

신경망의 민감도 분석을 이용한 귀납적 학습 기법의 변수 부분집합 선정

강부식

목원대학교 경영정보학과
(bookang@mokwon.ac.kr)

박상찬

한국과학기술원 산업공학과
(sangpark@kaist.ac.kr)

데이터로부터 학습하여 룰을 추출하는 귀납적 학습 기법은 데이터 마이닝의 주요 도구 중 하나이다. 귀납적 학습 기법은 불필요한 변수나 잡음이 섞인 변수를 포함하여 학습하는 경우 생성된 룰의 예측 성능이 떨어지고 불필요하게 룰이 복잡하게 구성될 수 있다. 따라서 귀납적 학습 기법의 예측력을 높이고 룰의 구성도 간단하게 할 수 있는 주요 변수 부분집합을 선정하는 방안이 필요하다. 귀납적 학습에서 예측력을 높이기 위해 많이 사용되는 부분집합 선정을 위한 포장 기법은 최적의 부분집합을 찾기 위해 전체 부분집합을 탐색한다. 이때 전체 변수의 수가 많아지면 부분집합의 탐색 공간이 너무 커져서 탐색하기 어려운 문제가 된다. 본 연구에서는 포장 기법에 신경망 민감도 분석을 결합한 귀납적 학습 기법의 변수 부분집합 선정 방안을 제시한다. 먼저, 신경망의 민감도 분석 기법을 이용하여 전체 변수를 중요도 순으로 순서화 한다. 다음에 순서화된 정보를 이용하여 귀납적 학습 기법의 예측력을 높일 수 있는 부분집합을 찾아 나간다. 제안된 방법을 세 데이터 셋에 적용한 결과 일정한 반복 회수 이내에 예측력이 향상된 부분집합을 얻을 수 있음을 볼 수 있다.

1. 서론

정보 기술의 활용에 따라 많은 양의 데이터가 컴퓨터 시스템 내에 축적이 되었고, 이런 데이터의 효과적 저장, 처리 및 활용이 더욱 중요해지고 있다. 데이터 마이닝은 데이터 내의 숨겨진 지식을 찾아 내기 위한 것으로 통계적 기법이나 신경망, 귀납적 학습 같은 기계학습(machine learning) 기법이 많이 사용되고 있다. C4.5(Quinlan, 1993)와 같은 귀납적 학습 기법은 데이터가 수치 값 혹은 명목 값이든 관계없이 처리가 가능하고 또한 결측 치도 처리가 가능하며

생성된 의사결정나무의 일반화 기능 등을 가지고 있을 뿐더러 룰(rule)도 생성할 수 있어 데이터로부터 룰을 추출하는 많은 영역의 판별 문제에 적용이 되고 있다. 이처럼 실무적으로 많이 사용되고 있는 기계학습의 대표적 기법인 C4.5나 CART(Breiman et al., 1984)를 적용할 경우, 사용되는 기계학습 데이터의 특성에 따라 생성된 룰의 예측력은 변하게 된다. 실제 세계에서 전체 데이터를 구성하는 변수 중에서 클래스(종속) 변수의 판별에 영향을 주는 주요 변수들은 일부인 경우가 많으며 불필요한 변수가 포함되어 룰이 생성될 경우 에러율이 증가하고 또한 룰

이 불필요하게 복잡해 지는 것으로 알려지고 있다(Liu and Motoda, 1998). 따라서 불필요한 혹은 잡음이 있는 변수를 제거하고 기계학습의 예측력을 높일 수 있는 주요 변수집합을 추출하기 위한 방안이 필요하다.

주요 변수집합을 찾아내기 위한 방안 중의 하나가 변수 부분집합 선정 기법으로 전체 변수의 부분집합을 모두 고려하여 최적의 예측율을 나타내는 부분집합을 선정하는 것이다. 그러나 고려해야 할 변수의 수가 많아지면 부분집합의 탐색 공간이 너무 커져서 풀기 어려운 문제가 된다. 본 연구에서는 귀납적 학습의 성능을 향상시키는 변수 부분집합을 선정하는 데 있어 일정한 반복 회수 이내에 선정할 수 있도록 하기 위한 휴리스틱 방법을 제안한다. 이 방안에서는 탐색 공간을 줄이기 위해 변수들 중에서 탐색할 순서를 미리 결정하고 이 순서에 따라 부분집합을 탐색하는 방안을 제시한다. 변수들의 탐색 순서는 클래스를 판별하는 변수의 영향도에 따라 순서를 정하는 데, 판별 문제에 효과적인 기법 중의 하나인 신경망을 이용하여 변수들의 중요도를 결정하게 된다.

다음 2절에서는 기계학습 방법의 예측율을 높이기 위한 부분집합 선정 방안에 대한 이론적 배경에 대해 살펴보고, 3절에서는 신경망을 이용한 주요 변수 선택 기법에 대해 살펴본다. 4절에서는 신경망과 결합된 귀납적 학습의 성능 향상을 위한 부분집합 선정 방안에 대해 제안하고, 5절에서는 제안된 방안을 캘리포니아 대학의 기계학습 레파지토리 데이터 셋(Merz and Murphy, 1996)에 적용한 실험 결과에 대해 서술한다. 마지막으로 6절에서는 연구의 결론 및 향후 연구 사항에 대해 살펴보도록 한다.

