

## 연결강도분석접근법에 의한 부도예측용 인공신경망 모형의 입력노드 선정에 관한 연구

이웅규

대구대학교 경영학과  
(wklee@biho.taegu.ac.kr)

손동우

대구대학교 경영학과  
(dwson@biho.taegu.ac.kr)

본 연구에서는 부도예측용 인공신경망의 입력노드 선정을 위한 휴리스틱으로 연결강도분석접근법을 제안한다. 연결강도분석은 학습이 끝난 인공신경망에서 입력노드와 은닉노드를 연결하는 연결가중치의 절대값 즉, 연결강도를 분석하여 입력변수를 선정하는 접근법으로, 선정기준에 따라 약체연결뉴론제거법과 강제연결뉴론선택법을 들 수 있다. 본 연구에서는 약체연결뉴론제거법, 강제연결뉴론선택법 그리고 이 두 기법을 통합한 통합 연결강도 모형을 제안하여 각각 의사결정트리 및 다변량판별분석에 의해 선정된 입력변수를 이용한 인공신경망 모형과 예측율을 비교한다. 실험 결과 본 연구에서 제안하고 있는 방법론이 의사결정트리나 다변량판별분석 기법 보다 높은 예측율을 보여 주었다. 특히 두 기법의 통합연결강도 모형의 경우에는 다른 단일 기법보다 높은 예측율을 보이고 있다.

### 1. 서론

부도예측(Bankruptcy Prediction)은 학문적인 면에서나 실무적인 면에서 오래 전부터 재무관리의 중요한 관심대상이 되어 왔었다(Altman, 1968; Altman 1982). 특히 1980년대 후반부터 이전까지 통계적 기법에 의존해 왔던 부도예측은 인공신경망에 의한 연구가 활발히 이루어지면서 부터, 재무관리 분야와 함께 정보시스템 분야에서도 매우 중요한 연구과제로 인식되어져 왔다.

부도예측에서 사용되는 입력변수는 Altman의 모형 이래 일반적으로 다섯 가지의 재무비율을 기준으로 한 모형이 널리 사용되어 왔으나

(Altman, 1968), 최근 10여년 간의 인공신경망에 의한 연구에서는 다섯 가지 재무비율에 국한하지 않고 매우 다양한 설명변수를 입력변수로 사용하고 있다(Zhang et. al., 1999). 특히 다변량판별분석이나 의사결정트리 또는 자기형상화지도(Self-Organizing Feature Map)와 같은 방식으로 입력변수를 선정하는 복합적(Hybrid) 접근 방식에 대한 타당성을 검증한 연구결과도 있다(Lee et. al., 1996).

본 연구에서는 부도예측용 인공신경망의 입력노드 선정을 위한 새로운 휴리스틱으로 연결강도 분석접근법을 제안한다. 연결강도분석은 학습이 끝난 인공신경망에서 입력노드와 은닉노드를 연

\* 이 논문은 2000 학년도 대구대학교 학술연구비 지원에 의한 논문임

결하는 연결가중치의 절대값 즉, 연결강도를 분석하여 입력변수를 선정하는 접근법으로 선정기준에 따라 약체연결뉴론제거법과 강제연결뉴론선택법을 들 수 있다(이응규, 손동우, 2000b). 본 연구에서는 약체연결뉴론제거법, 강제연결뉴론선택법 이외에 이 두 기법을 통합한 통합 연결강도 모형을 제안하여 각각 의사결정트리 및 다변량판별분석에 의해 선정된 입력변수를 이용한 인공신경망 모형과 예측율을 비교하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 부도예측을 위한 통계적 기법과 인공지능 기법을 살펴보고, 인공신경망 기법에서의 변수 선정을 위한 기법과 그 한계점을 논의한다. 3절에서는 본 연결강도분석 접근법의 개념과 방법을 소개하고 4절에서는 실험설계에 대해서 기술하고 5절에서는 본 연결강도분석에 의한 모형과 의사결정트리 및 다변량판별분석에 의한 모형 구축에 대해 기술하고 6절에서는 실험결과에 대한 분석을 하며 마지막으로 7절에서는 본 연구의 기여도 및 향후 연구과제에 대해 논의하도록 한다.

## 2. 문헌고찰

부도(Bankruptcy)란 기업이 파산적 지급불능 상태(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 현금흐름지표 등을 통해 미리 예측하려는 부도예측(Bankruptcy Prediction)은 실무적 차원에서 뿐 아니라 학계에서도 오랜 기간 동안 중요 관심사

였다.

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

그러나, 1980년대 후반부터 다층퍼셉트론(Multi-Layered Perceptrons) (Hertz et. al., 1991)에 의한 인공신경망의 가능성이 새롭게 인정되면서부터 인공신경망에 의한 부도예측 연구가 활발히 진행되었고, 실제로 통계적 기법보다 뛰어난 예측력을 보여주고 있음이 입증되고 있다(조홍규, 1999; Zhang et. al., 1999; Wilson and Sharda, 1994; Tsukuda and Baba, 1994; Fletcher and Goss, 1993; Udo, 1993; Raghupathi et. al., 1991; Salchengerger et. al., 1992; Tam and Kiang 1991; Odom and Sharda, 1990).

