

# T-S-K 模糊逻辑算法在抚河水文预报中的应用

朱冰<sup>1</sup>, 赵兰兰<sup>1</sup>, 李萌<sup>2</sup>

(1.水利部水文局,北京 100053;2.南水北调中线干线工程建设管理局,北京 100053)

**摘要:**目前我国水文机构存有大量的水文数据,如何充分有效地利用智能算法对这些数据进行分析 and 挖掘,降低预测的复杂度并提高水文预测的精度,是水文预报的一个重要的发展方向。数据挖掘中的 T-S-K 模糊逻辑算法从大量数据中提取未知、可操作、有相关性的水文模式和关系用于水文预报。本文以抚河上下游 3 个重要水文站历史数据为实例,建立 T-S-K 模糊逻辑算法水文预报,分析结果表明该算法所建模型误差较小,预测精度高,可满足实际水文预报需求。

**关键词:**T-S-K;数据挖掘;洪水预报

中图分类号:P338

文献标识码:A

文章编号:1000-0852(2015)03-0053-06

## 1 引言

随着计算机和自动化数据采集工具的广泛应用,人类在领域里所累积的数据正以指数方式增长,如何对这些海量数据进行实时的和深层次的分析,以揭示隐藏在这些数据背后更有用的信息,更好的利用这些数据变得非常重要。目前的数据库系统可以高效地实现数据的录入、查询、统计等功能,但无法发现数据中隐藏的关系和规则,无法根据现有的数据来预测未来的发展趋势,于是数据挖掘(Data Mining)技术应运而生。数据挖掘简单来说就是运用基于计算机的方法,从数据中获得有用知识的整个过程<sup>[1]</sup>。它不仅是面对特定数据的简单检索查询调用,而且要对这些数据进行微观、中观乃至宏观的统计、分析、综合和推理,以指导实际问题的求解,企图发现事件间的相互关联,甚至利用已有的数据对未来的活动进行预测<sup>[2]</sup>。

水文预报是据已知的信息对未来一定时期内的水文状态作出定性或定量的预测<sup>[3]</sup>。而目前占据主流的物理模型,是以物理机理为基础,即计算降雨量、土壤含水量、上下游流量水位情况等多种物理量值,通过模拟产汇流机制来预报水情。这种物理模型一般要有明确的物理关系,目前应用广泛但仍然存在一定问题:物理模型较复杂,很难收集到建立模型所需的所有按时间空间分布的细节数据,例如地形数据、河道特性数

据、流域土壤特性分布和降雨径流数据,而不同流域特性千差万别,所用水文模型一般很难将各种复杂情况都考虑充分。数据挖掘技术在水文领域的应用,是从大量的、不完全的、有噪声的、模糊的、随机的水文及其相关数据中,提取隐含其中有用的水文信息和知识的过程<sup>[4]</sup>。数据挖掘从历史水文入手,利用各种智能算法对历史数据从各个角度进行分析,通过仿真能力对水文现象成因进行模拟,在一定程度上进行模型自我调节,从而提高水文预报的准确度。目前,神经网络、决策树、模糊理论(Fuzzy Logic System)等数据挖掘技术在水文预报方面都有着不同程度的应用,本文以抚河上下游 3 个重要水文站历史数据为实例,采用数据挖掘中 T-S-K (Takagi-Sugeno-Kang) 模糊逻辑算法,建立上下游流量关系,从而进行水文预报。

## 2 T-S-K 模糊逻辑算法

### 2.1 模糊逻辑及 T-S-K 模糊逻辑算法的基本概念

1965 年美国数学家 L. Zadeh 首先提出了模糊集合的概念,标志着模糊数学的诞生。经典逻辑坚持所有事物(陈述)都可以用二元项(0 或 1,黑或白,是或否)来表达,而模糊逻辑用真实度替代了真值。模糊逻辑建立在二值逻辑基础上的原有的逻辑与数学难以描述和处理现实世界中许多模糊性的对象。模糊数学与模糊逻辑实质上是要对模糊性对象进行精确的描

