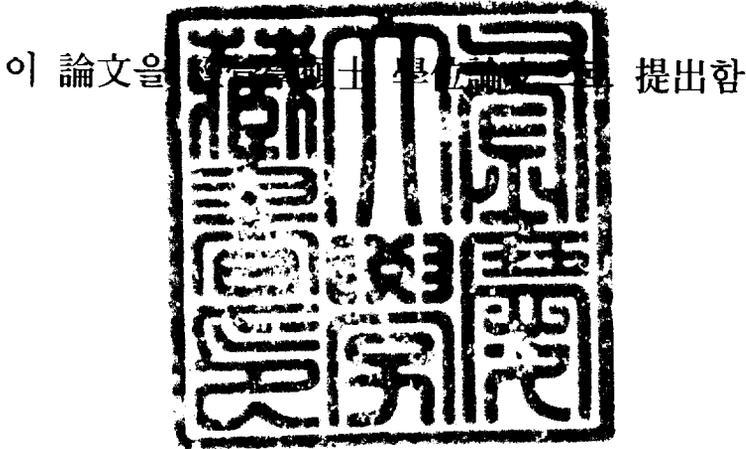


經營學碩士學位論文

인공신경망을 이용한 한국벤처기업의
도산예측에 관한 연구

指導教授 韓 載 昊



2002年 8月

釜慶大學校大學院

經 營 學 科

趙 芝 鎬

趙芝鎬의 經營學碩士學位論文을 認准함

2002年 6月 29日

主 審 工學博士 廉 昌 善



委 員 經營學博士 金 泰 容



委 員 工學博士 韓 載 昊



목 차

I. 서론	1
1. 연구배경 및 목적	1
2. 연구의 체계	3
II. 이론적 배경	5
1. 벤처기업	5
2. 판별분석법	6
3. 인공신경망의 이론적 배경	7
3.1 인공지능의 정의	7
3.2 인공신경망	10
3.2.1 다층퍼셉트론	11
3.2.2 역전파학습	14
4. 인공신경망과 판별분석의 차이점	18

Ⅲ. 선행 연구	19
1. 다변량관별분석을 이용한 예측 모형	19
2. 인공신경망을 이용한 예측 모형	21
Ⅳ. 연구방법론	27
1. 표본기업의 선정	27
1.1 분석대상기업 선정	27
1.2 검증용 표본의 선정	28
2. 도산예측변수의 선정	30
3. 실증분석 방법	31
Ⅴ. 실증분석	33
1. T-검정 결과	33
2. 다변량관별분석	36
2.1 동시적 방법	36
2.2 단계적 선택법	37

3. 인공신경망	39
3.1 신경망 아키텍처의 선정	39
3.2 t-검정의 유의한 변수를 이용한 예측	44
3.3 선택변수를 이용한 예측	45
4. 종합 비교	47
VI. 결 론	49
<참 고 문 헌>	51
Abstract	56

표 목차

<표 1> 인공신경망을 이용한 도산예측에 관한 선행연구의 비교	26
<표 2> 표본 기업수	29
<표 3> t-검정 결과 유의한 변수들	33
<표 4> 유의한 도산예측변수의 정의	35
<표 5> 판별분석의 동시적 방법 결과	37
<표 6> 판별분석의 단계적 선택법 결과	39
<표 7> 인공신경망의 t-검정 변수 사용 결과	45
<표 8> 인공신경망의 선택적 변수 사용 결과	46
<표 9> 전체결과	47

그림 목차

[그림 1] 의사결정자의 추론과정	8
[그림 2] 다층퍼셉트론의 아키텍처	12
[그림 3] 역전파학습 알고리즘의 학습 과정	17
[그림 4] 네트워크의 은닉층의 수 설정	40
[그림 5] 조기정지법에 사용할 시험표본의 설정	41
[그림 6] 학습회수 설정 및 조기정지법 설정	42
[그림 7] 은닉층 노드수 설정 및 이전함수 설정	43
[그림 8] 인공신경망 추정용표본 입력	44

I. 서론

1. 연구배경 및 목적

기업도산에 대한 예측은 도산 가능성이 잠재하고 있는 기업을 사전에 예측하여 그 충격을 최소화시킬 수 있는 시간적 여유와 대책을 강구하도록 해줌으로서 회피가능한 손실을 부담하지 않도록 도와줄 수 있기 때문에 매우 중요하다. 기업도산의 원인은 매우 다양하기 때문에 몇몇 자료에 의해 도산을 예측하기란 어렵다. 하지만 기업도산을 100% 정확히는 예측하지 못하더라도 가능한 정확히만 예측할 수 있다면 사후손실을 고려할 때 그 효과는 매우 크다 할 수 있다.

기업의 도산은 기업간의 경쟁이나 경영자의 잘못된 투자결정 등으로 인해 일정비율 발생해왔다. 그러나, IMF관리체제 이후의 구조조정과 최근의 세계경제의 불황으로 인한 국내 경기의 동반 불황의 영향으로 도산기업이 급속하게 증가함에 따라 도산기업의 예측문제가 중요한 문제로 부각되었다. 기업의 도산은 경영자, 투자자, 채권자, 종업원 등 기업의 직접적 이해관계자에게 많은 손실을 안겨줄 뿐만 아니라 사회 전반적으로도 실업 및 경기침체 등의 사회적 문제를 수반하기 때문에 기업에 대한 정확한 도산예측은 매우 중요한 문제라 할 수 있다.