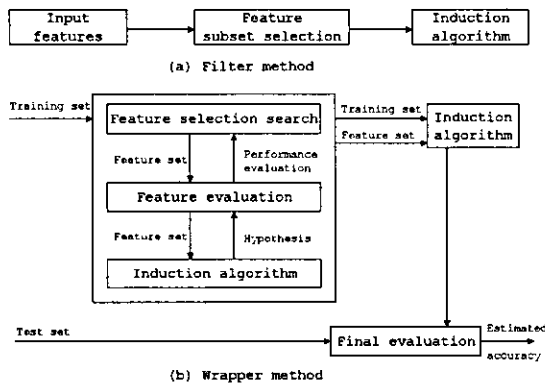
2. 기계학습 기법의 부분집합 선정 방법

기계학습 기법의 예측력을 높이기 위한 노력 중 하나는 불필요하거나 잡음이 섞인 변수를 제거함으로써, 즉 판별하는데 있어 충분한 최소한의 변수로 구성된 주요 변수집합을 추출함으로써 학습된 의사결정나무 혹은 신경망의 일반화 성능을 높이고 생성되는 룰이 간단해 지도록 하는 것이다. 주요 변수를 추출하기 위해 주로 사용하는 방법에는 변수 생성(feature construction), 변수 추출(feature extraction) 및 부분집합 선정(subset selection) 기법이 있다(Liu and Motoda, 1998).

- 변수 생성: 이 기법은 주어진 데이터의 양 혹은 정보가 부족한 경우 추가로 새로운 변수를 생성하여 추가하는 방법이다(Wnek and Michalski, 1994). 전체 변수의 수에 생성된 변수를 추가함으로써 탐색 공간은 더 커지게 된다.
- 변수 추출: 이 기법은 함수의 맵핑 방법을 통해 원래의 데이터로부터 새로운 변수를 추출하는 방법이다(Wyse et al., 1980). 원래 변수의 수 보다 새로 생성된 변수의 수는 적으며 탐색 공간은 그만큼 작아지게 된다.
- 부분집합 선정: 변수의 변형은 일어나지 않으며 전체 변수로부터 종속 변수에 영향을 주는 주요 변수만을 뽑는 방법이다(Kohavi and John, 1998; Blum and Langley, 1997; Dash and Liu, 1997; Wang et al., 1998). 따라서 탐색 공간은 그 만큼 줄어들게 된다.

세 가지 방법 중에서 이 논문에서는 변수 부분집합 선정 기법에 대해 관심이 있으며 많은 응용 분야에서 판별 예측에 견고한 특성을 나타내는

C4.5의 예측력을 높이기 위한 부분집합 선정 기법에 대해 살펴본다. 부분집합 선정 알고리즘은 크게 2가지로 나누어 지는데 기계학습 기법의 적용 이전의 전처리 작업으로 주요 부분집합을 선정하느냐 아니면 기계학습 기법과 결합된 형태에서 주요 변수 부분집합을 선정하느냐에 따라 <그림 1>과 같이 필터(filter) 기법과 포장(wrapper) 기법으로 구분한다(Kohavi and John, 1998).



<그림 1> 주요 변수 부분집합 선정을 위한 필터 기법과 포장 기법

필터 기법은 데이터의 전처리 작업을 통해 주요 변수를 선택하고 이 선정된 변수들을 이용하여 기계학습 기법에 적용하는 것이다. 일반적으로 필터 기법이 계산의 효율성이 높으나 사용하려는 기계학습 기법과 독립적으로 주요 변수를 선택함으로써 실제로 해당 기계학습 기법을 적용 시 성능은 포장 기법보다 떨어지게 된다(Kohavi and John, 1998). FOCUS(Almuallim and Dietterich, 1994)와 Relief(Kira and Rendell, 1992) 같은 알고리즘이 있다. FOCUS 알고리즘은 학습 데이터에서 가설(hypothesis)을 세우는데 충분한 가장 작은 변수의 집합을 찾기 위해

너비우선탐색(breadth-first search)을 한다. 소모적(exhaustive) 탐색으로 부분집합을 찾아 나감으로 변수의 수가 많아지는 경우 계산 시간이 많이 걸리게 된다. Relief는 목표 클래스를 판별하는데 필요한 변수들의 적정도를 각 변수에 부여한다. 일정 적정도를 갖는 모든 변수를 찾아 나가는 데, 클래스 판별을 위한 변수들이 서로 간에 높은 상관 관계를 갖는 경우 상관 관계가 높은 변수들 모두를 선정함으로써 꼭 필요한 변수만을 선정하지는 못하고 있다(Kohavi and John, 1998). 포장 기법은 주요 변수의 선정 시에 적용하려는 기계학습 기법을 결합시켜 주요 변수를 선정함으로써 기계학습 기법의 예측력을 더욱 높이기 위한 것이다. 포장 기법에서는 전체 변수들의 모든 부분집합에 대해 탐색하여 예측율을 가장 좋게 하는 최적 부분집합을 찾는 것이다.