收稿日期:2014-07-05

作者简介:朱冰(1986-),女,浙江绍兴人,工程师,从事水文情报预报工作。E-mail:zhubing@mwr.gov.cn

述和处理。L. Zadeh 为了建立模糊性对象的数学模型,把只取 0 和 1 二值的普通集合概念推广为在[0, 1]区间上取无穷多值的模糊集合概念,并用“隶属度”这一概念来精确地刻画元素与模糊集合之间的关系。正因为模糊集合是以连续的无穷多值为依据的,所以,模糊逻辑可看做是运用无穷连续值的模糊集合去研究模糊性对象的科学<sup>[5]</sup>。模糊逻辑在水文预报的优点是能更好地描述水文物理量及产生过程的各个方面的特征。目前模糊逻辑应用最为广泛的推理方法是 Mamdani 模糊算法<sup>[6]</sup>和 T-S-K 模糊逻辑算法<sup>[7]</sup>,前者主要分为 3 个阶段:模糊化、模糊逻辑推理、反模糊化,即通过 Mamdani 模糊规则得到的是模糊集,必须通过反模糊化,从而得到一个数值型的清晰结果。T-S-K 模糊逻辑算法和前者的不同在于他将第二和第三阶段合二为一,即在模糊规则的输出部分是数值型的清晰结果。由于河道洪水从成因上分析是由于上游各支流洪水传播到出口河段加上区间降雨而组合得到的,和 T-S-K 模糊逻辑算法的结构更为类似<sup>[8]</sup>。

### 2.2 T-S-K 模糊逻辑算法基本步骤

T-S-K 模糊逻辑算法的洪水预报基本步骤如下:

(1)数据预处理,将历史数据的河道历史流量或水位数据分析并处理成符合预报的数据集,再通过隶属函数将清晰数据转换为模糊隶属度;

(2)建立模糊规则集及模型参数计算,基于历史数据及规则分析、优化建立规则集,并通过模糊隶属度、最小二乘法进行参数率定;

(3)生成预报结果,利用确定了的模糊规则集和率定出的参数,根据上游河道流量或水位数据得出下游河道流量或水位数据,最终生成预报结果。

### 2.3 T-S-K 模糊逻辑算法的计算方法

模糊规则库是由一组基于常识的规则组成的。规则格式是 IF ... THEN, 包括输入和输出之间的所有可能的模糊关系。多输入-单输出的 T-S-K 模糊逻辑算法的规则组成如下:

Rr: IF ( $x_1$ =模糊子集 1)AND/OR ( $x_2$ =模糊子集 2) AND/OR ... THEN

$$output=f(x_1, x_2, \dots, x_n)$$

规则之间权重计算一般通过 AND 或 OR 逻辑关系来整合每条规则的隶属度,采用 AND 逻辑关系,可以选取 minimum 或 product 算法;采用 OR 逻辑关系,可以选取 maximum 或 probability 算法<sup>[9]</sup>。

例如双输入-单输出,输入和输出之间关系线性

的 T-S-K 模糊逻辑算法计算如下:

(1)建立规则集。

$$R1: \text{if } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } z_1=p_1x+q_1y+r_1$$

Also

$$R2: \text{if } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ then } z_2=p_2x+q_2y+r_2$$

$A_1, A_2, B_1, B_2$  是两个输入变量的模糊子集。

(2)计算权重。利用隶属关系得出两个规则对应的隶属度分别是  $\omega(x), \omega(y)$ , 利用逻辑关系得出权重, 计算如下, 算法计算流程见图 1:

$$\omega_1=(\text{Min or Product})(\omega_1(x), \omega_1(y))$$

$$\omega_2=(\text{Min or Product})(\omega_2(x), \omega_2(y))$$

(3)采用重心法得出最终结果。

$$z = \frac{\omega_1 z_1 + \omega_2 z_2}{\omega_1 + \omega_2}$$

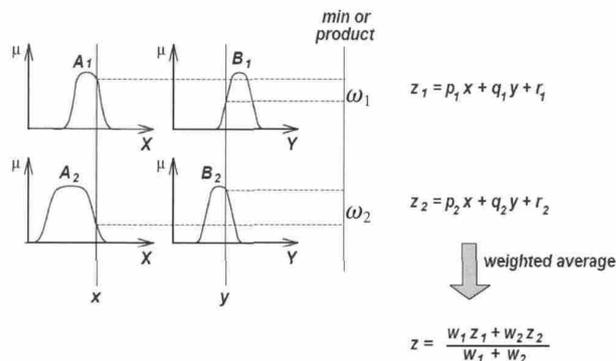


图 1 双输入单输出双规则 T-S-K 模糊逻辑算法计算流程

Fig.1 The calculation process of T-S-K fuzzy logic algorithm with double input and single output rules

