지능형 정보검색을 위한 신경망 설계

The Optimal Design of a Neural Network
For Intelligent Information Retrieval

김 성 희 중앙대학교 문헌정보학과 강사
박 상 찬 한국과학기술원 경영과학과

Kim, Seonghee Instructor, Chung-Ang University
Park, Sang-Chan, Dept. of Management Science, KAIST

초록

이 논문은 지능형 정보검색을 위한 신경망 시스템을 구축하는데 있어서 신경망을 어떻게 디자인하는 것이 가장 이상적인지에 관해 기술한다. 구체적으로 말하면, 신경망 위상 (Network Topology) 와 학습매개변수 (Learning Parameter)들이 신경망 시스템 성능에 어떠한 영향을 미치는지에 대해 문헌조사를 통해 검토하고 있다. 그 결과 신경망 위상과 학습매개변수는 정보검색을 위한 신경망 시스템 효율성에 강하게 영향을 미치고 있으므로 신경망 설계시 이 요소들을 신중히 고려해서 결정해야 한다.

1. 서론

현재 대부분의 정보검색 시스템은 질의용어와 색인용어 사이의 일치(match) 또는 불일치(no match)에 기초를 둔 불리언 논리에 의존하고 있다. 이런 불리언정보검색의 단점은 이미 기존의 연구를 통해 잘 알려져 있다. 그 단점은 첫째,질의용어와 문헌간이 부분일치할 경우에는 많은 적합한 문헌이 검색되지않을 수 있다. 둘째, 검색된 문헌은 적합성정도에 따라 순위를 매길 수 없다. 셋째로는 질의어나 문헌에서의 상대적인 개념의 중요성을 고려하지 않는다.

그리하여 이런 단점을 극복하기위해, 부분일치기법(Partial matching technique)을 정보검색 시스템에 도입해 왔다. 예를 들면, Salton과 그의 동료(1975)는 벡터공간모델을 개발해 왔고, Croft(1986)는 확률모델을 제안해 왔다. 이들의 실험적이고 이론적인 조사는 현존하는 불리언정보검색의 단점을 보완할 수 있다는 것을 증명하여왔으나 이들 모델 역시 상징적이며 본문일치 수준에서 머무는 키워드 검색에 기초하고 있으며, 검색과정에서 의미론적이고 문맥적인 정보를 무시하고 있다. (Watter, 1989).

한편 1980년이후, 이런 단점을 해결하기 위해 학자들은 정보검색영역에 신경망(Neural Network)을 제안해 왔다. 신경망은 인간두뇌의 신경조직을 모방하여 인간과 유사한 사고와 학습을 하고자하는 연구분야이다. Mozer(1984), Belew(1986), Wilkinson and Hignston(1992)은 신경망이 정보검색 시스템에 이용될 수 있다는 것을 보여줬다. 그러나 이들 연구들은 어떻게 구체적으로 신경망이

정보검색시스템의 성능을 향상시키기 위해 디자인 해야하는지를 전혀 보여주지 않았다. 그러므로 본 논문에서는 신경망을 이용한 정보검색 시스템을 구현하기위해 어떤 요소들이 고려되어져야 하는지를 살펴보고자 한다.

2. 지능형 정보검색을 위한 신경망설계

신경망은 인간의 뇌 그리고 신경세포가 반응하는 것과 유사하게 설계된 회로이다. 이는 많은 수의 소자를 네트웍으로 연결하고, 각 소자들 사이의 연결의 세기 (connection strength)로 정보를 표현하고 기억한다. 이는 인간두뇌와 같이 비결정적인 특성을 가지고 있으므로 약간 틀리거나 비슷한 입력을 인식할 수 있다. 음성인식, 문자인식, 영상처리, 자연어 이해등의 분야에 주로 이용되고 있다. 가장 대표적인 신경망으로는 역전파 (Back-propagation) 신경망이다. 이것은 입력층 (Input layer), 출력층 (outpu layer) 그리고 이들 사이에 1개 이상의 은닉층 (Hidden layer)으로 구성되어 있다 (그림1).

