

최적류사표본렬탐색(NN)법에 기초한 일류출 예측모형의 계절적특성

조 명 봉

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《물자원을 적극 보호하고 효과적으로 리용하여야 합니다.》(《김정일선집》 증보판 제21권 193페이지)

최근에 하천류출과정의 예측문제들에 인공신경망, 카오스리론, 유전계획, 진화계산 등 비선형체계해석수법들을 적용하려는 시도들이 적지 않게 제기되고있다.

이러한 비선형체계해석방법들에 의한 하천 일류출예측모형의 개발은 큰 전망성을 가지고있으나 아직 세계적으로 그 적용실례가 드물고 예측방법도 완성되어있지 않다.

한편 최근에 최적류사표본렬탐색(NN: Nearest-Neighbor)법을 하천류출예측에 적용할수 있다는 연구결과들이 발표되였다.[3, 4, 7, 8]

NN법의 기초원리가 간단명료하고 예측공정이 복잡하지 않으며 자료조건이 비교적 현실적인것으로 하여 일류출예측에 이 방법을 능히 적용할수 있는 가능성이 최근에 제기되고있다.[1, 2, 5, 6]

이로부터 논문에서는 NN법에 기초하는 일류출예측모형들의 계절적특성에 대한 문제를 논의하였다.

1. NN법에 기초하는 일류출예측모형과 방법

최적류사표본렬탐색(NN)법의 구체적인 모형과 방법론은 다음과 같다.

우선 일정한 기간 연속적으로 관측된 일흐름량자료계렬을 준비한다.

흐름량자료로서는 수문관측소에서 관측한 일평균흐름량값들로 이루어진 자료계렬이다.

$$q_1, q_2, q_3, \dots, q_{i-1}, q_i, q_{i+1}, \dots, q_{n-2}, q_{n-1}, q_n$$

여기서 $q_i (1 \leq i \leq n)$ 는 i 시각의 일흐름량자료, n 은 자료계렬의 길이이다.

자료가 준비되면 예측을 위한 자료표본렬(Pattern)이 있어야 한다.

자료표본렬은 자료계렬을 임의로 M 개씩 토막내어 만들며 이렇게 되면 모두 $(n-M+1)$ 개의 자료표본렬이 만들어진다. 이 자료표본렬을 예측에 필요한 계산에 리용하므로 특징벡토르라는 수학적 개념으로 부르기로 한다.

흐름량자료를 M 개 리용하여 현재의 류출현상의 상태를 반영하는 자료표본렬을 특징벡토르로 다음과 같이 표시할수 있다.

$$X(n) = (q(n), q(n-1), \dots, q(n-M+1)) \quad (1)$$

여기서 $X(n)$ 은 n 시각의 특징벡토르, $q(n)$ 은 n 시각의 흐름량자료, M 은 특징벡토르에 들어있는 흐름량요소의 개수($1 \leq M \leq n$).

이제 현재 시각 n 에 이르기까지의 흐름량시계열 $|q(i)| (1 \leq i \leq n)$ 에 기초하여 시각 $n+1$ 의 흐름량 $q(n+1)$ 을 예측하는 경우 즉 1일전 예측공정을 보자.

n 시각의 특징벡토르를 식 (1)과 같이 구성한 다음 시각 $i (1 \leq i \leq n-1)$ 마다 $X(n)$ 과 같은 구조인 이전의 특징벡토르 $X(i)$ 를 만든다.

$$X(i) = (q(i), q(i-1), \dots, q(i-M+1)) \quad (2)$$

다음 현재의 특징벡토르 $X(n)$ 과 유사한 이전의 특징벡토르 $X(i)$ 들을 찾아야 한다.

이때 유사성정도를 특징짓는 량으로서 유클리드거리를 이용하였다.

이전의 특징벡토르 $X(i)$ 와 현재의 특징벡토르 $X(n)$ 사이의 유클리드거리를 $\|d\|_{in}$ 으로 표시하면 다음과 같이 쓸수 있다.

$$\|d\|_{in} = \sqrt{(q(i) - q(n))^2 + (q(i-1) - q(n-1))^2 + \dots + (q(i-M+1) - q(n-M+1))^2} \quad (3)$$

계산된 유클리드거리들을 그 값이 작은 순서로 배열한다.

$$\|d\|_{in}(1) < \|d\|_{in}(2) < \dots < \|d\|_{in}(j-1) < \|d\|_{in}(j) < \|d\|_{in}(j+1) < \dots < \|d\|_{in}(n-M+1) \quad (4)$$

여기서 $j (j=1, n-M+1)$ 는 크기순서첨수이다.

다음 $j=k$ 되게 즉 유클리드거리값이 작아지는 순서로 k 개의 특징벡토르 $X(i, j)$ 들을 선택한다.

$$\begin{cases} X(i, 1) = (q(i, 1), q(i-1, 1), \dots, q(i-M+1, 1)), q(i+1, 1) \\ X(i, 2) = (q(i, 2), q(i-1, 2), \dots, q(i-M+1, 2)), q(i+1, 2) \\ \vdots \\ X(i, j) = (q(i, j), q(i-1, j), \dots, q(i-M+1, j)), q(i+1, j) \\ \vdots \\ X(i, k-1) = (q(i, k-1), q(i-1, k-1), \dots, q(i-M+1, k-1)), q(i+1, k-1) \\ X(i, k) = (q(i, k), q(i-1, k), \dots, q(i-M+1, k)), q(i+1, k) \end{cases} \quad (5)$$

여기서 j 는 크기순서첨수이며 $q(i+1, j)$ 는 흐름량자료계열에서 매 특징벡토르뒤에 놓이는 1일후의 흐름량값들이다.

이 특징벡토르들로부터 예측을 진행한다.