포장 기법은 주요 변수의 선택에서부터 해당 기계학습 기법과 결합하여 실시함으로써 최종 생성된 룰의 성능은 높일 수 있으나 주요 변수를 선정하기 위해 모든 변수들의 부분집합을 탐색하는 경우 계산 시간은 필터 기법에 비해 훨씬 증가하게 된다. 불행히도 포장 기법은 전체 변수의 모든 부분집합을 탐색하는 경우 복잡도가 $O(2^n)$ 이 되기 때문에 고려해야 할 변수의 수가 많은 경우 풀기 어려운(NP-complete) 알고리즘이 된다. 실제로는 최적의 부분집합을 미리 알지 못하므로 미리 정해 둔 종료 조건을 만족한 경우 탐색을 중지하며, 탐색된 부분집합에서 예측율이 가장 좋은 부분집합을 선정하게 된다.

기계학습 기법에서 예측율이 높은 부분집합을 선정하기 위한 연구를 살펴보면, Yang and Honavar(1998)는 신경망의 예측율을 높이는 부분집합 선정을 위하여 유전자 알고리즘을 이용하여 탐색함으로써 최적 부분집합을 찾는 방안을

제시하였다. Wang and Sundaresh(1998)는 불필요한 변수를 제거하여 최적의 부분집합을 선정하는 대신 데이터 셋의 불필요한 인스턴스를 제거하여 의사결정나무 기법 등의 예측력을 높이는 방법을 제안하였다. Kohavi and John(1998)은 변수의 적정성(relevance)을 도입하여 포장 기법에서 부분집합을 찾기 위한 탐색 방법을 제안했다. 전체 변수집합에서 각 변수를 뺀 경우의 클래스를 판별하는 수행 성능의 차이에 따라 적정 혹은 비적정 변수로 나누고, 적정 변수만을 가지고 최선우선(best-first) 탐색(Ginsberg, 1993)을 하는 방법이다. 변수의 적정도를 파악하는 과정에서만 변수의 수만큼 반복된 계산이 필요하고, 다음 최선우선탐색에서 원하는 부분집합을 선정할 때까지 반복된 계산이 필요하다. Wang et al.(1998)은 조건적 적절도(conditional relevance)에 근거한 휴리스틱 방법을 제안하여 C4.5의 적용 시 예측율이 높아지면서 사용되는 변수의 수는 주는 것을 실험을 통해 밝혔지만 이 방법론의 복잡도 역시 $O(N^2)$ 으로 N 이 커지게 되면 복잡도가 크게 증가하게 된다. 따라서 변수의 수가 많은 문제에서 일정 시간 안에 받아들일 만한 예측율을 가진 부분집합을 찾는 것이 필요하게 된다.

변수의 부분집합 선택을 위해서 다양한 통계적 기법이나 신경망 기법, 그리고 휴리스틱 기법 등이 적용될 수 있다(Weiss et al., 1991). 신경망은 분석 대상 데이터의 사전 정보를 알지 못해도 신경망의 학습 기능을 이용하여 데이터 처리를 할 수 있다는 장점 때문에 여러 분야에서 판별 및 예측을 위한 기법으로 많이 사용되고 있다. 신경망도 학습된 신경망의 일반화 오류를 줄이기 위해 주요 변수 선정이 중요한 이슈 중의 하나로 이는 3절에서 자세하게 다루어진다. 따라서 학습 기능을 갖는 신경망에서의 주요 변수 선정 방안

은 귀납적 학습을 위한 변수 부분집합 선정에 도움이 될 수 있을 것이다. Setiono and Liu(1998)는 적정한 신경망 네트워크 구조를 찾는 기법을 통해 찾아진 신경망의 주요 변수를 C4.5에 적용하면 원래의 전체 변수를 가지고 생성한 의사결정나무에 비해 간단한 의사결정나무를 구성할 수 있음을 필터 기법으로 보여 주었다. 본 연구에서는 C4.5의 예측력을 높이는 부분집합의 선정을 위한 포장 기법에서 포장 기법 사용 시의 탐색 공간을 줄이기 위한 방안으로 신경망의 민감도 분석을 이용하는 방법에 대해 살펴보고자 한다.

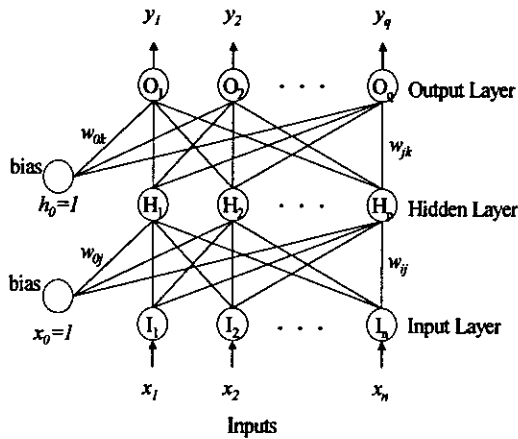
3. 신경망을 이용한 주요 영향 변수 선택 기법

신경망은 많은 데이터를 처리할 수 있고, 데이터간의 복잡한 비선형 관계에 대해서도 학습할 수 있는 능력 때문에 많은 분야에서 적용되고 있다. 신경망 중에서 역전파 알고리즘(Rumelhart et al., 1986)은 감독 학습을 하는 신경망에서 일반적으로 가장 많이 사용되는 학습 알고리즘이다. <그림 2>는 하나의 은닉 층을 가진 신경망의 한 예를 나타내고 있다.