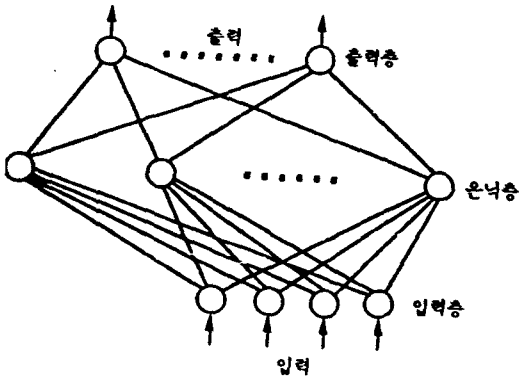

그림 1 역전파 신경망

역전파 신경망은 그동안 가장 광범위하게 사용해 왔고 또 정보검색시스템에 있어서도 가장많이 사용해온 신경망 구조이다. 예를 들면 Mital and Gedeon (1991), Gersho and Reiter (1990),그리고 Wilkinson and Hingston (1992) 이 모두 정보검색을 위해 역전파신경망 구조를 사용했다. 그러므로 다음 부분은 주로 역전파 신경망 구조가 정보검색에 응용될 때 고려해야 할 사항을 중심으로 기술 되고 있다.

2.1 네트웍 표현 (Network representation)

일반적으로 네트웍 표현은 층 사이즈 (Layer size)를 기준으로 해서 기술 되는데, 정보검색은 문헌, 색인용어, 질의용어와 함께 다루어진다. 즉, 정보검색 시스템은 일반적으로 질의용어와 적합한 문헌을 색인용어를 통해서 매치 (match)시킨다. 그러므로 검색도메인의 한쪽은 질의용어이고 다른 쪽은 문헌이며 그중간에는 색인용어가 있다. 그러므로 신경망용어로 말하면 이런 정보검색 시스템은 3개층으로 구성된 네트웍이다. (그림 2)

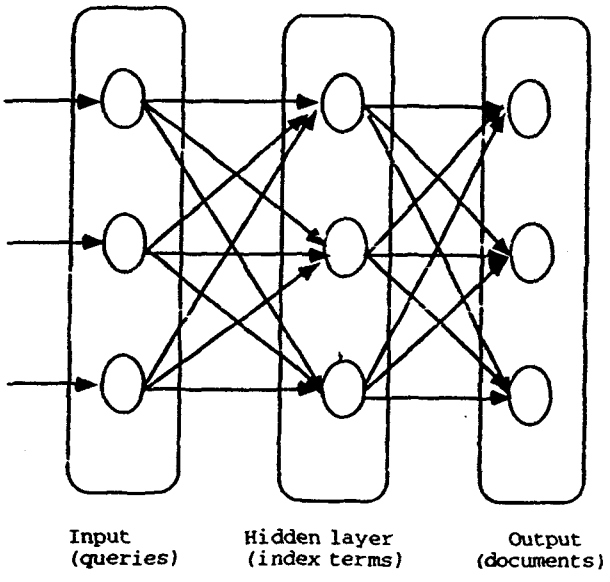

그림 2 Three layer network

이런 3층 (three layer)으로된 신경망을 구성한 연구자로는 Kwok(1989; ,1990) 과 Wilkinson and Hingston (1991; 1992)를 들 수 있다.

또한 3층 네트웍과는 달리 Mital and Gedeon

(1991)은 정보검색을 위해 2층 (two layer)신경망을 사용하였으며 Mori (1990)는 7층(seven layer) 신경망조직을 사용하였다. 이 7층 신경망에서 첫째층은 키워드층이라 했고, 마지막층은 문헌층이라 불렀으며, 나머지 5층은 은닉층 (Hidden layer)이라고 하였다. 이들은 왜 7층을 사용했는 지에 대해서는 구체적으로 설명하지 않았다. 이 결과들이 시사하는 것은 결국 정보검색을 위해 신경망조직을 디자인하는데 이상적인 층의 수는 없다는 것이다. 이러한 신경망에 있어서 가장효율적인 층의 수를 찾기위해 앞으로 더 많은 연구가 필요하다.