우리 경제는 기존산업의 성장한계, 둔화되는 GDP성장률, 무역수지가 악화되는 등 어려운 상황에 직면하고 있으며, 이는 경기 순환적인 요인과 엔화약세, 교역조건 악화 등 외생적인 요인뿐만 아니라

오랫동안 누적된 구조적인 경쟁력 약화에 기인한다. 그동안 대기업 중심의 대규모 생산방식으로는 경제활력 회복 및 지속적인 경제발전에 한계가 있으므로, 기술과 지식 집약적인 벤처기업의 육성을 통하여 우리산업의 구조조정을 촉진하고 경쟁력을 제고할 필요가 있다. 이러한 배경에서 등장한 국내의 많은 벤처기업들은 세계적인 벤처기업의 경영실적 악화로 인하여 도산되거나 대규모 기업들만이 생존하는 정글의 법칙에 영향을 받아 M&A 등을 통하여 정리되고 있는 실정이다. 또한 이런 현상을 반영이라도 한 듯 코스닥지수의 거품 현상으로 인한 2001년까지 투자자, 종업원 등의 피해가 컸고 그 피해는 지금도 계속되고 있는 실정이다. 이러한 국내 벤처기업의 현실에서 벤처기업의 특성인 고위험·고수익의 불완전한 시장환경에서 재무적 정보를 이용하여 도산을 예측하는 것은 벤처기업의 직·간접적인 이해관계자에게 정말 중요한 사항이라고 말할 수 있겠다.

그러나, 기존의 연구에 의하면 도산예측은 매우 복잡한 문제가 수반되어 정확한 예측이 어렵다고 보고 있다. 이에 따라 정확한 도산예측을 위하여 통계적인 방법과 인공지능의 공학적 방법 등 다양한 방법이 시도되었다. 이러한 도산예측 중 국내 도산예측에 대한 연구들은 특정 업종별로 구분치 않고 전 산업의 회사들을 대상으로 한 연구가 주를 이루고 있는 실정이다. 그러나 한국경제의 구조조정의 촉진제 역할을 하는 벤처기업을 대상으로 한 도산예측연구는 전무한 상태에 있다.

따라서 본 논문은 많은 재무비율 중 통계적으로 유의한 재무비율만을 선택하여 국내 벤처기업을 대상으로 도산과 관련된 벤처기업

고유의 변수를 도출하고자 한다. 한편 본 연구는 80년대 후반부터 도산예측연구에서 높은 분류예측율을 나타내고 있는 인공신경망을 이용한 도산예측모형과 기존의 도산예측에 주로 사용되어 오던 통계적 모형인 판별분석모형과의 예측 성능을 비교하는 데 그 목적이 있다.

2. 연구의 체계

일반적으로 기업도산에 대한 개념은 비교적 광범위하게 해석되어 “기업이 기술적 지급불능이나 법적 도산에 이르게 되는 재무 및 경영활동상 어려움을 겪는 상태(박순식·김병주, 2000)” 정도로 이해되고 있다. 도산에 관해 Beaver(1966)는 법적도산, 사채의 연체, 우선주 배당금의 미지급, 은행의 부도로 정의하였고, Altman(1968)은 도산법에 도산신청을 낸 기업으로 설정하였다. 국내에서 도산예측에 대한 연구는 주로 상장기업을 대상으로 이루어지고 있으며, 도산율 부도발생, 회사정리, 금융기관 관리대상업체, 자본 완전잠식, 영업활동정지 및 폐업 등으로 정의되고 있다.

본 연구에서는 동일비율의 도산기업과 정상기업을 선정한다. 또한 표본기업에 대응하는 기업을 어떻게 선정하느냐에 따라 예측율이 달라질 수 있으므로 좀 더 현실적인 가정을 위해 우량기업을 대응기업으로 선택하지 않고 동종업종의 매출액을 고려하여 도산기업과 유사한 환경의 기업을 선정했다.

연구 대상기업 선정에는 1) 한국신용평가정보(주) 2001년판

KOSDAQ CD, 2) 기업재무총람, 3) 한국상장회사협의회 기업정보 웨어하우스 TS2000 등을 사용하였으며, 선정된 기업의 재무비율 자료는 한국상장회사협의회 기업정보 웨어하우스 TS200를 사용하여 수집하였다.

기존의 도산예측에 대한 인공지능적 접근방식과 통계적 접근방식의 기존문헌을 연구하며, 변수(재무비율)를 설정 및 자료를 취합한 뒤 자료의 전처리과정을 실시한다. 결측치(Missing Value) 및 이상치(Outliers)는 제거하고, 인공 신경망의 입력을 위한 변수는 정규화 과정을 실시한다.

전처리 과정 후, 사용할 분류기법인 인공신경망을 이용하여 도산·정상을 분류한다. 본 연구에서 사용된 모형의 성능을 평가하기 위해 기존의 통계적 모형인 다변량판별분석모형과 인공신경망의 도산예측의 성능을 비교한다. 다변량판별분석에는 SAS Version 8.1을 사용했으며, 인공신경망의 도산예측에는 신경망 설계 프로그램인 NeuroDimension사에서 만든 NeuroSolutions Version 4.18를 사용했다.

II. 이론적 배경

1. 벤처기업

벤처기업이 무엇인가에 대해 명확히 확립된 개념은 없으나, 일반적으로 신사업자, 기술집약형 기업 등으로 표현하여 국가별로 달리 정해오고 있다. 외국의 벤처기업 정의를 살펴보면, 미국의 경우 「중소기업투자법」에서 “위험성이 크나 성공할 경우 높은 기대수익이 예상되는 신기술 또는 아이디어의 독립기반 위에서 영위하는 신생기업(new business with high risk-high return)”으로 규정하고 있다. 일본에서는 「중소기업의 창조적 사업활동 촉진에 관한 임시조치법」에서 “중소기업으로서 R&D 투자비율이 매출액의 3%이상인 기업, 창업 후 5년 미만인 기업”으로 분류하고 있다. OECD에서는 “R&D 집중도가 높은 기업” 또는 “기술혁신이나 기술적 우월성이 주요 요인인 기업”으로 정의하고 있다.¹⁾ 우리나라에서는 그동안 벤처기업을 위한 특별한 법규정 없이 “창업투자회사와 신기술금융사가 출자한 총액이 자본금의 10%를 넘는 기업”이란 애매한 개념을 사용해왔다. 1997년 7월 제정된 「벤처기업육성에 관한 특별조치법」에서는 벤처기업을 다음과 같이 4가지로 구분하였다. 첫째, 중소기업창업투자회사, 중소기업창업투자조합, 신기술사업금융업자 및 신기술사업투자조합의 투자총액이 당해 기업 자본금의 100분의 10 이상인 기업, 둘째, 연간 총매출액 대비 연구개발비의 비율이 100분

1) 박용구·강신겸(1997), “벤처기업 육성을 위한 입지지원방안”, 삼성경제연구소 연구보고서.

