

연결강도판별분석에 의한 부도예측용 신경망 모형의 입력노드 설계 : 강제연결뉴론 선정 및 약체연결뉴론 제거 접근법 (Link Weight Discrimination Analysis based Design of Input Nodes in ANN Models for Bankruptcy Prediction: Strong-Linked Neurons Selection and Weak-Linked Neurons Elimination Approach)

이용규(wklee@biho.taegu.ac.kr) 손동우(dwson@biho.taegu.ac.kr)
대구대학교 경영학과

요약

본 연구에서는 부도예측용 인공신경망 모형의 입력노드를 선정하기 위한 방법론으로 연결강도판별분석(Link Weight Discrimination Analysis)에 의한 약체뉴론제거법(Weak-Linked Neuron Elimination)과 강제뉴론선택법(Strong-Linked Neurons Selection)을 제안한다. 연결강도판별분석이란 적절한 학습이 끝난 인공신경망 모형에서 입력노드와 연결되는 가중치의 합에 대한 절대값인 연결강도 판별식(Link Weight Discrimination)에 의해 해당 입력노드가 출력노드에 미치는 영향정도를 분석하는 것이다. 한편 강제연결뉴론선택법은 선처리를 통해 얻어진 학습된 인공신경망의 입력노드 가운데서 연결강도판별식이 큰 뉴론만을 본처리의 입력노드로 선정하는 것인데 비해 약체연결뉴론제거법은 연결강도판별식이 일정 값 즉, 연결강도 판별임계치(Link Weight Discrimination Cut off Value) 보다 낮은 입력노드를 제외하고 나머지 입력노드만을 본처리의 입력노드로 선정하는 것이다. 본 연구에서는 강제연결뉴론선택법과 약체연결뉴론제거법을 각각 정형적인 방법론으로 정립하고 이 방법론에 의해 부도예측용 인공신경망을 구축하여 각각의 모형을 의사결정트리에 의해 선정된 인공신경망 모형 및 선처리 과정을 거치지 않은 인공신경망 모형과 성능을 비교, 분석하여 본 연구에서 제안한 방법론의 타당성을 제시하였다.

주제어: 연결강도판별분석, 약체연결뉴론제거법, 강제연결뉴론제거법, 부도예측

1. 서론

본 연구에서는 부도예측용 인공신경망 모형의 입력노드를 선정하기 위한 방법론으로 연결강도판별분석(Link Weight Discrimination Analysis)에 의한 약체뉴론제거법(Weak-Linked Neuron Elimination)과 강제뉴론선택법(Strong-Linked Neurons Selection)을 제안한다.

연결강도판별분석이란 적절한 학습이 끝난 인공신경망 모형에서 입력노드와 연결되는 가중치의 합에 대한 절대값인 연결강도 판별식(Link Weight Discrimination)에 의해 해당 입력노드가 출력노드에 미치는 영향정도를 분석하는 것이다(손동우, 이용규, 2000; 이용규, 손동우, 2000).

한편 강제연결뉴론선택법은 선처리를 통해 얻어진 학습된 인공신경망의 입력노드 가운데서 연결강도판별식이 큰 뉴론만을 본처리의 입력노드로 선정하는 것인데 비해 약체연결뉴론제거법은 연결강도판별식이 일정 값 즉, 연결강도 판별임계치(Link Weight Discrimination Cut off Value) 보다 낮은 입력노드를 제외하고 나머지 입력노드만을 본처리의 입력노드로 선정하는 것이다.

본 연구에서는 강제연결뉴론선택법과 약체연결뉴론제거법을 각각 정형적인 방법론으로 정립하고 이 방법론에 의해 부도예측용 인공신경망을 구축하여 각각의 모형을 의사결정트리에 의해 선정된 인공신경망 모형 및 선처리 과정을 거치지 않은 인공신경망 모형과 성능을 비교, 분석한다.

본 논문의 구성은 다음과 같다. 2절에서는 부도예측을 위한 통계적 기법과 인공지능 기법을 살펴보고, 인공신

경망 기법에서의 변수 선정을 위한 기법과 그 한계점을 논의한다. 3절에서는 본 연결강도판별분석 접근법의 이론적 배경과 방법론 그리고 그 예를 보이고, 4절에서는 연결강도판별분석 접근법에 의해 선처리된 인공신경망 모형과 선처리 없는 인공신경망 모형 및 의사결정트리에 의해 선처리된 인공신경망 모형의 성능을 비교분석하도록한다. 마지막으로 5절에서는 본 연구의 기여도 및 향후 과제에 대해 논의한다.

2. 문헌고찰

부도(Bankruptcy)란 기업이 파산적 지급불능상태(Bankrupt Insolvency)에 있거나 법원에 의해 파산선고가 내려진 상태를 일컫는 것으로 기업의 이해관계자는 물론이고 일반소비자나 국민경제 전체에 걸쳐 막대한 비용을 초래하게 된다 (박정식, 신동령, 1999). 따라서 부도에 대한 부실증후 (Symptoms of Failures)를 재무비율이나 현금흐름시표 등을 통해 미리 예측하려는 부도예측(Bankruptcy Prediction)은 실무적 차원에서 뿐 아니라 학계에서도 오랜 기간 동안 중요 관심사였다.

초기의 부도예측은 다변량 판별분석(Altman, 1968)이나 로짓회귀분석(Logit Regression), 프로빗분석(Probit Analysis) 등의 회귀분석(Regression) (Ohlson, 1980; Zmijewski, 1984)과 같은 통계적 기법이 주로 이용되어 왔었다.

