

基于联系隶属度的城市需水量预测模型

汪明武, 蒋辉, 张立彪, 赵奎元

(合肥工业大学 土木与水利工程学院, 安徽 合肥 230009)

摘要: 城市需水量预测是生态城市规划与管理的基础,但受诸多不确定因素影响,是一个复杂的预测难题。为能定量统一表达预测年份需水量各影响因素间及与历史数据间的交叉、交融的确定和不确定关系,在此应用有序聚类理论与集对分析的耦合方法,提出了基于联系隶属度概念的城市需水量预测模型。该模型首先基于城市需水量历史数据进行最优分割聚类,应用联系隶属度对预测年份需水量的影响因子与历史数据关系进行同异反分析,并构建相似模型预测相应年份的城市需水量。实例应用及与其他方法对比的结果表明,该模型应用于城市需水预测是有效可行的。

关键词: 需水量; Fisher 最优分割法; 联系隶属度; 集对分析; 需水量预测

中图分类号: TU991.31

文献标识码: A

文章编号: 1672-643X(2015)01-0012-04

Forecast model of urban water demand based on connectional membership degree

WANG Mingwu, JIANG Hui, ZHANG Libiao, ZHAO Kuiyuan

(School of Civil and Hydraulic Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: The forecast of urban water demand is the basis of plan and management of ecological city. However, it is a complex problem affected by various uncertainty factors. In order to express the integration and blending relationships of certainty and uncertainty among influence factors and between samples and historical data as a whole, the paper used an orderly clustering theory and coupling method of set pair analysis to present a forecast model of urban water demand based on connectional membership degree. In the model, the ordered historical data of urban water demand of corresponding years was first clustered according to Fisher's optimal partition method, then the identical - discrepancy - contrary analysis was carried out to express the relationship between impact factors of one year and historical data. Moreover, it structured a similar model to forecast city water demand of corresponding year. The results from the practical example and comparative analyses with other methods show it is reasonable and effective for application of the model to prediction of city water demand.

Key words: water demand; fisher optimal partition method; connectional membership degree; set pair analysis; forecast of urban water demand

1 研究背景

随着我国经济的高速发展,城市人口的持续增长和生活水平的不断提高,城市用水需求量也随之大幅度增长,导致城市水资源供需矛盾日益加剧,城市缺水问题已成为城市化建设面临的难题,故合理准确预测城市需水量是水资源战略制定的关键基础

资料,且直接影响到给水系统调度决策的可靠性和实用性以及城市水资源的可持续利用和社会经济的可持续发展^[1]。然而,城市需水量预测受到城市规模和社会经济发展水平等诸多不确定因素的影响与控制,是一个复杂的系统预测难题^[2]。人们为提高需水量预测的准确性,提出了基于回归分析^[3]、投影寻踪^[4]、神经网络^[5]、灰色系统^[6]等的预测方法,

收稿日期:2014-09-16; 修回日期:2014-11-09

基金项目:国家自然科学基金项目(41172274)

作者简介:汪明武(1972-),男,安徽歙县人,博士,教授,博士生导师,主要从事智能与不确定性分析理论与应用方面的教学与科研工作。

取得了有益的进展,但这些预测方法不能统一考虑需水量影响因素间交叉、交融的动态确定和不确定关系,常导致预测结果同实际情况存在一定差异。而近期发展起来的集对分析耦合方法则可很好地解决此难题,也为城市需水量预测提供了新思路。本文采用最优分割聚类算法和集对分析耦合理论,探讨了可充分利用历史数据和考虑影响因素间的不确定关系及转化态势的城市需水量预测新模型,以期提高城市需水量预测的精度和可靠性。

2 基于联系隶属度的预测模型

2.1 基本原理

基于最优分割算法的联系隶属度预测模型基本原理为:首先对预测对象的历史数据进行排序和最优分割聚类,求得相应的聚类区间;然后应用联系隶属度的概念构造待预测对象各影响因子与历史数据聚类类别间的同异反关系,进而构建基于预测样本与历史分类间关系的预测模型,结合权重计算相应的综合联系隶属度,以预测研究对象的演化趋势。相应的预测流程如图 1 所示。