의 5이상인 기업, 셋째, 특허권 및 실용신안권, 의장권 등의 기술을 주된 부분으로 하여 사업화하는 기업, 넷째, 기타 공업기반기술개발 사업의 성과를 사업화하거나 신기술의 사용 또는 지식을 집약하는 사업 등으로서 대통령이 정하는 사업을 영위하는 기업 등이 여기에 해당된다.

벤처기업의 정의는 국가별로 다르지만 모두 공통적으로 신기술을 포함하고 있다. 벤처의 핵심적 특성은 기업가 정신(entrepreneurship), 첨단기술(high technology), 신생기업(new firm), 위험선호(risk taking) 등을 들 수 있다.²⁾ 다시 말하면, 소수의 기술창업인이 기술혁신의 아이디어를 상업화하기 위해 설립한 신생기업이며, 개인투자가 엔젤 혹은 벤처캐피탈로부터 자금을 조달하며 높은 위험부담이 있으나 성공할 경우 높은 기대이익이 예상되는 기업을 말한다.

2. 판별분석법

판별분석(discriminant analysis)은 둘 이상의 모집단이 있을 때 관찰치가 지니는 특성에 따라 관찰치를 선형적 그룹 중의 하나로 분류 또는 예측하는 통계적 기법이다.(Altman, 1968) 이 기법은 종속변수가 주로 남성 혹은 여성, 도산 혹은 비도산과 같은 질적 형태로 나타나는 문제의 구분 또는 예측에 사용된다. 도산예측의 경우 모집단이 둘(도산기업군과 비도산기업군)인 경우로서 재무비율이나

2) 김정호·김완표(1999), “벤처 붐의 조성파 NASDAQ 진출”, 삼성경제연구소 연구보고서.

주식시장 정보 등이 관찰치에 해당된다. 이러한 판별분석은 1936년에 피셔(R.A Fisher)에 의해 개발되었으며, 주로 생물학이나 행동과학 등에서 질적 성격을 갖는 변수의 분류에 많이 이용되었으나, 오늘날에는 기업의 도산예측, 개인이나 기업의 신용평가, 발행대상채권의 등급예측, 생산관리, 시장분석 등에 광범위하게 이용되고 있다.

이와 같은 판별분석의 사용 목적은, 첫째, 둘 또는 그 이상의 선형적 집단의 평균 판별평점이 통계적으로 유의성이 있을 만큼 차이가 있는지 없는지를 결정하기 위하여, 둘째, 도출된 판별함수를 이용하여 개발대상의 판별평점을 계산하여 그 대상이 어느 특정 집단에 속하는지를 결정하기 위해서, 셋째, 집단간의 차이를 가장 잘 설명해 주는 독립변수를 결정하는데 있다. 판별분석은 대체로, ① 판별함수의 도출단계, ② 판별함수의 검증단계, ③ 해석단계를 거쳐 행해진다.³⁾

3. 인공신경망의 이론적 배경

3.1 인공지능의 정의

인공지능(Artificial Intelligence : AI)이란 인간의 기억이나 경험에 의하여 가지게 되는 선형적 체험 등의 지적능력을 학습을 통하여 컴퓨터에 지식베이스로 구축하고 구축된 지식베이스를 이용하여 주어진 자료를 추론하고 그 결과를 예측하고 설명하는 기능 즉, 인간

3) 김충련(1997), SAS라는 통계상자, 데이터플러스.

의 휴리스틱(heuristic) 측면을 컴퓨터가 가지게 됨을 말하는 것이다.4) 현재 인공지능에서 가장 주목을 받고 있는 분야는 귀납적 학습 방법과 인공신경망이라고 볼 수 있다.

의사결정과정을 살펴보면 [그림 1]과 같이 요약할 수 있는데, [그림 1]에서 의사결정자의 추론과정과 인공지능을 비교하여 볼 때 인공지능은 특정 문제에 대한 사례를 투입자료로 활용하여 시스템 내에서 처리과정을 거쳐 지식베이스를 구축하게 되며 구축된 지식베이스를 이용하여 다른 사례에 대한 예측을 한다는 점이 두드러지는 차이점이라 할 수 있다.

[그림 1] 의사결정자의 추론과정



출처: Vonk. E, L. C. Jain, R. P. Johnson(1997), Automatic Generation of NeuralNetwork Architecture Using Evolutionary Computation, World Scientific Publishing Co. Pte. Ltd, p.25.

지식습득(Knowledge Acquisition)을 위하여 기계학습에서 주로 사용하고 있는 방법은 귀납적 학습방법이다. 귀납적 학습방법은 시간과 비용이 많이 드는 비효율성을 가지는 연역적 학습방법을 개선하고자 제안된 것으로 현재 전문적 지식을 학습하는 가장 효율적인 방법으로 사용되고 있다. 즉, 이 방법은 기존의 사례에서 지식을 학

4) Lippmann, R(1987), "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, April, 1987, pp.4-22.