그러나, 1980년대 후반부터 다층퍼셉트론(Multi-Layered Perceptrons) (이재규 외 1998; Berry and Linoff, 1997; Hertz et. al., 1991)에 의한 인공신경망의 가능성이 새롭게 인정되면서부터 인공신경망에 의한 부도예측 연구가 활발히 진행되었고, 실제로 통계적 기법

보다 뛰어난 예측력을 보여주고 있음이 입증되고 있다 (조흥규, 1999 ; Tam and Kiang 1991 ; Fletcher, 1993 ; Tsukuda, 1994 ; Udo, 1993 ; Wilson, 1994)

이와 같이 통계적 기법에 비해 다층퍼셉트론 구조의 인공신경망이 부도예측에 널리 사용되고 있는 이유는 비선형성을 갖는 분류문제에 뛰어난 적응력을 보이면서도 입력노드에 대한 통계적 가정을 요구하고 있지 않기 때문인 것으로 알려져 있다 (Berry and Linoff, 1997).

다층퍼셉트론 이외에도 자기조직화 형상화지도 (Self-Organization Features Map) (Serrano-Cinca, 1996)와 의사결정트리(Sung, et al, 1999 ; Messier and Hansen, 1988) 등도 부도예측에 적용하여 좋은 예측율을 보여 주고 있다.

인공신경망에 의한 모형화과정에서 가장 중요한 고려사항 중의 하나는 적절한 입력노드의 선정인데, 그 이유는 지나치게 많은 입력노드는 많은 학습시간을 요구할 뿐 아니라 적절치 못한 해(Solution)를 도출할 가능성이 높기 때문이다. 따라서 적절한 입력노드의 선정을 위해 전문가의 직관이나 분석에서부터 통계적 분석, 의사결정트리 등 다양한 기법이 제안되고 있다 (Berry and Linoff, 1997).

부도예측용 인공신경망 모형의 입력노드 선정기법으로는 다변량 판별분석, 자기형상화지도, 의사결정 트리 등을 들 수 있는데 이들 기법을 이용하여 선처리를 실시한 인공신경망 모형이 그렇지 않은 경우보다 높은 예측력을 보여 주고 있는 것으로 알려져 있다(Lee, et al, 1996).

그러나, 기존의 입력변수 선정기법은 나름대로의 한계점을 보유하고 있는데, 다변량 판별분석의 경우 독립변수들의 집단이 다변량 정규분포를 이루면서 공분산이 같아야 하는 가정이 필요하기 때문에 이에 대한 검증이 필요로 한다(채서일, 1999 ; Tam and Kiang, 1992). 의사결정트리의 경우 독립변수와 종속변수의 관계를 쉽게 설명할 수 있는 IF~THEN 규칙을 생성하는 장점이 있어서 결과를 쉽게 이해하고 설명하고자 할 때 적합한 방법이지만(Messier and Hansen, 1988), 독립변수가 취할 수 있는 값이 많아질 경우에는 오류가 발생할 가능성이 많다는 약점을 가지고 있다(Berry and Linoff, 1997). 자기조직화 형상지도에 의한 변수 선정방법은 자율학습(Unsupervised Learning)을 채택하고 있기 때문에 자칫 부도예측과 상관이 없는 변수 분류를 초래할 가능성이 있다 (이재규 외 1998; Berry and Linoff, 1997; Hertz et. al., 1991).

한편 최근 유전자알고리즘과 인공신경망의 결합에 의한 연구가 활발한데 부도예측에서도 유전자알고리즘에 의한 입력변수 선정을 시도하여 기존의 통계적 기법보다 더 좋은 예측율을 보이고 있다(Shin and Hong, 2000).

3. 연결강도판별분석 접근법

3.1 연결강도판별

다층퍼셉트론의 구조를 갖는 인공신경망에서 역전파 알고리즘을 이용하여 학습을 끝냈다는 것은 입력노드와 은닉노드를 연결하는 가중치의 값들과 은닉노드와 출력노드를 연결하는 가중치의 값을 구했음을 의미한다.

n개의 입력노드와 m개의 은닉노드 및 1개의 출력노드를 갖는 인공신경망이 있다고 하자. 그리고 입력노드는 x_i 은닉노드는 h_j 출력노드는 z 라고 하자.

x_i : 입력노드 ($i = 1, 2, \dots, n$)

h_j : 은닉노드 ($j = 1, 2, \dots, m$)

z : 출력노드

ω_{ij} : 학습이 끝난 상태에서 입력노드 x_i 와 은닉

노드 h_j 를 연결하는 가중치($i = 1, 2, \dots, n, j = 1, 2, \dots, m$)

ξ_j : 학습이 끝난 상태에서 은닉노드 h_j 와 출력

노드 z 를 연결하는 가중치($j = 1, 2, \dots, m$)

$f(\cdot)$: 노드에서의 변환 함수

따라서 z 와 h_j 및 x_i 와의 관계는 다음과 같다.

$$z = \sum_{j=1}^m \xi_j f(h_j) = \sum_{j=1}^m \xi_j f\left(\sum_{i=1}^n \omega_{ij} f(x_i)\right)$$

여기서 $f(\cdot)$ 는 시그모이드(Sigmoid) 함수, 또는 하이퍼탄젠트 함수와 같은 단조증가함수를 취하게 되므로 (이재규외, 1998; Berry and Linoff, 1997; Hertz et. al., 1991), 각 z 는 ω_{ij} 의 단조증가함수가 될 수 있기 있고 입력노드인 x_i 와 연결되는 모든 ω_{ij} ($j=1,2,\dots,m$)의 합은 x_i 가 z 를 설명할 수 있는 정도로 간주될 수 있다. 여기서 다음과 같은 정의를 한다.

정의: 적절한 훈련이 끝난 인공신경망 모형에서 각 입력노드와 연결되는 가중치들의 합에 대한 절대값을 연결강도판별식이라 한다. 즉, x_i 의 연결강도판별식 y_i 는 다음과 같이 정의된다.

$$y_i = \left| \sum_{j=1}^m \omega_{ij} \right| \text{ (for } i=1, 2, \dots, n)$$

위 정의에 근거하여 다음과 같은 세 가지 명제를 제안한다.