如果规则集总有  $n$  条规则,则最后的清晰输出为:

$$z_0 = \frac{\sum_{i=1}^n \omega_i * z_i}{\sum_{i=1}^n \omega_i}$$

式中:  $\omega_i$  是每条规则对应的隶属度;  $z_i$  是每条规则得出的输出;  $z_0$  是最终预报结果。

## 3 在抚河流域的应用

### 3.1 抚河流域简介

抚河是鄱阳湖水系五大河流之一,发源于广昌、石城、宁都三县交界处的灵华峰东侧里木庄,干流自南向北流经抚州市,在进贤县汇入鄱阳湖,流域总面积  $1.68 \times 10^4 \text{ km}^2$ 。下面将以抚河流域的李家渡、娄家村、廖家湾三个重点水文站作为对象(如图 2),研究基于 T-S-K 模糊逻辑算法的河道洪水预报。

下游李家渡水文站的水是由上游娄家村和廖家



图2 抚河流域河系图

Fig.2 The Fuhe river system

湾两个水文站的洪水传播产生，由于3站之间流域面积较小，区间降雨对下游李家渡水文站洪水贡献不大，因此忽略此区间的降雨对李家渡水文站的影响。选取1960~1975年共16a的3站日流量数据为率定数据建立预报模型，使用1977~1979年共3a的3站日流量数据对模型进行检验。娄家村和廖家湾水文站到李家渡水文站的洪水传播时间均小于24h，以下计算将利用上游两站日流量预报24h后即第2天的下游李家渡水文站日流量。

### 3.2 数据预处理

将两个上游水文站的日流量的率定数据分别进行聚类分析，利用聚类软件STATA中k-means算法将两站流量聚类为超低(VL)、低(L)、中(M)、高(H)、超高(VH)流量5类。聚类中心如表1。

表1 两站流量的聚类中心

Table1 The clustering center values of the two inputs

聚类中心	娄家村 / m <sup>3</sup> ·s <sup>-1</sup>	廖家湾 / m <sup>3</sup> ·s <sup>-1</sup>
超低(VL)	55	95
低(L)	244	317
中(M)	578	692
高(H)	1187	1377
超高(VH)	2381	2740

以得到的聚类中心为参考，分别将两站流量划分超低、低、中、高、超高五个模糊子集，其流量隶属函数将采用三角以及梯形函数，隶属函数及函数图如图3、图4所示。

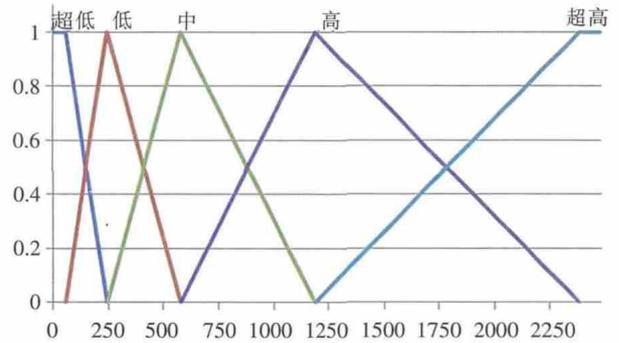


图3 娄家村隶属度函数图

Fig.3 The membership degree function for the Loujiachun station

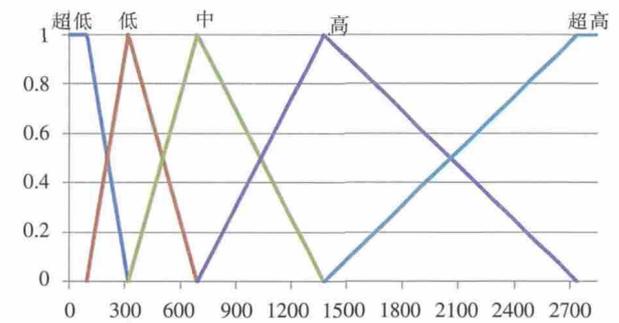


图4 廖家湾隶属度函数图

Fig.4 The membership function for the Liaojiawan station

### 3.3 建立规则集及参数计算

#### 3.3.1 建立模糊规则集

本次预报由2个输入，5种模糊状态组成，模糊规则集共有52条模糊规则，并且考虑下游李家渡的流量是由上游两个水文站流量叠加而成的，所有假定输出和输入之间是线性关系。具体规则集如下：