즉 일흐름량자료계열에서 추출된 매개의 특징벡토르 $X(i, j)$ 들의 뒤로 련속되는 k 개의 1일후의 흐름량값 $q(i+1, j)$ 들의 평균값이 바로 $n+1$ 시각의 예측흐름량 $q(n+1)$ 이다.

$$q(n+1) = \frac{1}{k} \sum_{j=1}^k q(i+1, j) \quad (6)$$

이 값은 예측하는 날의 다음날 예측값이므로 1일전 예측값이 된다. 2일전 또는 그 이상의 예전기를 가지는 예측에서도 우와 마찬가지로 전개할수 있다.

2. 결과 및 분석

논문에서는 3개의 연구지역(룡림, 갈현, 대흥)들에 배치된 수문관측소에서 관측한

15년간의 일흐름량자료계렬을 리용하였다.

15년간의 자료가운데서 10년(3 650일)간의 자료는 예측모형의 최적파라미터결정을 위한 모형작성자료이며 5년(1 825일)간은 모형검토자료이다.

년중 계절에 따라 수문학적특성에서 차이가 있으므로 여기서는 갈수기(10~3월)와 풍수기(4~9월), 장마철(7~9월)에 리용할수 있는 예측모형을 개별적으로 작성하고 그 결과를 분석하였다. 예측모형들에 필요한 최적파라미터추정방법은 시행착오법을 적용한다.

예측정확도가 그중 높은 2개의 최적파라미터조합($M=2, k=100$ 과 $M=3, k=100$)을 리용하였을 때의 예측예견기별상대오차는 다음과 같다.(표 1-3)

1일전 예측결과(예견기가 1일일 때)

표 1. 해당 시기에 따르는 지점별상대오차(%)

지점	$M=2, k=100$			$M=3, k=100$		
	풍수기	갈수기	장마철	풍수기	갈수기	장마철
통림지점	7.4	7.9	9.6	8.0	8.2	10.3
갈현지점	13.0	12.7	13.5	12.4	11.3	13.7
대흥지점	7.6	6.4	8.5	8.5	6.4	9.2

2일전 예측결과(예견기가 2일일 때)

표 2. 해당 시기에 따르는 지점별상대오차(%)

지점	$M=2, k=100$			$M=3, k=100$		
	풍수기	갈수기	장마철	풍수기	갈수기	장마철
통림지점	15.0	10.2	17.5	13.4	10.1	16.8
갈현지점	22.3	21.9	21.9	23.5	21.1	22.6
대흥지점	15.0	7.8	16.0	15.4	7.9	16.2

3일전 예측결과(예견기가 3일일 때)

표 3. 해당 시기에 따르는 지점별상대오차(%)

지점	$M=2, k=100$			$M=3, k=100$		
	풍수기	갈수기	장마철	풍수기	갈수기	장마철
통림지점	20.1	11.5	23.0	17.5	11.3	21.5
갈현지점	30.8	29.9	28.2	32.3	25.7	29.8
대흥지점	18.7	9.5	19.4	18.7	9.0	19.4

표 1-3까지의 예측결과를 종합하여보면 첫째로, 풍수기와 갈수기, 장마철에 따르는 예측모형의 계절별최적파라미터값들은 세가지 경우에 큰 차이가 없다는것이다. 계절별모형들에서 가장 좋은 예측결과를 얻을수 있는 최적파라미터조합은 $M=3, k=100$ 이라는것을 알수 있다. 일부 경우에 파라미터값이 조금 차이나는것이 있지만 그에 따르는 예측오차의 편차는 작기때문에 무시할수 있다.

둘째로, 계절별예측모형들은 일정한 차이가 있다. 특히 갈수기와 장마철의 차이가 크며 풍수기와 장마철의 차이는 작다. 갈수기와 장마철예측모형들의 상대오차의 차이는 최대 17%정도, 최소 2~3%정도이다. 즉 갈수기모형들은 오차가 작다.

맺 는 말

최적류사표본렬탐색(NN)법에 기초한 일류출예측모형은 계절에 따라 명백한 차이를 가지며 특히 갈수기와 장마철모형은 따로 작성리용하는것이 좋다.

최적류사표본렬탐색(NN)법에 기초한 일류출예측모형의 최적파라미터는 계절적인 차이가 거의 없으며 최적파라미터조합은 $M=3$, $k=100$ 이다.

참 고 문 헌

- [1] 김일성종합대학학보(자연과학), 58, 8, 158, 주제101(2012).
- [2] 리서진; 기상과 수문, 1, 18, 주제103(2014).
- [3] G. Galeati; Hydrol. Sci. J., 35, 79, 1999.
- [4] M. Karlsson, et al.; Water Resour. Res., 23, 7, 1300, 1987.
- [5] B. Sivakumar et al.; Journal of Hydrology, 265, 225, 2002.
- [6] 藤原洋一 等; J. Japan Soc. Hydro. & Water Resource, 10, 1, 33, 2003.
- [7] 近森秀高 等; 水資源学会誌, 15, 2, 164, 2002.
- [8] 多田明夫 等; 農土論集, 210, 75, 2000.

주제 104(2015)년 8월 5일 원고접수

Seasonal Characteristics of Daily Flow Forecast Model based on the Nearest-Neighbor Method

Jo Myong Bong

In this paper, I considered the seasonal characteristics of daily flow forecast model based on the nearest-neighbor method.

The daily flow forecast model based on the nearest-neighbor method shows obvious differences according to seasons and especially, the dry season and the rainy season models may be separately made and used.

However, the optimal parameters of the daily flow forecast model based on the nearest-neighbor method make few seasonal differences and those are $M=3$, $k=100$.

Key words: nearest-neighbor method, daily flow