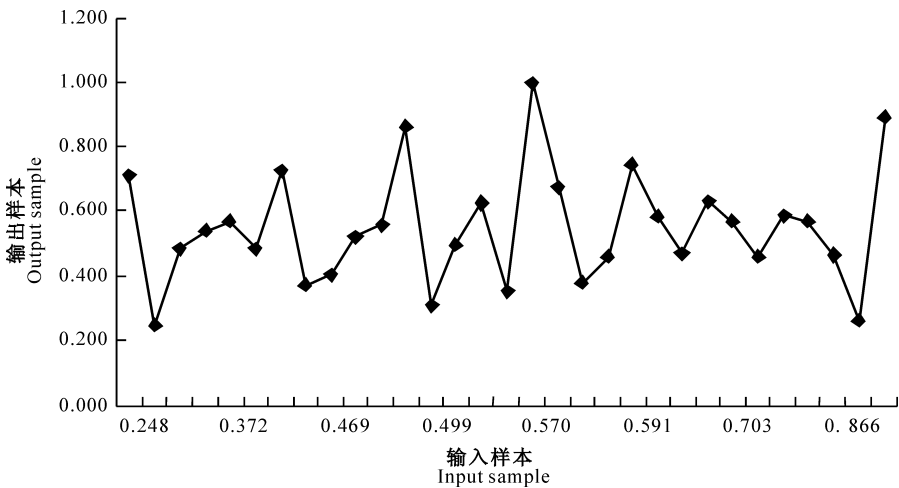


图 1 信息扩散近似推理训练样本的分布图

Fig. 1 Map of training sample used in information-diffusion approximate reasoning

4 结 论

本研究借助于信息扩散技术进行了年降水量的预测。在小样本条件下,信息不够完备,通过信息扩散技术,可以充分利用样本的位置信息,得出较为可靠的预测。使用信息扩散方法可以将样本点转换成模糊集,部分弥补了由于数据的不完备性所造成的信息空白,并可以将矛盾模式转换成兼容模式。通过与人工神经网络模型和传统线性自回归模型比较,发现信息扩散近似推理模型能够在信息不完备的条件下,取得较好的预测结果。

[参考文献]

[1] Sugimoto S, Nakakita E, Ikebuchi S, et al. A stochastic approach to short-term rainfall prediction using a physically

由于矛盾模式之间的冲突,权重修正不能正常进行,最终导致神经网络不能收敛。降水量数据常常出现强烈的矛盾模式,前一年降水大小近似,而后一年的降水可能有较大差别。然而,信息扩散近似推理能够使样本光滑化(如训练样本分布图,图 1),能从根本上解决上述问题。

based conceptual rainfall mode [J]. Journal of Hydrology, 2001;242:137-155.

[2] Wong K W, Wong P M, Gedeon T D, et al. Focus: Rainfall prediction model using soft computing technique [J]. Soft Computing, 2003, 7(6): 434-438.

[3] 陈森法, 张文红, 张建坤, 等. 短期降雨预测的随机微分方程 [J]. 系统工程学报, 2004, 19(3): 239-243.

Chen S F, Zhang W H, Zhang J K, et al. Stochastic differential equations on short-term rainfall prediction [J]. Journal of System Engineering, 2004, 19(3): 239-243. (in Chinese)

[4] 王立坤, 付强, 杨广林, 等. 季节性周期预测法在建立降雨预报模型中的应用 [J]. 东北农业大学学报, 2002, 33(1): 67-71.

Wang L K, Fu Q, Yang G L, et al. Applying the method of season forecast to building up the model of forecasting rainfall [J]. Journal of Northeast Agricultural University, 2002, 33(1): 67-71. (in Chinese)

[5] 贾海峰, 郑耀泉, 丁跃元, 等. 灰色-时序组合预测模型及其在年降水量预测中的应用 [J]. 系统工程理论与实践, 1980(8):

122-126.

Jia H F, Zheng Y Q, Ding Y Y, et al. Grey time series combined forecasting model and its application in annual precipitation [J]. Systems Engineering Theory & Practice, 1980 (8): 122-126. (in Chinese)

[6] 黄国如, 芮孝芳. 流域降雨径流时间序列的混沌识别及其预测研究进展 [J]. 水科学进展, 2004, 15(2): 255-260.