<그림 2>에서 신경망의 출력 노드 k 의 출력 값 y_k 는 식(1)과 같다.

$$y_k = f \left(\sum_{j=1}^p f \left(\sum_{i=1}^n x_i w_{ij} + w_{0j} \right) w_{jk} + w_{0k} \right) \quad (1)$$

신경망을 이용한 주요 변수 선택 방법은 주로 신경망의 적정한 네트워크 구조를 찾기 위한 목적으로 제시되고 있다. 큰 네트워크를 가진 신경망이 학습은 잘되지만 학습 데이터에 맞춰 과도하게



<그림 2> 하나의 은닉 층을 가진 신경망의 구조
(I: 입력 노드, H: 은닉 노드, O: 출력 노드)

학습됨으로써, 일반화를 할 경우 오류가 커지게 됨으로 적절한 혹은 가능한 적은 네트워크로 구성된 신경망을 구성하는 것이 바람직하다 (Ponnappalli et al., 1999; Karnin, 1990; Reed, 1993; Engelbrecht, 2001). 적절한 신경망을 찾기 위한 연구에는 벌칙 항목(penalty term)을 목적 함수에 추가하여 학습하는 방법, 신경망 네트워크 건설(network construction) 및 네트워크 가지치기(network pruning) 접근 방법 등이 있다. 신경망의 목적함수에 벌칙 항목(penalty term)을 추가하여 학습하는 조정(regularization) 접근 방법은 목적함수를 $\epsilon = \epsilon T + \lambda \epsilon C$ 로 나타내는 데, ϵT 는 신경망의 출력 값과 목표 값과의 차이를 나타내며 신경망의 원래 목적 함수와 관련되어 있고, ϵC 는 벌칙 항목을 나타낸다(Girosi et al., 1995; Reed, 1993). λ 는 벌칙 항목의 영향도로 적절한 값을 결정하여 사용한다. 일반적으로 구현하기 쉬운 장점이 있으나, 적절한 λ 값을 결정해야 하는 문제가 있으며 학습 시간을 증가시키고 지역 최적화로 빠지는 경향이 있다. 네트워크 건설 기법은 은

닉 노드의 수가 1인 작은 네트워크를 가지고 학습을 시작하여 미리 정한 신경망의 예측에 도달할 때까지 점차 은닉 노드의 수를 증가시켜 가는 방법이다(Hirose et al., 1991; Zhang and Morris, 1997; Setiono and Liu, 1998). 이 기법은 언제 새로운 노드를 추가하고 네트워크 건설을 언제 멈출 것이며 새로운 노드의 부가 시 어느 부분에 연결할 것인지 등과 관련된 효과적 기준의 설정이 필수적이다. 기준에 따라 과도하게 학습되는 문제점과 학습 시간이 길어지는 문제 등이 나타날 수 있다. 네트워크 가지치기 방법은 큰 구조의 신경망으로 학습을 시작하여 점차 불필요한 파라미터를 가지치기 해 나간다. 제거할 파라미터의 대상에는 신경망의 가중치, 은닉 노드, 입력 노드가 포함된다. 각 파라미터에 대해 적정성을 평가하여 불필요한 파라미터를 제거함에 따라 큰 구조를 가진 신경망이 점차 적절한 구조를 가진 신경망으로 수렴된다. 네트워크 가지치기 알고리즘을 위한 여러 연구가 제시되고 있는데 많이 연구되고 있는 기법 중의 하나가 신경망의 민감도를 이용하는 방법이다(Engelbrecht, 2001). 신경망의 민감도 계산을 이용하여 신경망의 적정 구조를 찾아가는 방법에는 노드의 적정성(relevance)을 계산하여 가장 중요하지 않은 노드를 학습 과정에서 제거(Mozer and Smolensky, 1989)하거나 신경망의 가중치 민감도(sensitivity)를 계산하여 민감도가 가장 작은 가중치를 제거(Karnin, 1990) 혹은 신경망의 가중치에 대한 에러의 이차 미분 값을 추정하여 가중치의 돌출도(saliency)를 계산하여 가중치를 제거(Le Cun et al., 1990)하는 방법 등이 있다(Reed, 1993).

민감도 방법 중 가장 많이 언급되고 있는 방법 중의 하나가 Karnin(1990)의 방법이다. Karnin은 신경망의 노드들을 연결하고 있는 가중치를 제거

할 경우의 에러 함수의 민감도를 계산한다. 신경망을 학습한 후에 각 가중치를 제거할 경우의 에러 함수의 값과 제거하지 않을 경우의 에러 함수 값의 차이를 이용하여 각 가중치의 민감도를 계산한다. 일정 기준 이하의 민감도를 가진 가중치를 제거하고 학습을 다시 하여 이전 신경망의 성능과 비교하여 일정 기준을 만족하면 계속 가중치를 제거해 나가는 작업을 반복한다. Engelbrecht(2001)는 신경망의 에러 함수 민감도와 신경망 출력 값의 민감도는 개념상 같음을 보였다. 따라서 복잡한 신경망의 에러 함수(목적 함수) 대신 목적 값을 이용한 민감도 분석이 효율적임을 제시했다. 또한 에러 함수의 민감도는 에러 함수와 신경망의 최적화 알고리즘에 의존적이지만 출력 값의 민감도는 에러 함수와 최적화 알고리즘에 독립적으로 신경망의 어느 알고리즘이나 적용 가능함을 보였다.