습(Learning From Example : LFE)하여 의사결정트리 형태로 구성된 지식베이스를 구축하고 구축된 지식베이스를 활용하여 현재 및 미래의 의사결정에 이용하는 방법이다.⁵⁾ 귀납적 학습방법의 판별·예측에 대한 정확성 여부는 지식베이스의 효율성에 의하여 결정되므로 얼마나 정확한 알고리즘을 가진 지식학습방법을 사용하는가가 매우 중요하게 된다. 현재까지 개발된 귀납적 학습방법의 지식학습 방법으로는 ID3를 비롯하여 ACLS, CRIS, ID4, ID5 등의 다양한 방법이 있다.

인공신경망은 단순처리단위(Simple Process Unit)인 노드(nodes)에 연결가중치(Connection Weight)를 부여하여 각 단위간에 상호작용을 가능하게 하는 네트워크이다. 이러한 네트워크는 전통적인 처리방식인 직렬식 처리를 개선하여 병렬식 분산처리(Parallel Distribution Process : PDP)를 실현하여 인간의 휴리스틱한 분석기법을 구현하고자 하는 데 목적이 있다. 인간의 신경세포는 인공신경에서는 노드, 즉 처리단위로서 표현되고, 인간의 신경망은 연결가중치로 표현된다.⁶⁾ 이러한 인공신경망의 네트워크는 정보의 처리속도를 빠르게 할 뿐만 아니라 정보에 담긴 다중적 의미를 파악하는데 도움이 된다. 인공신경망의 지식학습 방법으로는 역전파학습(Back-propagation Learning)방법이 가장 널리 알려져 있다.

5) 최신재(1995), “기업도산예측을 위한 인공신경망과 귀납적 학습방법의 통합적 적용”, 성균관대학교 대학원, 박사논문.

6) Lippmann, R(1987), “An Introduction to Computing with Neural Nets”, IEEE ASSP Magazine, April, 1987, pp.4-22.

3.2 인공신경망

인공신경망은 외부의 정보입력에 대하여 역동적으로 정보를 처리하기 위하여 단순하지만 상호연결성이 높은 수많은 처리단위로 구성되어 있는 컴퓨터 시스템이라고 정의할 수 있다.⁷⁾ 이 정의로 볼 때 인공신경망을 구성하는 주요 구성요인으로는 처리단위와 상호연결성을 들 수 있다. 처리단위는 생체내의 신경세포와 비슷한 것으로서 수많은 입력신호(input signal)를 받고 이 입력신호에 근거하여 단일의 출력신호를 발생시킨다. 하나의 처리단위에서 발생한 출력신호는 다른 수많은 처리단위로 가중치화된 상호연결성을 통하여 입력신호로서 전송된다.

인공신경망은 서로 연결되어 있는 다수의 처리단위들로 구성되어 있기 때문에 다음과 같은 견고성과 학습성을 갖고 있다(Lippmann, 1987).

첫째, 견고성(Robustness) : 인공신경망은 일반적으로 오류극복성을 갖는데 이는 곧 특정 몇몇 처리단위에 오류가 발생하여도 인공신경망의 전체적인 기능이 크게 영향을 받지 않음을 의미한다. 따라서, 인공신경망의 성과는 오류의 정도가 증가됨에 따라 점차적으로 감소하는 추세를 보임으로써 급격히 변하는 환경이나 예측치 못했던 환경에서 인공신경망이 안정적으로 가늠하는데 크게 공헌한다.

7) Nelson, M. C. and Illingworth, W. T(1991), "A Practical Guide to Neural Nets", Addison-Wesley, Reading, MA.

둘째, 학습성(Learnability) : 인공신경망은 주어진 학습 입출력 자료로부터 숨겨진 규칙성(Regularities)을 찾아낼 수 있다. 이러한 규칙성은 인공신경망내에 분산된 처리단위에 각각 저장되며 이는 결국 귀납적학습방법의 지식베이스에 해당된다.

이와 같은 인공신경망의 지식은 수리적 가중치로 표현되고 학습 기법에 의하여 어느 정도 자동적으로 얻을 수 있다. 이러한 특징은 결국 인공신경망을 구성하고 있는 다수의 처리단위와 이들 처리단위끼리의 상호연결성(interconnections)에 기인한 것이다.

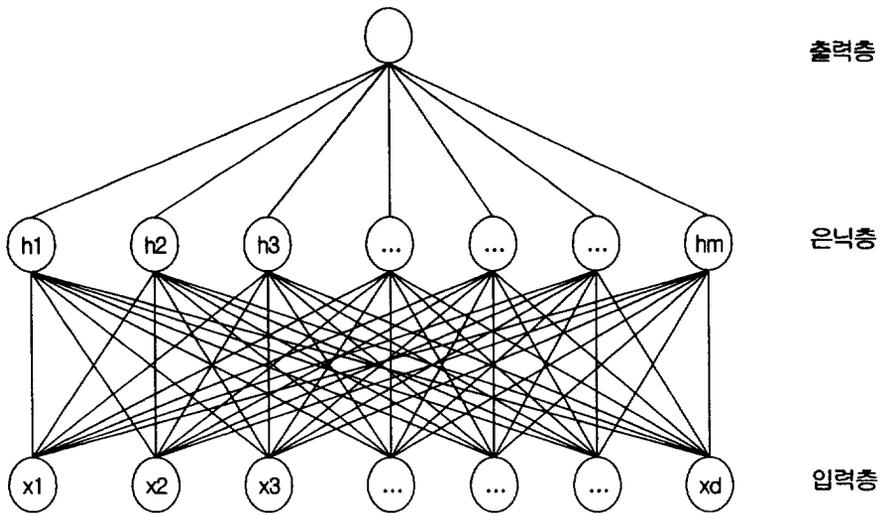
인공신경망의 추론결과는 0에서 1까지의 실수로 나타나게 된다. 이를 인공신경망에서는 퍼지화(fuzzification)라고 한다. 예를 들면 도산기업의 추론 결과는 “도산 가능성(0.20), 비도산 가능성(0.80)”으로 나타나게 되며, 이 때의 값 0.20와 0.80이 해당 도산 및 비도산 가능성의 퍼지값이 된다.