연결강도판별 명제: 적절한 훈련이 끝난 인공신경망 모형에서 출력노드가 값은 각 입력노드와 연결되는 가중치들의 합에 대한 절대값 즉, 연결강도판별식의 값에 의존한다. 즉, 연결강도판별식이 큰 입력노드는 연결강도판별식의 값이 작은 입력노드에 비해 출력값에 대한 설명력이 높다.

약체연결뉴론 명제: 연결강도판별식의 값이 작은 입력노드를 제거하고 새롭게 훈련을 한 인공신경망 모형은 그렇지 않은 경우보다 출력값에 대해 더 높은 설명력을 갖는다.

강체연결뉴론 명제: 연결강도판별식의 값이 큰 입력노드만으로 새롭게 훈련을 한 인공신경망 모형은 그렇지 않은 경우보다 출력값에 대해 더 높은 설명력을 갖는다.

예를 들어 그림-1에서 보는 바와 같이 네 개의 입력노드와 세 개의 은닉노드 그리고 한 개의 출력노드로 이루어진 인공신경망의 경우 각 입력노드가 출력노드를 설명하는 설명력은 그림-2의 실선이 나타나고 있는 입력노드와 은닉노드간 가중치의 합에 대한 절대값인 연결강도판별식에 의해 결정된다.

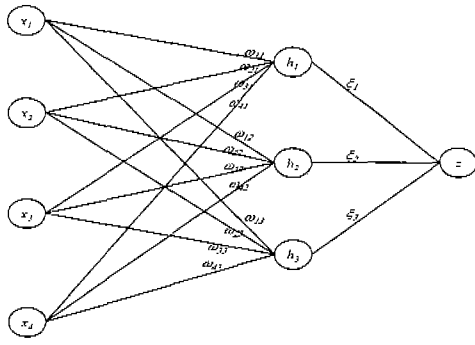


그림-1 학습된 인공신경망 예

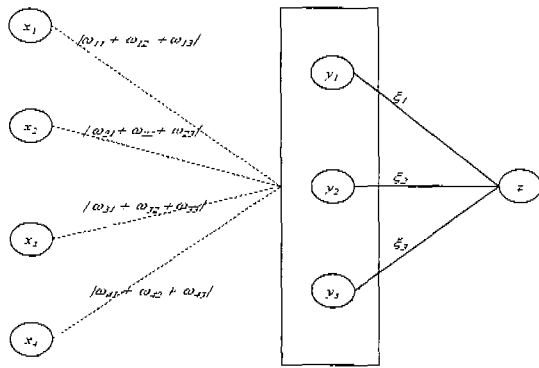


그림 2 연결강도판별식의 예

3.2 강제연결뉴런 선택법

3.2.1 강제연결뉴런 선택법의 절차

강제연결뉴런 선택법은 인공신경망의 각 연결강도가 큰 입력노드는 전체 출력값에의 영향이 강할 것이라는 가설하에 입력노드와 은닉노드간의 연결 가중치의 합에 대한 절대값인 연결강도 판별식의 값이 각 라운드별로 가장 큰 노드들만 선택해서 입력변수로 이용한다. 다시 말하면, 출력노드에 미치는 영향이 큰 변수들만 최종 인공신경망 모델의 입력변수로 선정하여 출력노드에 대한 설명력을 높이려는 입력노드 선정 휴리스틱이라고 할 수 있다.

인공신경망의 학습방법은 연결가중치의 값을 무작위 방식에 의해 초기화하여 이 초기화한 연결가중치를 점진적으로 향상시키는 방식으로 전개된다. 따라서 인공신경망에 의한 학습결과는 항상 무작위성을 내포하게 되어 있다. 이러한 무작위성을 반영하기 위해 강제연결뉴런 선택법에서는 위에서 제시한 과정을 일정 회수만큼 반복하여 여기에서 얻어진 선택대상 입력노드의 빈도(Frequency)를 구하고, 선택된 빈도수가 일정한 횟수 이상이 되는 입력노드들을 강제연결뉴런 선택법에 의한 선처리 과정을 거친 최적인공신경망 모델의 최종 입력노드로 이용한다.

강제연결뉴런 선택법을 이용하여 입력변수를 선정하는 과정은 다음과 같다.

제 0 단계) 실험회수 k 와 실험라운드 s 를 각각 1로 한다.

제 1 단계) 입력노드 x_i , 은닉노드 h_j , 출력노드는 z 로 이루어진 인공신경망에 의해 학습을 시행하여

최적화시킨다.

제 2 단계) 각각의 y_i 를 구하여 그 중 최대값을 선택대상 변수목록에 추가시킨다.

제 3 단계) s 값이 실험예정 회수 K 보다 작을 경우 s 를 1증가시키고, 선택대상 변수목록에 추가시킨 노드를 입력노드에서 제외한 후 제 1 단계로 돌아가고, 그렇지 않을 경우 제 4 단계로 간다.

제 4 단계) k 값이 실험예정 라운드 M 보다 작을 경우 k 의 값을 1 증가시키고 제 1 단계로 돌아가고 그렇지 않을 경우 중단한다.

위 알고리즘에서 나온 선택대상 변수목록의 원소들의 빈도수별로 정리하여 이 가운데 좋은 예측율을 갖는 경우를 선택하여 본 처리의 입력노드로 선정한다.

3.2.2 강제연결뉴런 선택법의 예

본 연구에서 실험한 연결강도판별분석 접근법 가운데 강제연결뉴런 선택법의 실험과정은 다음과 같다. 먼저 43개의 입력변수로 강제연결뉴런 선택법을 적용하여 입력변수를 선정하게 되는데, 여기서 실험예정 라운드 수 $M=10$, 각 라운드내 실험예정회수 $K=10$ 으로 하여 실험을 진행하였다. [표-1]은 제 1라운드에서 선정된 강제연결변수이다. 제 1라운드 수행결과, 1회에서 X49, 2회에서 X44, 3회에서 X19, 4회에서 X46 등, 10회에 걸친 실험에서 아래와 같이 10개의 변수가 강제연결 변수로 선정되었다(전체 선택대상변수는 부록-1 참고).

라운드 및 회수	선택대상변수
D ₁₁	X49
D ₂₁	X44
D ₃₁	X19
D ₄₁	X46
D ₅₁	X33
D ₆₁	X55
D ₇₁	X29
D ₈₁	X56
D ₉₁	X41
D ₁₀₁	X13

[표-1] 제1라운드의 회수별 선택대상변수

이와 같은 방법의 실험을 10라운드에 걸쳐 진행한 결과로 얻은 선택대상변수를 빈도수별로 [표-2]에서 정리하여 보여주고 있다.

빈도수	변수
0	X15, X21, X23, X25, X27, X28, X31, X34, X53, X59
1	X9, X10, X29, X36, X42, X43, X52, X60, X61, X67, X70, X73
2	X32, X33, X35, X38, X39, X62, X69
3	X14, X37, X66
4	X18, X41, X44, X47, X56
5	X50
6	X13
7	X46, X55
10	X19, X49

[표-2] 빈도수별 선택대상변수

3.3 약체연결뉴런 제거법

3.3.1 약체연결뉴런 제거법의 과정

약체연결뉴런 제거법은 인공신경망의 각 연결강도가 작은 입력노드는 전체 출력값에의 영향이 미미할 것이라는 가설하에 입력노드와 연결되는 가중치의 합에 대한 절대값인 연결강도판별식이 일정값 즉, 연결강도판별 임계치(Linked Weight Discrimination Cut off Value)보다 작은 경우 해당 입력노드를 제거함으로써 입력노드에 관한 최적 인공신경망 구조를 찾으려는 입력노드 선정방법이다 (이용규, 손동우, 2000; 손동우, 이용규 2000).

약체연결뉴런 제거법에서는 일단 인공신경망을 학습시켜 인공신경망을 최적 상태로 만든 다음 이 인공신경망에서 입력노드와 은닉노드간의 연결강도판별식의 값이 일정값 즉, 연결강도판별 임계치보다 작은 입력노드를 제거시키고 나머지 입력노드들만으로 새롭게 학습을 한다. 이와 같은 과정을 다음과 같은 경우가 나올 때까지 반복 실험을 하는데, 이를 정지규칙(Stopping Rule)이라고 한다. 첫째, 연결강도판별식이 연결강도판별 임계치보다 작은 입력노드가 나오지 않거나 둘째, 인공신경망의 성능이 더 이상 개선되지 않거나 셋째, 반복회수가 일정 회수에 도달할 때까지 반복한다.

약체연결뉴런 제거법을 이용하여 입력변수를 선정하는 과정은 다음과 같다.

제 0 단계) k와 s를 각각 1로 한다.

제 1 단계) 입력노드 x_i , 은닉노드 h_j , 출력노드는 z 로 이루어진 인공신경망에 의해 학습을 시행하여 최적화시킨다.

제 2 단계) 각각의 y_j 를 구하여 제거대상 변수목록에 첨부한다. 즉, 다음을 만족하는 y_j 의 D_{sk} 를 구한다

$D_{sk} = \{ y_j : y_j < \alpha \}$ (α : 연결강도판별 임계치)

제 3 단계) $D_{sk} = \emptyset$ 이거나 오류율이나 오분류율이 필요이상으로 커지고 있다고 판단될 경우 제 4단계로 가고 그렇지 않을 경우 입력 노드들 가운데서 D_{sk} 에 속하는 해당되는 x_j 를 제거하고 s 값을 1 증가시키면서 제 1 단계로 돌아간다.

제 4 단계) k 값이 실험예정회수 M 보다 작은 경우 k의 값을 1 증가시키고 제 1 단계로 돌아가고 그렇지 않을 경우 중단한다

위 알고리즘에서 나온 $\{D_{sk}\}$ 의 원소들의 빈도수를 조사하여 빈도수가 N보다 큰 제거대상변수를 골라 이들을 제외한 나머지 변수들만을 본 처리의 입력노드로 선정한다.

3.3.2 약체연결뉴런 제거법의 예

연결강도판별분석 접근법에서 제시하고 있는 또 하나의 변수 선정법인 약체연결뉴런 제거법에서는 판별값의 임계치 $\alpha=0.05$, 실험예정 라운드 수 $M=10$, 실험예정회수 $K \leq 10$ 으로 설정하여 실험을 진행하였다. 한 라운드 내에서 회수를 반복하면서 제거대상변수를 선정했으며, 제거대상변수가 없는 경우에 실험을 중단하고 다음 라운드의 실험을 진행하였다. [표-3]은 제 1라운드에서 선정된 제거대상변수와 그 가중치를 보여주고 있다. 제 1라운드, 1회의 실험결과 X15, X25를 비롯한 총 10개(D11), 2회에서는 X18, X19를 비롯한 4개(D21), 3회에서는 X10, X13을 비롯한 8개(D31)의 제거대상변수가 선정되는 등, 제1라운드의 실험에서는 총 32개의 변수가 제거대상변수로 선정되었으며, 8회에서는 해당변수가 없었기 때문에 제1라운드의 실험을 중단하였다.