$R_1$ : If ( $q_1$  is very low, and  $q_2$  is very low)

Then  $Q_1 = a_1 q_1 + b_1 q_2 + c_1$

$R_2$ : If ( $q_1$  is very low, and  $q_2$  is low)

Then  $Q_2 = a_2 q_1 + b_2 q_2 + c_2$

.....

$R_5$ : If ( $q_1$  is very low, and  $q_2$  is very high)

Then  $Q_5 = a_5 q_1 + b_5 q_2 + c_5$

$R_6$ : If ( $q_1$  is low, and  $q_2$  is very low)

Then  $Q_6 = a_6 q_1 + b_6 q_2 + c_6$

.....

$R_{20}$ : If ( $q_1$  is high, and  $q_2$  is very high)

Then  $Q_{20} = a_{20} q_1 + b_{20} q_2 + c_{20}$

$R_{21}$ : If ( $q_1$  is very high, and  $q_2$  is very low)

Then  $Q_{21} = a_{21} q_1 + b_{21} q_2 + c_{21}$

.....

$R_{24}$ : If ( $q_1$  is very high, and  $q_2$  is high)

Then  $Q_{24} = a_{24}q_1 + b_{24}q_2 + c_{24}$

$R_{25}$ : If ( $q_1$  is very high, and  $q_2$  is very high)

Then  $Q_{25} = a_{25}q_1 + b_{25}q_2 + c_{25}$

式中: $q_1$ 和 $q_2$ 是上游两个水文站流量值; $Q_i$ 是 $i$ 规则中下游李家渡水文站24h后的流量值,参数 $a_i$ 和 $b_i$ 是两个输入的权重;参数 $c_i$ 是误差修正系数。

经过分析的16a率定数据,由于上游两个水文站距离较近,发现在实际情况中,部分规则发生的概率极小甚至不存在。经过规则优化,将概率为0的规则删除,将概率较小的规则相整合后,模糊规则集中规则数量从25条优化至19条,参数从75个优化到57个。优化后规则如下:

$R_1$ : If ( $q_1$  is very low, and  $q_2$  is very low)

Then  $Q_1 = a_1q_1 + b_1q_2 + c_1$

$R_2$ : If ( $q_1$  is very low, and  $q_2$  is low)

Then  $Q_2 = a_2q_1 + b_2q_2 + c_2$

$R_3$ : If ( $q_1$  is very low, and  $q_2$  is middle or high)

Then  $Q_3 = a_3q_1 + b_3q_2 + c_3$

.....

$R_7$ : If ( $q_1$  is low, and  $q_2$  is high or very high)

Then  $Q_7 = a_7q_1 + b_7q_2 + c_7$

.....

$R_{13}$ : If ( $q_1$  is high, and  $q_2$  is very low or low)

Then  $Q_{13} = a_{13}q_1 + b_{13}q_2 + c_{13}$

.....

$R_{17}$ : If ( $q_1$  is very high, and  $q_2$  is low or middle)

Then  $Q_{17} = a_{17}q_1 + b_{17}q_2 + c_{17}$

$R_{18}$ : If ( $q_1$  is very high, and  $q_2$  is high)

Then  $Q_{18} = a_{18}q_1 + b_{18}q_2 + c_{18}$

$R_{19}$ : If ( $q_1$  is very high, and  $q_2$  is very high)

Then  $Q_{19} = a_{19}q_1 + b_{19}q_2 + c_{19}$

### 3.3.2 计算参数

T-S-K模糊逻辑算法最终预报结果公式可表示为:

$$Q_{com} = \frac{\omega_1}{\sum \omega} (a_1q_1 + b_1q_2 + c_1) + \frac{\omega_2}{\sum \omega} (a_2q_1 + b_2q_2 + c_2) \dots \dots + \frac{\omega_{19}}{\sum \omega} (a_{19}q_1 + b_{19}q_2 + c_{19})$$

式中: $\omega$ 是隶属度; $a, b, c$ 是参数。

当规则中只采用AND逻辑关系时,计算隶属度可以采用minimum或product算法:

$$\omega_k = \text{MIN}\{\omega_k(q_1), \omega_k(q_2)\}$$

或

$$\omega_k = \omega_k(q_1) \cdot \omega_k(q_2)$$

当规则中采用AND和OR逻辑关系时,计算隶属度可以采用minimum和maximum或product和probability算法:

$$\omega_k = \text{MIN}\{\omega_k(q_1), \text{MAX}\{\omega_{k1}(q_2), \omega_{k2}(q_2)\}\}$$

或

$$\omega_k = \omega_k(q_1) * [\omega_{k1}(q_2) + \omega_{k2}(q_2) - \omega_{k1}(q_2) * \omega_{k2}(q_2)]$$

式中: $\omega_k(q_2), \omega_k(q_1)$ 是两个输入对应规则 $k$ 的隶属度, $\omega_{k1}(q_2), \omega_{k2}(q_2)$ 是输入2在规则 $k$ 中对应OR逻辑关系两侧的模糊状态的隶属度。

最终计算57个参数,采用最小二乘法进行计算,公式如下

$$\text{Min} \sum_{i=1}^{19} [\sum_{t=1}^N (Q^{t+1} - a_i q'_1 - b_i q'_2 - c_i)^2]$$

式中: $N$ 是历史记录的数据; $q'_1, q'_2$ 表示输入1、输入2在 $t$ 时刻的流量值; $Q^{t+1}$ 表示输出在 $t+1$ 时刻的流量值。

当采用minimum和maximum计算隶属度时,参数如表2:

表2 采用minimum和maximum得出的模糊参数

Table2 The parameters of the minimum and maximum operator

规则号	参数		
	$a$	$b$	$c$
1	1.88	0.36	-21.74
2	-0.38	0.40	85.66
3	1.95	1.61	-135.3
		.....	
18	4.37	1.83	-9362.4
19	1.01	0.22	2163.27

模糊集如下:

$R_1$ : If ( $q_1$  is very low, and  $q_2$  is very low)

Then  $Q_1 = 1.88 * q_1 + 0.36 * q_2 - 21.74$

$R_2$ : If ( $q_1$  is very low, and  $q_2$  is low)

Then  $Q_2 = -0.38 * q_1 + 0.40 * q_2 + 85.66$

$R_3$ : If ( $q_1$  is very low, and  $q_2$  is middle or high)

Then  $Q_3 = 1.95 * q_1 + 1.61 * q_2 - 135.3$

.....

$R_{18}$ : If ( $q_1$  is very high, and  $q_2$  is high)

Then  $Q_{18} = 4.37 * q_1 + 1.83 * q_2 - 9362.4$

$R_{19}$ : If ( $q_1$  is very high, and  $q_2$  is very high)

Then  $Q_{19} = 1.01 * q_1 + 0.22 * q_2 + 2163.27$

当采用product和probability计算隶属度时,参数

如表 3:

表3 采用product和probability得出的模糊参数

Table3 The parameters of the product and probability operator

规则号	参数		
	a	b	c
1	1.79	0.40	-20.70
2	3.75	0.45	-103.41
3	28.1	2.5	-2161
	.....		
18	4.62	0.09	-7051.29
19	1.23	-0.29	3360.29

模糊集如下:

$R_1$ : If ( $q_1$  is very low, and  $q_2$  is very low)

Then  $Q_1=1.79*q_1+0.40*q_2-20.70$

$R_2$ : If ( $q_1$  is very low, and  $q_2$  is low)

Then  $Q_2=3.75*q_1+0.45*q_2-103.41$

$R_3$ : If ( $q_1$  is very low, and  $q_2$  is middle or high)

Then  $Q_3=28.1*q_1+2.5*q_2-2161$

.....

$R_{18}$ : If ( $q_1$  is very high, and  $q_2$  is high)

Then  $Q_{18}=4.62*q_1+0.09*q_2-7051.29$

$R_{19}$ : If ( $q_1$  very high, and  $q_2$  is very high)

Then  $Q_{19}=1.23*q_1-0.29*q_2+3360.29$

### 3.4 预报结果验证

使用 1977~1979 年共 3a 的日流量数据组成的验证资料带入上述规则集中,可得到预报结果。预报结果和实测流量过程对比见图 5, 采用两种不同算法的散点图见图 6。

由于水文预报所选的性能评价指标需具有专业性、综合性、普适性、实用性。本文将选用以下平均绝对误差(MAE)、均方根误差(RMSE)、相对均方根误差(RRMSE)、效率系数(EF)、确定性系数(CD)五种评定指标分别对率定数据和验证数据进行综合评定,具体数值如表 4。

从预报结果和评价指标来看,基于 T-S-K 模糊逻辑算法的抚河流域洪水预报的率定期和验证期的确

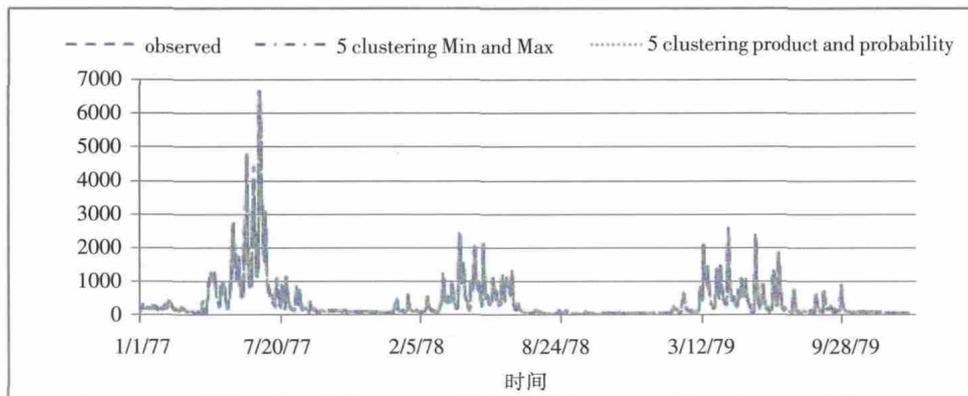


图 5 1977~1999 年预报结果和实测流量过程对比图

Fig.5 The time series plots of the observed and forecasting discharge during 1977-1979

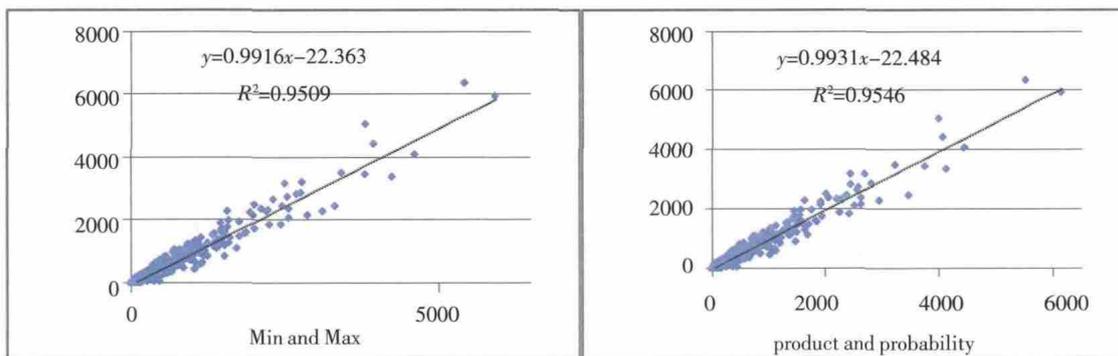


图 6 预报结果散点图

Fig.6 The scatter plots of the observed and calculated discharge during 1977-1979

表4 预报结果评价指标表  
Table 4 The performance indicators

	Min and Max		Product and probability		最优结果
	率定数据	验证数据	率定数据	验证数据	
MAE	75.113	74.579	74.961	73.664	0
RMSE	148.517	133.085	148.028	128.143	0
RRMSE	0.375	0.413	0.374	0.398	0
EF	0.949	0.949	0.949	0.953	1.0
CD	1.054	1.032	1.053	1.031	1.0

定性系数在 0.94 以上,而算法对抚河流域的水文响应也比较及时,算法模拟预报的洪水流量起涨点与实测洪水流量起涨点较吻合,洪峰出现时间、量级也基本一致,满足水文预报的实际需求。

#### 4 结论

本次研究选取抚河流域上下游三个重点水文站 19a 历史资料,通过设定上下游流量线性关系,选取 k-means 算法进行数据聚类,选用三角及梯形方程建立隶属函数,建立 T-S-K 模糊逻辑规则集,采用最小二乘法率定参数,最终建立 T-S-K 模糊逻辑算法在抚河流域的河道水文预报模型。研究表明,T-S-K 模糊逻辑算法水文预报建立的河道水文预报模型,误差较小,预测精度高,对计算机性能要求小,可满足实际水文预报需求。但由于 T-S-K 模糊逻辑算法对历史资料的数据量有较强的依赖性,本算法可运用于历史数据齐全、但计算机资源较匮乏的地区。

#### 参考文献:

- [1] Kantardzic Mehmed. Data Mining Concepts, Models, Methods, and Algorithms (Second Edition) [M]. WILEY, A JOHN WILEY & SONS, INC., 2011:150-240.
- [2] 包为民. 普通高等教育“十五”国家级规划教材--水文预报(第四版)[M].北京:中国水利水电出版社,2009. (BAO Weiming. Hydrological Forecast (4th edition) [M]. Beijing: China WaterPower Press, 2009. (in Chinese))
- [3] 艾萍,倪伟新. 我国水文数据挖掘技术研究的回顾与展望[J]. 计算机工程与应用, 2003,3(28):13-17. (AI Ping, NI Weixin. Review and preview of the research on hydrological data mining technology in China [J]. Computer Engineering and Applications, 2003,3(28):13-17. (in Chinese))
- [4] 赵德齐. 模糊数学 [M]. 北京:中央民族大学出版社,1995:5-12. (ZHAO Deqi. Fuzzy Mathematics [M]. Beijing: Minzu University of China Press, 1995:5-12. (in Chinese))
- [5] Mamdani E. H., Assilian S.. An experiment in linguistic synthesis with a fuzzy logic controller [J]. International Journal of Man Machine Studies, 1975,7(1):1-13.
- [6] Takagi T, Sugeno M.. Fuzzy identification of systems and its application to modeling and control [J]. IEEE Transactions on Systems, Man and Cybernetics,1985,15(1):116-132.
- [7] 张弛. 数据挖掘技术在水文预报中的应用及水文预报发展趋势研究[J]. 水文, 2007,27(2):74-77. (ZHANG Chi. Application of data mining technology in hydrological forecasting and research on development trend of hydrological forecasting[J]. Journal of China Hydrology, 2007,27(2):74-77. (in Chinese))
- [8] Jang J.-S. Rogerm. Fuzzy Inference Systems Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence [M]. Prentice-Hall, Inc., 1997:73-91.

### Hydrological Forecasting based on T-S-K Fuzzy Logic System in Fuhe River Basin

ZHU Bing<sup>1</sup>, ZHAO Lanlan<sup>1</sup>, LI Meng<sup>2</sup>

(1. Bureau of Hydrology, Ministry of Water Resources of China, Beijing 100053, China;

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

**Abstract:** A large amount of hydrological data stores in hydrological institutions in our country. In order to improve the accuracy of hydrological forecasts and to lower the complexity of the forecast, a important research direction is how to use intelligent algorithm to analyze the data. T-S-K fuzzy logic system is one of the opinions. The Takagi-Sugeno-Kang (T-S-K) fuzzy system is adopted in this study to establish the relationship between the observed downstream discharges and the observed upstream discharges in the Fu River basin. The result shows that the quality of the modle built by T-S-K fuzzy logic system is good, the forecast accuracy is high, the forecast model quite meet the practical application requirements.

**Key words:** T-S-K; data mining; flood forecast