3.2.1 다층퍼셉트론(Multi Layer Perceptron)

가장 널리 사용되는 신경망 모형은 다층퍼셉트론과 이의 학습 알고리즘인 역전파(back-propagation) 알고리즘이다. 다층퍼셉트론은 입력층과 출력층 사이에 하나 이상의 은닉층이 존재하는 신경망이다. 다층퍼셉트론은 입력층, 은닉층, 출력층 방향으로 연결되어 있으며, 각 층내의 연결과 출력층에서 입력층으로의 직접적인 연결은 존재하지 않는 전방향(feedforward) 네트워크이다.

다층퍼셉트론은 비선형 함수적 사상(mapping)을 이용하여 일련의 입력값들을 일련의 출력값으로 변환하는 틀로 간주할 수 있다. ([그림 2] 참조)

[그림 2] 다층퍼셉트론의 아키텍처



d개의 입력변수, l개의 출력변수, 그리고, m개의 은닉층의 노드가 존재한다고 할 때, 은닉층의 j번째 노드의 총입력값은 (식 1)과 같이 바이어스(bias)를 포함하여 각 입력변수들의 선형결합을 통해 계산된다.

$$net_j = \sum_i W_{ji} X_i \quad (\text{식 1})$$

여기서는 W_{ji} 는 입력층의 노드 i 와 은닉층의 노드 j 를 연결하는 웨이트를 나타낸다. 그리고 바이어스 X_0 는 항상 1의 값을 갖는 상수항이다. 은닉층의 노드 j 의 출력값(Z_j)은 앞에서 계산된 총입력값에 전이함수를 취하여 (식 2)와 같이 계산된다.

$$Z_j = g(\text{net}_j) \quad (\text{식 2})$$

이때 함수 g 로는 전이함수가 사용될 수 있는데, 일반적으로 로지스틱 시그모이드(logistic sigmoid)함수와 하이퍼볼릭탄젠트 시그모이드(hyperbolic tangent sigmoid)함수가 많이 사용되고 있다. 시그모이드함수는 S자 형태를 갖는 함수로 주어진 계산단위의 총입력값을 $[0, 1]$ 과 $[-1, 1]$ 사이의 값으로 전이시키는 함수이다.

출력층 노드의 총입력값은 이와 같이 계산된 은닉층의 노드들의 출력값을 다시 가중합산하여 구해지는데 (식 3)과 같이 나타낸다.

$$\text{net}_k = \sum_j W_{kj} Z_j \quad (\text{식 3})$$

이 값은 다시 전이함수를 통해 변환되어 출력값을 산출한다.

$$Y_k = \tilde{g}(\text{net}_k) \quad (\text{식 4})$$

(식 4)에서, 함수 g 대신 \tilde{g} 를 사용한 이유는 은닉층의 전이함수와 출력층의 전이함수가 서로 다를 수 있기 때문이다.

다층퍼셉트론의 각 노드들을 연결하는 웨이트(weight)들은 초기에는 무작위로 할당되며, 학습을 통해 개선되는 방식을 취한다. n개의 학습 예가 있다고 할 때, 이 중에서 표본 p의 입력값과 출력값의 쌍인 (X_p, T_p) 가 뽑아지고, 이 중 X_p 는 입력층으로 전달된다. 그러면, 입력벡터들은 은닉층으로 전달되고, 은닉층의 벡터들은 다시 출력층으로 전달되어 출력벡터 Y_p 가 계산된다. 성과는 목표출력값인 T_p 와 실제 네트워크의 출력값인 Y_p 를 비교함으로써 측정되는데, 성과척도로는 오차자승(sum of squared error)이 많이 사용된다.

$$Err = \sum_p Err_p = \frac{1}{2} \sum_p \sum_k (Y_{pk} - T_{pk})^2 \quad (\text{식 5})$$

(식 5)에서 p는 예(exemplar pattern)를 가리키는 첨자이고, k는 출력층의 계산단위를 가리키는 첨자이다. 출력층에서 계산된 오차는 출력값이 계산된 방향의 반대방향으로 역전파되고, 가중치들은 오차의 기여 정도에 따라 수정된다.

3.2.2 역전파학습(Backpropagation Learning)

인공신경망의 지식베이스인 연결가중치(connection weights)는 역전파학습과 같은 감독학습(supervised learning)에 의해서 결정되거나 또는 경쟁학습과 같은 비감독학습(unsupervised learning)에 의해서 결정된다. 역전파학습은 인공신경망의 학습방법으로 가장 많이 이용되며, 그 이유는 역전파학습이 갖는 넓은 응용력과 높은 일반화

능력에 기인한다.⁸⁾

역전파학습을 이용한 삼중 전향신경층(three layer feedforward) 인공신경망에서 입력층(input layer)은 환경과 상호반응하며 외부입력을 받아 인공신경망에 전달하는 역할을 한다. 또한, 출력층(output layer)은 주어진 외부입력에 대한 적절한 출력을 내보내는 역할을 하며, 입력층과 출력층 사이의 신경망을 은닉층(hidden layer)이라고 하는데 주어진 입력으로부터 특성을 추출하여 출력층으로 보내는 기능을 한다.

