

연결강도분석을 이용한 통합된 부도예측용 신경망모형

이웅규(Woongkyu Lee)

대구대학교 경영학과 053-850-6272 woong3041@empal.com

임영하(Young Ha Lim)

한국통신 정보통신본부

Abstract

This study suggests the Link weight analysis approach to choose input variables and an integrated model to make more accurate bankruptcy prediction model. the Link weight analysis approach is a method to choose input variables to analyze each input node's link weight which is the absolute value of link weight between an input nodes and a hidden layer. There are the weak-linked neurons elimination method, the strong-linked neurons selection method in the link weight analysis approach. The Integrated Model is a combined type adapting Bagging method that uses the average value of the four models, the optimal weak-linked-neurons elimination method, optimal strong-linked neurons selection method, decision-making tree model, and MDA.

As a result, the methods suggested in this study - the optimal strong-linked neurons selection method, the optimal weak-linked neurons elimination method, and the integrated model - show much higher accuracy than MDA and decision making tree model. Especially the integrated model shows much higher accuracy than MDA and decision making tree model and shows slightly higher accuracy than the optimal weak-linked neurons elimination method and the optimal strong-linked neurons selection method.

I. 서 론(연구목적 및 문헌고찰)

오늘날과 같이 급변하는 환경 하에서 기업의 부실화는 기업의 규모나 그 기업이 속한 산업에 관계없이 발생하는 현상이다. 특히, 최근 대기업들의 잇단 도산은 이제 대기업은 도산이라는 위협으로부터 안전하다는 일반인들의 인식마저도 바꾸어 놓고 있다.

기업의 부실화는 금융기관, 투자자, 채권자, 종업원, 거래당사자 등의 직접적인 이해관계자들에 대한 경제적 손실뿐만 아니라 실업, 경기침체, 성장둔화 등의 사회적 비용을 수반하기 때문에 주요 관심사가 되고 있다. 특히 기업에 자금을 대출해주는 금융기관, 기업에 투자하는 투자자들에게 있어서 기업의 부실예측은 매우 절실하고 중요한 문제라 할 수 있다.

기업의 부실화와 관련된 연구는 크게 기업 부실화의 원인과 징후에 대한 연구와 부실기업 예측에 관한 연구로 구분될 수 있다. 부실화 원인과 징후에 대한 연구는 도산이라는 현상을 체계적으로 이해하는데 매우 중요한 분야이나 이를 일률적으로 정형화 할 수 없는 요소가 많아 과학적인 연구방법을 적용하기에 많은 어려움이 따른다. 부실기업 예측 연구는 크게 기업 부실을 잘 설명해주는 변수들에 대한 연구와 이들 변수를 이용한 부실예측 모형의 구축에 관한 연구로 구분될 수 있다.

기업의 부실예측 연구는 한편으로는 기업에 자금을 대출해주는 금융기관, 기업에 투자하는 투자자, 거래처, 그리고 종업원들은 물론 경영자들을 포함하여 기업의 생사에 관계되는 모든 이해관계자들의 의사결정에 매우 유용한 정보를 제공할 수 있을 뿐만 아니라, 해당기업의 위험신호를 조기에 발견하고 적절한 행동을 수행하게 하는 신호로서의 목적에도 활용될 수 있다.

기업 도산 현상은 특정 시점에 갑자기 나타나는 것이 아니고 기업의 내·외적 요인들이 복합적으로 작용하여 차츰 도산화 과정을 걸어간다고 볼 수 있다. 또한 기업의 도산 가능성에 대한 징후 또는 정보는 사전에 노출되는 경우가 많다. 즉 어느 경우에도 먼저 도산을 야기하는 원인이 있을 것이며, 이 원인이 발생한 얼마 후 도산 가능성을 예측할 수 있는 징후가 나타나기 시작한다. 이러한 기업 도산의 근본적인 원인들은 표면적으로는 판매부진, 자기자본부족, 생산저하, 고정자산 과다투자, 부실채권 과다 등의 재무적 도산 징후로 나타나기도 하고, 종업원의 사기 저하와 같은 비재무적인 도산 징후로 나타나기도 한다.

그러므로 위와 같은 재무적·비재무적 징후를 이용하여 기업 도산 과정을 사전에 예측하거나 잠재적 도산 가능성을 평가할 수 있다면 도산 가능성의 조기 경고를 통해 경영정책의 변화나 재무구조의 재조직 등과 같은 예방 대책을 적극적으로 강구함으로써 도산에 따른 직·간접적 비용과 비효율성을 최소화 할 수 있다.

이러한 경제·사회적으로 중대한 의미를 지닌 기업 도산에 관한 연구는 재무관리분야와 함께 정보시스템분야에서 매우 중요한 연구과제로 인식 되어져 왔다. 1960년대부터 미국 등 선진국을 중심으로 본격화하여 1966년 에 비버(Beaver)에 의해 1954년부터 1964년 사이에 도산한 79개 기업과 79개 정상 기업을 대상으로 두 집단의 재무비율 평균의 차이를 분석하였고,

1968년에 알트만(Altman)은 판별분석을 이용하여 1946년부터 1965년 사이의 33개의 도산 기업과 33개의 정상 기업을 대상으로 기업 도산 예측을, Ohlson(1980), Martin(1977), Zimijewski(1984), Zavgren(1988) 등은 확률모형인 로짓 분석(Logit Analysis)과 프로빗 분석(Probit Analysis)을 이용하여 기업 도산 예측을 연구하였다. 국내에서도 1980년대에 들어와 전춘옥(1984) 등이 판별분석 등 통계적 기법을 사용하여 선형 결합하는 기업 도산 예측 모형에 관한 실증적 연구를 하였다. 그러나 이러한 기업 도산 예측 모형이 독립변수에 대한 정규성(정규분포), 각 그룹간의 공분산을 가져야 하는 통계적 제약 및 환경의 변화에 따른 예측력 하락 등의 문제점이 있는 것으로 지적되고 있다.

이의 대안의 하나로써 등장한 인공신경망을 이용한 기업 도산 예측모형은 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 기법보다 뛰어난 예측력을 보여주고 있음이 입증되고 있다.

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 알려지고 있다.

또한 최근의 연구들은 인공신경망과 기존 통계기법의 성과를 비교, 분석하는 연구에서 나아가 다양한 방법론들을 통해 모형의 예측력을 향상시키기 위한 방안들을 제시하고 있다. 이재식과 한재홍(1995)은 비재무정보를 입력변수에 반영함으로써 재무정보만을 이용한 모형에 비해 10% 정도 증가된 예측률을 보이고 있다. 이재식과 차봉근(1996)은 인공신경망 모형의 최적의 변수를 선정하기 위하여 유전자 알고리즘을 사용하고 있다. 이견창 등(1994)의 연구에서는 인공신경망과 다변량 판별분석(MDA: Multivariate Discrimination Analysis), 귀납적 학습방법(ID3: Iterative Dichotomizer 3), 자기 조직화 형상지도(SOFM: Self Organizing Feature Map) 등의 입력변수 선정 기법과의 통합 방법론을 적용하였는데, 실증분석 결과 향상된 예측성적을 보이고 있다.

인공신경망에 의한 모형화 과정에서 가장 중요한 고려사항 중의 하나는 적절한 입력노

드의 선정인데, 그 이유는 지나치게 많은 입력노드는 많은 학습시간을 요구할 뿐 아니라, 또 입력변수들 가운데 노이즈가 섞이게 되는 경우에는 적절치 못한 해(Solution)를 도출할 가능성이 높기 때문에 적절한 입력노드의 선정을 위해 전문가의 직관이나 분석에서부터 통계적 분석, 의사결정트리 등 다양한 기법이 제안되고 있다.

또한 기존의 많은 연구들이 여러 가지 분류기법을 적용하여 그 중에서 하나의 기법이 주어진 문제에 대해 좋은 성과를 보인다는 결과를 발표 하고 있다. 그러나 많은 연구결과들이 특정한 문제에 대한 동일한 결론을 도출하지 못하고, 어떤 문제에 가장 적합한 기법이 무엇인지에 대해서는 아직 논란의 여지가 남아 있으며, 대부분 입력변수 선정 방법과 인공신경망과의 통합에 대한 연구를 수행하고 있다.

이에 이 연구에서는 연결강도분석법에 의해 선정된 입력변수를 이용한 인공신경망 모형을 구축하여 기존의 전통적 방법에 의한 입력변수 선정 모형과 예측률을 비교한다. 또한 기업 부도 예측에 있어서 많은 변인들이 매우 다양하게 적용되는데 이들의 변인들을 단일 모형으로 만 평가하는 것보다, 입력변수 군을 달리하는 다수의 인공신경망 모형의 결과를 통합함으로써 예측력 향상에 기여할 수 있는 방법론을 제시한다.

II. 연결강도분석 접근법

2.1 기존 방법론 고찰

인공신경망에 의한 모형화 과정에서 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 통계적분석, 의사결정트리 등 다양한 기법이 연구되고 있다

부도예측용 인공신경망 모형의 입력노드 선정기법으로는 다변량 판별분석, 자기형상화지도, 의사결정트리 등을 들 수 있는데 이들 기법을 이용하여 선 처리를 실시한 인공신경망 모형이 그렇지 않은 경우보다 높은 예측력을 보여 주고 있는 것으로 알려져 있다.

그러나, 기존의 입력변수 선정기법은 나름대로의 한계점을 보유하고 있는데, 다변량 판별분석의 경우 독립변수들의 집단이 다변량 정규분포를 이루면서 공분산이 같아야 하는 가정이 필요하기 때문에 이에 대한 검증을 필요로 한다. 의사결정 트리의 경우 독립변수와 종속변수의 관계를 쉽게 설명할 수 있는 IF~THEN 규칙을 생성하는 장점이 있어서 결과를 쉽게 이해하고 설명하고자 할 때 적합한 방법이지만, 독립변수가 취할 수 있는 값이 많아질 경우에는 오류가 발생할 가능성이 많다는 약점을 가지고 있다. 그 외에 자기 조직화 형상지도에 의한 변수 선정방법은 자율학습(Un-supervised Learning)을 채택하고 있기 때문에 자칫 부도예측과 상관이 없는 변수 분류를 초래할 가능성이 있다. 한편 최근 유전자 알고리즘과 인공신경망의 결합에 의한 연구가 활발한데 부도예측에서도 유전자 알고리즘에 의한 입력변수 선정을 시도하여 기존의 통계적 기법보다 더 좋은 예측률을 보이고 있다.

2.2 연결강도분석 개념

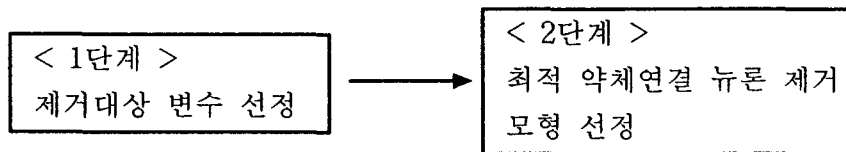
다층퍼셉트론의 구조를 갖는 인공신경망에서 역전파 알고리즘을 이용하여 학습을 끝

냈다는 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 출력 값에 영향을 미치는 정도로 생각할 수 있다. 즉, 연결된 각 가중치에 대한 절대값의 합이 정(+)의 값을 가질 경우 해당 입력노드는 전체 출력 값에 정의 영향을 미친다고 볼 수 있고, 부(-)의 값을 가질 경우는 부의 영향을 미친다고 볼 수 있으며, 그 값이 미미할 경우에는 출력 값에 대한 설명력이 약하다고 판단될 수 있다

이러한 개념을 기본으로 하여 다층퍼셉트론 구조의 인공신경망 모형에서 입력노드를 선정하는 휴리스틱이 연결강도분석 접근법이다. 연결강도분석 접근법은 입력변수 수만큼의 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 약체연결 뉴런 제거법과 강제연결 뉴런 선택법으로 구분된다.

2.3 약체연결 뉴런 제거법

약체연결 뉴런 제거법은 인공신경망의 각 연결강도가 작은 입력노드는 전체 출력 값에의 영향이 미미할 것이라는 가설 하에 입력노드와 연결되는 가중치의 합에 대한 절대값인 연결강도판별식의 일정값 즉, 연결강도판별 임계치(Linked Weight Discrimination Cut-off Value)보다 작은 경우 해당 입력노드를 제거함으로써 입력노드에 관한 최적 인공신경망 구조를 찾으려는 입력노드 선정 방법이다. 약체연결 뉴런 제거법은 [그림 2-1]에서 보는 바와 같이 크게 제거대상 변수 선정과 최적 약체연결 뉴런 제거 모형 선정으로 나뉜다.



[그림 2-1] 약체연결 뉴런 제거법

2.3.1 제거대상 변수 선정

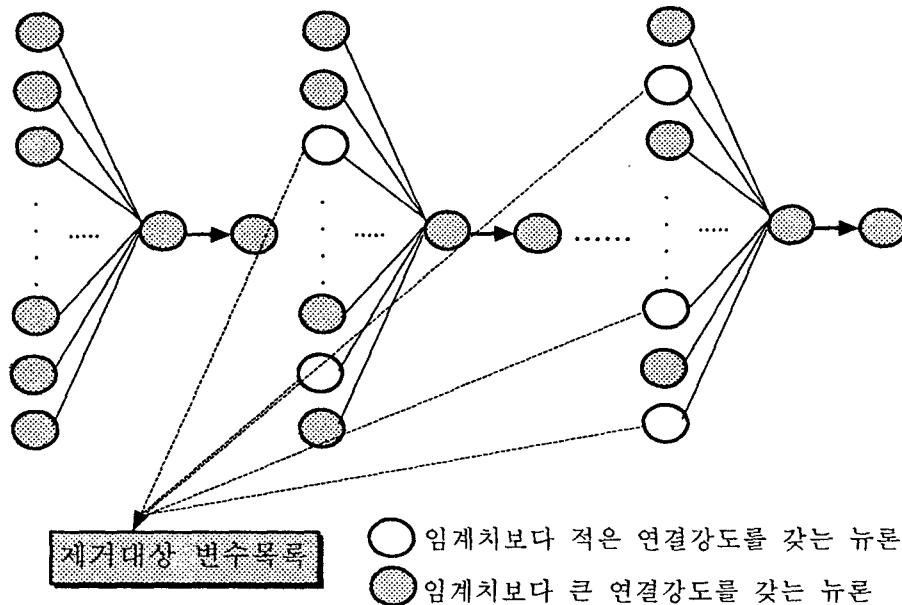
제거대상 변수 선정은 [그림 2-2]에서 보는 바와 같이 인공신경망 모형을 학습시킨 후 연결강도가 일정한 값 즉, 연결강도 임계치 보다 낮은 입력노드를 제외하고 나머지를 가지고 새롭게 학습을 한다. 이와 같은 과정을 다음과 같은 세 가지 가운데 한 가지가 나올 때까지 계속 반복한다.

첫째, 연결강도 값이 연결강도 임계치 보다 작은 입력노드가 나오지 않는다.

둘째, 인공신경망의 성능이 더 이상 개선되지 않는다.

셋째, 반복회수가 일정 회수에 도달했다.

위 방식에 의해 선정된 변수는 제거대상 변수 목록에 첨가하게 된다.



[그림 2-2] 약체연결 뉴런 제거법에서 제거대상 변수 선정

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 이 방법론에서는 은닉노드의 수가 항상 한 개로 고정되어 있기 때문에 최적의 결과를 얻기 위해서는 한 번의 라운드만으로는 적절한 결과를 얻어지지 않을 수 있다. 무작위성과 고정된 은닉노드 수의 문제를 해결하기 위해서 위에서 제시한 과정을 일정 라운드만큼 반복하여 제거대상이 되는 변수가 전체 라운드를 통해 몇 회나 나왔는지에 대한 빈도 수를 정리하는 제거대상 변수 목록을 작성하게 된다.

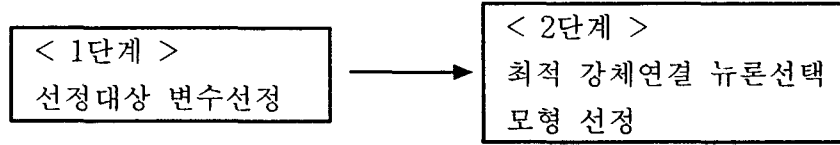
2.3.2 최적 약체연결 뉴런 제거모형 선정

제거대상 변수 목록이 작성되면 제거대상 변수의 빈도 수를 기반으로 하여 약체연결 뉴런 제거 모형인 WL(Weak Linked)(x)들을 구성한다. 즉, x개 이상 나온 변수들을 제외한 변수만으로 만들어진 인공신경망 모형을 말한다. 가령 WL(8)인 경우 8번 이상 제거대상 변수로 나온 변수를 일컫는 것이다. 각 WL(x)는 서로 다른 샘플에 일정 횟수만큼 반복하여 각각의 성능에 대한 평균값을 구한 다음 그 가운데서 가장 좋은 모형을 최적 약체연결 뉴런 제거 모형(WL*)으로 선정한다.

2.4 강제연결 뉴런 선택법

강제연결 뉴런 선택법은 인공신경망의 각 연결강도가 큰 입력노드는 전체 출력 값에의 영향이 강할 것이라는 가설 하에 입력노드와 은닉노드 간의 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 입력변수

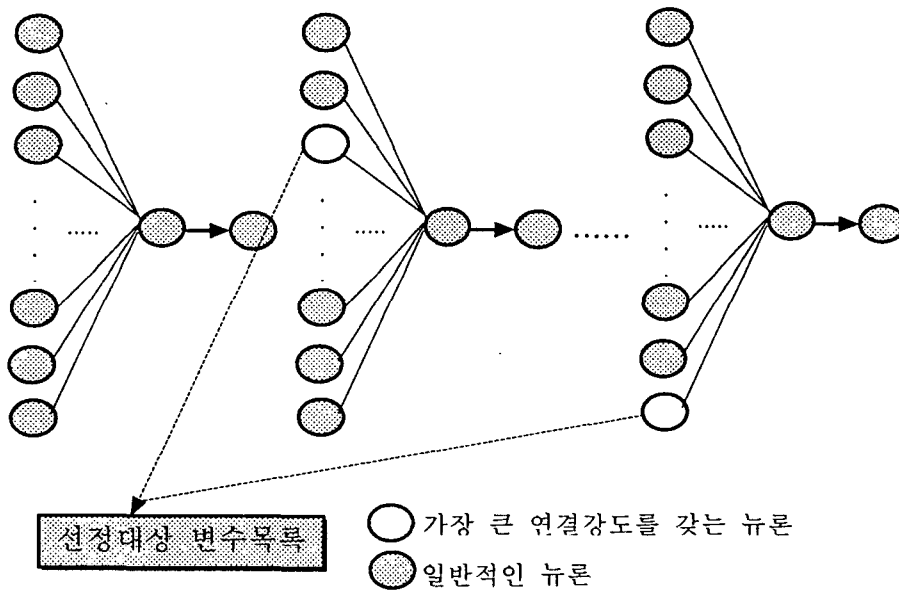
로 선정하여 출력노드에 대한 설명력을 높이려는 입력노드 선정 휴리스틱이라고 할 수 있다. 강제연결 뉴론 선택법의 경우에도 약체연결 뉴론 제거법에서와 유사하게 [그림 2-3]에서 보는 바와 같이 선정대상 변수 선정과 최적 강제연결 뉴론 선택 모형 선정의 두 단계로 나뉘어 진다.



[그림 2-3] 강제연결 뉴론 선택법

2.4.1 선정대상 변수 선정

약체연결 뉴론 제거법에서와 같이 대상이 되는 변수의 수만큼의 입력노드와 하나의 은닉노드로 이루어진 인공신경망 모형을 이용하는데 연결강도가 낮은 뉴론을 제거하는 대신에 연결강도가 높은 뉴론을 선정하는 방식으로 전개된다. [그림 2-4]에서 보는 바와 같이 약체연결 뉴론 제거법과 마찬가지로 인공신경망 모형을 학습시킨 후 연결강도가 가장 높은 값을 선정하여 이를 제외한 나머지를 가지고 새롭게 학습을 한다. 이때 인공신경망의 성능이 더 이상 개선되지 않거나 반복 횟수가 일정 횟수에 도달할 때까지 학습을 반복하여 여기서 선정된 변수들은 선정대상 변수 목록에 포함한다. 약체연결 뉴론 제거법에서와 마찬가지로 이와 같은 라운드를 일정 횟수만큼 반복한다



[그림 2-4] 강제연결 뉴론 선택법에서 선정대상 변수 선정

2.4.2 최적 강제연결 뉴론 선택 모형 선정

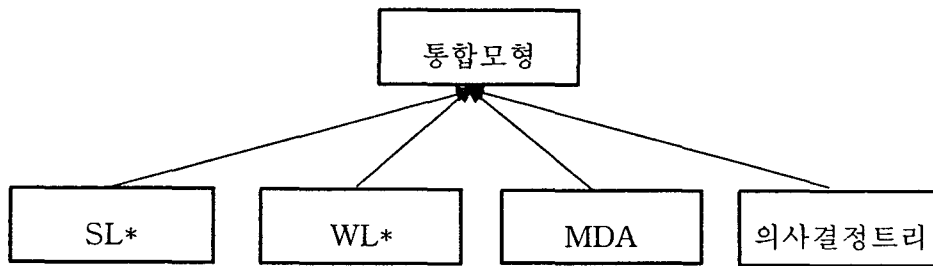
역시 약체연결 뉴론 제거법에서와 유사하게 x번 이상 선정대상변수 목록에 들어간 변수만으로 입력노드를 구성한 모형 SL(Strong Linked)(x)들을 대상으로 하여 각각의 성능에 대한 평균값을 구한 다음 그 가운데서 가장 좋은 모형을 최적 강제연결 뉴론 선택 모형 (SL*)을 선정한다.

2.5 통합모형

기존의 많은 연구들이 여러 가지 분류기법을 적용하여 그 중에서 하나의 기법이 주어진 문제에 대해 좋은 성과를 보인다는 결과를 발표하였다. 그러나 많은 연구결과들이 특정한 문제에 대한 동일한 결론을 도출하지 못하였고, 어떤 문제에 가장 적합한 기법이 무엇인지에 대해서는 아직 논란의 여지가 남아 있다. 인공신경망 기법은 결과가 불안정성(Unstability)을 갖는 경우가 있다. 이러한 경우에 똑같은 데이터를 사용하더라도 초기 값이 다르거나 데이터의 학습 순서가 다르면 서로 다른 결과를 얻을 수 있다. 그리고 현실 세계에서는 훈련 데이터가 무한히 존재하는 것이 아니라 매우 제한된 패턴 공간만이 관찰 가능하고, 잡음이 있는 유한의 데이터가 주어졌을 때 서로 다른 모형은 일반적으로 서로 다른 일반화(혹은 서로 다른 의사결정 경계)를 갖게 된다. 그래서 분류 문제를 다룰 때 여러 인공신경망 모형을 학습시켜 좋은 모형을 선택한다. 단일 모형을 선택하는 것이 반드시 이상적인 것은 아니다. 비교적 덜 성공적인 결과를 내는 모형에 잠재할 수 있는 귀중한 정보가 낭비될지도 모르기 때문이다. 단일 인공신경망 모형을 선택하여 정보를 잃을 위험을 회피하기 위해 모든 이용 가능한 모형들의 결과를 종합하여 의사결정을 수행 할 수 있다. 이러한 접근 방법은 잡음이 많거나, 학습 데이터의 수가 한정되어 있거나, 고차원적인 패턴 등으로 인해 어려운 문제의 경우 특히 효과적으로 활용될 수 있다.

기업 도산 예측에 있어서 기업이 도산에 이르기까지는 많은 변인들이 매우 다양하게 작용한다. 이러한 변인들을 한 모형에 모두 반영하는 것은 가능하지도 바람직하지도 않다. 모형 구축에 있어서 너무 많은 입력변수를 사용한 경우 인공신경망의 수렴(Convergence)과 일반화(Generalization) 등에 부담을 준다. 또한, 재무변수들간에 존재하는 다중공선성(Multi-collinearity) 문제를 유발할 위험을 항상 내포한다. 이를 극복하기 위하여 여러 분류기법을 적용·비교를 통한 최적 기법에 대한 탐색방법으로 분류기법들을 통합하는 방법론을 제시하고자 한다. 입력변수가 상이한 모형의 경우 비록 구축 방법론이 동일하다 할지라도 상이한 모형으로 볼 수 있다.

이 연구에서는 동일한 기업에 대하여 모델간에 상이한 출력 값을 보이는 경우에 있어서 최종 예측 값을 도출하는 통합 방법론을 제시한다. 즉, 기업 도산 예측모형 연구 중 통계적인 방법으로 가장 널리 활용하고 있는 MDA 방법과 이해 가능한 규칙을 만들어내는 의사결정트리 방법 그리고 인공신경망의 입력노드와 은닉노드의 연결가중치의 절대값인 연결강도를 기반으로 하는 연결강도분석 접근법인 최적 약체연결 뉴런 제거 방법 및 최적 강제연결 뉴런 선택 방법을 이용한 모형을 통합하고자 한다. 위의 네 가지 모형을 앙상블(Ensemble)기법에 의해 통합하는 것이다. 앙상블 기법은 여러 모형의 출력 값을 혼합하여 최적의 값을 구하는 기법으로 크게 세 가지로 구분될 수 있는데 학습용 데이터에서 나온 가장 낮은 오차를 갖는 모형 값만을 이용하는 Bumping 기법, 각 모형의 결과 값의 평균을 이용하는 Bagging 기법 그리고 각 모형의 값의 가중치를 고려한 Balancing 기법 등이 있다.¹⁶⁾ 휴리스틱 한 방법을 이용한 약체연결 뉴런 제거법과 강제연결 뉴런 선택법 등 네 가지 모형에 대한 사전 경험치나 선형적 지식을 가질 수 없기 때문에 Balancing 기법을 적용하기는 곤란하고, 또 Bumping 기법의 경우 네 모형의 통합 효과를 도모하기 어렵기 때문에 단순 평균법인 Bagging 기법에 의해 [그림 2-5]와 같이 통합하도록 한다.



[그림 2-5] 모형별 통합방법

각 모형의 분류 값이 상이하게 나타나는 것은 동일한 기업에 대해서 다른 입력변수들을 적용함으로써 한 기업을 다른 각도의 시각으로 보는 것이다. 이러한 상이한 분류 값들을 통합하여 적용하는 것은 단일 모형의 한계점을 여러 모형의 통합을 통하여 완화시켜주는 것으로 각각의 모형만을 사용하는 기존의 방법보다 더 우수한 결과를 제시할 수 있을 것이다. 이 연구에서는 각 모형에 대한 사전 지식이나 경험치가 없기 때문에 각 모형에 대한 가중치가 동일하다는 가정 하에 각 모형의 값들을 더하여 평균값을 구하는 Bagging 기법을 적용하도록 하였다.

III. 실험설계

3.1 사용변수 및 표본기업

일반적으로 인공지능망에 의한 부도예측에는 Altman(1968)이 제안한 다섯 가지의 재무비율(순운전자본구성비율, 이익잉여금구성비율, 총자산영업이익률, 장부가치, 총자산회전율)을 사용하고 있지만, Raghupathi 등(1998)은 13개의 재무비율 변수를, Salchenberger 등(1992)은 29개의 변수 그리고 Tam과 Kiang(1992)은 19개의 변수를 사용하고 있다. 한편 Shin과 Hong(2000)은 재무비율 이외에 현금흐름지표를 변수에 포함하고 있다. 본 연구에서는 기존 국내외 연구문헌들을 검토한 결과를 토대로 도산 예측에 대한 기업 도산을 잘 나타내주는 자료로서 자료의 입수가 용이하고 객관적인 통계처리가 가능한 <표 3-1>과 같이 43개의 재무비율과 현금흐름지표를 대상으로 하여 적절한 입력변수를 선정한다.

표본기업은 1997년부터 1999년 사이 475개의 국내기업의 제조업체 (부도기업 238개, 건전기업 237개)를 대상으로 하였는데 이 가운데 무작위로 추출한 100개(부도기업 56개, 건전기업 44개)는 스코어링을 위해 사용하였고 나머지 375개를 변수 선정과 학습(Training) 및 검증(Validation)을 위해 사용하였다.

스코어링을 위해 선정된 100개를 제외한 나머지 375개 가운데서 다시 100개기업(건전기업 50개, 부도기업 50개)을 무작위로 추출하여 제거대상 변수 목록 및 선정대상 변수 목록을 만드는데 사용하였다.

또 최적 약체연결 뉴런 제거 모형(WL*)과 최적 강체연결 뉴런 선택 모형(SL*)을 구하기 위해 375개의 데이터를 다시 학습용 150개, 검증용 150개 그리고 테스트용 75개를 10회에 걸쳐 무작위 추출하는 방식으로 모형을 구축하고, 각 모형들의 적중률의 평균값을 구하여 각각의 최적 모형을 구하였다. 다변량 판별분석과 의사결정트리에 의한 모형의 경우에도 375개의 데이터를 이용하여 모형을 구축하였다.

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

3.2 비교대상 모형

이 연구에서는 위에서 제시한 두 가지의 연결강도분석법인 최적 약체연결 뉴론 제거 모형과 최적 강체연결 뉴론 선택 모형 그리고 C4.5 알고리즘에 의한 의사결정트리와 다변량 판별분석의 설명변수 선택방식인 단계별 선택법(Stepwise Selection)에 의해 입력노드를 선정한 인공신경망 모형과 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. 이 연구에서는 데이터 세분화의 기준이 되는 알고리즘으로 C4.5를 선택했다. C4.5는 각 클러스터 내의 정보 무질서 (Information Disorder)의 정도인 엔트로피(Entropy)를 줄이려는 방식으로 세분화를 전개하는 알고리즘이다.¹⁾

다변량 판별분석의 단계별 선택법은 가장 높은 상관계수 값을 갖는 변수를 설명변수로 채택하고 이에 대한 유의도를 Wilk's Lambda에 의해 유의성을 확인하는 방식으로 단계별로 진행된다.¹⁾

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

- 첫째, 최적 약체연결 뉴론 제거모형에 의해 선 처리된 모형 (WL* + ANN)
- 둘째, 최적 강체연결 뉴론 선택모형에 의해 선 처리된 모형 (SL* + ANN)
- 셋째, 의사결정트리(C4.5)에 의해 선 처리된 모형 (C4.5 + ANN)
- 넷째, 다변량 판별분석에 의해 선 처리된 모형 (MDA + ANN)

<표3-1> 전체변수내역

번호	지표	변수	변수내역
1	규모	X1	매출액
2	규모	X2	자기자본
3	생산성	X3	총자본투자효율
4	생산성	X4	부가가치율
5	성장성	X5	매출액증가율
6	성장성	X6	재고자산증가율
7	성장성	X7	총자산증가율
8	수익성	X8	금융비용 대 매출액비율
9	수익성	X9	대출효율성계수
10	수익성	X10	매출액순이익률
11	수익성	X11	매출원가율
12	수익성	X12	손익분기점률
13	수익성	X13	순금융비용 대 매출액비율
14	수익성	X14	이자보상배율
15	수익성	X15	자기자본순이익률
16	수익성	X16	총자본경상이익률
17	수익성	X17	총자본순이익률
18	안정성	X18	고정장기적합률의 역
19	안정성	X19	단기부채 대 총차입금
20	안정성	X20	당좌비율
21	안정성	X21	매출채권 대 매입채무
22	안정성	X22	순운전자본비율
23	안정성	X23	유동비율
24	안정성	X24	유동부채 대 총자본
25	안정성	X25	유보액 대 총자산비율
26	안정성	X26	자기자본비율
27	안정성	X27	차입금의존도
28	안정성	X28	총차입금 대 매출액
29	추세	X29	금융비용부담금률 증가율
30	활동성	X30	매입채무회전율
31	활동성	X31	순운전자본 대 매출액
32	활동성	X32	운전자금 대 회전율
33	활동성	X33	재고자산회전율
34	활동성	X34	총자본회전율
35	현금흐름지표	X35	(영업활동후 CF/차입금합계)*100
36	현금흐름지표	X36	영업활동후 CF - 운전자금의 순증
37	현금흐름지표	X37	영업활동후 CF/금융비용
38	현금흐름지표	X38	(현금영업이익/금융수입)/금융비용
39	현금흐름지표	X39	(이자지급후 CF/부채총계)*100
40	현금흐름지표	X40	이자지급후 CF/금융비용
41	현금흐름지표	X41	(외부자금조달전CF/차입금합계)*100
42	현금흐름지표	X42	외부자금조달전 CF/금융비용
43	현금흐름지표	X43	외부자금조달후 CF/금융비용

사용된 소프트웨어는 인공지능경망과 의사결정트리의 경우 엔터프라이즈 마이너 (Enterprise Miner) 3.0이 이용되었고 다변량 판별분석의 경우 SAS 6.12가 이용되었다.

또한 통합모형은 최적 약체연결 뉴론 제거법, 최적 강제연결 뉴론 선택법, 의사결정트리법, 그리고 다변량 판별분석(MDA)법에 의하여 선 처리하고, 이들을 각각 인공신경망 방법에 후리스틱하게 결합하여 선정된 모형의 예측값들의 평균을 이용한 Bagging 기법을 적용하여 구축한 모형을 이용하였다.

IV. 연구모형구축

4.1 최적 약체연결 뉴론 제거 모형

43개의 입력변수로 약체연결 뉴론 제거법을 적용하여 입력변수를 선정하였다. 총 라운드 수 (R)는 10회로 하였고 라운드별 제한 횟수(K)도 10회로 제한하였으며 연결강도 임계치(α)는 0.05로 하여 실험을 하였다. 한 라운드 내에서 회수를 반복하면서 제거대상변수를 선정했으며, 제거대상변수가 없는 경우에 실험을 중단하고 다음 라운드의 실험을 진행하였다. 라운드마다 연결강도 값이 연결강도 임계치(α)보다 작은 입력노드가 나오지 않은 경우, 인공신경망의 성능이 더 이상 개선되지 않은 경우, 반복회수가 10회에 도달했을 경우 중 하나가 만족될 때까지 계속 반복하였다.

<표 4-1>는 첫 번째 라운드에서 선정된 제거변수와 각각의 연결강도를 보여 주고 있다. 1회의 실험결과에서는 X5, X10을 비롯한 총 10개(D11), 2회에서는 X6, X7을 비롯한 4개(D21), 3회에서는 X2, X3을 비롯한 8개(D31)의 제거대상 변수가 선정되는 등, 제1라운드의 실험에서는 총 32개의 변수가 제거대상 변수로 선정되었으며, 8회에서는 해당변수가 없었기 때문에 제1라운드의 실험을 중단하였다. 같은 방식으로 총 10 라운드에 걸친 실험을 하여 <표 4-2>와 같은 제거대상 변수 목록을 작성하였다.

<표 4-1> 약체연결 뉴론 제거법의 1라운드 결과

1차		2차		3차		4차		5차		6차		7차	
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	X19	0.0031	X18	0.0072	X12	0.0234	X11	0.0269
X10	0.0052	X7	0.0256	X3	0.0186	X30	0.0257	X28	0.0345	X33	0.0408		
X13	0.0456	X16	0.0359	X4	0.0452	X36	0.0348	X41	0.0182				
X14	0.0089	X31	0.0101	X8	0.0413	X43	0.0308						
X15	0.0262			X9	0.0137								
X17	0.0279			X20	0.0326								
X34	0.0381			X25	0.0058								
X37	0.0063			X32	0.0337								
X38	0.047												
X40	0.0244												

제거대상 변수 목록에 10회, 9회, 8회, 7회, 6회, 5회, 4회, 2회, 1회에 걸쳐서 제거대상 변수로 나타난 변수를 각각 누적하여 제거한 후 나머지 변수를 입력변수로 약체연결 뉴론 제거 모형인 각각의 WL(x) 모형을 얻었다.

즉, WL(x)는 x회 이상 제거대상 변수로 선정된 변수를 제외한 변수들을 입력노드로 하여 구축된 모형이다. WL(x) 모형들의 은닉노드의 수는 각 입력노드의 수와 같도록 구조를 설

<표 4-2> 제거대상 변수목록

빈도수	제거대상 변수
10	X36
9	X15, X19, X28, X42
8	X10, X14, X35, X38, X41
7	X5, X9, X11, X16, X18, X22, X31, X32, X371, X39
6	X2, X3, X7, X8, X12, X17, X23, X33, X34, X40
5	X1, X6, X20, X25, X30, X43
4	X4, X13, X26, X27
2	X21
1	X24, X29

정하여 10회에 걸쳐 실험을 하였다. 이 때 앞에서 언급한 바와 같이 375개의 데이터를 10회에 걸친 무작위 추출을 통해 학습용, 검증용 및 테스트용으로 구분하여 <표 4-3>와 같은 결과를 얻었다.

<표 4-3>에서 보는 바와 같이 WL(6)가 가장 높은 예측률(81.2%)을 보여 WL(6)를 최적 약체연결 뉴런 제거 모형(WL*)으로 선정하였다.

<표 4-3> 약체연결 뉴런 제거 모형들의 평균 예측률

구분	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
평균값	58.2	68.0	71.4	72.0	81.2	78.4	69.2

최적 약체연결 뉴런 제거 모형(WL*)에서 선정된 입력변수는 <표 4-4>와 같다.

<표 4-4> 최적 약체연결 뉴론 제거 모형의 입력변수

번호	지표	변수	변수내역
1	규모	X1	매출액
2	생산성	X4	부가가치율
3	성장성	X6	재고자산증가율
4	수익성	X13	순금융비용 대 매출액비율
5	안정성	X20	당좌비율
6	안정성	X21	매출채권 대 매입채무
7	안정성	X24	유동부채 대 총자본
8	안정성	X25	유보액 대 총자산비율
9	안정성	X26	자기자본비율
10	안정성	X27	차입금의존도
11	추 세	X29	금융비용부담금률 증가율
12	활동성	X30	매입채무회전율
13	현금흐름지표	X43	외부자금조달후 CF/금융비용

4.2 최적 강체연결 뉴론 선택 모형

최적 약체연결 뉴론 제거 모형과 유사하게 라운드 수 (R)는 10회로 라운드별 제한 횟수 (K)도 10회로 하여 진행하였다. 라운드마다 인공신경망 모형의 성능이 더 이상 개선되지 않거나, 반복회수가 10회에 도달할 때까지 계속 반복하였다. 1라운드 수행결과, 1회에서 X29, 2회에서 X26, 3회에서 X7, 4회에서 X27, 5회에서 X16 등, 10회에 걸친 실험에서 <표 4-5>과 같이 10개의 변수가 선정대상 변수로 선정되었다.

<표 4-5> 제1라운드 회수별 선정대상 변수

라운드 및 회수	선정대상변수
D11	X29
D12	X26
D13	X7
D14	X27
D15	X16
D16	X33
D17	X13
D18	X34
D19	X23
D110	X3

같은 방식으로 총 10라운드에 걸친 실험한 결과는 <표 4-6>와 같은 선정대상 변수 목록을 얻었다.

<표 4-6> 강제연결 뉴론 선정대상변수 내역

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

이와 같은 방법의 실험을 10라운드에 걸쳐 진행한 결과로 얻은 선택대상 변수를 빈도 별로 <표 4-7>와 같이 정리하였다.

<표 4-7> 빈도별 선정대상 변수목록

회수 라운드	1회	2회	3회	4회	5회
1	X29	X26	X7	X27	X16
2	X7	X3	X33	X26	X29
3	X7	X29	X33	X34	X30
4	X33	X29	X39	X23	X30
5	X30	X3	X7	X29	X33
6	X3	X7	X29	X19	X27
7	X7	X29	X6	X15	X30
8	X20	X30	X21	X29	X4
9	X7	X29	X16	X27	X33
10	X29	X27	X7	X1	X15
회수 라운드	6회	7회	8회	8회	10회
1	X33	X13	X34	X23	X3
2	X23	X4	X27	X31	X28
3	X36	X25	X38	X39	X40
4	X7	X22	X34	X27	X18
5	X27	X34	X20	X4	X38
6	X41	X26	X37	X43	X20
7	X22	X39	X28	X23	X2
8	X7	X33	X24	X6	X26
9	X6	X28	X18	X41	X3
10	X42	X28	X6	X21	X3

<표 4-7>에서 보는 바와 같이 10회에 걸쳐서 선정대상 변수로 선정된 변수는 X7, X29 두 개의 변수이고, 9회와 8회 선정된 변수는 없었으며, 7회 선정대상 변수로 선정된 것은 X27, X33이며 6회 선정된 변수는 X3이다. 이를 기반으로 x회 이상 선정대상 변수 목록에 포함된 변수만으로 입력노드를 선정한 SL(x) 즉, 강제연결 뉴런 선택 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 4-8>와 같은 결과를 얻었다.

<표 4-9>에서 보는 바와 같이 SL(6)가 가장 높은 예측률(67.6%)을 보여 SL(6)를 최적 강제연결 뉴런 선택 모형(SL*)으로 선정하였다.

<표 4-8> 강제연결 뉴론 선택 모형들의 평균 예측률

구분	SL(5)	SL(6)	SL(7)	SL(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
평균값	66.8	67.6	64.4	54.4

최적 강제연결 뉴론 선택 모형(SL*)에서 선정된 입력변수는 <표 4-9>와 같다.

<표 4-9> 최적 강제연결 뉴론 선택 모형의 입력변수

번호	지표	변수	변수내역
1	생산성	X3	총자본투자효율
2	생산성	X7	총자산증가율
3	안정성	X27	차입금의존도
4	추세	X29	금융비용분담금률증가율
5	활동성	X30	매입채무회전율
6	활동성	X33	재고자산회전율

4.3 의사결정트리 및 다변량 판별분석에 의한 모형

본 연구에서는 의사결정트리의 경우 데이터 세분화의 기준이 되는 알고리즘으로 C4.5를 적용하여 <표 4-10>에서 보는 바와 같이 X25, X13, X38, X4 등 4개의 변수가 선정되었고, 다변량 판별분석의 단계별 선택법은 가장 높은 상관계수 값을 갖는 변수를 설명변수로 채택하고 이에 대한 유의도를 Wilk's Lambda에 의해 유의성을 확인하는 방식으로 단계별로 진행하여 <표 4-11>와 같이 다변량 판별분석의 경우 X3, X5 등 16개의 변수가 선정되었다.

<표 4-10> 의사결정트리 모형의 입력변수

번호	지표	변수	변수내역
1	생산성	X4	부가가치율
2	수익성	X13	순금융비용 대 매출액비율
3	안정성	X25	유보액 대 총자산비율
4	현금흐름지표	X38	(현금영업이익/금융수입)/금융비용

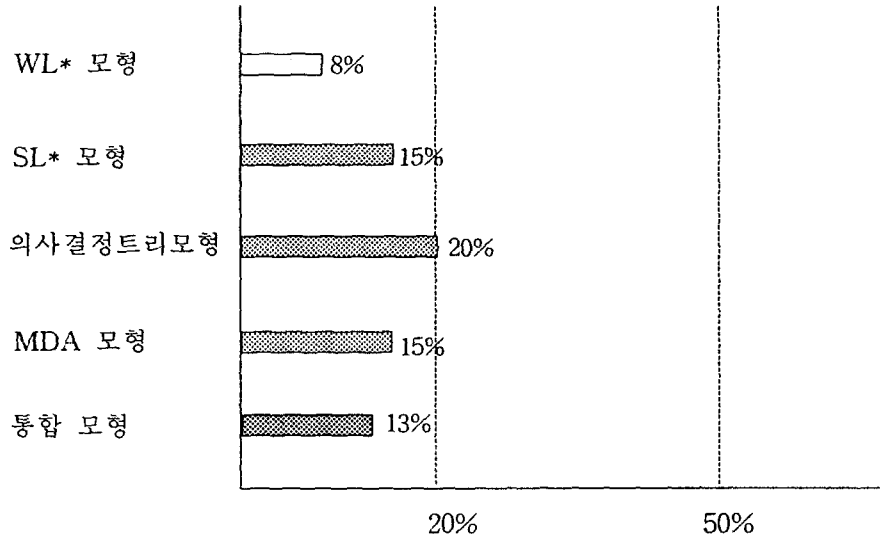
<표 4-11> 다변량 판별분석 모형의 입력변수

번호	지표	변수	변수내역
1	규모	X1	매출액
2	생산성	X3	총자본투자효율
3	생산성	X4	부가가치율
4	성장성	X5	매출액증가율
5	성장성	X6	재고자산증가율
6	성장성	X7	총자산증가율
7	수익성	X11	매출원가율
8	수익성	X13	순금융비용 대 매출액비율
9	수익성	X15	자기자본순이익률
10	안정성	X18	고정장기적합률의 역
11	안정성	X20	당좌비율
12	안정성	X23	유동비율
13	안정성	X24	유동부채 대 총자본
14	안정성	X26	자기자본비율
15	활동성	X31	순운전자본 대 매출액
16	활동성	X32	운전자금 대 회전을

V. 실험결과의 분석

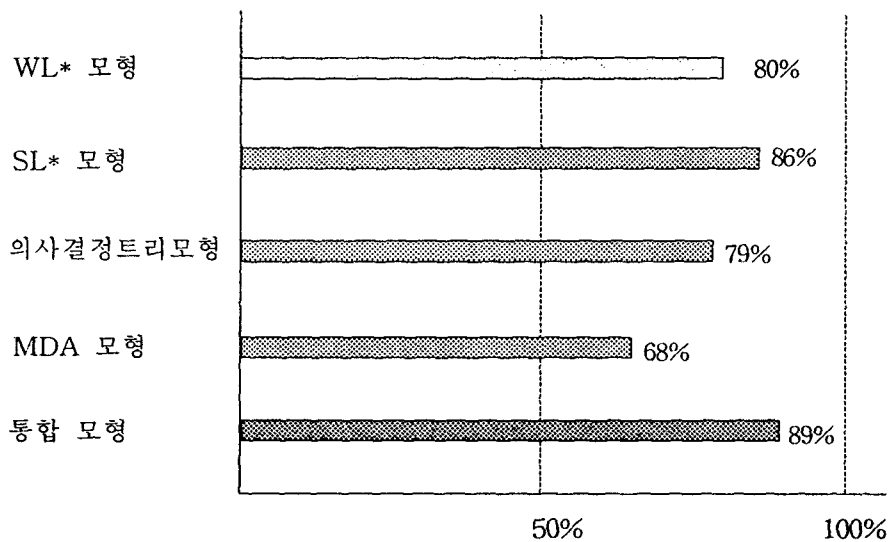
각 모형의 예측률을 얻기 위하여 이제까지 훈련이나 검증에 전혀 사용되지 않은 데이터 100개를 스코어링용으로 사용하여 결과를 도출하였다. 모형의 출력 값은 [0, 1]사이의 값으로 출력되며 분류기준값 (Cut-off Value)은 '0.5'로 설정하여 구분하였다. 건전 기업을 '1'로 부도 기업을 '0'으로 구분하고 각 모형에 적용하여 분류정확도(Accuracy)를 분석하였다. 즉 건전 기업을 건전 기업으로, 부도 기업을 부도 기업으로 인식하였을 경우는 정확도가 '1'이고, 건전 기업을 부도 기업으로, 부도 기업을 건전 기업으로 잘못 인식한 경우는 '0'으로 하여 정확도를 분별하였다. 분류정확도는 실제의 부도 기업과 건전 기업을 올바르게 분류한 경우의 수를 전체 관측치 수로 나누어 준 값으로서 모형의 전체적인 예측률을 나타낸다. 각 모형별 예측률은 최적 약체연결 뉴런 제거 모형(WL*)이 평균 80%, 최적 강제연결 뉴런 선택 모형(SL*)이 86%, 의사결정트리에 의한 모형이 79%, 다변량 판별분석에 의한 모형이 68%로 나타났다. 또한 통합 모형은 각 모형의 결과 값을 통합하여 평균을 구하여 얻은 모형으로 89%의 예측률을 보이고 있다. [그림 5-1]과 같이 각 모형의 예측률을 비교한 결과 MDA 모형 예측률이 낮았고, 그 다음이 의사결정트리 모형, 최적 약체연결 뉴런 제거 모형, 최적 강제 연결 뉴런 선택 모형 순서로 나타났다. 이는 연결강도분석을 이용하여 입력변수를 선정하는 방법이 기존의 두 방법을 이용하여 인공신경망 모형을 결합한 방법보다 예측률이 좋은 것으로 나타났고, 통합모형과의 차이는 의사결정 트리 모형보다는 10%, MDA 모형보다는 29%나 높은 예측률의 결과를 보였고, 연결강도분석에 의한 최적 약체연결 뉴런 제거 모형보다는 9%, 최적 강제연결 뉴런 선택 모형보다는 3%의 향상을 기할 수가 있었다. 또한 모형예측 평가에 있어서 중요한 것은 'Type I 오류'와 'Type II 오류'의 관계이다.

'Type I 오류율'이란 부도 기업을 건전 기업으로 잘못 예측할 수 있는 비율을 의미하고, 'Type II 오류율'은 'Type I 오류'의 반대로 건전 기업을 부도 기업으로 잘못 예측할 수 있는

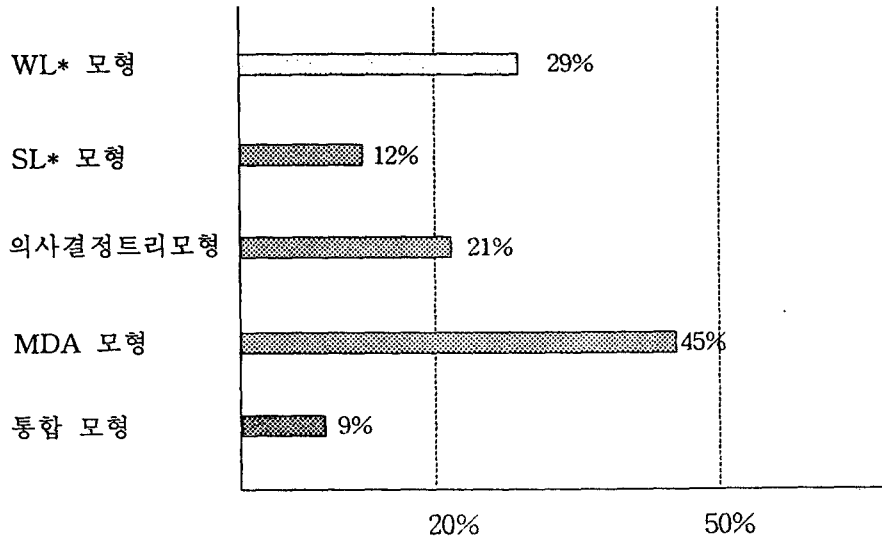


[그림 5-1] 모형별 예측률 비교

비율을 의미한다. 일반적으로 'Type I 오류'는 'Type II 오류'보다 훨씬 더 큰 손실을 끼친다. 즉 건전 기업을 부도 기업으로 분류하는 것보다 부도 기업을 건전 기업으로 분류했을 경우 주가의 폭락이나 채권의 회수가 불가능하게 되는 더 큰 손실을 가져오게 되기 때문이다. [그림 5-2]와 같이 MDA 모형, 의사결정트리모형, 최적 약체연결 뉴론 제거모형, 최적 강체연결 뉴론 선택 모형 등 네 가지 모형에 대하여 'Type I 오류율'이 각각 45%, 21%, 29%, 12%로 나타나고 'Type II 오류율'은 15%, 20%, 8%, 15%로 나타났다.



(a) 모형별 Type I 오류율



(b) 모형별 Type II 오류율

[그림 5-2] 모형별 Type I 및 Type II 오류 비교

최적 강제연결 뉴런 선택 모형을 제외한 세 모형은 'Type II 오류율' 보다 'Type I 오류율'이 높게 나타났다. 'Type I 오류율'도 예측률 비교와 같이 MDA에 의해 입력변수를 선정한 후 인공신경망 모형을 적용한 모형이 가장 예측률이 좋지 않고, 그 다음이 의사결정트리 모형, 최적 약체연결 뉴런 제거 모형, 최적강제연결 뉴런 선택 모형 순서로 나타났다. 그리고 통합 모형은 'Type I 오류율' 9%, 'Type II 오류율' 13%로 'Type I 오류율'이 훨씬 낮게 나타났다. 이는 네 가지 모형을 통합한 모형이 예측률 및 'Type I 오류율'에 있어서 다른 네 가지 모형보다 훨씬 좋은 실증적 결과 값을 보여주고 있다.

각 모형별 예측률 간에 통계적으로 유의하게 차이가 있는지를 검증하기 위하여 본 연구에서는 쌍체검증(Z-검증)을 사용하였다. 그 결과는 <표 5-1>와 같다. 연결강도분석에 의한 최적 강제연결 뉴런 선택 모형(SL*)은 MDA 모형과 의사결정트리 모형에 통계적으로 유의한 예측력 차이를 보이나, 최적 약체연결 뉴런 제거 모형(WL*)은 MDA 모형에만 유의한 차이를 보이고 있다. 또한 통합모형은 기타 다른 모형과 통계적으로도 매우 유의한 결과를 보이고 있다.

이상의 결과를 종합하여 보면 본 연구에서 제안한 최적 강제연결 뉴런선택 모형과 최적 약체연결 뉴런 제거 모형이 MDA 모형이나 의사결정트리 모형에 비해 향상된 예측률을 보이고 있으며, 또한 통합모형도 다른 모형의 결과보다 통계적으로 유의하게 향상된 예측률을 보이고 있다는 것을 알 수 있고, 'Type I 오류율'에서도 다른 모형보다도 향상된 실증적 결과 값을 보여주고 있다.

〈표 5-1〉 모형간의 쌍체검증(Z-검증) 결과

모형별	WL* 모형	SL* 모형	통합모형
MDA 모형	2.05372**	2.91457*	3.71995*
의사결정트리 모형	0.1922	1.25581***	2.00813*
WL* 모형	-	-	1.91661**
SL* 모형	-	-	0.58778

* 통계적 유의수준 1%

** 통계적 유의수준 5%

*** 통계적 유의수준 10%

VI. 결 론

기업의 부실화는 금융기관, 채권자, 종업원 등 직접적인 이해관계자들 뿐만 아니라, 실업, 경기침체 등 사회적으로도 중요한 관심사가 되고 있다. 이러한 기업의 부실화와 관련된 연구는 재무관리분야와 함께 정보시스템분야에서도 매우 중요한 연구과제로 인식되어 왔다. 1960년대부터 미국 등 선진국을 중심으로 다변량 판별분석 등 통계적 방법을 이용한 연구를 시작으로 1980년대 후반부터는 인공신경망 기법을 이용한 연구가 활발히 진행되어 왔다. 그러나 인공신경망에 의한 모형화 과정에서 가장 중요한 고려사항 중의 하나는 적절한 입력노드의 선정인데, 그 이유는 지나치게 많은 입력노드는 많은 학습시간을 요구할 뿐 아니라, 또 입력변수들 가운데 노이즈가 섞이게 되는 경우에는 적절치 못한 해(Solution)를 도출할 가능성이 높기 때문에 적절한 입력변수 선정을 위한 다양한 기법이 제안되고 있다.

이에 이 논문에서는 입력변수 선정을 위하여 연결강도분석 접근법을 제안하고, 좀 더 정확한 부도 예측 모형을 구축하기 위하여 통합모형을 제시하였다. 또한 연결강도분석에 의한 최적 약체연결 뉴런 제거법과 최적 강제연결 뉴런선택법에 의해 입력노드를 선정한 인공신경망 모형이 C4.5에 의한 의사결정트리나 다변량 판별분석에 의해 입력노드를 선정한 인공신경망 모형보다 더 나은 예측률을 보이고 있음을 실증하였으며, 위의 네 가지 모형을 결합한 통합모형이 각각의 모형보다 예측률이 향상됨을 실증하였다.

이 연구에서 제시한 통합모형은 최적 약체연결 뉴런 제거법, 최적 강제연결 뉴런 선택법, 의사결정트리법, 다변량 판별분석(MDA)법에 의하여 선 처리하고, 이들을 각각 인공신경망 방법에 휴리스틱하게 결합하여 선정된 모형의 예측값들의 평균을 이용한 Bagging 기법을 적용하여 결합한 형태이다.

표본기업은 1997년부터 1999년 사이 475개의 국내기업 제조업체(부도기업 238개, 건전기업 237개)를 대상으로 하였다. 실험 샘플링은 이 가운데 무작위로 추출한 100개(부도기업 56개, 건전기업 44개)는 스코어링을 위해 사용하였다. 나머지 375개를 변수선정과 학습(Training) 및 검증(Validation)을 위해 사용하였는데 이 중 다시 100개 기업(건전기업 50개, 부도기업 50개)을 무작위로 추출하여 제거대상 변수목록 및 선정대상 변수목록을 만드는데 사용하였다. 사용변수는 43개의 재무비율과 현금흐름지표를 대상으로 하여 적절한 입력변수를 선정하였다. 특히 이 연구에서는 입력변수 선정을 위하여 연결강도분석을 이용하여 선 처리하였는데 연결강도분석이란 학습이 끝난 인공신경망에서 입력노드와 은닉노드와 연결된 가중치의 절대값인 연결강도를 분석하여 입력변수를 선정하는 접근법으로 약체연결 뉴런 제

거법, 강제연결 뉴런 선택법이 있다. 이 연구에서 실증분석은 먼저 입력변수 선정에 있어서 전통적인 방법에 의하여 선 처리하여 인공신경망과 결합한 모형의 예측률, 연결강도분석 방법에 의하여 선 처리하여 인공신경망과 결합한 모형의 예측률을 비교하였으며, 또한 이들 모형의 예측률과 통합모형의 예측률을 비교하였다.

비교 결과 최적 강제연결 뉴런 선택 모형 및 최적 약체연결 뉴런 제거 모형이 MDA 모형과 의사결정트리 모형보다는 월등한 예측률 향상을 보이고 있다. 이는 기업 부도 예측 모형에 있어서 인공신경망 모형이 우수하다는 기존연구들과 동일한 결과를 보여주고 있다. 또한 통합모형은 의사결정 트리 모형과 MDA 모형보다는 상당히 높은 예측률의 결과를 보였고, 연결강도분석에 의한 최적 약체연결 뉴런 제거 모형 및 최적 강제연결 뉴런 선택 모형보다는 약간 향상된 예측률을 보여주고 있으며, 부도 기업을 건전 기업으로 잘못 예측할 수 있는 비율 'Type I 오류'와 건전 기업을 부도 기업으로 잘못 예측할 수 있는 비율 'Type II 오류'의 관계에 있어서도 좋은 결과를 나타내고 있다. 특히 'Type I 오류'에서 의사결정 트리 모형과 MDA 모형보다는 상당히 좋은 결과를 보였고, 연결강도분석에 의한 최적 약체연결 뉴런 제거 모형 및 최적 강제연결 뉴런 선택 모형보다는 다소 향상된 결과를 보였다. 이러한 실증적 결과는 통계적으로도 유의수준 5%에서 유의하다는 결과를 보여주고 있다.

이상 이 연구가 제시하는 공헌을 정리하면 결국 입력변수 선정에 있어서 연결강도분석에 의한 최적 약체연결 뉴런 제거 모형과 최적 강제연결 뉴런 선택 모형으로 선 처리한 모형이 기존의 의사결정트리 모형이나, MDA 모형으로 선 처리된 모형보다 향상된 예측률을 보이고, 또한 통합모형이 다른 네 가지 모형보다는 좋은 예측률을 보여준다는 사실을 실증적으로, 통계적으로 검증한 것이다.

이 연구의 한계와 과제로는 우선 강제연결 뉴런과 약체연결 뉴런을 정의하는 방법론적인 보완을 들 수 있다. 특히 약체연결 뉴런을 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 번째로는 통합모형에 대하여 예측값들의 평균을 이용하는 기법 이외에 유전자 알고리즘 등 또 다른 방법론으로의 통합모형을 만들어 좀 더 일반화할 수 있어야 할 것이다.

세 번째로는 보다 방대한 데이터와 다양한 데이터로 실험하여 이 연구의 결과를 좀 더 일반화해 볼 필요성도 있다. 이 연구에서 사용한 데이터 셋의 규모를 좀 더 확장시키든지, 혹은 우리나라가 아닌 다른 나라에의 데이터를 적용할 경우 좋은 결과를 얻을 수 있을 것으로 보인다.

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

다섯 번째로는 이 연구에서 제안하고 있는 방법론을 부도예측이 아닌 다른 방법론에의 적용이다. 채권평가와 같이 분류(Classification)에 적용할 경우 흥미로운 연구 결과를 얻을 수 있을 것으로 기대된다.

여섯 번째, 최적 약체연결 뉴런 제거 모형이나 최적 강제연결 뉴런 선택 모형 그리고 의사결정트리나 다변량 판별분석에 의한 모형에서 선정된 입력변수에 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.

참 고 문 헌

- [1] 이건창, "기업 도산 예측을 위한 통계적 모형과 인공지능 모형간의 예측력 비교에 관한 연구 : MDA, 귀납적 학습방법, 인공신경망," 한국경영과학회지, 제18권, 제2호, 1993, pp.57-81.
- [2] 이건창, 한인구, 김명중, "통계적 모형과 인공지능 모형을 결합한 기업신용평가 모형에 관한 연구," 한국경영과학회지, 제21권, 제1호, 1994, pp.81-100.
- [3] 이건창, 김명중, 김혁 "기업도산예측을 위한 귀납적 학습지원 인공신경망 접근 방법," 경영학연구, 제23권, 제3호, 1994, pp.109-144.
- [4] 이용규, 손동우, "부도예측용 인공신경망모형의 최적 입력노드 설계: 연결강도판별분석 접근," 2000 춘계정기학술대회, 한국지능정보시스템학회, 2000.
- [5] 채서일, "사회과학 조사방법론," 학현사, 1999, 2판.
- [6] Alici, Y., "Neural networks in corporate failure prediction: The UK experience," in: A.P.N. Refenes, Y. Abu-Mostafa, J. Moody, A. Weigend (Eds.) Neural Networks in Financial Engineering, World Scientific, Singapore, 1996, pp.393-406.
- [7] Altman, Edward I., "Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy," Journal of Finance, Sep., 1968, pp. 589-609.
- [8] Altman, Edward I., Marco, G., Varetto, F., "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks(the Italian experience)," Journal of Banking and Finance, Vol.18, 1994, pp.505-529.
- [9] Berry, Michael J. A. and Gordon, Linoff, "Data Mining Techniques: For Marketing, Sales, and Customer Support," John Wiley and Sons, 1997.
- [10] Blum, M., "Failing Company Discriminant Analysis," Journal of Accounting Research, Vol.12, 1974, pp.1-25.
- [11] Boritz, J. E., Kennedy, D. B., A. de Miranda e Albuquerque, "Predicting corporate failure using a neural network approach," Intelligent Systems in Accounting, Finance and Management, Vol.4, 1995a, pp.95-111.
- [12] Boritz, J. E., Kennedy, D. B., 'Effectiveness of neural network types for prediction of business failure," Expert Systems with Applications, Vol.9, No.4, 1995b, pp.503-512.
- [13] Coats, P. K., Fant, L. F., "Recognizing financial distress patterns using a neural network tool," Financial Management, 1993, pp.142-155.
- [14] Deakin, E. B., "A Discriminant Analysis of Predictors Business Failure ," Journal of Accounting Research, Vol.10, 1972, pp.167-179.
- [15] Dun & Bradstreet Inc., "The Failure Record Through 1961," A Comprehensive Failure Study, New York, 1974.

- [16] Fletcher, D. and Goss, E., "Forecasting with Neural Networks: an Application using Bankruptcy Data," *Information and Management*, 24, 3, 1993, pp.159-167.
- [17] Heskes, T., "Balancing between Bagging and Bumpling," In Mozer, M. C., M. I. Jordan, and T. Petsche, (eds.) *Advances in Neural Information Processing Systems 9*, The MIT Press, 1996.
- [18] Jo, H., Han, I. and Lee, H., "Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis," *Expert Systems with Applications*, Vol.13, No.2, 1997, pp.97-108.
- [19] Kohonen, T., "Self-Organizing Map," *Proceedings of the IEEE* 78, No.9, 1990.
- [20] Lacher, R. C., Coats, P. K., Sharma, S. C., Fant, L. F., "A neural network for classifying the financial health of a firm," *European Journal of Operations Research* 85, 1995, pp.53-65.
- [21] Lee, K. C., Han, I., Kwon, Y., "Hybird neural network models for bankruptcy predictions," *Decision Support Systems*, Vol.18, 1996, pp.63-72.
- [22] Lenard, M. J., Aiam, P., Madey, G. R., "The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision," *Decision Science* 26(2), 1995, pp.209-226.
- [23] Odom, M., Sharda, R., "A Neural network Model for Bankruptcy Prediction," In *Proceedings of the IEEE International Conference on Neural Network*, 1990, pp.163-168.
- [24] Piramuthu, S., Shaw, M. J., Gentry, J. A., "A classification approach using multi-layered neural networks," *Decision Support Systems*, Vol.11, 1994, pp.509-525.
- [25] Raghupathi, W., Schkade, L. L. 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.
- [26] Salchenberger, L. M., Cinar, E. M. and Lash, N. A., "Neural Networks: A New Tool for Predicting Thrift Failures," *Decision Sciences*, Vol.24, No 4, 1992, pp.899-916.
- [27] Serrano-Cinca, Carlos, "Self Organizing Neural Networks for Financial Diagnosis," *Decision Support Systems*, 17, 1996, pp.227-238.
- [28] Shin, K., Jo. H. and Han, I., "Integration Methodology of Multiple Techniques Using Genetic Algorithms: A Case of Corporate Failure Prediction," *경영과학 '96 추계학술대회*, pp.199-202.
- [29] Shin, Kyung-shik and Hong, Seung-hyun, "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.
- [30] Sung, Tae Kyoung., Chang, Namsik and Lee, Gunhee, "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.
- [31] Tam, Kar Yan and Kiang, Melody Y., "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", *Management Science*, Vol.38, No.7, July, 1992, pp.926-947.
- [32] Tsukuda, J., and Baba, S. L., "Predicting Japanese Corporate Bankruptcy in terms of Financial Data Using Neural Network," *Computers and Industrial Engineering*,

27(1-4), 1994, pp.445-448.

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

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

[35] Zhang, Guoqiang, Michael Y. Hu, Patuwo, B. Eddy and Daniel, C. Indro, "Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analyss", *European Journal of Operational Research*, Vol.116, 1999, pp.16-32.

[36] Zmijewski, Mafke, "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, Vol.22, 1984, pp.59-82.

Session 3C

정보시스템관리 2

좌장 : 김상진(경동대학교)

3C-1 IS 프로젝트 관리자의 지식과 기술
문용은(신라대학교 경영학부 교수)

3C-2 AHP를 이용한 GIS 구축방법 선호도 평가
서창교(경북대학교 경영학부 교수)
남형근(경북대학교 경영대학원 석사)

3C-3 사례기반추론을 이용한 신기술 가치평가 시스템개발에 관한 연구
박기남(동의대학교 인터넷비즈니스학과)
김창진(리더스앤컴퍼니 책임컨설턴트)