본 연구에서는 Karnin의 개념을 이용하여 주요 변수의 영향도를 결정하고자 한다. 다만 에러 함수의 민감도 대신 출력 값의 민감도를 사용한다. 그리고 우리의 관심은 입력 변수들의 중요도에 따라 순서화 하는 데 있기 때문에 신경망의 가중치를 제거하는 것은 하지 않고 학습된 신경망에서 입력 노드를 제거한 경우와 그렇지 않은 경우의 민감도만을 계산하도록 한다. 먼저 신경망을 학습하기 전에 입력 변수의 값은 동일한 범위를 갖도록 조정한다(식(2) 참조). 이 연구에서는 입력 변수의 범위는 [0,1]로, 목표 값의 범위는 [0.1, 0.9]로 정규화 한다. 조정된 값들은 단위가 다름에 따라 나타날 수 있는 탐색 공간에서의 규모의 문제를 해결함으로써 학습 시 빠르게 수렴시킬 수 있고 또한 입력 값의 분산에 따른 영향도 최소화시킬 수 있다.

$$F(i) = (NewRangeHI(i) - NewRangeLO(i)) / (Max(i) - Min(i))$$

$$Scaled = F(i) * original + NewRangeLO(i) - F(i) * Min(i) \quad (2)$$

신경망의 학습 후에 입력 변수의 민감도를 계산한다. 입력 변수 x_i 의 민감도는 Karnin(1990)이 제안한 개념에 따라 다음과 같이 계산된다.

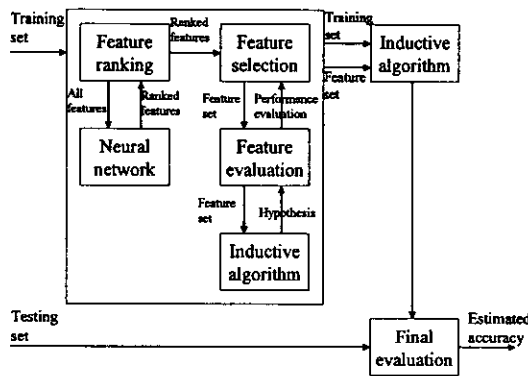
$$S_i = \frac{1}{m} \frac{1}{q} \sum_L \sum_{k=1}^q \left(\frac{|O_k^0 - O_k^i|}{O_k^0} \right) \quad (3)$$

(3)식에서 O_k^0 는 각 학습 데이터의 신경망의 출력 값을, O_k^i 는 입력 변수 x_i 의 값이 제거된 경우의 신경망의 출력 값이다. L은 학습 데이터를 나타내고 m은 학습 데이터의 수를, q는 출력 노드의 수를 나타낸다. 이 연구에서 입력 변수 x_i 의 제거는 x_i 의 값을 0으로 하여 계산한다. 즉 각 학습 데이터에 원래의 입력 변수 x_i 의 값이 있을 경우와 $x_i=0$ 일 경우의 신경망의 출력 값의 차를 이용하여 입력 변수 x_i 의 민감도를 계산한다. 민감도가 클수록 신경망에 대한 영향도가 크다고 할 수 있다.

4. 포장 기법에서의 신경망을 이용한 변수 부분집합 선택

포장 기법에서의 전체 부분집합에 대해 탐색하는 경우 변수의 수가 늘어나게 되면 풀기 어려운 문제가 된다. 따라서 탐색 공간을 줄이면서 효과적으로 탐색하기 위한 방법이 필요하다. 이 절에서는 신경망의 민감도 분석을 이용하여 각

변수들의 중요도를 찾고 중요도의 순서에 따라 탐색함으로써 일정한 반복 회수 이내에 탐색이 끝나도록 하는 휴리스틱 방법(<그림 3> 참조)에 대해 제안한다.



<그림 3> 신경망을 이용한 포장 기법의 구성

포장 기법의 적용 시 변수 부분집합을 선정하기 위한 귀납적 학습 방법으로 C4.5가 많이 사용 되는 데(Wang and Sundaresh, 1998; Kohavi and John, 1998; Wang et al., 1998; Setiono and Liu, 1998), 그 이유 중의 하나는 C4.5가 실제 데이터 베이스에 좋은 성능을 보이는 알고리즘으로 일반적으로 변수 부분집합 선정으로 성능을 개선하는 것이 어렵다고 기대되기 때문이다(Kohavi and John, 1998). 따라서 제안된 포장 기법에서의 기계학습 기법도 C4.5를 사용하여 변수 부분집합을 선정하기로 한다. 또한 입력 변수의 중요도를 이용하여 변수 부분집합을 선택하기 위해 다음 2가지의 선택 조건을 채택한다.

- 선택 조건1: 기준 에러율 θ 가 주어졌을 때, 변수 부분집합의 에러율 ϵ_m 가 θ 보다 적거나 같은 부분집합 중에서 가장 작은 에러율 ϵ_m 를 가진

부분집합을 선택한다.

- 선택 조건2: 두개 이상의 부분집합이 선택 조건1을 만족했을 때, 부분집합에 속한 변수의 수가 가장 적은 것을 선택한다.