2.2. 노드의 수(the number of node)

네트웍표현에서 층의 사이즈가 결정된 다음에는 각층의 노드수가 결정되어야 한다. 각층에 노드는 한단어 또는 문헌으로 표시된다. 검색효율성은 노드의 수에 달려 있지만, 누구도 각층의 이상적인 노드의 수를 결정하지 못했다. Belew (1984)는 대략 5,000개의 노드로 형성되는 1600개의 문헌으로부터 구성된 네트웍을 사용했다. Mital과 Gedeon(1991)은 은닉층 없이 입력과 출력층으로만 구성된 네트웍을 사용했는데 여기서 569 노드수를 포함하는 19개 문헌만을 사용했다. Wilkinson과 Hingston (1992)는 9,000 노드를 사용했다. 이들 연구들에서는 입력, 출력 그리고 은닉층에 있어서 구체적인 노드수는 언급하지 않았다. 그러나, Cherkassky and Vassilas (1989)는 재현률의 질(quality)은 네트웍의 은닉노드수에 의해 강하게 영향을 받는다고 하였다. 이들의 연구결과는 이전의 역전파 연구 (Burr, 1986)의 결과와 다음과 같은 면에서 일치한다.

- 너무작은 은닉노드수는 현재 데이터의 본질적인 특성을 파악할 수 없다.
- 너무많은 노드수로 설계된 신경망은 입력 데이터의 작은 변화에 대해 과민하게 반응함으로써 불필요한 분류를 시도한다.

이들 연구결과가 시사하는 것은 구조화된 접근이 이상적인 노드 수를 결정하기위해서는 새로운 방법이 개발될 필요가 있다는 것이다.

2.3 학습 또는 훈련 매개 변수 (Learning or Training Parameter)

이상의 네트웍 위상 (network topology)이 결정이 되고나면 이 시스템이 성공적으로 구현되는 것을 돕기위해 두개의 매개변수 즉 러닝 레이트 (Learing rate) 와 모멘텀 (momentum) 이 사용된다. 이 두개의 상수는 시스템의 훈련시간(training times)을 단축시켜주며 시스템이 성공적으로 수렴할 수 있도록 도와준다. 러닝 레이트는 가중치변화를 계산하기위해 사용된 비례상수이고 모멘텀은 현가중치 세트 총체 (aggregation)에 대한 비례상수이다. 이는 일반적으로 가중치에 더해진다. 이상적인 러닝 레이트와 모멘텀은 각 데이터의 특성에 따라 다르고 따라서 미리 정해질 수 없기 때문에 이들 매개변수의 선택은 경험적으로 결정되어야 한다 (Cherkassky and Vassilas, 1989). 예를 들면, 보통 러닝 레이트와 모멘텀의 값은 0.01 에서 0.9 사이에 존재한다. Wilkinson and Hingston (1992)에 따르면 정보검색을 위한 효율성결과에 모멘텀과 러닝 레이트가 아주 강하게 영향을 미친다. 다시말하면 만일 러닝 레이트와 모멘텀이 제대로 선택되지 않는다면 그 시스템은 성공적으로 구현될 수 없거나 또는 훈련시간이 매우 길어질 가능성이 높아서 그 효율성은 아주 떨어진다는 것이다. 연구결과를 좀더 살펴보면, ' Wilkinson and Hingston (1992)은 러닝 레이트와 모멘텀은 정보검색 효율성에 영향을 미친다고 주장했는데 그 연구에서 그들은 모멘텀을 0.5 그리고 러닝 레이트를 0.05 사용 하였다. Gersho and Reiter (1990)은 역전파 학습 알고리즘을 이용한 다층 신경망 네트웍을 정보검색 시스템에 테스트하고 구현했는데 여기서는 러닝 레이트를 0.6 그리고 모멘텀을 0.9를 사용하였다. 이들 결과들은 주어진 데이터에 대해 신경망을 구성할 경우 학습매개변수는 그 시스템 효율성에 중요한 영향을 미치지만 그 구체적인 값을 이끌어낼 공식적인 방법이 없다는 것을 암시하고 있다. 그러므로 정보검색 시스템에 있어서 신

경망을 설계할 때에는 그 상황에 맞게 학습매개변수를 신중히 선택해야 한다는 것이다.

