

지지벡터기구를 이용한 월 강우량자료의 Downscaling 기법

Downscaling Technique of the Monthly Precipitation Data using Support Vector Machine

김성원* · 경민수** · 권현한*** · 김형수****

Abstract

The research of climate change impact in hydrometeorology often relies on climate change information. In this paper, neural networks models such as support vector machine neural networks model (SVM-NNM) and multilayer perceptron neural networks model (MLP-NNM) are proposed statistical downscaling of the monthly precipitation. The input nodes of neural networks models consist of the atmospheric meteorology and the atmospheric pressure data for 2 grid points including 127.5°E/35°N and 125°E/35°N, which produced the best results from the previous study. The output node of neural networks models consist of the monthly precipitation data for Seoul station. For the performances of the neural networks models, they are composed of training and test performances, respectively. From this research, we evaluate the impact of SVM-NNM and MLP-NNM performances for the downscaling of the monthly precipitation data. We should, therefore, construct the credible monthly precipitation data for Seoul station using statistical downscaling method. The proposed methods can be applied to future climate prediction/projection using the various climate change scenarios such as GCMs and RCMs.

Key words : NCEP, Climate Change, Downscaling, SVM-NNM

1. 서 론

최근에 시나리오의 구축과 모의/예측을 위하여 수문기상학의 범주에서 축소기법이 광범위하게 응용되는 것으로 나타나고 있다. 그 응용범위로는 지역적인 강우 (Kim et al., 2004), 저빈도 강우사상 (Wilby, 1998) 및 유출 (Amell et al., 2003) 등이 있다. GCMs의 축소기법을 위하여 제시되는 기법은 광범위하게 두 개의 범주로 구분되며, 동역학적 축소기법과 통계학적 축소기법으로 구성된다. 동역학적 축소기법에서는 RCMs이 GCMs에 포함되어 있다. RCMs의 단점은 복잡한 설계와 높은 계산 비용이다. 게다가 RCMs의 경우는 지역을 확장하는데 유연하지 않다 (Crane and Hewitson, 1998). 통계학적 축소기법은 GCMs의 대규모 특성이 강우, 기온, 그리고 유출과 같은 지역적인 규모의 변수로의 전이와 같은 경험적인 관계의 유도를 포함하고 있다.

본 연구의 목적은 신경망모형을 기본으로 하는 축소기법 (Downscaling)을 적용하여 서울기상대의 월 강우량자료의 모형화를 실시하는 것이다. 또한 신경망모형은 지지벡터기구 신경망모형 (SVM-NNM)과 다층 퍼셉트론 신경망모형 (MLP-NNM)을 각각 적용하였다. 본 연구에서 신경망모형을 이용한 축소기법을 적용시키기 위하여 입력층의 입력변수로는 미국 국립해양대기청 (National Oceanic and Atmospheric Administration, NOAA)에서 관측된 NCEP 자료를 이용하였으며, 입력변수의 구성은 선행연구 (김성원 등, 2009)를 통하여 최적의 결과치를 나타낸 127.5°E/35°N 와 125°E/35°N 격자점 및 500mb 와 850mb 기압에 따라서 대기기상변수와 기압변수로 구성하였다. 본 연구를 통하여 신경망을 이용한 월 강우량자료의 모형화를 위한 축소기법의 개발이 가능한 것으로 판단된다.

2. 신경망모형

2.1 지지벡터기구 신경망모형 (Support Vector Machine Neural Networks Model, SVM-NNM)

지지벡터기구 신경망모형 (SVM-NNM)은 다음 두 가지 개념에 의하여 운영되는 새로운 개념의 분석기법이다. 첫 번째는 고차원공간으로의 자료의 전환이며, 이것은 복잡한 결정 평면들을 가진 복잡한 문제를 선형 관별식의 함수로 사용할 수 있는 단순한 문제로 전이 할 수 있다. 두 번째는 SVM-NNM의 경우 분류에 대한 좀 더 많은 정보를 제공할 수 있기 때문에 결정 평면 근처에 있는 단지 그들 입력을 이용하고, 훈련의 개념에 의하여 동기를 가지게 된다. SVM-NNM 적용에 있어서 첫 번째의 단계는 고차원공간으로 전이하는 것이다. 일반적으로 이것은 각 자료에 가우시안을 설정하는 반경