위 선택 조건에서는 전체 변수를 가지고 C4.5를 실행했을 경우 생성된 일반화 의사결정나무(pruned decision tree)의 시험 데이터에 대한 에러율로 설정한다. 선택 조건 2에 의하면 물을 구성하는 변수의 수를 줄임으로써 물을 간단하게 할 수 있게 된다. 제안된 휴리스틱 포장 기법의 알고리즘을 정리하면 다음과 같다.

- Step 1: 모든 입력 변수에 대해 식(2)를 이용하여 변수 값을 구간 $[0,1]$ 로 조정한다. 클래스 변수의 값은 구간 $[0.1, 0.9]$ 로 조정한다.
- Step 2: 신경망을 이용하여 학습한다.
- Step 3: 학습된 신경망에서 식(3)을 이용하여 입력 변수들의 민감도를 계산하고, 민감도가 큰 순서로 v_1, v_2, \dots, v_n 으로 순서화한다.
- Step 4: 기준 에러율 θ 를 정한다. 여기에서는 전체 입력 변수를 가지고 C4.5를 실행했을 경우에 생성되는 일반화 의사결정나무의 시험 데이터에 대한 에러율로 정한다.
- Step 5: 초기 값으로 $L=1, R=n$ (n 은 입력 변수의 수), $t=\text{integer}(L+(R-L)/2), m=1$ 로 정한다.
- Step 6: 부분집합 $\{v_1, v_2, \dots, v_t\}$ 를 C4.5로 평가하고 일반화 의사결정나무의 시험 데이터에 대한 에러율 ϵ_m 를 구한다.
- Step 7: 만약 ϵ_m 이 θ 보다 적거나 같으면 $R=t-1, \theta = \epsilon_m$ 이고, 그렇지 않으면 $L=t+1$ 이다.
- Step 8: 만약 $L > R$ 이면, 가장 작은 ϵ_m 값을 갖고

를 구성하는 변수의 수가 가장 적은 부분집합을 주요 변수 집합으로 선택한다. 그렇지 않으면, $t = \text{integer}(L + (R - L) / 2)$, $m = m + 1$ 을 계산하고 Step 6, 7, 8을 반복한다.

위 알고리즘에서 포장 기법은 $\log_2 N$ (N 은 변수의 수) 만큼 반복하여 실행하게 된다.

5. 실험 결과

4절에서 제시한 휴리스틱 포장 기법의 검증을 위해 캘리포니아 대학의 기계학습 데이터 레퍼지토리(Merz and Murphy, 1996) 중에서 <표 1>에 보인 것처럼 세 데이터 셋에 대해 적용을 해 보았다. 각 데이터 셋은 클래스 변수가 비슷하게 분포되도록 고려하면서 전체 데이터의 처음 인스턴스부터 차례로 3개의 부분 데이터 셋 묶음으로 구분하고 이중 두 묶음은 학습 데이터로 한 묶음은 시험 데이터로 사용하였다. 각 묶음이 한 번씩 시험 데이터가 되도록 3-겹 교차 검증(3-fold cross validation)을 하여 실험을 하였다.

<표 1> 실험 대상 데이터 셋

데이터 셋	입력 변수 수	클래스 변수 수	인스턴스 수
Wine	13	1	178
Ionosphere	34	1	351
Vote	16	1	435

C4.5의 경우 입력 변수가 명목 값이거나 혹은 수치 값인 경우 그리고 결측 치가 있는 경우에도 처리 가능하나 신경망의 경우 입력 변수의 값이

수치 값으로 이루어져야 한다. 따라서 모든 입력 변수가 y , n 값을 갖는 Vote 데이터 셋의 경우 y 는 1, n 은 0, 결측 치는 0.5로 하여 신경망을 학습하였다. 신경망의 학습에서 입력 노드 수는 입력 변수의 수만큼 출력 노드는 클래스 변수의 수만큼 구성하고 은닉 노드의 수는 시험 데이터에 대한 어려움이 일정 기준보다 적은 최소의 노드 수를 갖도록 학습하였다. 어려움의 일정 기준은 큰 네트워크를 가진 신경망의 예측율로 설정하였고, 이 예측율보다 떨어지지 않으면서 은닉 노드의 수를 최소로 갖는 네트워크를 가지고 학습하였다. 실험 결과는 <표 2>와 같다.

<표 2>에서 θ 는 전체 변수를 C4.5로 평가한 결과 생성된 일반화 의사결정나무의 시험 데이터에 대한 어려움을 나타낸다. ϵ 는 4절의 선택 조건을 만족하는 부분집합에 대해 C4.5로 평가한 결과 생성된 일반화 의사결정나무의 시험 데이터에 대한 어려움을 표현한다. Vote 데이터 셋과 Ionosphere 데이터 셋처럼 클래스가 고르게 분포되어 있는 경우는 전체 인스턴스를 차례로 3등분하여 시험 데이터(1/3)와 학습 데이터(2/3)를 만들고, Wine 데이터 셋처럼 클래스 값에 따라 정렬되어 있는 경우에는 각 클래스별로 인스턴스를 3등분하고 각 클래스의 1/3씩을 모아 시험 데이터를 만들고 나머지는 학습 데이터로 사용한다. 이와 같은 실험을 각 인스턴스가 한번은 시험 데이터가 되도록 3번 반복한 결과가 <표 2>에 표시되어 있다.