은닉층의 수와 처리단위규모의 결정은 주관적으로 결정되어 인공신경망의 응용에 따라 달라지며 따라서 그 타당성은 실험을 통하여 확인하여야 한다. 그러나 일반적으로 은닉층의 처리단위는 인공신경망의 성과에 커다란 영향을 주지는 않는다고 알려져 있다. 만약, 입력자료가 특성추출(feature extraction)이 용이하지 않는 자료로 구성되어 있으면 그러한 입력자료로부터 고차원의 특성을 추출하기 위해서는 여러 개의 은닉층이 요구된다. 반면에 입력자료가 이미 어느 정도 고차원의 특성치를 나타내고 있으면 하나 또는 두 개 정도의 은닉층만 있어도 거의 모든 형태의 문제해결 공간을 구성할 수 있다.⁹⁾

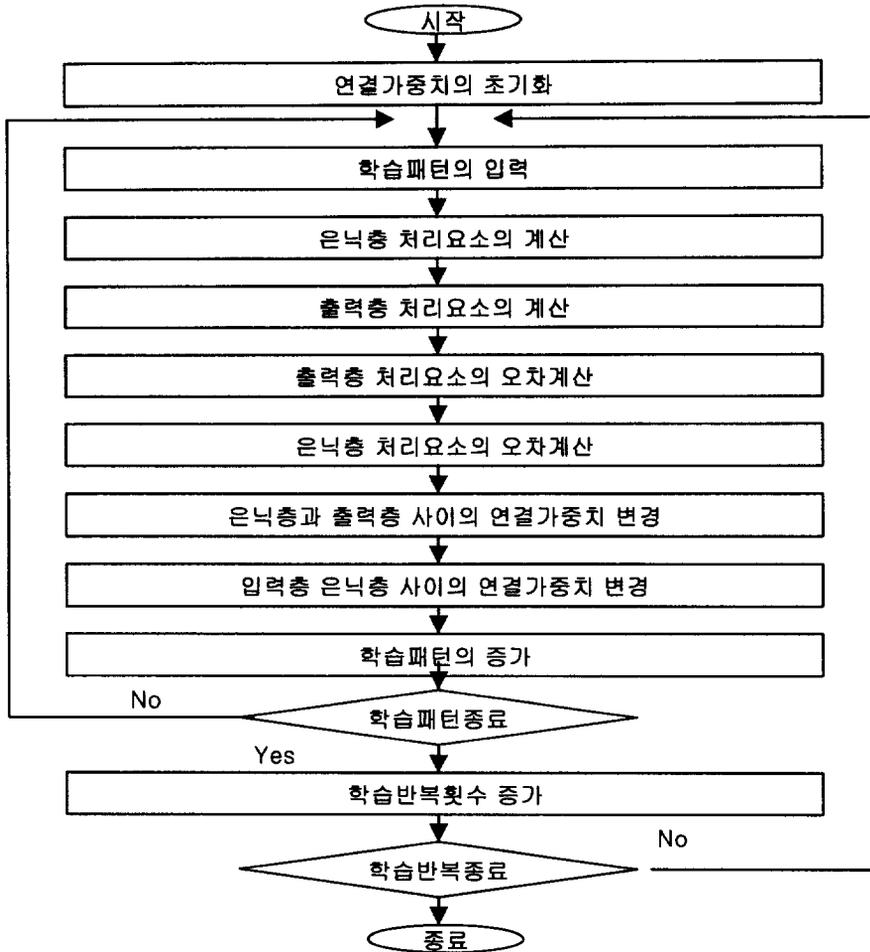
출력처리단위에서 실제값과 인공신경망의 계산값 사이의 차이를 구하는 일반적으로 차이의 제곱합을 구하여 이를 시스템 오류(system error)라고 한다. 이 같은 과정을 전향적 단계(forward pass)라고 한다. 다음 단계로서는 이러한 시스템 오류를 줄이기 위

8) Odom, M. D. and Sharda, R(1990), "A Neural Networks by Soft Weight-sharing", Neural Computation, 4(4), pp.473-493.

9) Lippmann, R(1987), "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, April, 1987, pp.4-22.

한 과정으로서 이를 위해 연결가중치를 조정하는데 이를 역전파 학습방법이라 한다. 역전파학습은 시스템 오류를 작게 하기 위하여 델타규칙(Delta Rule)을 사용하는데, 이러한 가중치의 변화는 사례가 증가될수록 시스템의 오류를 지속적으로 감소시켜 결국에는 가중치의 변화가 거의 없게 된다. 이러한 역전파 과정은 출력층에서 입력층으로 진행되므로 후향적단계(backward pass)라고 한다. 역전파학습 알고리즘의 학습과정은 [그림 3]과 같다.

[그림 3] 역전파학습 알고리즘의 학습 과정



출처: Vonk. E, L. C. Jain, R. P. Johnson(1997), Automatic Generation of Neural Network Architecture Using Evolutionary Computation, World Scientific Publishing Co. Pte. Ltd, p.58.

4. 인공신경망과 판별분석의 차이점

공학적 기법의 인공신경망과 통계적 기법인 판별분석의 차이점은 다음과 같다.

첫째, 인공신경망은 독립변수의 분포에 아무런 가정이 없는 반면에 판별분석은 독립변수의 분포가 정규분포라는 것을 가정하고 있다.

둘째, 두 방법은 서로 다른 함수를 가진다. 즉, 인공신경망은 독립변수와 종속변수간에 비선형함수를 가정하는 반면에 판별분석은 독립변수와 종속변수간에 선형함수를 가정하고 있다.

셋째, 인공신경망은 기존의 지식에 새로운 유형의 사례를 학습할 때 지식베이스가 확장 견고하게 되어 일반성과 안정성을 가지게 된다. 따라서, 인공신경망은 시간이 경과하여 새로운 유형의 사례가 학습될수록 그 성과가 향상되는 반면에 판별분석은 새로운 사례의 학습에도 불구하고 지식베이스의 확장이 불가능할 뿐만 아니라 일반화가 되고, 안정성 있는 지식베이스를 가질 수 없다.

넷째, 인공신경망은 다수의 연결가중치로 상호연결되어 있어 미래환경의 변화를 시스템 자체에서 조정하여 환경변화에 대한 적응능력을 가지고 있다. 따라서 판별분석에 비하여 적응능력을 가지고 있다.