반복 회수	제거변수	가중치	반복 회수	제거변수	가중치
D ₁₁	X15	0.0397	D ₃₁	X14	0.0452
	X25	0.0052		X21	0.0413
	X29	0.0456		X23	0.0137
	X31	0.0089		X37	0.0326
	X32	0.0262		X43	0.0058
	X34	0.0279		X53	0.0337
	X56	0.0381		X36	0.0031
D ₄₁	X61	0.0063	X50	0.0257	
	X62	0.047	X60	0.0348	
	X67	0.0244	X73	0.0308	
	D ₂₁	X18	0.0092	D ₅₁	X35
X19		0.0256	X47		0.0345
X33		0.0359	X69		0.0182
X52		0.0101	D ₆₁		X28
D ₃₁	X10	0.008		X55	0.0408
	X13	0.0186	D ₇₁	X27	0.0269

[표-3] 제1라운드의 회수별 제거대상변수

[표-4]에서는 위와 같은 실험을 10라운드에 걸쳐 반복회수를 최대 10회까지 설정하고 약체연결변수를 선정한 결과를 토대로 하여 제거대상변수로 선정된 변수들을 빈도수별로 정리하여 보여주고 있다.

빈도수	제거대상변수
10	X60
9	X32, X36, X47, X70
8	X25, X31, X59, X62, X69
7	X15, X23, X27, X33, X35, X39, X52, X53, X61, X66
6	X10, X13, X19, X21, X28, X34, X41, X55, X56, X67
5	X9, X18, X37, X43, X50, X73
4	X14, X29, X44, X46
2	X38
1	X42, X49

[표-4] 빈도수별 제거대상변수

4. 비교분석

4.1 데이터와 입력변수

본 연구는 238개의 부도기업과 237개의 건전기업을 이루어진 전체 475개 기업의 13개 재무변수(부록-2참고)를 이용하여 실험을 진행하였다. 인공신경망 모형에서 학습 및 검증용으로 425개 기업의 데이터를 선정하여 이용하였으며, 그 가운데 85%를 학습(Training)에 사용하였고 15%를 학습에 대한 검증(Validation)에 이용하였다. 나머지 50개 기업의 데이터는 학습을 거친 인공신경망 모형의 부도예측 적중율을 확인(Scoring)하는데 사용하였고 마이닝 도구는 엔터프라이즈 마이너(Enterprise Miner)3.0을 이용하였다.

4.2 실험방법

강체연결뉴런 선택법에 의해 선정된 입력변수를 인공신경망 모델에 적용하여 얻은 모델(SL)과 약체연결뉴

론 제거법에 의해 선정된 입력변수를 인공신경망 모델에 적용하여 얻은 모델(WL)을 인공신경망을 독립적으로 수행한 순수 인공신경망 모형(NN) 및 의사결정 트리기법으로 입력변수 선정과정을 거친 후 인공신경망 모델에 적용한 모형(DT)들의 예측율을 측정하여 그 예측율을 각각 비교하고 그 차이가 유의한지를 통계적 기법을 이용하여 검증하였다.

SL(x)는 강제연결뉴론 선택법에 의해 선정된 입력 변수들을 인공신경망 모형에 적용한 실험인데, 먼저 은닉층의 노드수를 1개로 고정시킨 후 학습을 통해서 모형을 최적화 한 다음, 입력노드의 연결강도 판별값을 각각 확인하여 가장 큰 값을 가지는 노드는 선택대상변수 목록에 추가하고 다음회의 실험에서는 입력변수에서 제거하였다. 10라운드 10회동안 100회의 강제연결뉴론을 선정하였고 이를 빈도수별로 정리하여 각 빈도마다 SL(x)모형을 만들어서 그 성능을 10라운드 반복실험하여 측정하였다. 그 가운데 성능이 가장 좋은 SL(x)를 SL모델로 선정하였으며, 선정된 SL모델을 다시 40라운드에 걸쳐 시드를 변환하면서 예측율을 측정하였다.

WL(x)는 약체연결뉴론 제거법에 의해 선정된 입력 변수들을 인공신경망 모형에 적용한 실험이다. 먼저 은닉층의 노드수를 1개로 고정시킨 후 학습을 통해서 모형을 최적화 한 다음, 각각의 입력노드가 가지는 연결강도 판별값을 확인하여 연결강도 판별값이 임계치인 α 값보다 작은 입력노드는 제거대상변수 목록에 추가하고 다음회의 실험에서는 입력노드에서 제거하였다. 한 라운드내에서의 반복실험과정에서 α 값보다 작은 입력노드가 나타나지 않을 경우, 다음 라운드의 실험을 진행하였다. 반복회수 최대 10회, 10라운드에 걸쳐 실험을 진행, 그 결과를 빈도수별로 정리하여 각 빈도마다 WL(x)모형을 만들어서 그 성능을 10라운드 반복실험하여 측정하였다. 그 가운데 성능이 가장 좋은 WL(x)를 WL모델로 선정하였으며, 선정된 WL모델을 다시 40라운드에 걸쳐 시드를 변환하면서 예측율을 측정하였다.

다음으로 DT는 의사결정 트리를 이용하여 입력변수 선정을 위한 선처리를 거쳐서 선정된 입력변수들을 인공신경망 모형에 적용하여 실험을 실시했다. 선정된 변수 전체 43개의 변수 가운데 X25, X50, X52를 제외한 나머지 40개의 변수가 제거대상변수로 선정되었고 X25, X50, X52가 입력변수로 선정되었다. 인공신경망에 적용시 은닉층의 노드수는 입력층의 노드수와 같게 하였다.

NN은 43개의 변수를 그대로 인공신경망 모델에 적용하여 실험을 진행했으며 은닉층의 노드수는 입력층의 노드수를 따랐다.

4.3 실험결과

4.3.1 강제연결뉴론 선택법

강제연결뉴론 선택법을 이용하여 변수선정 실험을 한 결과 10회에 걸쳐서 선택대상변수로 선정된 변수는 X19, X49 두 개의 변수이고, 9회와 8회 선정된 변수는 없었으며, 7회 선택대상변수로 선정된 것은 X46, X55이며 6회 선정된 변수는 X13이었다. 실험결과 강제연결변수로 선정된 변수를 빈도수별로 정리하여 10회, 7회, 6회, 5회에 걸쳐서 선택대상변수로 선정된 변수를 누적하여 인공신경망의 입력변수로 적용한 SL(x)별 예측율의 결과는 [표-5]와 같다.

모델 라운드	SL(5)	SL(6)	SL(7)	(10)
1	74	66	54	56
2	68	68	56	56
3	60	60	62	54
4	72	66	60	54
5	64	72	64	56
6	54	74	60	52
7	62	66	68	48
8	74	64	70	54
9	70	62	76	56
10	70	*78	74	58

[표-5] SL(x)별 예측율

실험결과 SL(6)의 10라운드에서 가장 좋은 예측율(78%)을 보여주었기 때문에 SL(6)이 SL모델로 선정되었다.

4.3.2 약체연결뉴론 제거법

약체연결뉴론 제거법을 이용하여 변수선정 실험을 한 결과 10회에 걸쳐서 제거대상변수로 선정된 변수는 X60이고, 9회에 걸쳐 선정된 변수는 X32, X36 등 네 개의 변수였다. 각 빈도수별 제거대상변수는 [표-4]에서 보여주고 있다.

빈도수별로 제거대상변수를 정리하여 10회, 9회, 8회, 7회, 6회, 5회, 4회, 2회, 1회에 걸쳐서 제거대상변수로 나타난 변수를 각각 누적하여 제거한 후 나머지 변수를 입력변수로 인공신경망 모델에 적용하여 각각의 WL(x)모형을 얻었고, 그 모델별 예측율을 측정하여 가장 성능이 좋은 WL(x)를 WL모델로 선정하였다.

각각의 WL(x)별 예측율은 [표-6]과 같이 나타났다.

모델 라운드	WL (10)	WL (9)	WL (8)	WL (7)	WL (6)	WL (5)	WL (4)
1	58	64	68	74	80	80	74
2	60	70	68	70	82	74	68
3	56	70	72	72	82	78	72
4	58	66	70	72	82	80	70
5	58	68	74	74	78	78	68
6	60	68	74	74	82	78	66
7	60	68	76	76	82	72	70
8	56	66	70	66	*84	82	68
9	58	72	72	72	82	82	66
10	58	68	70	70	78	80	70

[표-6] WL(x)별 예측율

실험결과 WL(6)의 8라운드에서 가장 좋은 예측율(84%)을 보여주었기 때문에 WL(6)이 WL모델로 선정되었다.

4.3.3 모델별 예측율

위와 같이 WL모델과 SL모델을 최종 결정한 후 네 가지 모델별을 각각 40라운드씩 시드를 변환하여 반복실험을 진행했다. [표-7]은 각 모델별 예측율을 요약해서 나타내고 있다.

	NN	DT	SL	WL
최대	62	82	*88	86
최소	50	46	68	76
평균	57.15	70.8	76.45	80

[표-7] 모델별 예측율

	NN	DT	SL	WL
NN	-	2.0057*	2.384**	2.2896*
DT	-	-	0.4937	0.3398
SL	-	-	-	0.1591
WL	-	-	-	-

*유의수준 25%, **유의수준 1%

[표-8] T검증결과

위의 결과와 같이 각 모델별 예측율 측정결과 SL의 예측율이 가장 우수했으면 NN의 예측율이 가장 낮았다.

그림-3은 전체 실험결과(부록-2 참고), 예측율의 추이를 그래프로 나타낸 것이다.

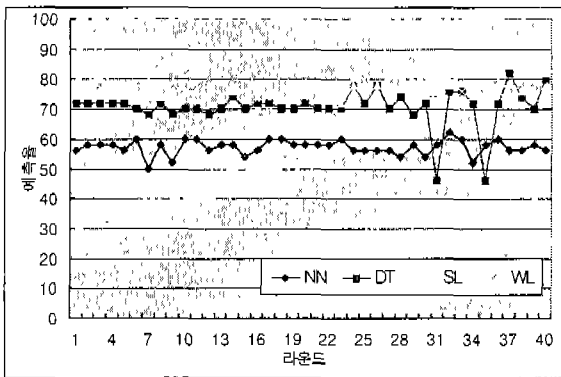


그림-3 모델별 예측율 추이

그림-3에서 보는 바와 같이 각 모델별 예측율 추이에서 연결강도판별에 의한 인공신경망 모델들의 예측율이 순수인공신경망은 물론이고 의사결정트리에 의해 선처리된 인공신경망 모델보다 전반적으로 높은 예측율을 보이고 있음을 알 수 있다.

4.4 실험결과분석

지금까지의 실험을 통해 각 모델의 예측결과를 살펴보면, [표-6]에서 보는 바와 같이 순수 인공신경망 모델의 최대 예측율은 62%, 최소 예측율은 50%, 평균 예측율은 57.15%로 나타났으며, DT모델은 최대 예측율이 82%, 최소 예측율이 46%, 평균 예측율은 70.8%로 나타났다. 다음으로 본 연구에서 제안하고 있는 연결강도판별분석에 의해 입력변수 선정과정을 거친 인공신경망 모델 가운데 강체연결 뉴런 선택법을 이용한 SL모델의 최대 예측율은 88%로 네 가지 비교모델 가운데 가장 높게 나타났으며, 최소 예측율은 68%, 평균 예측율은 76.45%로 각각 나타났다. 약체연결뉴런 제거법을 이용한 WL모델은 최대 예측율이 86%로 SL보다 약간 낮게 나타났으나, 최소 예측율은 76%, 평균 예측율이 80%로, 최소 예측율과 평균 예측율은 네 가지 모델 가운데서 가장 높게 나타났다.

위의 실험결과가 통계적으로 유의한 차이가 있는지를 검증하기 위해 각 모델별로 최대 예측율의 차이를 Z-test로 검증하였다. [표-8]은 각 모델별 Z-test의 결과를 보여주고 있다.

Z-Test의 결과 NN과 DT, SL, WL간에는 상당한 유의적인 차이가 있음을 나타내고 있지만 DT와 SL, WL간에는 통계적으로 유의적인 차이를 보여주지는 못하고 있다.

따라서 실험결과는 다음과 같이 요약될 수 있다. 먼저, 본 연구에서 제안하고 있는 연결강도 판별분석 접근법에 의해 입력노드 선정과정을 거친 인공신경망 모델의 예측율이 NN이나 DT모델보다 최대 예측율 뿐만 아니라 전반적으로 높게 나타났다. 두 번째, SL모델의 최대 예측율(88%)이 가장 높게 나타났다. 세 번째, 비교적 SL이나 WL의 예측율에 근접하는 DT은 SL이나 WL보다 편차가 심하게 나타나고 최소 예측율이 현저하게 낮다. 네 번째, 입력노드 선정과정을 거친 모델이 그렇지 않은 모델보다 예측율이 높다는 선행연구들의 결과를 실증적으로 입증하고 있다.

5. 결론

본 연구에서는 부도예측용 인공신경망 모형의 입력노드를 선정하기 위한 방법론으로 연결강도판별분석에 의한 약체뉴런제거법과 강체뉴런선택법을 제안하였고, 이들 방법에 의해 입력노드를 선정한 인공신경망 모형이 의사결정트리에 의해 입력노드를 선정한 인공신경망 모형이나 선처리과정을 거치지 않은 인공신경망 모형보다 더 높은 예측율을 보이고 있음을 실증하였다.

향후 연구과제로는 우선 강체연결뉴런과 약체연결뉴런을 정의하는 방법론적인 보완을 들 수 있다. 특히 약체연결뉴런을 결정하는 임계치에 대해 민감도 분석을 통해 최적의 임계치를 도출하는 것이나 강체연결뉴런선택에서 선정된 변수와 약체연결뉴런제거에서 선택된 변수를 통합하여 하나의 모형을 설정하는 것도 새로운 연구과제가 될 수 있을 것이다.

두 번째의 연구과제로는 본 연구에서 제안하고 있는 방법론을 부도예측이 아닌 다른 방법론에의 적용이다. 채권평가와 같이 분류(Classification)에 적용할 경우 흥미로운 연구 결과를 얻을 수 있을 것으로 기대된다.

세 번째로는 보다 방대한 데이터와 다양한 데이터로 실험하여 본 연구의 결과를 좀 더 일반화해 볼 필요성도 있다. 본 연구에서 사용한 데이터셋의 규모를 좀 더 확장시킨다면, 혹은 우리나라가 아닌 다른 나라의 데이터를 적용할 경우 좋은 결과를 얻을 수 있을 것으로 보인다.

네 번째로는 본 연구의 한계와도 연관되는 문제인데 좀 더 최적화된 인공신경망 모형을 통해 좀 더 나은 타당성을 얻어 보는 것도 좋은 연구과제가 될 것으로 보인다.

마지막으로 다변량 판별분석이나 회귀분석(Regr)과 같은 통계적 기법들, 그리고 자기조직형상화 지도, 유전자 알고리즘(Genetic Algorithm) 등과 같은 인공지능 기법들과도 그 성능을 비교해 볼 필요가 있을 것으로 보인다.

참고문헌

Altman, Edward I., "Financial Ratios, Discriminant Analysis and Prediction of Corporate

Bankruptcy", Journal of Finance, Sep.1968, pp.589-609.

Berry, Michael J. A. and Gordon Linoff, 'Data Mining Techniques: For Marketing, Sales, and Customer Support', John Wiley and Sons, 1997.

Fletcher, D. and E. Goss, "Forecasting with Neural Networks: an Application using Bankruptcy Data", Information and Management, 24. 3, 1993, pp.159-167

Hertz, John, Anders Krogh and Richard G. Palmer, 'Introduction to the Theory of Neural Computation', Addison Wesley, 1991.

Lee, Kun Chang, Ingoo Han and Youngsig Kwon, "Hybrid neural network models for Bankruptcy Predictions", Decision Support Systems, 18, 1996, pp.63-72.

Messier, William F. Jr. and James V. Hansen, "Inducing Rules For Expert System Development: An Example Using Default and Bankruptcy Data", Management Science, Vol.34, No.12, December. 1988, pp.1403-1415.

Ohlson, J. "Financial Ratios and The Probabilistic Prediction of Bankruptcy", Journal of Accounting Research, 18. 1, 1980, pp.109-131.

Serrano-Cinca, Carlos, "Self Organizing Neural Networks for Financial Diagnosis", Decision Support Systems, 17, 1996, pp.227-238.

Shin, Kyung-shik and Seung-hyun Hong, "A GA-based Input Selection Approach for Neural Networks Modeling: Application to Bankruptcy Prediction", Proceedings of the 5th Asia-Pacific Decision Sciences Institute Conference, Tokyo, Japan, 2000.

Sung, Tae Kyoung., Namsik Chang and Gunhee Lee, "Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction," Journal of Management Information Systems, Vol.16, No.1, Summer.1999, pp.63-85.

Tam, Kar Yan And Melody Y Kiang., "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", Management Science, Vol.38, No.7, July.1992, pp.926-947.

Tsukuda, J., and S. L. Baba, "Predicting Japanese Corporate Bankruptcy in terms of Financial Data Using Neural Network", Computers and Industrial Engineering, 27(1-4), 1994, pp.445-448

Udo, G., "Neural Network Performance on the Bankruptcy Classification Problem", Computers and Industrial Engineering, 25(1-4).1993, pp.377-380.

Wilson, Rick L. and Ramesh Sharda, "Bankruptcy Prediction using Neural Networks", Decision Support Systems, 11. 1994., pp. 545-557.

Zmijewski, Mafke, "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", Journal of Accounting Research, Vol.22 1984,

박정식, 신동령, '경영분석', 다산출판사, 1999.

손동우, 이용규, 약체연결뉴론 제거법에 의한 부도예측용 인공신경망 모형에 관한 연구, 2000년도 춘계학술대

회 발표논문집, 한국정보시스템학회, 2000.

이용규, 손동우, 부도예측용 인공신경망모형의 최적 입력노드 설계: 연결강도판별분석 접근, 2000 춘계정기학술대회, 한국지능정보시스템학회, 2000.

이재규, 최형림, 김현수, 서민수, 주석진, 지원철, '전문가 시스템 원리와 개발', 법영사, 1998.

조홍규, "도산예측을 위한 인공지능 방법과 통계적 방법의 통합 방법론," 한국과학기술원(박사학위논문) 1999.

채서일, '사회과학 조사방법론', 학현사, 1999. , 2판.

부록-1 전체선택대상변수

회수 라운드	1회	2회	3회	4회	5회
1	X49	X44	X19	X46	X33
2	X19	X13	X55	X44	X49
3	X19	X49	X55	X56	X50
4	X55	X49	X66	X41	X50
5	X50	X13	X19	X49	X55
6	X13	X19	X49	X36	X46
7	X19	X49	X18	X32	X50
8	X37	X50	X38	X49	X14
9	X19	X49	X33	X46	X55
10	X49	X46	X19	X9	X32
회수 라운드	6회	7회	8회	8회	10회
1	X55	X29	X56	X41	X13
2	X41	X14	X46	X52	X47
3	X60	X43	X62	X66	X67
4	X19	X39	X56	X46	X35
5	X46	X56	X37	X14	X62
6	X69	X44	X61	X73	X37
7	X39	X66	X47	X41	X10
8	X19	X55	X42	X18	X44
9	X18	X47	X35	X69	X13
10	X70	X47	X18	X38	X13

부록2. 전체변수내역

번호	지표	변수	변수내역
1	규모(1)	x9	매출액
2	규모(1)	x10	자기자본
3	생산성(2)	x13	총자본투자효율
4	생산성(2)	x14	부기가치율
5	성장성(3)	x15	매출액증가율
6	성장성(3)	x18	재고자산증가율
7	성장성(3)	x19	총자산증가율
8	수익성(4)	x21	금융비용대 매출액비율
9	수익성(4)	x23	대출효율성계수
10	수익성(4)	x25	매출액순이익률
11	수익성(4)	x27	매출원가율
12	수익성(4)	x28	손익분기점율
13	수익성(4)	x29	순금융비용대 매출액비율
14	수익성(4)	x31	이자보상배율
15	수익성(4)	x32	자기자본순이익률
16	수익성(4)	x33	총자본경상이익률
17	수익성(4)	x34	총자본순이익률
18	안정성(5)	x35	고정장기적합율의 역
19	안정성(5)	x36	단기부채대 총차입금
20	안정성(5)	x37	당좌비율
21	안정성(5)	x38	매출채권대 매입채무
22	안정성(5)	x39	순운전자본비율
23	안정성(5)	x41	유동비율
24	안정성(5)	x42	유동부채대 총자본
25	안정성(5)	x43	유보액대총자산비율
26	안정성(5)	x44	자기자본비율
27	안정성(5)	x46	차입금의존도
28	안정성(5)	x47	총차입금대 매출액
29	주세(6)	x49	금융비용부담금율 증가율
30	활동성(7)	x50	매입채무회전을
31	활동성(7)	x52	순운전자본대 매출액
32	활동성(7)	x53	운전자금대 회전을
33	활동성(7)	x55	재고자산회전을
34	활동성(7)	x56	총자본회전을
35	현금흐름 지표(8)	x59	(영업활동후 CF/차입금합계)*100
37	현금흐름 지표(8)	x61	영업활동후 CF/금융비용
38	현금흐름 지표(8)	x62	(현금영업이익/금융수입)/금 융비용
39	현금흐름 지표(8)	x66	(이자지급후 CF/부채총계)*100
40	현금흐름 지표(8)	x67	이자지급후 CF/금융비용
41	현금흐름 지표(8)	x69	(외부자금조달전 CF/차입금합계)*100
42	현금흐름 지표(8)	x70	외부자금조달전 CF/금융비용
43	현금흐름 지표(8)	x73	외부자금조달후 CF/금융비용

부록-3 모델별 예측율

라운드	NN	DT	SL	WL
1	56	72	76	80
2	58	72	78	82
3	58	72	78	78
4	58	72	78	82
5	56	72	78	84
6	60	70	78	78
7	50	68	76	80
8	58	72	74	*86
9	52	68	76	80
10	60	70	76	82
11	60	70	76	78
12	56	68	78	80
13	58	70	74	76
14	58	74	76	80
15	54	70	*88	78
16	56	72	74	82
17	60	72	76	*86
18	60	70	74	78
19	58	70	74	82
20	58	72	78	80
21	58	70	74	82
22	58	70	76	84
23	60	70	72	78
24	56	78	78	80
25	56	72	78	80
26	56	78	78	82
27	56	70	78	80
28	54	74	78	80
29	58	68	76	*86
30	54	72	78	80
31	58	46	76	80
32	*62	76	78	80
33	60	76	68	76
34	52	72	76	76
35	58	46	78	76
36	60	72	76	80
37	56	*82	72	76
38	56	74	76	78
39	58	70	74	78
40	56	80	82	76
최대	62	82	88	86
최소	50	46	68	76
평균	57.15	70.8	76.45	80

*: 각 모델별 최대 예측율