Wine 데이터 셋에서 전체 변수 13개를 가지고 C4.5를 적용한 결과 3겹 교차의 평균 어려움은 16.27%이었고, 생성된 의사결정나무에서 생성된 룰을 구성하는 변수는 4개로 나타났다. 휴리스틱 포장 기법에서 변수의 민감도에 따라 C4.5을 적용해 본 결과 룰을 구성하는 변수의 개수는 3.67

<표 2> 휴리스틱 포장 기법의 세 데이터 셋에 대한 적용 결과

데이터 셋	신경망 구성 (I-H-O)	C4.5 결과 (전체 변수)				C4.5 결과 (휴리스틱 포장 기법)			
		θ	θ 평균	룰 구성 변수 수	평균	ϵ	ϵ 평균	룰 구성 변수 수	평균
Wine 1	13-2-1	21.7%	16.27%	3	4	11.7%	12.33%	3	3.67
Wine 2	13-2-1	15.0%		5		15.0%		5	
Wine 3	13-2-1	12.1%		4		10.3%		3	
Ionosphere 1	34-3-1	16.2%	12.27%	5	8	16.2%	11.1%	5	6.33
Ionosphere 2	34-3-1	10.3%		7		10.3%		6	
Ionosphere 3	34-3-1	10.3%		12		6.8%		8	
Vote 1	16-3-1	2.1%	4.13%	5	4	2.1%	3.9%	5	2.67
Vote 2	16-3-1	3.4%		4		3.4%		3	
Vote 2	16-3-1	6.9%		3		6.2%		1	

개로 비슷했으나 에러율은 12.33%로 예측율이 향상된 결과를 나타냈다. Ionosphere 데이터 셋의 경우 에러율은 12.27%에서 11.1%로, 룰 구성 변수의 수는 8에서 6.33으로 줄어 든 것을 알 수 있다. Vote 데이터 셋에서는 에러율이 4.13%에서 3.9%로, 룰 구성 변수 수는 4에서 2.67로 줄어 들었다. 휴리스틱 포장 기법을 적용한 결과 룰을 구성하는 변수의 수가 줄고 C4.5의 시험 데이터에 대한 예측 성능이 높아지는 것을 알 수 있다.

6. 결론

C4.5와 같은 귀납적 학습 기법을 적용하여 판별 작업을 하려는 경우 잡음이 포함된 변수 혹은 불필요하게 많은 변수들은 귀납적 학습 결과 생성된 룰을 복잡하게 하고 또한 새로운 사례에 대한 예측력을 감소시키게 된다. 따라서 결과를 판별하는데 필요한 변수들만을 선정하여 귀납적 학습 기법의 예측력을 높이는 것이 필요하다. 이

연구에서는 신경망을 이용하여 C4.5의 예측력을 높이기 위한 포장 기법에 대해 살펴보았다. 일반적인 포장 기법의 경우 전체 변수들의 모든 부분 집합을 탐색하게 되면 많은 탐색 시간이 걸림에 따라 현실적으로 적용하는 데 어려움이 있다. 이 연구에서는 신경망의 민감도에 대한 연구를 적용하여 포장 기법에서 탐색할 공간을 줄이면서 C4.5의 예측력을 높일 수 있는 방안을 제시하였다. 제안된 휴리스틱 포장 기법을 3개의 데이터 셋에 적용해 본 결과 C4.5의 예측력이 높아지고 룰을 구성하는 변수의 수도 줄었다. 즉 탐색 공간의 복잡도를 $O(\log_2 N)$ 으로 줄이면서 C4.5의 수행 성능을 높이는 변수 부분집합을 추출할 수 있음을 실험에서 보여 주고 있다.

그러나 제안된 기법이 일반화되기 위해서는 더 많은 다양한 형태의 데이터 셋에 대한 실험을 통한 일반화가 필요하리라 생각한다. 그리고 선정된 부분집합이 최적 부분집합과 어느 정도의 차이가 있는 지도 밝혀야 한다. 또한 데이터가 신경망에 의해 학습이 된 후 이 기법을 적용할

수 있는 한계도 있다. 다만 많은 데이터 셋이 복잡한 연관 관계에 의해 서로 얽혀 있는 경우에도 신경망의 학습 기능으로 입력 변수들의 중요도를 쉽게 계산할 수 있는 점과 신경망의 중요 주제 중 하나인 주요 변수 선정 기법은 귀납적 학습의 변수 부분집합 선정에도 도움이 될 수 있을 것이다. 따라서 제안된 방법처럼 신경망의 민감도 분석 같은 주요 변수 선정 방안과 귀납적 학습 기법의 변수 부분집합 선정을 위한 포장 기법의 연결은 데이터 마이닝을 위한 전처리 작업에 적절히 적용될 경우 예측력 높은 지식을 추출하는데 있어 유용하게 기여할 것으로 기대된다.