III. 선행 연구

1. 다변량판별분석을 이용한 예측 모형

1960년대 후반 통계적인 분석기법이 발전하게됨에 따라 기업도산 예측에 있어서도 다변량 통계분석 기법들이 도입되기 시작했다.

도산예측에 가장 많이 사용되는 통계적 모형은 다변량판별분석, 로짓분석¹⁰⁾ 및 프로빗분석¹¹⁾ 등을 들 수 있다. 다변량판별분석은 Altman에 의해 개발되어 이후에도 많은 연구가 시행되었다.

Altman(1968)은 1946년부터 1965년까지 파산법에 의해 파산신청을 낸 33개 제조기업과 짝짓기 표본추출 기법을 적용하여 업종이 같고 규모가 비슷한 33개의 건전기업을 선정하였다. 그는 직관에 의해 선정된 22개의 재무비율을 도산 1년 전의 자료를 이용하여 가장 예측정확도가 높은 5개의 재무비율을 산출하였다. 프로파일분석, 변수상호간의 상관관계분석, 판별함수를 사용하여 예측정확도를 산출한 결과 도산기업의 경우 96%, 정상기업의 경우 79%의 정확성을 보여 비교적 높은 판별력을 보여주었다. 그는 이 모형을 Z-score 모형으로 불렀으며, 판별식을 통해 계산된 Z값이 2.675를 밑돌면 도산 그룹에 속하는 것으로 예측을 하였다. 그러나 변수선정에 있어 이론적 근거를 제시하지 못하고 시행착오법에 의존하고 있다는 점이 문

10) 로짓분석(로지스틱 회귀분석)은 판별분석과 같은 용도의 분석기법으로 판별분석의 다변량정규분포의 가정이 불명확할 때 사용하는 방법이다.

11) 독립변수가 명목척도 또는 서열척도와 같이 정성적인 척도와 등간척도가 섞여 있으면서 변수들이 다변량 정규분포를 한다고 할 때 사용하는 분석기법이다.

제점으로 지적되고 있다.

Altman, Haldeman, & Narayanan(1977)은 1968년 연구의 Z-score 모형을 변형하여 Zeta 모형을 구축하였으며, Zeta 모형의 예측력이 더 우수함을 보였다. 1969년에서 1975년까지의 자료를 이용하여 제조업 외에 도·소매업을 포함시켜 표본 수를 증가시켰다. 27개의 주요 재무변수 중 7가지 변수를 독립변수로 선정하고, 선형 및 2차 함수식을 활용하여 도산예측모형을 개발한 결과, 도산 5년 전에도 예측력이 높음을 보여주었다.

다변량판별분석은 방법론적·통계학적 문제점으로 인하여 그 현실적인 유용성에 제약을 받고 있다. 로짓분석과 프로빗분석은 다변량판별분석의 정규분포의 가정에 대한 문제점을 해결하기 위한 통계적 모형으로 사용되어 왔다.

임영규(1995)는 지금까지 연구됐던 기업부실화 예측연구들의 문제점을 개선하고 기업부실화를 보다 정확히 예측할 수 있는 모형 개발에 중점을 두었는데, 재무제표정보 및 주식시장정보를 이용한 부실화예측모형 등 2개의 개발모형과 이들 두 정보를 동시에 이용한 통합모형, 이중부실화예측모형 등 모두 4개의 모형을 개발한 후 이들 네 모형의 유용성을 분석했고, 판별예측력을 높이기 위해서는 다중공선성 문제에도 불구하고 통합모형을 사용하는 것이 유리하며, 부실예측력을 높이기 위해서는 이중부실화예측모형을 사용하는 것이 합리적이라고 제시하였다.

정헌웅(1998)은 1996년 3월 은행연합회에서 개발하여 사용하고 있는 '기업신용평가표'중 재무상태를 나타내는 16개 재무비율을 이용하여 다변량판별분석 방법으로 기업도산예측에 관한 연구를 하였던

데 1989~1995년 사이에 도산한 40개 기업을 대상으로 16개의 재무제표변수 중 자기자본비율, 현금흐름/총부채, 매출액경상이익률, 총자본순이익률, 영업자산회전을 등 5개 변수가 기업부실에 유의한 변수로 채택되었으며, 도산예측모형의 예측력은 도산 당해연도와 가까울수록 높은 예측력을 나타냈다.

강종만·홍성희(1999)는 부실예측에 효과적이라고 알려져 있는 Z점수모형과 ZETA모형, 로짓모형을 사용하여 각 모형의 부실예측성과를 비교·평가하였다. 1991~1997년까지 7년의 기간동안 상장제조업체를 대상으로 65개 부실기업과 이에 대응하는 65개 정상기업을 선정하여 144개의 재무회계변수 뿐만 아니라 9개의 비재무적 변수를 망라하여 이들 중에서 부실예측에 가장 유의한 변수를 선정하였는데 8개의 재무회계변수와 추가초과수익률과 총위험의 2개의 비재무회계변수가 추출되었고 결과는 다변량판별분석을 이용한 Z점수모형이 각각 86.2%(부실기업), 84.6%(우량기업)의 판별력을 보여 가장 우수한 것으로 나타났다.

2. 인공신경망을 이용한 예측 모형

기업부실예측의 정확성 및 객관성을 높이기 위해 다중판별분석, 프로빗, 로짓분석 등의 통계학적 기법들이 전통적으로 이용되어왔으나, 1980년대 후반부터 인공지능기법의 하나인 귀납적 학습방법, 신경망모형 등이 부실예측에 응용되기 시작하였다. 인공신경망 모형이

통계적 방법보다 우수한 예측력을 보였다는 연구가 많으나, 아직 많은 의문점이 남아 있는 실정이다. 향후 이론적 보완 및 실무적용 과정의 어려움을 극복하고, 예측력 향상을 위한 방법의 하나로 통계적 기법과의 결합 등도 고려되어야 할 점이다. 인공지능을 이용한 예측 기법은 주로 새로운 기법을 기업부실예측이라는 영역에 적용하려는 노력으로 도산예측에 있어서의 이론적 측면보다는 주로 방법론 자체와 판별분석이나 로짓분석과 같은 다른 예측기법과의 비교가 주를 이루고 있다.