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 자기조직 형상화지도(Self-Organization Features Map) (Serrano-Cinca, 1996)와 의사결정트리(Sung, et al, 1999 ; Messier and Hansen, 1988) 등도 부도예측에 적용되어 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. 따

라서 적절한 입력노드의 선정에 대해 전문가의 직관이나 분석에서부터 통계적 분석, 의사결정트리 등 다양한 기법이 제안되고 있다(Berry and Linoff, 1997).

부도예측용 인공신경망 모형의 입력노드 선정 기법으로는 다변량판별분석, 자기형상화지도, 의사결정트리 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 선정에 시도하여 기존의 통계적 기법보다 더 좋은 예측율을 보이고 있다(Shin and Hong, 2000; 이재식, 차봉근, 1999).

### 3. 연결강도분석 접근법

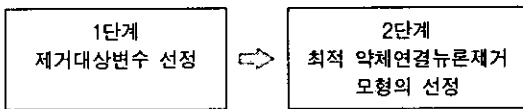
#### 3.1 연결강도분석

다층퍼셉트론의 구조를 갖는 인공신경망에서 역전파 알고리즘을 이용하여 학습을 끝냈다는 것은 입력노드와 은닉노드를 연결하는 가중치의 값들과 은닉노드와 출력노드를 연결하는 가중치에 대한 최적 값을 구했음을 의미한다. 따라서 학습이 끝난 상태의 인공신경망에서 각 입력노드와 연결되는 가중치들의 합은 해당 입력노드가 이 인공신경망의 출력값에 영향을 미치는 정도로 생각할 수 있다. 즉, 연결된 각 가중치에 대한 합이 정(+)의 값을 가질 경우 해당 입력노드는 전체 출력값에 정의 영향을 미친다고 볼 수 있고, 부(-)의 값을 가질 경우는 부의 영향을 미친다고 볼 수 있으며, 그 값이 미미할 경우에는 출력값에 대한 설명력이 약하다고 판단될 수 있다. (이용규, 손동우, 2000a).

본 연구에서는 이러한 개념을 기본으로 하여 다층퍼셉트론 구조의 인공신경망 모형에서 입력노드를 선정하는 휴리스틱으로 연결강도분석접근법을 제안한다. 연결강도분석접근법은 입력변수 수 만큼의 입력노드와 한 개의 은닉노드 그리고 한 개의 출력노드로 구성된 인공신경망을 반복적으로 학습시키면서 입력노드와 은닉노드의 연결가중치의 절대값인 연결강도를 기반으로 하여 최적 입력노드를 선정하는 것으로 접근방식에 따라 다시 약체연결뉴론제거법과 강제연결뉴론선택법 그리고 두 방법으로부터 선정된 모형을 통합한 통합 연결강도 모형이 있다.

### 3.2 약체연결뉴론제거법

약체연결뉴론제거법은 <그림 1>에서 보는 바와 같이 크게 제거대상변수 선정과 최적 약체연결뉴론제거모형 선정으로 나뉜다.



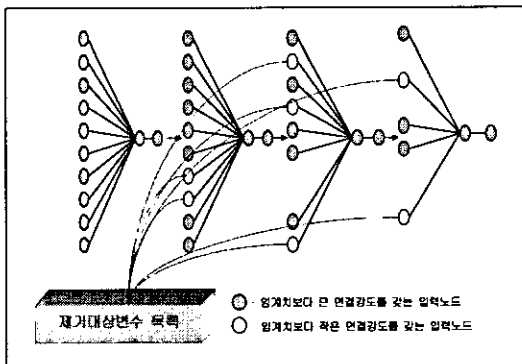
<그림 1> 약체연결뉴론제거법

#### 3.2.1 제거대상 변수 선정

제거대상변수 선정은 <그림 2>에서 보는 바와 같이 인공신경망모형을 학습시킨 후 연결강도가 일정한 값 즉, 연결강도 임계치보다 낮은 입력노드를 제외하고 나머지를 가지고 새롭게 학습을 한다. 이와 같은 과정을 다음과 같은 세 가지 가운데 한 가지가 나올 때까지 계속 반복한다.

첫째, 연결강도 값이 연결강도 임계치보다 작은 입력노드가 나오지 않는다.

둘째, 학습이 끝난 인공신경망의 성능이 더 이상 개선되지 않는다.



<그림 2> 약체연결뉴론제거법에서 제거대상변수 선정

셋째, 반복회수가 일정 회수에 도달했다.

위 방식에 의해 선정된 변수는 제거대상변수 목록에 첨가하게 된다.

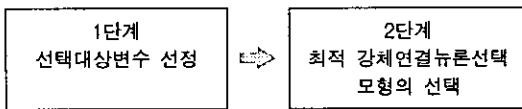
인공신경망의 학습방법은 연결가중치의 값을 무작위 방식에 의해 초기화하여 이 초기화한 연결가중치를 점진적으로 향상시키는 방식으로 전개되기 때문에 인공신경망에 의한 학습결과는 항상 무작위성을 내포하게 되어 있다. 한편 본 방법론에서는 은닉노드의 수가 항상 한 개로 고정되어 있기 때문에 최적의 결과를 얻기 위해서는 위에서 제시한 과정을 한 번 수행하는 것으로는 적절한 결과를 얻어지지 않을 수 있다. 무작위성과 고정된 은닉노드 수의 문제를 해결하기 위해서 <그림 2>에서 보여 주는 제거대상변수 과정을 하나의 라운드로 보고 이와 같은 라운드를 여러 번 반복하여 제거대상이 되는 변수가 전체 라운드에서 몇 회나 나왔는지에 대한 빈도수를 정리하는 제거대상변수 목록을 작성하게 된다.