* 정회원, 공학박사, 수자원개발기술사, 동양대학교 철도토목학과 부교수 ; E-mail : swkim68@phenix.dyu.ac.kr

** 정회원, 인하대학교 환경토목공학부 박사과정 ; E-mail : gigiata@inha.ac.kr

*** 정회원, 공학박사, 한국건설기술연구원 선임연구원 ; E-mail : hkwon@kict.re.kr

**** 정회원, 공학박사, 인하대학교 환경토목공학부 부교수 ; E-mail : sookim@inha.ac.kr

기초함수 (Radial Basis Function, RBF)를 사용하여 이루어진다. 그리하여 특성공간 (Feature Space)은 자료의 수만큼 증가하게 된다. 그러나 일반적으로 반경기초함수는 마지막 결과치를 계산하기 위하여 가우시안의 선형조합을 훈련시키기 위하여 역전파 훈련알고리즘 (BackPropagation Training Algorithm)을 이용한다. 그러나 본 연구에서 SVM-NNM은 훈련과정을 위하여 대규모의 극한 분류 (Margin Classifier)의 개념을 사용한다. 이것은 입력공간으로 부터의 분류의 용량을 분리하고 동시에 양호한 일반화를 제공한다. 이것은 분류를 위한 이상적인 조합이다. 훈련알고리즘은 반경기초함수에 확장된 아다트론 알고리즘 (Adatron Algorithm)을 기초로 하고 있다. 아다트론 알고리즘은 핵함수 (Kernel Function)에 의하여 내부공간에서 패턴의 내적을 교체함으로써 반경기초함수로서 쉽게 확장되어 질 수 있다 (Vapnik, 1992, 1995; Principe et al., 2000). 본 연구에 적용된 SVM-NNM은 ϵ -SVM 회귀모형이고, kernel 함수는 RBF 함수이다. 또한 단계간격 (Step size)은 0.01, 최대 반복횟수는 10,000회 및 임계치는 0.001로 설정하였다. 다음 그림 1은 본 연구에 적용된 SVM-NNM 구조를 나타낸 것이다.

2.2 다층퍼셉트론 신경망모형 (Multilayer Perceptron Neural Networks Model, MLP-NNM)

일반적으로 MLP-NNM은 입력층, 출력층 및 그 사이에 1개 혹은 그 이상의 은닉층으로 구성되어 있으며, 각 층에서의 노드는 그 다음 층의 노드와 상호 긴밀하게 연결되어 있다. 또한 본 연구에서 MLP-NNM을 적용하는 과정은 훈련 및 테스트과정으로 구성되어 있다. 일반적으로 훈련과정을 통하여 MLP-NNM의 최적 매개변수를 결정하며, 결정된 매개변수를 이용하여 MLP-NNM의 테스트과정을 수행한다 (Haykin 2003). 본 연구에서 MLP-NNM은 학습율 0.1, 모멘텀 0.1, 초기 연결강도 0.3, 최대 반복횟수는 10,000회 및 임계치는 0.001로 설정하였다. 또한 본 연구에서는 TurboProp 역전파 훈련알고리즘을 이용하였다. 다음 그림 2는 본 연구에 적용된 MLP-NNM 구조를 나타낸 것이다.

3. 연구자료 및 범위

3.1 연구자료

본 연구에서 서울기상대의 월 강우량 자료시계열의 모형화를 위하여 축소기법을 적용할 대기기상 및 기압변수는 NOAA의 NCEP 자료로부터 선정하였다. 그리고 서울기상대의 경도 및 위도값에 해당하는 126.58°E/37.32°N이 소속되어 있는 격자망 (Grid Box)의 4개 격자점 중에서 선행연구 (김성원 등, 2009)를 통하여 최적의 결과치를 나타낸 127.5°E/35°N 와 125°E/35°N 격자점 및 500mb 와 850mb 기압에 따라서 대기기상변수와 기압변수에 대한 자료를 구성하였다. 본 연구에서 선정된 신경망모형의 입력층변수로는 각 격자점별로 NOAA에서 관측된 NCEP 자료를 사용하였으며, 출력층 변수는 서울기상대의 월 강우량 자료시계열을 사용하였다.

3.2 연구범위

본 연구에서 서울기상대의 월 강우량 자료시계열의 모형화를 위하여 축소기법의 적용은 NCEP의 기압 500mb 및 850mb에서 격자점 127.5°E/35°N 및 125°E/35°N의 월 강우량 자료시계열의 모형화에 적용하였다. 따라서 각 기압 및 격자점에서의 다양한 입력변수의 조합을 통하여 최적의 신경망모형을 선정하고자 한다.