Huang G R, Rui X F. Study advances in diagnosis of chaotic behaviour and its prediction for rainfall and streamflow time series in watershed [J]. Advances In Water Science, 2004, 15 (2): 255-260. (in Chinese)

[7] 黄崇福. 自然灾害风险分析的信息矩阵方法[J]. 自然灾害学报, 2006, 15(1): 1-10.

Huang C F. Information matrix method for risk analysis of natural disaster [J]. Journal of Natrual Disasters, 2006, 15 (1): 1-10. (in Chinese)

[8] 黄崇福. 自然灾害风险分析的基本原理 [J]. 自然灾害学报, 1999, 8(2): 21-29.

Huang C F. Basic principle of risk analysis of natural disasters [J]. Journal of Natrual Disasters, 1999, 8(2): 21-29. (in Chinese)

[9] 黄崇福. 自然灾害风险评价理论与实践[M]. 北京: 科学出版社, 2004: 111-115.

Huang C F. Risk assessment of natural disaster theory & practice [M]. Beijing: Science Press, 2004: 111-115. (in Chinese)

[10] 冯利华. 基于信息扩散理论洪水风险分析[J]. 信息与控制, 2002, 31(2): 164-165.

Feng L H. Flood risk analysis based on theory of news spread [J]. Information and Control, 2002, 31(2): 164-165. (in Chinese)

nese)

[11] 杨 宁, 石 硕, 冯 源, 等. 基于信息扩散近似推理实现配方产品的感官指标相关性分析方法: 中国, 200510042132 [P]. 2006-09-06.

Yang N, Shi S, Feng Y, et al. Sense index correlation analysis method of formulation products based on information-diffusion approximate reasoning: China, 200510042132 [P]. 2006-09-06. (in Chinese)

[12] 段春青. 基于现代智能技术的灌区水资源优化调度研究[D]. 西安: 西安理工大学, 2007.

Duan C Q. Research of water optimal operation in irrigation area base on advanced intelligence technique [D]. Xi'an: Xi'an Technology University, 2007. (in Chinese)

[13] 谢季坚, 刘承平. 模糊数学方法及其应用 [M]. 2 版. 武汉: 华中科技大学出版社, 2000: 37-41.

Xie J J, Liu C P. Fuzzy mathematics and its application [M]. 2nd ed. Wuhan: Huazhong University of Science and Technology Press, 2000: 37-41. (in Chinese)

[14] 邓辉文, 郭美云. 模糊集合隶属函数确定中的哲学思考[J]. 西南师范大学学报: 人文社会科学版, 2002, 28(5): 29-32.

Deng H W, Guo M Y. The philosophical consideration in establishing the membership function of a fuzzy set [J]. Journal of Southwest China Normal University: Philosophy & Social Sciences Edition, 2002, 28(5): 29-32. (in Chinese)

[15] 陈世权. 模糊集合论及其处理的狭义模糊性问题 [J]. 辽宁工程技术大学学报: 自然科学版, 2001, 20(5): 586-589.

Chen S Q. Fuzzy sets theory and its relevant issue dealt with narrow fuzziness [J]. Journal of Liaoning Technical University: Natural Science Edition, 2001, 20(5): 586-589. (in Chinese)

(上接第 212 页)

[13] 赵志刚. 求解双层规划模型的粒子群优化算法 [J]. 系统工程理论与实践, 2007(7): 92-98.

Zhao Z G. Particle swarm optimization for Bi-level programming problem [J]. Systems Engineering-Theory & Practice, 2007(7): 92-98. (in Chinese)

[14] 傅忠云. 粒子群优化算法在计算低压试验电路功率因数的应用 [J]. 福州大学学报: 自然科学版, 2006, 34(1): 76-79.

Fu Z Y. Computation of parameters of lower voltage experimental circuits based on particle swarm optimization algorithm [J]. Journal fo Fuzhou University: Natural Science Edition, 2006, 34(1): 76-79. (in Chinese)

[15] 刘 羿, 陈增强, 袁著祉. 一种改进粒子群优化算法及其在投资规划中的应用 [J]. 数学的实践与认识, 2007, 37(11): 82-87.

Liu Y, Chen Z Q, Yuan Z Z. An improved particle swarm optimization and its application in investment operation [J]. Mathematics in Practice and Theory, 2007, 37(11): 82-87. (in Chinese)