3. 결론

이상의 내용을 요약하면 다음과 같다.

지능정보검색을 위해 신경망을 설계하는데 있어서 가장 많이 쓰이는 네트웍은 역전파 (back-propagation) 신경망이다.

신경망 위상 (Neural Network Topology)는 그 시스팀성능에 강하게 영향을 미치므로 신경망 시스팀을 성공적으로 구현하고 최적의 성능을 유지하기 위해 층의 사이즈 (Layer size) 와 각 층의 노드수 (The number of node)는 신중히 고려해서 결정해야 한다.

또한 학습시간을 최소화하고 네트웍을 성공적으로 훈련시키기 위해서는 학습 매개변수도 이상의 여러 요인들과 더불어 최적의 값을 선택할 수 있도록 시스팀 설계시 고려해야 한다.

참고문헌

Belew,R.K.(1986).Adaptive Information Retrieval: Machine Learning In Associative Networks, Ph.D. Dissertation, University of Michigan, Ann Arbor.

Belew, R.K. (1987). A connectionist approach to conceptual information retrieval, Proc. First International Conference on Artificial Intelligence and Law, ACM, PP.116-125.

Cherkassky, V. and Vassilas, N. (1989). Performance of Back Propagation Networks for Associative Database Retrieval, Proc. International Joint Conference on Neural Networks, Volume 1.

Croft, W.B. (1986). "Boolean queries and Term dependencies in probabilistid retrieval models," Journal of the Americal Society for Information Science, Vol.37 (2), PP.71-77 „

Kwok, K.L. (1989). A Neural Network for Probabilistic Information Retrieval, 12th International Conference on Research and Development in Information Retrieval, Cambridge, Massachusetts, PP.21-30.

Kwok, K.L. (1990). Application of Neural Network to Information Retrieval,international Joint Conference on Neural Networks: Volume II;PP.623-626.

Rumelhart, D.E. and McClelland, J.L.(1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol.I:Foundations, Bradford,Cambridge, MA.

Mori, H., Cheng Long Chung, Yousuke Kinoe, and Yoshio Hayashi(1990). An Adaptive Document Retrieval System Using Neural Network, International Journal of Human-Computer interaction, 1990, Vol.2(3), PP.267-280.

Mozer, M. (1984). Inductive Information Retrieval Using Parallel Distributed Computation, Technical Report, ICS, UCSD, La Jolla, CA.

Rumelhart, D.E.; Hinton, G.E., and Williams, R.J. (1985). Learning Internal Representations by Error Propagation, Technical Report, ICS, UCSD, La Jolla, CA.New Tool For Predicting Thrift Failures, Decision Science, Vol.23, PP.899-916.

Salton, G., Fox, E.A., and Wu, H. (1983). Extended Boolean Information Retrieval, Communications of the ACM, vol.26(11), PP.1022-1036.

Salton, G., Wong, A., and Yang, C.S. (1975). A vector space model for automatic indexing., Communications of the ACM, 18, PP.613-620.

Watters, C.R. (1989). Logic framework for information retrieval, Journal of The American Society for Information Science, Vol. 40, PP. 311-324.

Wilkinson, R. and Hingston, P. (1992). Incorporating The Vector Space Model in a Neural Network used for Document Retrieval, Library HI Tech, Vol.10, PP.69-75.

Wilkinson, R. and Hingston P. (1991). Using the Cosine Measure in a Neural Network for Document Retrieval, 14th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval, 1991, PP.202-210.