4. 월 강우량 자료시계열의 모형화

4.1 훈련과정 (Training Performance)

일반적으로 낮은 해상도의 대기기상변수와의 관계성을 기초로 하여 국지적인 강우를 추정 혹은 예측하기 위하여 신경망모형을 적용한 연구가 수행된 경우는 있다. 그러나 강우는 유출현상과는 다르게 강우사상의 무작위성 즉 간헐성 (Intermittency)과 강우심도의 변화성 (Variability)으로 인하여 양호한 결과치를 나타내지는 못하였다 (Olsson et al., 2004). 본 연구에서는 기압 500mb 및 850mb에서의 격자점 127.5°E/35°N 및 125°E/35°N의 대기기상 및 기압변수로서 축소기법을 이용하여 서울기상대의 1961년 1월 1일부터 1980년 12월 31일까지의 월 강우량 자료시계열을 모형화 하였다. 따라서 훈련자료의 수는 총 204개의 자료시계열로 구성되었다. 다음 그림 3(a)-(d)는 기압 500mb 및 850mb의 격자점 127.5°E/35°N 및 125°E/35°N에서 SVM-NNM과 MLP-NNM의 훈련결과와 실측 월 강우량 자료시계열을 나타낸 것이다. 그림 3(a)-(d)와 통계분석 결과에 의하면 훈련결과 SVM-NNM의 결과치가 MLP-NNM의 결과치보다 양호한 것으로 분석되었다.

4.2 테스트과정 (Test Performance)

본 연구에서 월 강우량 자료시계열의 모형화에 대한 테스트과정의 자료는 총 36개의 월 강우량 자료시계열로 구성되었다. 다음 그림 4(a)-(d)는 기압 500mb 및 850mb의 격자점 127.5°E/35°N 및 125°E/35°N에서 SVM-NNM과 MLP-NNM의 테스트결과와 실측 월 강우량 자료시계열을 나타낸 것이다. 그림 4(a)-(d)와 통계분석 결과에 의하면 테스트결과 SVM-NNM의 결과치가 MLP-NNM의 결과치보다 양호한 것으로 분석되었다. 따라서 전반적으로 훈련 및 테스트결과를 비교하면 월 강우량 자료시계열의 모형화의 경우 SVM-NNM이 MLP-NNM보다 이주 양호한 것으로 분석되었다. 다음 표 1은 기압 500mb 및 850mb의 격자점 127.5°E/35°N 및 125°E/35°N에서 SVM-NNM과 MLP-NNM의 테스트과정의 통계분석 결과를 나타내고 있다.

5. 결 론

본 연구에서는 NOAA의 NCEP 대기기상 및 기압변수자료와 서울기상대의 월 강우량 자료시계열의 상관성을 검토하기 위하여 신경망모형을 이용한 축소기법을 적용하였다. 본 연구를 위하여 SVM-NNM과 MLP-NNM을 적용하였다. 월 강우량 자료시계열의 모형화를 위하여 신경망모형의 훈련 및 테스트결과 SVM-NNM이 MLP-NNM보다 양호한 결과를 나타내고 있다. 따라서 이용 가능한 다양한 변수는 신경망모형을 이용한 강우시계열의 모형화를 위하여 반드시 필요한 것으로 판단되며, 최적 입력층변수의 선택을 위하여 입력층변수의 불확실성분석에 대한 지속적인 연구가 필요한 것으로 판단된다.

감사의 글

본 연구는 국토해양부의 건설기술혁신사업의 일환인 「기후변화 대비 물 안보 확보 방안」 과제의 연구비 지원에 의해 수행되었습니다.

참고문헌

1. 김성원, 경민수, 김병식, 김형수 (2009). 일 강우량 Downscaling을 위한 신경망모형의 적용. 2009년 한국수자원학회 학술발표회 논문집 투고.
2. Arnell, N.W., Hudson, D.A., and Jones, R.G. (2003). Climate change scenarios from a regional climate model: estimating change in runoff in southern Africa. *Journal of Geophysical Research - Atmospheres*, 108(D16) : AR 4519.
3. Crane, R.G., Yarnal, B., Barron, E.J., and Hewitson, B. (2002). Scale interactions and regional climate: examples from the Susquehanna River Basin. *Human and Ecological Risk Assessment*, Vol. 8, No. 1, pp. 147-158.
4. Haykin, S. (2003). *Neural networks : A comprehensive foundation*, 4th Indian Reprint, Pearson Education, Singapore.
5. Kim, M.K., Kang, I.S., Park, C.K., and Kim, K.M. (2004). Superensemble prediction of regional precipitation over Korea. *Journal of Climatology*, Vol. 24, No. 6, pp. 770-790.
6. Olsson, J., Uvo, C.B., Jinno, K., Kawamura, A., Nishiyama, K., Koreeda, N., Nakashima, T., and Morita, O. (2004). Neural networks for rainfall forecasting by atmospheric downscaling. *Journal of Hydrologic Engineering*, ASCE, Vol. 9, No. 1, pp.1-12.
7. Principe, J.C., Euliano, N.R., and Lefebvre, W.C. (2000). *Neural and adaptive systems; fundamentals through simulation*, John Wiley & Sons, Inc., NY.
8. Wiley, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J., and Wilks, D.S. (1998). Statistical downscaling of general circulation model output: a comparison of method. *Water Resources Research*, Vol. 34, pp. 2995-3008.
9. Vapnik, V.N. (1992). Principle of risk minimization for learning theory. *Adv. Neural Inf. Processing Syst.* Vol. 4, pp. 831-838.
10. Vapnik, V.N. (1995). *The Nature of statistical learning theory*, Springer Verlag, NY.

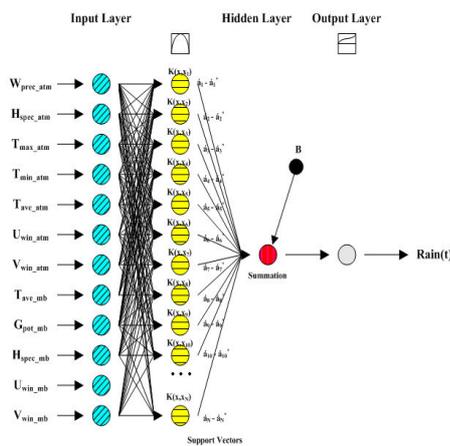


그림 1. SVM-NNM 의 구조

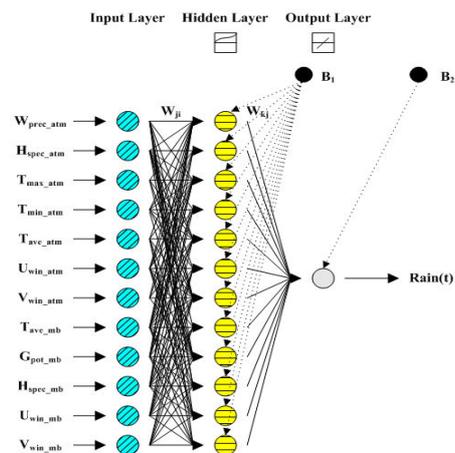


그림 2. MLP-NNM 의 구조

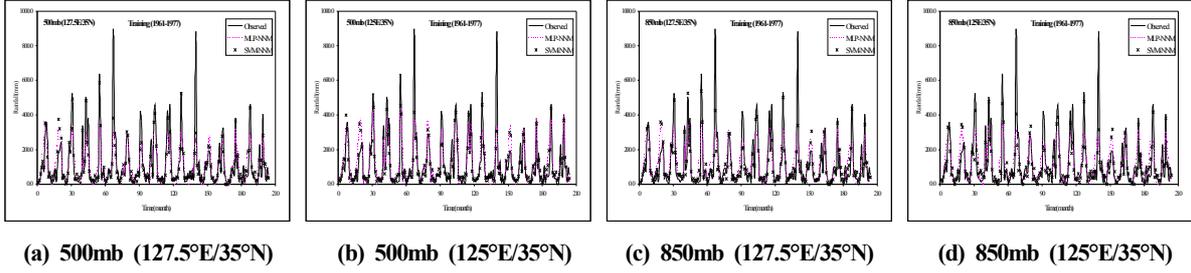


그림 3. 격자점에서 신경망모형의 훈련결과

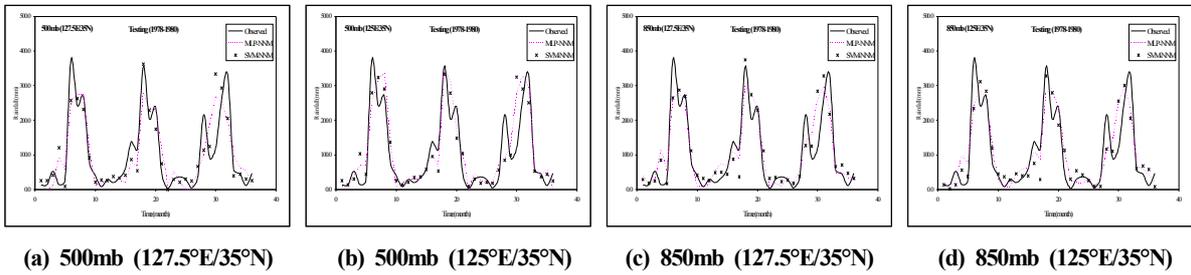


그림 4. 격자점에서 신경망모형의 테스트결과

표 1. 신경망모형의 테스트에 대한 통계분석 결과

Latitude/Longitude	Statistical Index	Testing Performance			
		500mb		850mb	
		MLP-NNM	SVM-NNM	MLP-NNM	SVM-NNM
127.5°E/35°N	CC	0.743	0.852	0.847	0.853
	RMSE	99.075	77.268	79.106	57.680
	E	0.542	0.723	0.708	0.722
	AARE	0.158	0.101	0.123	0.339
125°E/35°N	CC	0.766	0.856	0.814	0.854
	RMSE	95.465	58.325	89.278	57.031
	E	0.578	0.715	0.630	0.728
	AARE	0.150	0.187	0.163	0.401