김일성종합대학학보 (자연과학)
주체104(2015)년 제61권 제1호
JOURNAL OF KIM IL SUNG UNIVERSITY (NATURAL SCIENCE)
Vol. 61 No. 1 JUCHE104(2015).

NN법에 의한 일흐름량예측정확도에 주는 파라메터들의 영향평가

조명봉, 리서진

위대한 령도자 **김정일**동지께서는 다음과 같이 교시하시였다.

《강하천들에 언제와 갑문을 비롯한 여러가지 시설물을 건설하고 잘 관리하여야 큰물피해를 막을수 있을뿐아니라 나라의 풍치도 돋굴수 있습니다.》(《**김정일**선집》 증보판 제11권 39페지)

우리는 NN(Nearest Neighbor)법에 의한 일흐름량예측때 합리적인 파라메터를 찾기 위한 연구를 하였다.

NN법은 원래 통계적류형(Pattern)인식법의 하나이지만 시계렬예측에도 효과적으로 리용될수 있는 방법이다.

NN법에 의한 시계렬예측에서는 현재의 시계렬류형에 류사한 과거의 시계렬류형에 기초하여 예측한다.[1] 즉 예측에 필요한 요소들로 이루어진 일정한 류형들을 만든 다음 현재의 류형과 류사한 과거류형을 찾아 거기에 기초하여 예측한다. 류사성을 특징짓는 척도는 예측분야에 따라 각이하게 정할수 있다.

NN법에 의한 류출예측에서도 이와 마찬가지로 현재의 사건과 류사한 과거의 류출사건을 과거자료로부터 선택하고 그것들의 평균값을 예측값으로 한다. 여기서는 류형을 흐름량예측에 리용되는 요소들로 이루어진 벡토르로 하고 류사성척도는 유클리드거리로 하였다.

NN법에 의한 일평균흐름량예보의 기본공정은 다음과 같다.

현재 시각 n에 이르기까지의 흐름량시계렬 $q(i), (1 \le i \le n)$에 기초하여 시각 $n+1$의 류량 $q(n+1)$를 예측하는 경우를 보자.

우선 최근의 흐름량자료(M개)를 리용하여 현재의 류출현상의 상태를 반영하는 다음과 같은 고유벡토르 $X(n)$을 형성한다.

$$X(n) = [q(n),\ q(n-1),\ \cdots,\ q(n-M+1)]$$

여기서 M은 흐름량의 수$(1 \le M \le n)$, $X(n)$은 n시각의 고유벡토르, $q(n)$은 n시각의 흐름량이다.

다음으로 시각 $i(1 \le i \le n-1)$마다 $X(n)$과 같은 구조인 과거의 고유벡토르 $X(i)$를 형성한다.

$$X(i) = [q(i),\ q(i-1),\ \cdots,\ q(i-M+1)]$$

$X(i)$와 $X(n)$사이의 유클리드거리 $\|d\|_{in}$은 다음식으로 계산된다.

$$\|d\|_{in} = \sqrt{[q(i)-q(n)]^2 + [q(i-1)-q(n-1)]^2 + \cdots + [q(i-M+1)-q(n-M+1)]^2}$$

다음 계산된 유클리드거리들중에서 그 값이 작은 순서로 k 개의 고유벡토르 $X(i)$ 들을 선택하여 k 개의 $q(i+1)$ 들의 평균값을 $n+1$ 시각의 예측흐름량 $q(n+1)$ 로 한다.

$$q(n+1)=\frac{1}{k}\sum_{i\in S(X,\ n)}q(i+1)$$

여기서 $S(X, n)$ 은 고유벡토르 $X(i)$ 들과 $X(n)$ 과의 유클리드거리가 작은 순서로 k 개 선택하여 꺼낸 고유벡토르 $X(i)$ 들에 대응하는 시각첨수 i 의 모임이다.