참고문헌

- Almuallim, H. and T. Dietterich, Learning Boolean Concepts in the Presence of Many Irrelevant Features, *Artificial Intelligence*, Vol.69, (1994), 279~305.
- Blum, A.L. and P. Langley, "Selection of Relevant Features and Examples in Machine Learning", *Artificial Intelligence*, Vol.97, (1997), 245~271.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth International Group, Belmont, California, 1984.
- Dash, M. and H. Liu, "Feature Selection Methods for Classifications", *Intelligent Data Analysis*, Vol.1, (1997), 131~156.
- Engelbrecht, A.P., "A New Pruning Heuristic Based on Variance Analysis of Sensitivity Information", *IEEE Transactions on Neural Networks*, Vol.12, No.6, Nov. (2001), 1386~1399.
- Ginsberg, M.L. *Essentials of Artificial Intelligence*, Morgan Kaufmann, 1993.
- Girosi, F., M. Jones, and T. Poggio, "Regularization Theory and Neural Network Architectures", *Neural Computation*, Vol.7, (1995), 219~269.
- Hirose, Y., K. Yamashita, and S. Hijiya, "Backpropagation Algorithm which Varies the Number of Hidden Units", *Neural Networks*, Vol.4, (1991), 61~66.
- Karnin, E.D., "a Simple Procedure for Pruning Back-Propagation Trained Neural Networks", *IEEE Transactions on Neural Networks*, Vol.1, No.2, June (1990), 239~242.
- Kohavi, R. and G.H. John, "The Wrapper Approach", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda (Eds.), Klumer Academic Publishers, Dordrecht, Netherlands, 1998, 33~50.
- Le Cun, Y., J.S. Denker, and S.A. Solla, "Optimal Brain Damage", in *Advances in Neural Information Processing (2)*, D.S. Touretzky, (Ed.), 1990, 598~605.
- Liu, H. and H. Motoda, "Less is More", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda (Eds.), Klumer Academic Publishers, Dordrecht, Netherlands, 1998, 3~12.
- Merz, C.J. and P.M. Murphy, *UCI Repository of Machine Learning Databases*, University of California, Irvine, CA, 1996, (<http://www.ics.uci.edu/mlearn/MLRepository.html>).
- Mozer, M.C. and P. Smolensky, "Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment", in *Advances in Neural Information Processing (1)*, D.S. Touretzky, (Ed.), 1989, 107~115.

- Ponnappalli, P.V.S., K.C. Ho, and M. Thomson, "A Formal Selection and Pruning Algorithm for Feedforward Artificial Neural Network Optimization", *IEEE Transactions on Neural Networks*, Vol.10, No.4, July (1999), 964~968.
- Quinlan, J.R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, California. 1993.
- Reed, R. "Pruning Algorithms-A Survey", *IEEE Transactions on Neural Networks*, Vol.4, (1993), 740~747.
- Rumelhart, D.E., G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation", In *Parallel distributed Processing*, Vol. 1, D.E. Rumelhart et al., (Eds.), Cambridge, MA, MIT Press. 1986, 318~362.
- Setiono, R. and H. Liu, "Feature Extraction via Neural Networks", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda, (Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, 1998, 191~204.
- Wang, H., D. Bell, and F. Murtagh, "Relevance Approach to Feature Subset Selection", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda, (Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, 1998, 85 ~100.
- Wang, K. and S. Sundaresh, "Selecting Features by Vertical Compactness of Data", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda, (Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, 1998, 71~84.
- Weiss, S.M., *Computer Systems that Learn*, Morgan Kaufmann Publishers, San Francisco, CA, 1991.
- Wnek, J. and R.S. Michalski, "Hypothesis-Driven Constructive Induction in AQ17-HCI: a Method and Experiments", *Machine Learning*, Vol.14, (1994), 139~168.
- Wyse, N., R. Dubes, and A.K. Jain, "a Critical Evaluation of Intrinsic Dimensionality Algorithms", In *Pattern Recognition in Practice*, E.S. Gelsema and L.N. Kanal., (Eds.), Morgan Kaufmann Publishers, 1980. 415~425.
- Yang, J. and V. Honavar, "Feature Subset Selection Using a Genetic Algorithm", In *Feature Extraction, Construction, and Selection: A Data Mining Perspective*, H. Liu and H. Motoda, (Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, 1998, 117~136. Zhang J. and A. Morris, a Sequential Learning Approach for Single Hidden Layer Neural Networks, *Neural Networks*, Vol.11, (1997), 65~80.

Abstract

Feature Subset Selection in the Induction Algorithm using Sensitivity Analysis of Neural Networks

Boo-sik Kang*
Sang-chan Park**

In supervised machine learning, an induction algorithm, which is able to extract rules from data with learning capability, provides a useful tool for data mining. Practical induction algorithms are known to degrade in prediction accuracy and generate complex rules unnecessarily when trained on data containing superfluous features. Thus it needs feature subset selection for better performance of them. In feature subset selection on the induction algorithm, wrapper method is repeatedly run it on the dataset using various feature subsets. But it is impractical to search the whole space exhaustively unless the features are small.

This study proposes a heuristic method that uses sensitivity analysis of neural networks to the wrapper method for generating rules with higher possible accuracy. First it gives priority to all features using sensitivity analysis of neural networks. And it uses the wrapper method that searches the ordered feature space. In experiments to three datasets, we show that the suggested method is capable of selecting a feature subset that improves the performance of the induction algorithm within certain iteration.

Key words: feature subset selection, wrapper method, neural networks, induction algorithm

* Dept. of Business Administration and Information, Mokwon University

** Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology