

人工免疫算法在洪水分类中的应用*

张 灵^{1,2}, 陈晓宏², 翁 毅³, 刘青娥²

(1. 广州大学地理科学学院, 广东 广州 510006;

2. 中山大学水资源与环境研究中心, 广东 广州 510275;

3. 中山大学地理科学与规划学院, 广东 广州 510275)

摘 要: 在总结洪水分类研究的基础上, 提出了一种新的洪水分类方法, 以人工免疫网络对洪水样本进行免疫学习和记忆, 提取表征洪水强度的有用特征得到抗体库和相似度矩阵, 利用最小生成树方法, 依据抗原与记忆集的亲和度确定洪水的分类。以宜昌站 12 场典型洪水过程和广东石狗站 17 场典型洪水过程为例进行了洪水聚类分析, 结果表明: 所提算法有效提取了同类型洪水的模糊特征和规律, 去除了不必要的信息冗余, 较好地地将同类洪水聚集在了一起; 与进化粒子群优化算法相比, 该法有更快的收敛速度。

关键词: 洪水分类; 人工免疫; 克隆选择; 最小生成树

中图分类号: P333.2 **文献标识码:** A **文章编号:** 0529-6579 (2008) 05-0110-05

一般根据洪水强度的大小, 把洪水分成特大洪水、大洪水、中等洪水和小洪水等不同等级。准确可靠的洪水分类是洪水分类预报、洪水资源化调度和洪水灾害评估的重要依据。但由于洪水强度受洪峰水位、洪峰流量、洪水历时、洪水总量、洪水重现期等要素影响, 具有较强的不确定性和模糊性, 因此对历史洪水过程进行聚类分析, 寻找同类型洪水的模糊特征和规律, 以消除不确定性因素的影响, 具有重要的指导意义。

按照聚类准则的不同, 聚类算法可分为基于模糊关系的聚类算法和基于目标函数的聚类算法两种, 其中后者把聚类问题转化为约束优化问题。目前洪水分类采用的大多是后者, 如马寅午等^[1-2]提出了按流量均值对历史洪水进行分类的聚类分析方法; 闵骞^[3]根据地震、风力等级的数学结构, 提出了基于洪峰水位和警戒水位的洪水等级划分公式; 吕宏军等^[4]构建了洪水强度的灰色聚类模型; 许成武等^[5]基于洪水重现期对嘉陵江洪水进行了分级。但这些方法均需预先给出洪水类型所对应的洪水要素阈值, 而由于洪水的地域性限制, 规定统一的洪水要素分类指标值比较困难^[6]。同时由于洪水过程的复杂性, 采用常用洪水要素度量洪水的结果有时不太客观、合理。针对这一问题, 近年来在洪水分类研究中引入了遗传算法^[7]、投影寻踪^[8-9]等进化算法。这些方法直接由历史洪水样本数据驱动得到洪

水的自然分类结果, 在一定程度上克服了传统方法的不足。但由于进化算法固有的缺陷, 在进化过程中目标函数常常出现局部极值, 甚至出现退化解和无解的情况^[10], 最终导致迭代次数过多和聚类准确率不是很高。

人工免疫系统 (Artificial Immune System, AIS) 是模仿自然免疫系统功能的一种人工智能方法, 它通过实现外界物质的自然防御机理的学习技术, 提供了噪声忍耐、自学习、自组织、记忆等进化学习机理, 结合了分类器、神经网络和机器推理等系统的优点^[10-11]。该算法具有良好的整体收敛性能和全局搜索能力, 不需要任何先验知识。目前 AIS 已在模式识别^[12]、数据挖掘^[13]、图象识别^[14]等领域得到了成功的应用, 但将 AIS 用于洪水分类的研究尚不多见。基于此, 本文尝试引入人工免疫系统的克隆选择算法, 以期解决洪水分类中的先验性、局部极值、鲁棒性等实际难点。

1 基于人工免疫算法的洪水分类

1.1 人工免疫算法

生物免疫系统抵御外部入侵使其机体免受病原侵害的应答反应称为免疫。诱导免疫系统产生免疫应答的物质称为抗原, 能与抗原进行特异性结合的免疫细胞称为抗体, 免疫应答主要由分布在生物体内部的免疫细胞实现。免疫算法主要模拟生物免疫

* 收稿日期: 2007-11-26

基金项目: 国家自然科学基金资助项目 (50579078); 广东省自然科学基金资助项目 (04009805)

作者简介: 张灵 (1974 年生), 男, 博士研究生; 通讯联系人: 翁毅; E-mail: onlyzsu@yahoo.com.cn

系统中有关抗原处理的核心思想, 其中克隆选择算法描述了获得性免疫的基本特性: 只有成功识别抗原的免疫细胞才得以增殖, 经历变异后的免疫细胞分化为浆细胞和记忆细胞两种。克隆选择算法的具体实现过程可参阅文献[15]。

1.2 基于人工免疫算法的洪水分类

应用人工免疫算法对洪水进行聚类分析, 首先将需要解决的问题抽象成符合免疫系统处理的抗原形式, 抗原识别则对应为问题的求解。因此, 本文将洪水样本数据作为免疫系统的抗原, 应用人工免疫算法得到抗体库 M 和各抗体之间的相似度矩阵 S , 并用最小生成树方法根据抗原与记忆集的亲和度确定洪水的分类。主要步骤如下:

① 定义抗原: 输入洪水样本数据

$$X = \{x_1, x_2, \dots, x_n\}$$

作为抗原, 其中 $x_1 \sim x_n$ 为洪水样本中最高水位、洪峰流量、3 日洪量等属性, 按式对洪水样本进行归一化。

$$x_{ij} = \frac{x_{ij} - \text{Min}x_{ij}}{\text{Max}x_{ij} - \text{Min}x_{ij}} \quad i, j = 1, \dots, n \quad (1)$$

② 初始化: 随机初始化抗体群体作为初始网络细胞。

③ 模式识别与匹配: 按式 (2) 计算抗原 Ag_i 和每个抗体 Ab_i 之间的亲和力 a_{ij} 。按 a_{ij} 降序排列, 选择 n 个前抗体作为网络细胞进行克隆和变异。

$$a_{ij} = (1 + \|Ab_i - Ag_i\|)^{-1} \quad (2)$$

④ 克隆选择: 采用实数变异机制对抗体进行随机变异操作, 实现亲和力的成熟。选择具有最高亲和力的抗体作为记忆细胞。

⑤ M 抑制: 按式 (2) 计算抗体 M 库中各记忆细胞的相似度, 淘汰相似度低于阈值的抗体, 实现不同克隆集的网络抑制。

⑥ 迭代: 返回步骤③, 进行多次克隆选择直到亲和力达到阈值或达到最大迭代次数。

⑦ 洪水分类: 根据最后得到的抗体库 M 和相似度矩阵 S , 利用最小生成树方法对记忆细胞进行分类, 根据抗原与记忆集的亲和度 (抗原与类的重心之间的距离) 实现洪水的分类。

2 应用实例

3.1 宜昌站洪水分类

文献[8]利用投影寻踪和粒子群优化算法对湖北省宜昌站 1931 - 1998 年间的 12 场典型洪水过程

进行了分类研究; 本文采用人工免疫算法对其样本也进行了分类。算法中参数设置如下: 抗原个数 $n = 12$, 初始抗体个数 $N = 10$, 克隆选择率 $\beta = 10\%$, 克隆乘数因子 $a = 1$, 亲和力阈值 $\sigma_d = 1$, 免疫抑制阈值 $\sigma_s = 0.075$, 最大迭代次数 $N_c = 14$ 。最终本文算法将 1931、1954、1981、1998 年 4 场洪水分为特大洪水, 将 1935、1958、1966、1974、1980、1982 年 6 场洪水分为大洪水, 将 1969 年、1983 年洪水分为中等洪水。限于篇幅, 本文仅给出两种分类方法的结果对比。

表 1 人工免疫算法计算过程

Tab. 1 Clustering calculation of Artificial Immune

迭代 次数	平均 亲和力	网络细 胞个数	迭代 次数	平均 亲和力	网络细 胞个数
1	0.7073	9	8	0.7168	8
2	0.7021	8	9	0.7231	7
3	0.7152	9	10	0.7183	9
4	0.7116	8	11	0.7173	9
5	0.7146	7	12	0.7143	8
6	0.7089	8	13	0.7073	8
7	0.7061	9	14	0.7129	9

表 1 是洪水样本的人工免疫计算过程。可以看出, 网络细胞个数小于初始抗体个数 10, 说明免疫系统将整个数据信息记忆到有限个记忆细胞中, 去除了不必要的信息冗余, 有效缩减了数据规模。从最终结果来看, 分类结果与文献[8]基本一致, 说明本文算法能较好地将同一类型的洪水聚集在一起。而从收敛速度来看, 本文算法迭代至第 9 代时, 记忆细胞和抗原之间的平均亲和力最大, 说明此时的数据分类结构最为理想; 而文献[8]运行 20 次后才得到最大目标函数值, 说明本文算法收敛的更快, 更适用于洪水分类。

2.2 广东石狗站洪水分类

石狗站位于广东省四会境内, 是绥江干流上的主要水文站。本文以 1955 - 2004 年石狗站 17 场典型洪水过程为例, 进行基于人工免疫算法的洪水分类研究 (表 2)。

将 17 个洪水样本按式 (1) 进行归一化后作为抗原, 并随机生成初始抗体, 即可用人工免疫算法进行洪水样本的免疫识别和记忆。算法中参数设置与宜昌站计算实例相同, 克隆选择过程见如图 1。

表 2 石狗站历史洪水样本

Tab. 2 The historical flood data at Shigou hydrological station during 1955 - 2004

洪号	最高水位/m	洪峰流量/($\text{m}^3 \cdot \text{s}^{-1}$)	3 日洪量/亿 m^3	7 日洪量/亿 m^3	15 日洪量/亿 m^3
19550722	16.92	4480	11.3	24.3	43.1
19610902	14.82	2340	5.6	11.7	21.0
19620518	15.35	2700	6.9	15.6	31.7
19640616	15.20	2700	6.8	13.8	25.0
19660613	15.35	2670	6.8	15.0	26.5
19700503	14.87	2250	5.6	11.7	22.1
19730529	15.44	2600	6.5	12.9	23.3
19750522	15.39	2540	6.5	14.0	25.4
19820514	16.28	3460	8.4	16.4	27.3
19830619	15.93	3320	8.3	17.1	32.9
19840428	15.61	2970	7.3	15.1	27.7
19850924	15.67	2920	6.9	16.0	33.4
19870523	15.11	2480	6.3	13.9	24.7
19890523	15.73	2890	7.2	15.0	28.4
19940725	16.05	3450	8.6	17.1	30.9
19951005	15.22	2340	6.0	12.9	23.4
19970705	14.79	2460	6.0	11.6	19.4

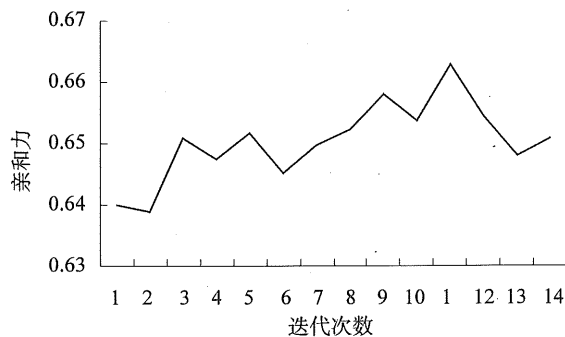


图 1 亲和力变化曲线

Fig. 1 The trend of affinity

从图 1 可以看出, 迭代到第 11 次时, 亲和力达到最大值, 意味着此时抗体库 M 同抗原之间的平均欧基里德距离最小, 从而得到较理想的数据分类结构。为直观起见, 将该数据结构用最小生成树“先末梢后主干、由远及近”的概念描述为图 2, 其中虚线表示可以剪除的连接。显然, 按照“物以类聚”的特性, 17 个抗体分成了 4 类。

采用重心法计算各抗原与抗体类重心之间的距离, 并与其年份对应, 从而得到石狗站历史洪水样本分类结果 (图 3)。17 个历史洪水样本序列中, 1955、1982、1983、1994 年 4 场洪水属于特大洪水, 对应于图 2 的 IV 类; 1962、1964、1984、1985、1989 年 5 场洪水为大洪水, 对应于图 2 的 III 类; 1966、1973、1975、1987、1997 年 5 场洪水属于中等洪水, 对应于图 2 的 II 类; 1961、1970、1995 年 3 场洪水则属于常遇洪水, 对应于图 2 的

I 类。

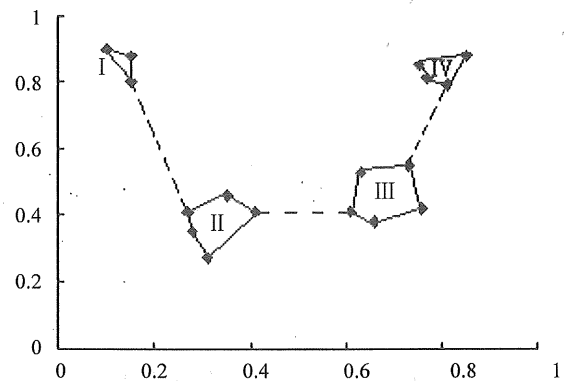


图 2 数据分类结果

Fig. 2 Network structure of clustered data

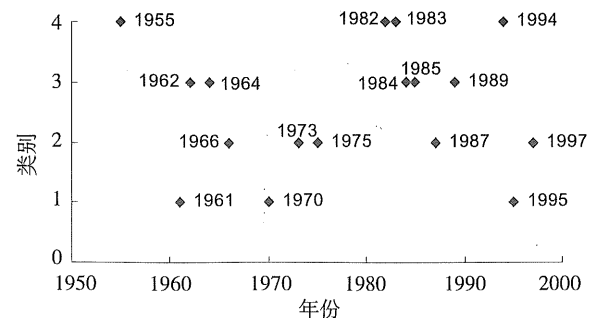


图 3 石狗站历史洪水样本分类散点图

Fig. 3 Clustered historical flood sample of Shigou hydrological station

图 3 说明, 本文基于人工免疫的聚类算法能较好地将同一类型的洪水聚集在一起, 得到的分类结

果与石狗站的历史洪水实际情况也较为吻合。以石狗站 1953 - 2004 年间 52 a 的实测径流资料进行排频计算, 所得结果和文中算法分类结果一致, 说明人工免疫聚类算法有效地提取了同类型洪水的模糊特征和规律, 在洪水分类中具有较高的应用价值。但需要指出的是, 该方法只能完成分类, 并不能判断出类的级别, 因此实际应用时要根据洪峰、洪量、水位等洪水要素的值来判断所分类别的等级。

3 小 结

由于洪水高度的随机性和复杂性, 准确的洪水分类具有重要的理论与实践意义。本文设计的基于人工免疫算法的洪水分类方法直接由历史洪水样本数据驱动, 不需预设洪水类型所对应的洪水要素阈值。该方法避免了采用某几个洪水要素指标度量洪水的局限性, 综合考虑了多因素的影响, 利用了样本序列中更多的洪水信息, 分类结果更客观、合理。该方法较好地解决了洪水分类中的局部极值、鲁棒性等问题, 其记忆和学习能力较好地消除了原始样本中不确定因素的影响。应用实例说明, 人工免疫聚类算法有效地提取了同类型洪水的模糊特征和规律, 去除了不必要的信息冗余, 而与投影寻踪和粒子群优化算法相比, 本文算法收敛速度更快, 表明该方法在洪水分类中具有较高的应用价值。同时, 本文聚类算法有较多的可调参数, 如何根据具体问题更合理地选择参数, 是下一步的研究方向。

参考文献:

- [1] 马寅午, 周晓阳, 尚金成, 等. 防洪系统洪水分类预测优化调度方法[J]. 水利学报, 1997(4): 1 - 8.
MA Yinwu, ZHOU Xiaoyang, SHANG Jincheng, et al. A new method for flood types prediction and optimal control [J]. Journal of Hydraulic Engineering, 1997(4): 1 - 8.
- [2] 马寅午, 周晓阳, 尚金成, 等. 长江防洪系统洪水分类预报调度研究[J]. 水电能源科学, 1996, 14(2): 84 - 92.
MA Yinwu, ZHOU Xiaoyang, SHANG Jincheng, et al. Study on flood classified prediction and dispatching method in Yangtze river flood control system [J]. International Journal Hydroelectric Energy, 1996, 14(2): 84 - 92.
- [3] 闵骞. 洪水的等级划分及其减灾意义[J]. 自然灾害学报, 1994, 3(1): 34 - 39.
MIN Qian. Classification of flood level and its meaning in disaster reduction [J]. Journal of Natural Disasters, 1994, 3(1): 34 - 39.
- [4] LU Hongjun, CHEN Yinchuan. The grey clustering method of the evaluation of flood severity [C]. // Proceedings of International Symposium on Floods and Droughts. Nanjing: Hohai University Press, 1999.
- [5] 许成武, 王文, 黎明. 嘉陵江流域洪水等级的建议划分标准[J]. 自然灾害学报, 2005, 14(3): 51 - 55.
XU Chengwu, WANG Wen, LI Ming. A suggested gradation standards of flood in Jialing river basin [J]. Journal of Natural Disasters, 2005, 14(3): 51 - 55.
- [6] 魏一鸣, 金菊良, 杨存建. 洪水灾害风险管理理论[M]. 北京: 科学出版社, 2002.
- [7] 金菊良, 金保明, 杨晓华, 等. 建立洪水灾情等级模型的实用方案[J]. 灾害学, 2000, 15(2): 1 - 6.
JIN Juliang, JIN Baoming, YANG Xiaohua, et al. A practical scheme for establishing grade model of flood disaster loss [J]. Journal of Catastrophology, 2000, 15(2): 1 - 6.
- [8] 董前进, 王先甲, 艾学山, 等. 基于投影寻踪和粒子群优化算法的洪水分类研究[J]. 水文, 2007, 27(4): 10 - 14.
DONG Qianjin, WANG Xianjia, AI Xueshan, et al. Study on flood classification based on project pursuit and particle swarm optimization algorithm [J]. Journal of China Hydrology, 2007, 27(4): 10 - 14.
- [9] 王顺久, 张欣莉, 侯玉, 等. 洪水灾情投影寻踪评估模型[J]. 水文, 2002, 22(4): 1 - 4.
WANG Shunjiu, ZHANG Xinli, HOU Yu, et al. Projection pursuit model for evaluation of flood events [J]. Journal of China Hydrology, 2002, 22(4): 1 - 4.
- [10] FORREST, PERELSON A S, ALLEN L, et al. Self-non-self discrimination in a computer [C]. // Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland, USA, 1994: 202 - 212.
- [11] DASGUPTA D. Artificial Immune Systems and Their Applications [M]. Berlin: Springer-Verlag, 1999.
- [12] CARTER J H. The Immune System As a Model for Pattern Recognition and Classification [J]. Journal of the American Medical Informatics Association, 2000, 7(3): 28 - 41.
- [13] TIMMIS J, NEAL M, HUNT J. An Artificial Immune System for Data Analysis [J]. Biosystem, 2000, 55: 143 - 150.
- [14] ZHONG Y F, ZHANG L P, HUANG B, et al. An unsupervised artificial immune classifier for multi hyperspectral remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(2): 420 - 431.
- [15] De CASTRO L N, Von ZUBEN F J. Clonal selection algorithm with engineering applications [C]. // Genetic and evolutionary computation conference. Nevada, USA, 2000: 36 - 37.

Application of Artificial Immune Network in Flood Classification

ZHANG Ling^{1,2}, CHEN Xiao-hong², WENG Yi³, LIU Qing-er²

(1. School of Geographical Sciences, Guangzhou University, Guangzhou 510006, China;

2. Center for Water Resources and Environment Research, Guangzhou 510275, China;

3. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: After summarizing flood classification research, a novel flood classifier employing artificial Immune network was proposed to carry out immune memory and learning of the flood sample. The useful characteristics that effectively represent the flood intensity are extracted to produce antibody repertoires and similarity matrix. Then minimum spanning tree is used to classify flood sample by calculating the affinity between antigen and memory cell sets. The proposed classifier was testified by 12 typical hydrographs of Yichang station and 17 typical hydrographs of Shigou station. The results show that the proposed classifier classified the similar flood into same class by extracting the fuzzy characteristics and removing redundant information. Comparing with Evolutionary Particle Swarm Optimization, the classifier fastens the convergence rate.

Key words: flood classification; artificial immune system; clonal selection algorithm; minimum spanning tree