같은 방법으로 예측흐름량 $q(i+2)$, $q(i+3)$ 도 계산된 유클리드거리들중에서 그 값이 작은 순서로 뽑은 k 개의 고유벡토르 $X(i)$ 들을 선택한 다음 각각 k 개의 $q(i+2)$, $q(i+3)$ 들의 평균값으로 한다.

NN법에 의한 일흐름량예보에서는 고유벡토르를 이루고있는 흐름량요소의 수 M 과 고유벡토르의 개수 k 의 값에 따라 예보결과가 달라지므로 우리는 이 두 파라메터의 합리적인 값을 얻어 예측정확도를 높이기 위한 실험을 하였다.

실험에서 모형작성에는 ㅈ강 ㄹ지점의 1981～1990년 일평균흐름량자료가, 모형검토에는 1991～1995년의 자료가 리용되였다.

파라메터들인 M 의 개수는 1～5로, k 는 1, 50, 100, 150, 200으로 변화시키면서 그것에 따르는 예측결과들을 얻었으며 예견기는 1일, 예측정확도는 상대오차(ε)로 평가하였다.

$$\varepsilon=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|Q_{C_i}-Q_{O_i}\right|}{Q_{O_i}}\times 100(\%)$$

여기서 Q_{C_i} 는 예측값, Q_{O_i} 는 관측값, N은 자료개수이다.

구체적으로 보면 매 년도별로 M 값을 고정하고 k 값을 변화시키는 방법으로 상대오차가 제일 작은 M 값을 찾아내고 그 값에 한하여 k 를 보다 더 세밀하게 변화시키면서 역시 상대오차가 제일 작은 값을 찾아내였다.

가장 합리적인 M 값을 찾기 위한 실험결과는 그림 1과 같다.

그림 1에서 보는바와 같이 가장 적합한 값은 $M=2$ 이다. 즉 이것은 예측하는데 리용되는 고유벡토르의 요소수가 2일 때가 제일 좋다는것으로서 2일전까지의 흐름량이 당일 흐름량에 큰 영향을 미친다는것을 보여준다.

M 을 2로 고정시키고 가장 합리적인 k 를 얻는 과정은 그림 2와 같다.

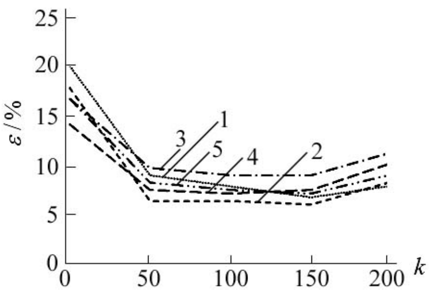

그림 1. k 에 따르는 상대오차변화
1－5는 M의 수

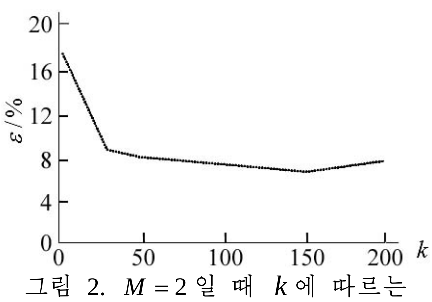

그림 2. $M=2$ 일 때 k 에 따르는 상대오차변화

그림 2에서 보는바와 같이 k 가 150근방일 때 예측정확도가 제일 높다.

맺 는 말

이 방법을 리용하면 NN법의 부족점을 극복할수 있다.

참 고 문 헌

[1] 藤原洋一 等; J. Japan Soc. Hydro. & Water Resour., 10, 1, 33, 2003.

주체103(2014)년 9월 5일 원고접수

Influence Assessment of Parameters on the Accuracy of Daily Flow Forecast by the Nearest Neighbor Method

Jo Myong Bong, Ri So Jin

We considered the results that experimented on the accuracy assessment of daily flow forecast according to the changes of parameters M, K by the nearest neighbor method. As the results, the optimal values of M and K are M=2 and k=150 for daily flow forecast in study areas.

Key words: nearest neighbor method, daily flow forecast