#### 3.2.2 최적 약체연결뉴론제거 모형 선정

제거대상변수 목록이 작성되면 제거대상변수의 빈도수를 기반으로 하여 약체연결뉴론제거 모형인 WL(x)들을 구성한다. 즉, x개 이상 나온 변수들을 제외한 변수만으로 만들어진 신경망모형을 말한다. 가령 WL(8)인 경우 8번 이상 제거대상변수로 나온 변수들을 제외한 변수들만을 입력 변수로 사용하여 만들어진 인공신경망 모형이다. 각 WL(x)는 서로 다른 샘플에 일정 횟수 만큼 반복하여 각각의 성능에 대한 평균값을 구한 다음 그 가운데서 가장 좋은 모형을 최적 약체연결 뉴론제거모형(WL\*)으로 선정한다.

### 3.3 강제연결뉴론선택법

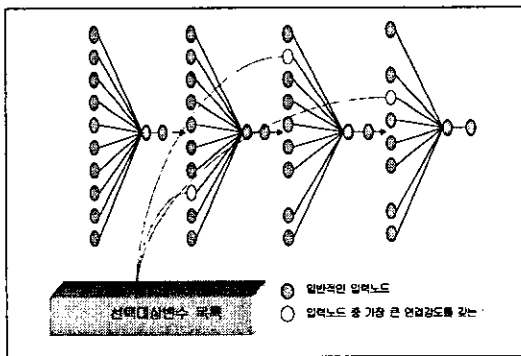
강제연결뉴론선택법의 경우에도 약체연결뉴론 제거법에서와 유사하게 <그림 3>에서 보는 바와 같이 선택대상변수 선정과 최적 강제연결뉴론선택 모형 선정의 두 단계로 나뉘어 진다.



<그림 3> 강제연결뉴론선택법

#### 3.3.1 선택대상변수 선정

약체연결뉴론제거법에서와 같이 대상이 되는 변수의 수 만큼의 입력노드와 하나의 은닉노드로 이루어진 인공신경망 모형을 이용하는데 연결강도가 낮은 뉴론을 제거하는 대신에 연결강도가 높은 뉴론을 선정하는 방식으로 전개된다. <그림 4>에서 보는 바와 같이 약체연결뉴론제거법과 마찬가지로 인공신경망모형을 학습시킨 후 연결강도가 가장 높은 값을 선정하여 이를 제외한 나머지를 가지고 새롭게 학습을 한다. 이때 인공신



<그림 4> 강제연결뉴론선택법에서 선택대상변수 선정

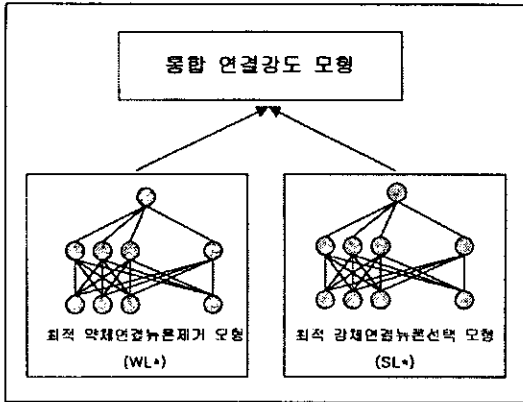
경망의 성능이 더 이상 개선되지 않거나 반복횟수가 일정횟수에 도달할 때까지 학습을 반복하여 여기서 선정된 변수들은 선택대상변수 목록에 포함한다. 약체연결뉴론제거법에서와 마찬가지로 이와 같은 라운드를 일정 횟수만큼 반복한다.

#### 3.3.2 최적 강제연결뉴론선택 모형 선정

역시 약체연결뉴론제거법에서와 유사하게 x번 이상 선택대상변수 목록에 들어간 변수만으로 입력노드를 구성한 모형  $SL(x)$ 들을 대상으로 하여 최적 강제연결뉴론선택모형( $SL^*$ )을 선정한다.

### 3.4 통합 연결강도분석 모형

통합 연결강도분석 모형은 최적 강제연결뉴론선택모형( $WL^*$ )과 최적 약체연결뉴론제거모형( $SL^*$ )을 앙상블(Ensemble)(Opiz and Shavlik, 1996) 기법에 의해 통합하는 것이다(<그림 5>). 앙상블 기법은 여러 모형의 출력값을 혼합하여 최적의 값을 구하는 기법으로 크게 세가지로 구분될 수 있는데 학습용 데이터에서 나온 가장 낮은 오차를 갖는 모형 값만을 이용하는 Bumping 기법, 각 모형의 결과값의 평균을 이용하는 Bagging 기법 그리고 각 모형의 값의 가중치를 고려한 Balancing 기법 등이 있다(Heskes, 1996). 본 연구에서는 약체연결뉴론제거법과 강제연결뉴론선택법에 대한 사전 경험치나 선험적 지식을 가질 수 없기 때문에 Balancing 기법을 적용하기는 곤란하고 또 Bumping 기법의 경우 두 기법의 통합 효과를 도모하기 어렵기 때문에 단순 평균법인 Bagging 기법에 의해 통합하도록 한다.



<그림 5> 통합 연결강도 모형

## 4. 실험설계

### 4.1 비교대상 모형

본 연구에서는 위에서 제시한 두 가지의 연결강도분석법인 최적 약체연결뉴론제거 모형과 최적 강제연결뉴론선택 모형 그리고 두 모형의 통합모형의 타당성을 비교하기 위해 C4.5 알고리즘에 의한 의사결정트리와 다변량판별분석의 설명변수 선택방식인 단계별 선택법(Stepwise Selection)에 의해 입력노드를 선정한 인공신경망모형과 비교하도록 한다.

의사결정트리는 분류(Classification) 문제에 적용할 수 있는 데이터마이닝 도구의 하나로 원래의 데이터를 클러스터별로 세분화해 나가는 과정을 재귀적(Recursive) 방식으로 진행해 나감으로 해서 하나의 나무(Tree) 형태의 결과물을 산출한다(Dahr and Stein, 1997). 본 연구에서는 데이터 세분화의 기준이 되는 알고리즘으로 C4.5를 선택했다. C4.5는 각 클러스터 내의 정보 무질서(Information Disorder)의 정도인 엔트로

피(Entropy)를 줄이려는 방식으로 세분화를 전개하는 알고리즘이다(Quinlan, 1992).

다변량판별분석의 단계별선택법은 가장 높은 상관계수 값을 갖는 변수를 설명변수로 채택하고 이에 대한 유의도를 Wilk's Lambda에 의해 유의성을 확인하는 방식으로 단계별로 진행된다(채서일, 1999).

따라서 본 연구에서는 다음과 같은 다섯 가지 모형을 대상으로 실험을 한다.

첫째, 최적 약체연결뉴론제거 모형(WL\*)

둘째, 최적 강제연결뉴론제거 모형(SL\*)

셋째, 통합 연결강도분석모형(WL\* + SL\*)

넷째, 의사결정트리(C4.5)에 의해 입력노드를 선정한 모형(C4.5+NN)

다섯째, 다변량판별분석에 의해 입력노드를 선정한 모형(MDA+NN)

한편 사용된 소프트웨어는 인공신경망과 의사결정트리의 경우 엔터프라이즈 마이너(Enterprise Miner) 3.0이 이용되었고 다변량판별분석의 경우 SAS 6.12가 이용되었다.

### 4.2 사용 변수

처음 부도예측을 통계적인 방법으로 시도한 Altman(1968)의 경우 다섯 가지의 재무비율(순운전자본구성비율, 이익잉여금구성비율, 총자산영업이익율, 장부가치, 총자산회전율)을 사용하고 있지만 인공신경망에 의해 부도예측을 시도하면 서부터 13개의 재무비율 변수(Raghupathi et. al., 1998) 29개의 재무비율 변수(Salchengenger et. al., 1992) 또는 19개의 재무비율 변수(Tam and Kiang, 1992)를 사용하는 등 매우 다양한 방식으로 접근하고 있다. 한편 Shin and Hong(2000)은 재무비율 이외에 현금흐름지표를 변수에 포함하

고 있다. 본 연구에서는 43개의 재무비율과 현금 흐름지표를(부록 참조) 대상으로 하여 적절한 입력변수를 선정한다.

### 4.3 부도여부의 결정

부도예측 문제는 기업의 부도 또는 건전 여부를 결정하는 일종의 분류 문제이기 때문에 최종 결과물의 값은 항상 0 또는 1과 같은 이진값을 갖어야 한다. 따라서 0에서 1사이의 소수를 갖는 인공신경망의 출력노드는 일정한 값을 기준으로 하여 0 또는 1의 값으로 전환시켜야 한다. 본 연구에서는 0.5를 기준으로 하여 0.5와 같거나 크면 1 즉, 건전으로 간주하고 0.5 보다 작으면 0 즉, 부도로 간주한다.

### 4.4 샘플링 방법

전체 475개의 국내 기업(부도기업 238개, 건전기업 237개)을 대상으로 하였는데 이 가운데 무작위로 추출한 100개(부도기업 56개, 건전기업 44개)는 스코어링을 위해 사용하였고 나머지 375개를 최적모형을 도출하기 위한 학습(Train), 검증(Validation) 및 테스트(Test)용으로 사용하였다.

스코어링을 위해 선정된 100개를 제외한 나머지 375개 가운데서 다시 100개 기업(건전기업 50개, 부도기업 50개)을 무작위로 추출하여 제거대상변수 목록 및 선택대상 변수 목록을 만드는 데 사용하였다.

또 최적 약체연결뉴론제거 모형(WL\*)과 최적 강체연결뉴론선택 모형(SL\*)를 구하기 위해 375개의 데이터를 다시 학습용 150개 검증용 150개 그리고 테스트용 75개를 10회에 걸쳐 무작위 추

출하는 방식으로 모형을 구축하고 각 모형들의 적중율의 평균값을 구하여 각각의 최적 모형을 구하였다. 다변량판별분석과 의사결정트리에 의한 모형의 경우에도 375개의 데이터를 이용하여 모형을 구축하였다.

최종적으로 앞에서 언급한 375개를 이용하여 다섯 가지 비교모형에 대한 학습 및 검증을 하였고 100개의 스코어링용 데이터를 이용하여 각 모형의 성능을 비교분석하였다.

## 5. 모형의 구축

### 5.1 최적 약체연결뉴론제거 모형(WL\*)

총 라운드 수(R)은 10회로 하였고 라운드별 제한횟수(K)도 10차로 제한하였으며 연결강도임계치( $\alpha$ )는 0.05로 하여 실험을 하였다.

<표 1>은 첫번째 라운드에서 선정된 제거변수와 각각의 연결강도를 보여 주고 있다. 1차의 실험결과에서는 X5, X10를 비롯한 총 10개, 2차에서는 X6, X7를 비롯한 4개, 3차에서는 X2, X3을 비롯한 2개의 제거대상변수가 선정되는 등, 제1라운드의 실험에서는 총 32개의 변수가 제거대상변수로 선정되었으며, 9차에서는 해당변수가 없었기 때문에 제1라운드의 실험을 중단하였다. 같은 방식으로 총 10 라운드에 걸친 실험을 하여 <표 2>와 같은 제거대상변수 목록을 작성하였다.

제거대상변수 목록에 10회, 9회, 8회, 7회, 6회, 5회, 4회, 2회, 1회에 걸쳐서 제거대상변수로 나타난 변수를 각각 누적하여 제거한 후 나머지 변수를 입력변수로 약체연결뉴론제거모형인 각각의 WL(x) 모형을 얻었다. 즉, WL(x)는 x회 이

<표 1> 약체연결뉴론제거법의 1 라운드 결과

1차		2차		3차		4차		5차		6차		7차		8차	
제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도	제거 대상	연결 강도
X5	0.0397	X6	0.0092	X2	0.008	X4	0.0452	X19	0.0031	X18	0.0072	X12	0.0234	X11	0.0269
X10	0.0052	X7	0.0256	X3	0.0186	X8	0.0413	X30	0.0257	X28	0.0345	X33	0.0408		
X13	0.0456	X16	0.0359			X9	0.0137	X36	0.0348	X41	0.082				
X14	0.0089	X31	0.0101			X20	0.0326	X43	0.0308						
X15	0.0262					X25	0.0058								
X17	0.0279					X32	0.0337								
X34	0.0381														
X37	0.0063														

<표 2> 제거대상변수 목록

빈도수	제거대상변수
10	X36
9	X15, X19, X28, X42
8	X10, X14, X35, X38, X41
7	X5, X9, X11, X16, X18, X22, X31, X32, X37, X39
6	X2, X3, X7, X8, X12, X17, X23, X33X34, X40
5	X1, X6, X20, X25, X30, X43
4	X4, X13, X26, X427
2	X21
1	X24, X29

<표 3> 약체연결뉴론제거 모형들의 평균 예측율

모형	WL(4)	WL(5)	WL(6)*	WL(7)	WL(8)	WL(9)
평균 예측율	68%	71%	80.4%	78.4%	78.7%	76.9%

\*: 최댓값

상 제거대상변수로 선정된 변수를 제외한 변수들을 입력노드로 하여 구축된 모형이다. WL(x) 모형들의 은닉노드의 수는 각 입력노드의 수와 같도록 구조를 설정하여 10회에 걸쳐 실험을 하여 <표 3>과 같은 결과를 얻었다. <표 3>에서

보는 바와 같이 WL(6)가 가장 높은 예측율(80.4%)을 보여 WL(6)를 최적 약체연결뉴론모형(WL\*)으로 선정하였다.

최적 약체연결뉴론제거 모형(WL\*)에서 선정된 입력변수는 <표 4>와 같다.



<표 4> 최적 약체연결뉴론제거

지표	변수명	변수내역
규모	X1	매출액
생산성	X4	부가가치율
생산성	X6	재고자산증가율
수익성	X13	순금융비용대 매출액비율
안정성	X20	당좌비율
안정성	X21	매출채권대 매입채무
안정성	X24	유동부채대 총자본
안정성	X25	유보액대총자산비율
안정성	X26	자기자본비율
안정성	X27	차입금의존도
추세	X29	금융비용부담금율 증가율
활동성	X30	매입채무회전율
현금흐름	X43	외부자금조달후 CF/금융비용

같은 방식으로 총 10라운드에 걸친 실험한 결과 <표 6>과 같은 선택대상변수 목록을 얻었다.

<표 6> 선택대상변수 목록

빈도수	변수
0	X5, X8, X9, X10, X11, X12, X14, X17, X32, X35
1	X1, X2, X13, X19, X24, X25, X31, X36, X37, X40, X42, X43
2	X15, X16, X18, X21, X22, X38, X41
3	X4, X20, X39
4	X6, X23, X26, X28, X34
5	X30
6	X3
7	X27, X33
10	X7, X29

## 5.2 최적 강체연결뉴론선택 모형(SL\*)

최적 약체연결뉴론제거모형과 유사하게 라운드 수 (R)은 10회로 라운드별 제한횟수 (K)도 10차로 하여 진행하였다. 1라운드 수행결과, 1회에서 X29, 2회에서 X26, 3회에서 X7, 4회에서 X27 등, 10회에 걸친 실험에서 <표 5>와 같이 10개의 변수가 선택대상변수로 선정되었다.

<표 5> 라운드 선택대상변수 목록

치수	1라운드 선택대상변수
1	X29
2	X26
3	X7
4	X27
5	X16
6	X33
7	X13
8	X34
9	X23
10	X3

<표 6>에서 보는 바와 같이 10회에 걸쳐서 선택대상변수로 선정된 변수는 X7, X29 두 개의 변수이고, 9회와 8회 선정된 변수는 없었으며, 7회 선택대상변수로 선정된 것은 X27, X33이며 6회 선정된 변수는 X3이다. 이를 기반으로 x회 이상 선택대상변수 목록에 포함된 변수만으로 입력노드를 선정한 SL(x) 즉, 강체연결뉴론선택 모형을 구성하여 WL(x)의 경우와 마찬가지로 입력노드 수와 은닉노드 수를 같게 하여 10회에 걸쳐 실험을 하였다. 역시 WL(x)의 경우에서와 마찬가지로 375개의 데이터를 10회에 무작위 추출을 하여 각기 샘플의 구성을 달리하여 실험하여 <표 7>과 같은 결과를 얻었다. <표 7>에서 보는 바와 같이 SL(6)가 가장 높은 예측율(74.6%)을 보여 SL(6)를 최적 강체연결뉴론선택 모형(SL\*)로 선정하였다.

<표 7> 강제연결뉴론선택 모형들의 평균 예측율

모형	WL(4)	WL(5)	WL(6)*	WL(7)	WL(8)	WL(9)
평균 예측율	68%	71%	80.4%	78.4%	78.7%	76.9%

\*: 최대값

최적 강제연결뉴론선택 모형(SL\*)에서 선정된 입력변수는 <표 8>과 같다.

<표 8> 최적 약체연결뉴론제거

지표	변수명	변수내역
생산성	X3	총자본투자효율
생산성	X7	총자산증가율
안정성	X27	차입금의존도
추세	X29	금융비용분담금율증가율
활동성	X33	재고자산회전율

### 5.3 의사결정트리 및 다변량판별분석에 의한 모형

<표 9>와 <표 10>에서 보는 바와 같이 의사결정트리의 경우 X25, X13, X38, X4 등 4개의 변수가 선정되었고, 다변량판별분석의 경우 X3, X5 등 16개의 변수가 선정되었다.

<표 9> 의사결정트리에 의해 선정된 변수

지표	변수명	변수내역
안정성	X25	유액대총자산비율
수익성	X13	순금융비용대 매출액비율
현금흐름 지표	X38	(현금영업이익/금융수입)/금융비용
생산성	X4	부가가치율

<표 10> 의사결정트리에 의해 선정된 변수

지표	변수명	변수내역
생산성	X3	총자본투자효율
생산성	X4	부가가치율
성장성	X5	매출액증가율
성장성	X6	재고자산증가율
성장성	X7	총자산증가율
수익성	X11	매출원가율
수익성	X13	순금융비용대 매출액 비율
수익성	X15	자기자본순이익율
안정성	X18	고정장기적합율의 역
안정성	X20	당좌비율
안정성	X23	유동비율
안정성	X24	유동부채대 총자본
안정성	X26	자기자본비율
활동성	X31	순운전자본대 매출액
활동성	X32	운전자금대 회전율
규모	X1	매출액

### 6. 실험결과에의 분석

각 모형의 예측율을 얻기 위해 이제까지 훈련이나 검증에 전혀 사용되지 않은 데이터 100개를 스코어링용으로 사용하였다. 각 모형별 예측율 및 쌍체검증 결과는 <표 11>과 같다. 여기서 통합 모형의 경우 최적 약체연결뉴론제거모형(WL\*)의 출력노드값과 최적 강제연결뉴론선택 모형(SL\*)의 출력노드값의 평균을 구하여 얻었다. 즉, 두 모형의 부도 여부의 판정값인 0과 1이 아니라 0에서 1사이의 출력노드값을 평균하여 그 값이 0.5보다 같거나 크면 건전, 0.5보다 작으면 부도로 판정하였다.

WL\*의 경우 86%, SL\*의 경우는 80%을 보임

<표 11> 모형별 예측율 및 모형간의 쌍체검증(z-검증) 결과

모형명(예측율)	WL*(86%)	SL*(80%)	WL*+SL*(89%)
의사결정트리(79%)	2.053727**	2.9646*	3.500354*
다변량판별분석(68%)	0.192199	1.255806***	1.779624**

\*: 통계적 유의수준 1%  
 \*\*: 통계적 유의수준 5%  
 \*\*\*: 통계적 유의수준 10%

으로써 다변량판별분석의 68% 보다 통계적으로 유의적 차이를 보여 주고 있다. 반면 의사결정트리(79%)의 경우 WL\*는 통계적으로 유의적 차이를 보낼 만큼 나은 예측율을 보이고 있으나 SL\*의 경우 통계적으로 유의성을 가지지 못할 만큼의 차이 밖에는 보여 주지 못하고 있다. 한편 통합모형의 경우 WL\*나 SL\*보다 월등한 좋은 예측율인 89%를 보여 주고 있으면서 상당히 낮은 유의수준에서 다른 기법들과 차이를 보여 주고 있다.

## 7. 결론

본 연구에서는 부도예측용 인공신경망 모형의 입력노드를 선정하기 위한 방법론으로 연결강도 분석에 의한 약체뉴론제거법과 강체뉴론선택법 그리고 통합 연결모형을 제안하였고, 이 들 방법에 의해 입력노드를 선정한 인공신경망 모형이 C4.5에 의한 의사결정트리나 다변량판별분석에 의해 입력노드를 선정한 인공신경망 모형보다 더 나은 예측율을 보이고 있음을 실증하였다. 특히 두 기법의 통합 연결모형은 다른 기법들보다 월등한 예측율을 보이고 있다.

향후 연구과제로는 우선 강체연결뉴론과 약체연결뉴론을 정의하는 방법론적인 보완을 들 수

있다. 특히 약체연결뉴론을 결정하는 임계치에 대해 민감도 분석을 통해 최적의 임계치를 도출하는 것이나 또는 연결강도분석을 일련의 의사결정시스템으로 구축하여 임계치 선택이나 모형 선택을 학습 기반으로 생성시키는 메커니즘을 개발하는 것은 중요한 연구과제가 될 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 나라에의 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. 본 연구에서는 대부분의 경우 입력노드와 은닉노드의 수를 같도록 모형의 구조를 만들었고, 또 연결강도 분석에서는 은닉노드의 수를 하나로 고정하였다. 이러한 방식은 사실 여러 번의 실험에 걸친 결과이긴 하지만 완전한 최적성을 보장할 수 없음을 인정한다.

다섯번째, 최적 약체연결뉴론제거 모형이나 최적 강제연결뉴론선택 모형 그리고 의사결정트리나 다변량판별분석에 의한 모형에서 선정된 입력 변수에 관한 재무관리나 회계적인 분석도 흥미로운 연구과제가 될 수 있다.

마지막으로 자기조직형상화 지도, 유전자 알고리즘(Genetic Algorithm) 등과 같은 인공지능 기법들과도 그 성능을 비교해 볼 필요가 있을 것으로 보인다.

## 참고문헌

- [1] 박정식, 신동령, '경영분석', 다산출판사, 1999.
- [2] 이종규, 손동우, "부도에측용 인공신경망모형의 최적 입력노드 설계: 연결강도판별분석 접근", 2000 춘계정기학술대회, 한국지능정보시스템학회, 2000a.
- [3] 이종규, 손동우, "연결강도판별분석에 의한 부도에측용 신경망 모형의 입력노드 설계: 강제연결뉴론 선정 및 약체연결뉴론 제거 접근법", 2000 추계정치학술대회, 한국지능정보시스템학회, 2000b.
- [4] 이재규, 최형림, 김현수, 서민수, 주석진, 지원철, '전문가시스템 원리와 개발', 법영사, 1998.
- [5] 이재식, 차봉근, "유전자 알고리즘을 이용한 인공신경망의 구조 설계", 한국경영과학회지, 24, 3, 1999.
- [6] 조홍규, "도산예측을 위한 인공지능 방법과 통계적 방법의 통합 방법론," 한국과학기술원(박사학위논문) 1999.
- [7] 채서일, '사회과학 조사방법론', 학현사, 1999., 2판.
- [8] Altman, Edward I, "Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy", Journal of Finance, Sep.1968, pp.589-609.
- [9] Altman, Edward I, "Accounting Implications of Failure Prediction Models", Journal of Accounting, Auditing and Finance 18, 1982, pp505-529.
- [10] Berry, Michael J. A. and Gordon Linoff, 'Data Mining Techniques: For Marketing, Sales, and Customer Support', John Wiley and Sons, 1997.
- [11] Dahr, Vasant and Roger Stein, Intelligent Decision Support Methods: The Science of Knowledge Work, Prentice Hall, Upper Saddle River, NJ, 1997.
- [12] Fietcher, D. and E. Goss, "Forecasting with Neural Networks: an Application using Bankruptcy Data", Information and Management, 24, 3, 1993, pp.159-167
- [13] Hertz, John, Anders Krogh and Richard G. Palmer, 'Introduction to the Theory of Nueral Computation', Addison Wesely, 1991.
- [14] Heskes, T., "Balancing between Bagging and Bumping", In Mozer, M. C., M. I. Jordan, and T. Petsche, (eds.) Advances in Neural Information Processing Systems 9, The MIT Press, 1996.
- [15] Lee, Kun Chang, Ingoo Han and Youngsig Kwon, "Hybrid Neural Network Models for Bankruptcy Predictions", Decision Support Systems, 18, 1996, pp.63-72.
- [16] 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.
- [17] Odom, M. and R. Sharda, "A Neural network Model for Bankruptcy Prediction", In Proceedings of the IEEE International Conference on Neural Network, 1990, pp 163-168.
- [18] Ohlson, J. "Financial Ratios and The Probabilistic Prediction of Bankruptcy",

- Journal of Accounting Research, 18. 1, 1980, pp.109-131.
- [19] Opitz D. W., and J. W. Shavlik, "Generating Accurate and Diverse Members of Neural-Network Ensemble", In Mozer, M. C., M. I. Jordan, and T. Petsche, (eds.) Advances in Neural Information Processing Systems 9, The MIT Press, 1996.
- [20] Quinlan, J.R., C4.5: Programs for Machine Learning, Morgan Kaufman, Mountain View, CA, 1992.
- [21] Raghupathi, W., L. L. Schkade, and B. S. Raju, "A Neural Network Approach to Bankruptcy Prediction", In Proceedings of the 24th Hawaii International Conference on System Science, 1991, pp 147-155.
- [22] Salchengerger, L. M., E. M. Cinar and N. A. Lash, "Neural Networks: A New Tool for Predicting Thrift Failures", Decision Sciences Vol 24, No 4, 1992, pp899-916.
- [23] Serrano-Cinca, Carlos, "Self Organizing Neural Networks for Financial Diagnosis", Decision Support Systems, 17, 1996, pp.227-238.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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
- [28] Udo, G., "Neural Network Performance on the Bankruptcy Classification Problem", Computers and Industrial Engineering, 25(1-4).1993, pp.377-380.
- [29] Wilson, Rick L. and Ramesh Sharda, "Bankruptcy Prediction using Neural Networks", Decision Support Systems, 11. 1994., pp. 545-557.
- [30] Zhang, Guoqiang, Michael Y. Hu, B. Eddy Patuwo, and Daniel C. Indro, "Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis", European Journal of Operational Research, Vol 116, 1999, pp16-32,
- [31] Zmijewski, Mafke, "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", Journal of Accounting Research, Vol.22 1984,

## 부록 (전체변수 내역)

지표	변수명	변수내역	지표	변수명	변수내역
X1	규모(1)	매출액	X23	안정성(5)	유동비율
X2	규모(1)	자기자본	X24	안정성(5)	유동부채대 총자본
X3	생산성(2)	총자본투자효율	X25	안정성(5)	유보액대 총자산비율
X4	생산성(2)	부가가치율	X26	안정성(5)	자기자본비율
X5	성장성(3)	매출액증가율	X27	안정성(5)	차입금의존도
X6	성장성(3)	재고자산증가율	X28	안정성(5)	총차입금대 매출액
X7	성장성(3)	총자산증가율	X29	추세(6)	금융비용부담금율 증가율
X8	수익성(4)	금융비용대 매출액비율	X30	활동성(7)	매입채무회전율
X9	수익성(4)	대출효율성계수	X31	활동성(7)	순운전자본대 매출액
X10	수익성(4)	매출액순이익률	X32	활동성(7)	운전자금대 회전율
X11	수익성(4)	매출원가율	X33	활동성(7)	재고자산회전율
X12	수익성(4)	손익분기점율	X34	활동성(7)	총자본회전율
X13	수익성(4)	순금융비용대 매출액비율	X35	현금흐름지표(8)	(영업활동후 CF/차입금합계)*100
X14	수익성(4)	이자보상비율	X36	현금흐름지표(8)	(영업활동후 CF/부채총계)*100
X15	수익성(4)	자기자본순이익률	X37	현금흐름지표(8)	영업활동후 CF/금융비용
X16	수익성(4)	총자본경상이익률	X38	현금흐름지표(8)	(현금영업이익/금융수입)/금융비용
X17	수익성(4)	총자본순이익률	X39	현금흐름지표(8)	(이자지급후 CF/부채총계)*100
X18	안정성(5)	고정장기집합율의 역	X40	현금흐름지표(8)	이자지급후 CF/금융비용
X19	안정성(5)	단기부채대 총차입금	X41	현금흐름지표(8)	(외부자금조달전 CF/차입금합계)*100
X20	안정성(5)	당좌비율	X42	현금흐름지표(8)	외부자금조달전 CF/금융비용
X21	안정성(5)	매출채권대 매입채무	X43	현금흐름지표(8)	외부자금조달후 CF/금융비용
X22	안정성(5)	순운전자본비율			

Abstract

## Selection of Input Nodes in Artificial Neural Network for Bankruptcy Prediction by Link Weight Analysis Approach

Woong-kyu Lee\*  
Dong-woo Son\*

Link weight analysis approach is suggested as a heuristic for selection of input nodes in artificial neural network for bankruptcy prediction. That is to analyze each input node's link weight-absolute value of link weight between an input node and a hidden node in a well-trained neural network model. Prediction accuracy of three methods in this approach, - weak-linked-neurons elimination method, strong-linked-neurons selection method and integrated link weight model - is compared with that of decision tree and multivariate discrimination analysis. In result, the methods suggested in this study show higher accuracy than decision tree and multivariate discrimination analysis. Especially an integrated model has much higher accuracy than any individual models.

---

\* Department of Business Administration, Daegu University