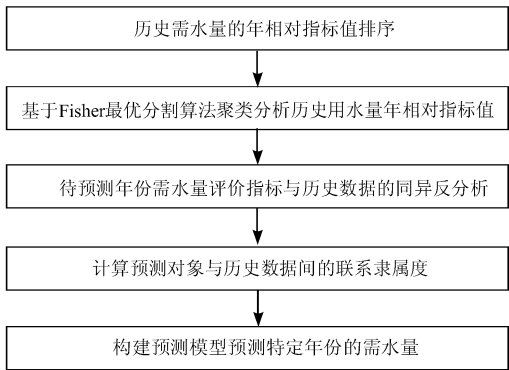


图 1 基于联系隶属度的预测流程图

2.2 基于 Fisher 最优分割的聚类分析

Fisher 最优分割法为有序样本分割聚类方法,它通过离差平方和(类直径)表示样本间的差异程度对数据进行处理分析和确定最优分类数^[7],具有多指标聚类、客观和最优的特点。若有序样本总数为 n ,样本集可分割成 m 个群 $G(m, n) = \{x_m, x_{m+1}, \dots, x_n\} (m < n)$, 则 Fisher 最优分割算法的递推计算模型可表达为:

$$L[P(n, 2)] = \min_{2 \leq m \leq n} \{D(1, m-1) + D(m, n)\} \quad (1)$$

$$L[P(n, k)] = \min_{k \leq m \leq n} \{D(m-1, k-1) + D(m, n)\} \quad (2)$$

$$D(m, n) = \sum_{t=m}^n (x_t - \bar{x})^2 \quad (3)$$

式中: $L[p(n, k)]$ 为损失函数; $D(m, n)$ 为群 $G(m, n)$ 的直径; x_t 和 $\bar{x}$ 分别为群 $G(m, n)$ 中指标实测值和均值。可见, Fisher 最优分割法先计算同类样本间的类直径,再计算损失函数,且获得的类直径越小,则同类样本之间的差异越小,各类别样本之间的差异就越大^[8]。一般当分类数固定时,应用 Fisher 最优分割时,可选择使误差函数达到最小值对应的分类情况作为最优分割。

2.3 基于联系隶属度的同异反分析

由最优分割算法可知,对样本与类间的聚类分析为确定性归类,而对于确定和不确定的系统定量描述则显得困难,且无法体现同类内的差异性和向相邻类别的转移特性。而近期发展起来的集对分析理论则能从同一性、差异性和对立性 3 个方面刻画事物,全面反映样本的动态演化趋势,从而可弥补聚类分析算法在分析不确定系统上的不足。在此引入集对分析耦合理论,从同异反 3 个方面分析样本和历史数据间的相似性关系。对于最优分割聚类划分获得的 N 个实数区间,由集对分析理论可知,待评样本与群 k 相应属性值间的同异反关系可描述见图 2。

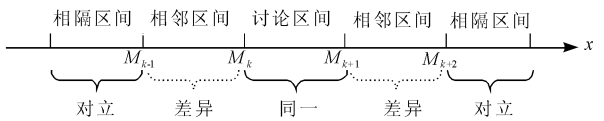


图 2 点与区间构成的集对同异反关系示意图

对于某个属性,当样本 x 处于讨论区间 k 内,则两者的集对关系为同一,相应的联系度 $\mu \in [0, 1]$; 当 x 处于讨论区间的相隔区间时,则两者的集对关系为对立,联系度 μ 取 -1 ; 当 x 处于讨论区间的相邻区间时,则研究对象与讨论区间的关系为差异,联系度 $\mu \in [-1, 0]$ 。然而,传统集对联系度的取值范围为 $[0, 1]$ ^[9-10], 不能满足类间转化描述要求。为满足质与量互变定理和对立统一原理,在此应用联系隶属度(connectional membership degree)来表达此类问题构成的集对关系^[11]。

$$\mu_{mn}^k = \begin{cases} - \left| \frac{x - M_k}{M_{k-1} - M_k} \right| & M_{k-1} \leq x \leq M_k \\ \frac{2(x - M_k)}{M_{k+1} - M_k} & M_k \leq x \leq \frac{M_k + M_{k+1}}{2} \\ \frac{2(M_{k+1} - x)}{M_{k+1} - M_k} & \frac{M_k + M_{k+1}}{2} \leq x \leq M_{k+1} \\ - \left| \frac{x - M_{k+1}}{M_{k+1} - M_{k+2}} \right| & M_{k+1} \leq x \leq M_{k+2} \\ -1 & \text{其他} \end{cases} \quad (4)$$

式中: μ_{mn}^k 为联系隶属度; $\mu_{mn}^k \in [-1, 1]$; M_{k-1} 、 M_k 、 M_{k+1} 和 M_{k+2} 为聚类区间的界限值。

2.4 预测模型

若基于城市需水量历史数据最优分割聚类可获得 $k(k = 1, 2, \dots, K)$ 类, 则待预测样本与历史城市需水量 k 类间的同异反关系可用集成联系隶属度 μ_m^k 来定量描述。

$$\mu_m^k = \sum_{n=1}^N \mu_{mn}^k q_n \tag{5}$$

式中: q_n 为城市需水量影响因素的权重。相应的, 对于某年的城市需水量预测模型可构建为:

$$f = \frac{\sum_{k=1}^K \bar{f}_k \mu_m^k}{\sum_{k=1}^K \mu_m^k} \tag{6}$$

式中: $\bar{f}_k$ 为历史第 k 类用水量中心值; f 为相应的预测值。

3 应用实例

为验证讨论模型的正确性和有效性, 现以上海市 1980 - 2011 年的年用水量及相关数据为例进行分析和验证^[12]。城市需水量受社会和经济等不确定性因素影响与控制。基于已有文献报导和以往研究经验, 并在选择指标过程中考虑了变量间相关性对预测精度的影响, 选择了易于获得和可靠性高, 且与需水量密切相关的户籍人口、服务业产值和轻工业产值指标作为评价指标, 因研究表明户籍人口、服务业产值和轻工业产值与城市需水量的相关分析系数大于 0.8^[2], 详细数据见表 1。以原始数据中相邻年份总用水量的比值作为表征用水量的年动态相对指标, 则相应的年动态指标值见表 2 所示。按式(1) ~ (3) 可求得不同分类下的损失函数如图 3 所示, 可见当 $k = 5$ 时曲线出现拐点, 表明实例历史需水量可最优聚类为 5 类, 相应的影响指标动态范围见表 3。

表 1 上海市 1980 - 2011 年用水量及影响因素

年份	总用水量/亿 m ³	户籍人口/万人	服务业产值/亿元	轻工业产值/亿元	年份	总用水量/亿 m ³	户籍人口/万人	服务业产值/亿元	轻工业产值/亿元
1980	9.73	1146.52	65.69	331.13	1996	23.11	1304.43	1292.11	2334.29
1981	10.17	1162.84	69.84	360.12	1997	23.32	1305.46	1592.74	2528.19
1982	10.09	1180.51	74.44	359.62	1998	24.12	1306.58	1855.36	2527.56
1983	10.21	1194.01	82.97	363.92	1999	23.51	1313.12	2129.60	2679.71
1984	10.64	1204.78	98.22	395.60	2000	24.00	1321.63	2486.86	2880.74
1985	11.54	1216.69	121.59	456.59	2001	24.15	1327.14	2728.94	2929.50
1986	11.71	1232.33	135.12	493.09	2002	24.40	1334.23	3038.90	3077.11
1987	12.71	1249.51	159.48	556.96	2003	25.82	1341.77	3404.19	3416.80
1988	12.61	1262.42	187.89	679.17	2004	27.35	1352.39	4097.26	3871.33
1989	12.92	1276.45	200.73	789.57	2005	28.65	1360.26	4776.20	4299.31
1990	13.32	1283.35	241.82	846.63	2006	29.19	1368.08	5508.48	4747.28
1991	13.53	1287.20	309.07	976.34	2007	30.34	1378.86	6821.11	5318.85
1992	14.1	1289.37	402.77	1132.75	2008	30.90	1391.04	7872.23	5839.41
1993	14.94	1294.74	579.03	1401.33	2009	30.47	1400.70	8930.85	5663.34
1994	16.27	1298.81	794.80	1890.12	2010	30.90	1412.32	9833.51	6692.35
1995	22.07	1301.37	1020.20	2092.89	2011	31.13	1419.32	11142.86	7420.92

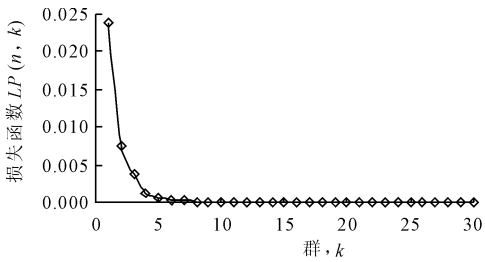


图 3 不同分割方案的损失函数

现以 2011 年的户籍人口、服务业产值和轻工业产值的年动态相对指标值 {1.005, 1.1332, 1.1089} 作为待测样本的观测值来预测该年的总用水量年动态相对指标值。由 2011 年各影响指标实测值代入式(4), 可求得该年各影响因子与历史数据聚类区间的联系隶属度, 其中由于第 5 类样本数只有一个, 故依照前 4 类区间宽度予以适当放大。评价指标权重的确定对评价结果有着重要影响, 但在此为便于

计算,以验证模型的可靠性和适用性,故假定 3 个指标对城市需水量的影响为均等,即相应权重取 1/3,则集成联系隶属度可由式(5)求得,并代入式(6)即可求得 2011 年的总用水量相对 2010 年的增长动态指标为 1.0053,计算结果见表 4。可见,预测结果较

实际值 1.0074 的相对误差为 -0.2156%,与其他预测方法对比结果也表明本文模型的预测精度相对较高(表 5)。另一方面,指标权重的合理确定对预测结果是有影响的,结合指标的实测值和重要性确定各权重尚需深入探讨。

表 2 排序后的用水量及影响因素动态相对指标值(1980-2010 年)

年份	总用水量	户籍人口	服务业产值	轻工业产值	年份	总用水量	户籍人口	服务业产值	轻工业产值
1999	0.9747	1.0050	1.1478	1.0602	1990	1.0310	1.0054	1.2047	1.0723
2009	0.9861	1.0069	1.1345	0.9698	1998	1.0343	1.0009	1.1649	0.9998
1988	0.9921	1.0103	1.1781	1.2194	2007	1.0394	1.0079	1.2383	1.1204
1982	0.9921	1.0152	1.0659	0.9986	1984	1.0421	1.0090	1.1838	1.0871
2001	1.0063	1.0042	1.0973	1.0169	1992	1.0421	1.0017	1.3032	1.1602
1997	1.0091	1.0008	1.2327	1.0831	1981	1.0452	1.0142	1.0632	1.0875
2002	1.0104	1.0053	1.1136	1.0504	1996	1.0471	1.0024	1.2665	1.1153
1983	1.0119	1.0114	1.1146	1.0120	2005	1.0475	1.0058	1.1657	1.1106
2010	1.0141	1.0083	1.1011	1.1817	2003	1.0582	1.0057	1.1202	1.1104
1986	1.0147	1.0129	1.1113	1.0799	2004	1.0593	1.0079	1.2036	1.1330
1991	1.0158	1.0030	1.2781	1.1532	1993	1.0596	1.0042	1.4376	1.2371
2008	1.0185	1.0088	1.1541	1.0979	1985	1.0846	1.0099	1.2379	1.1542
2006	1.0188	1.0057	1.1533	1.1042	1987	1.0854	1.0139	1.1803	1.1295
2000	1.0208	1.0065	1.1678	1.0750	1994	1.0890	1.0031	1.3726	1.3488
1989	1.0246	1.0111	1.0683	1.1626	1995	1.3565	1.0020	1.2836	1.1073

表 3 总用水量聚类结果及影响因素特征

类别	总用水量	样本数目	户籍人口	服务业产值	轻工业产值
$k = 1$	[0.97,1.00)	4	[1.0050,1.0152]	[1.0659,1.1781]	[0.9698,1.2194]
$k = 2$	[1.00,1.03)	12	[1.0008,1.0129]	[1.0683,1.2781]	[1.0120,1.1817]
$k = 3$	[1.03,1.06)	10	[1.0009,1.0142]	[1.0632,1.4376]	[0.9998,1.2371]
$k = 4$	[1.06,1.09)	3	[1.0031,1.0139]	[1.1803,1.3726]	[1.1295,1.3488]
$k = 5$	[1.09,1.36]	1	[1.0000,1.0040]	[1.2536,1.3136]	[1.0673,1.1473]

表 4 预测结果

影响因素	联系隶属度				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
户籍人口	0	0.6942	0.6165	0.3519	-0.2500
服务业产值	0.8004	0.6187	0.3739	-0.2449	-1.0000
轻工业产值	0.8854	0.8580	0.9195	-0.0939	0.9600
集成联系隶属度	0.5619	0.7236	0.6366	0.0044	-0.0967

表 5 与其他预测模型的误差对比

方 法	联系隶属度方法	SVM 模型 ^[2]	BP 模型 ^[2]	Elman 模型 ^[2]	加权平均集成模型 ^[2]
相对误差/%	-0.2156	-1.3655	6.9562	-7.0995	-0.4068

实例应用表明本文所述模型应用于城市需水量预测是有效可行的,该模型可充分利用需水量历史数据,并以此进行最优分割反映需水量变化规律以及同类样本间的差异程度,从同异反 3 个方面考虑待测样本和历史数据影响因素间的确定和不确定关系以及数据的动态可变性,克服了传统预测方法的不足,并具有预测精度高和误差变化平稳的优点。

4 结 语

本文基于 Fisher 最优分割算法聚类分析历史数据,应用联系隶属度分析预测年份影响因素与历史
(下转第 24 页)

- tributed rainfall-runoff model[J]. Journal of Hydrology, 2004,299(1-2):107-135.
- [15] Karvonen T, Koivusall H, Jauhaine M, et al. A hydrological model for predicting runoff from different land use areas[J]. Journal of Hydrology, 1999,217(3):253-265.
- [16] Bergstrom S, Graham L P. On the scale problem in hydrological modeling[J]. Journal of Hydrology, 1998,211(1):253-265.
- [17] Cammeraat E L H. Scale dependent thresholds in hydrological and erosion response of a semi-arid catchment in southeast Spain[J]. Agriculture, Ecosystems and Environment, 2004,104(2):317-332.
- [18] Tanaka G. Stochastic response characteristics of kinematic wave model[J]. Annual Journal of Hydraulic Engineering, JSCE, 2003,47:229-234.
- [19] Tanaka G, Fujita M, Kudo M. Comparison between the kinematic wave model and the storage routing function runoff model frequency characteristic and stochastic characteristic[J]. Journal of Hydraulic, Coastal and Environmental Engineering, 1999,614(II-46):21-36.
- [20] Cabral M C, Garrote L, Bras R, et al. A kinematic model of infiltration and runoff generation in layered and sloped soils[J]. Advances in Water Resources, 1992,15(5):311-324.
- [21] Yomoto A, Islam M N. Kinematic analysis of flood runoff for a small-scale upland field[J]. Journal of Hydrology, 1992,137(1-4):311-326.
- [22] Chua L H C, Wong T S W, Sriramula L K. Comparison between kinematic wave and artificial neural network models in event based runoff simulation for an overland plane[J]. Journal of Hydrology, 2008,357(3-4):337-348.
- [23] The Japan Society of Mechanical Engineerings. Fundamentals of computational fluid dynamics[M]. Tokyo: Corona Publishing Co., Ltd,1998:52-56.
- [24] Huang Jinbai, Hinokidan O, Yasuda H, et al. Effects of the check dam system on water redistribution in the Chinese Loess Plateau[J]. Journal of Hydrologic Engineering, 2013,18(8):929-940.
- [25] Tan Lu, Wang Zhongbo, Huang Jinbai, et al. Seasonal variation of moisture availability at the water-wind erosion crisscross region on the northern Loess Plateau, China[J]. Journal of Northeast Agricultural University (English edition), 2013,20(4):72-77.
- [26] 黄金柏,王 斌,桧谷治,等. 耦合融雪的分布式流域“降雨-径流”数值模型[J]. 水科学进展,2012,23(2):194-199.
- [27] Japanese River Society. River erosion control technical criterion 699 of MLIT (Ministry of Land, Infrastructure, Transport and Tourism)[M]. Tokyo: Sankaido publishing Co.,Ltd,1997:85-87.

(上接第 15 页)

数据间的关系,以预测相应的城市需水量,并将其应用于上海市的城市需水量的预测,应用结果及对比其他方法结果表明该模型应用于城市需水量预测是有效可行的,且预测精度相对较高,该模型能够充分挖掘分析样本与历史数据间的相似关系以及相应的确定和不确定关系,并在大样本预测中表现出明显的优越性,但模型在合理确定指标权重方面尚需深入研究和完善。

参考文献:

- [1] 雷鸣,张宏伟,闫静静. 城市用水量预测的 SIMULINK 仿真技术研究[J]. 中国给水排水,2010,26(15):54-57.
- [2] 崔东文. 加权平均集成神经网络模型在城市需水预测中的应用[J]. 水资源保护,2014,30(2):27-32+45.
- [3] 张雅君,刘全胜,冯萃敏. 多元线性回归分析在北京城市生活需水量预测中的应用[J]. 给水排水,2003,29(4):26-29.
- [4] 张 灵,陈晓宏,谢毅文. 投影寻踪在珠海市需水预测中的应用[J]. 人民长江,2008,39(6):41-43.
- [5] 崔东文,郭 荣. BP 神经网络模型与灰色 GM(1,1)模型在需水预测中的应用[J]. 水资源研究,2012,33(3):19-22.
- [6] 王 煜. 灰色系统理论在需水预测中的应用[J]. 系统工程,1996,14(1):60-64.
- [7] 刘克琳,王银堂,胡四一,等. Fisher 最优分割法在汛期分期中的应用[J]. 水利水电科技进展,2007,27(3):14-16+37.
- [8] 潘映默,王菲凤,张江山. 基于最优分割理论的酸雨监控点位的分类研究[J]. 环境科学与技术,2012,35(8):66-70+157.
- [9] 赵克勤. 集对分析及其初步应用[M]. 杭州:浙江科学技术出版社,2000.
- [10] Wang Mingwu, Chen Guangyi. A novel coupling model for risk analysis of swell and shrinkage of expansive soils[J]. Computers and Mathematics with Applications, 2011,62(7):2854-2861.
- [11] 汪明武,金菊良,周玉良. 集对分析耦合方法与应用[M]. 北京:科学出版社,2014.
- [12] 上海市统计局. 上海市统计年鉴 2012[M]. 北京:中国统计出版社,2012.