Odom and Sharda(1990)의 연구에서는 1975년에서 1982년 사이에 도산한 기업을 대상으로 65개의 도산기업과 64개의 건전기업을 표본으로 하였으며, Altman(1968)의 연구와 동일한 변수를 사용하였다. 그들은 신경망과 판별분석에서 예측력에 차이가 있는 지에 중점을 두고, 역전파 학습알고리즘을 약간 변형한 다층퍼셉트론을 다변량판별분석 모형과 비교하였으며, 표본을 세 가지로 나누어 비교한 결과 인공신경망 모형이 모두 예측력에서 우수함을 보여 도산예측에 있어 신경망 응용의 가능성을 보여주었다.

Liang, Chandler, Han & Roan(1991)의 연구에서는 프로빗분석과 ID3, 인공신경망을 이용하여 도산예측에 대한 실험을 하였다. 이들의 연구에서는 인공신경망의 성과가 가장 높고 프로빗분석과 귀납적학습방법의 결과는 정확성의 차이를 보이지 못하였다.

Berry and Treigueiros(1991)의 연구에서 그들은 다변량판별분석과 인공신경망을 이용하여 도산예측을 한 결과 인공신경망의 예측정확성이 다변량판별분석의 예측정확성보다 높음을 보고하고 있다.

Tam and Kiang(1992)의 연구는 1985년에서 1987년 사이에 59개

의 도산은행과 자산규모, 지점 수 등을 기준으로 짝지어진 59개의 건전은행을 대상을 선형판별분석, 로짓모형, kNN, ID3, 신경망 모형 등을 비교하였다. 여러 방식으로 실험을 한 결과, 신경망 모형이 다른 모든 모형에 비해 우수하다고 하였다. 신경망 모형은 설명력 차원에서는 제한이 있지만 다른 모형에 비해 예측정확성이 높고, 모형의 강건성이 존재하며, 시간에 따라 데이터가 추가되어 감에 따라 적응성을 보인다는 점에서 기존 도산예측 모형의 대안으로 활용될 수 있다고 하였다.

이건창·김명중·김혁(1994)은 1979년에서 1992년까지 한국증권거래소에 상장된 기업을 모집단으로 하여 81개의 도산기업을 추출하였고, 매출액과 업종을 기준으로 하여 85개의 건전기업을 선정하여 연도별로 세 그룹으로 누적적으로 표본을 선정하였다. 성장성, 수익성, 안정성, 현금흐름, 활동성, 기업신용율의 57개 변수 중 일부 변수를 이용하였다. 다변량판별분석, 귀납적 학습방법(ACLS), 인공신경망의 예측력을 비교하였는데, 인공신경망은 다시 다변량판별분석(MDA)에 의해 선택된 변수를 이용한 인공신경망 I 과 귀납적 학습방법에 의해 선택된 변수를 이용한 인공신경망 II로 구분하여 비교한 결과 인공신경망II > 인공신경망 I > 귀납적 학습방법 > MDA의 순으로 우수한 예측력을 보였으며, 특히 분석용 표본의 수가 많아짐에 따라 예측정확도의 차이가 더욱 커진다고 하였다.

Altman, Marco & Varetto(1994)은 1982년부터 1992년까지 이태리 기업을 모집단으로 하여 554개의 부실기업과 동수의 건전기업을 분석 대상으로 선정하였다. 그들은 처음에 판별분석 모형, 로짓 모형, 신경망 모형을 비교하였으나 로짓 모형의 예측력이 떨어져 판별

분석 모형과 신경망 모형의 예측력을 비교하였다. 그들은 인식력 측면에서는 신경망 모형이 매우 우수하였으나, 표본의 추정방법으로 볼 때에는 큰 예측력 차이를 보이지 않았으나, 신경망의 수용하기 힘든 행동패턴의 존재로 인해 판별분석이 전체적으로는 더 좋은 모형일 수 있음을 시사하였다. 그러나, 실험을 통해 명확한 결론을 내리지는 못했다. 앞으로 더 정밀한 연구가 필요하고 두 모형의 상호 보완적인 결합방법이 연구될 필요가 있다고 결론지었다. 그들은 신경망이 인식력과 유연성 측면에서는 강점이 있으나, 처리시간이 길고, 신경망 아키텍처의 식별을 위해 많은 실험이 필요하며, 블랙박스모형이라는 점과 신경망의 과도적합 문제를 지적하였다. 그들은 과도적합을 신경망의 큰 문제점으로 삼고 있으나, 최근 중점적으로 연구되고 있는 과도적합을 막고 일반화 능력을 제고하기 위한 방법들을 전혀 고려하지 못하였다는 한계를 안고 있다.

Shin, Jo & Han(1996)은 1993년에서 1995년 사이에 도산한 1,274개의 기업과 그에 대응하는 1,274개의 건전기업을 대상으로 판별분석(MDA), 로짓분석(Logit), 신경망 모형(NN)을 통해 분석하였다. 이중 90%의 데이터를 분석용으로 사용하고 10%의 데이터를 시험용으로 무작위 추출하는 과정을 3번 반복하여 실험한 결과, 예측력은 $NN > \text{Logit} > \text{DA}$ 으로 나타났으나, 신경망과 다른 모형과의 차이는 경미함을 보였다. 또한 GA(Genetic Algorithm)¹²⁾를 이용하여 세

12) 교배연산(crossover), 돌연변이(mutation)을 통해 생성된 새로운 다음 세대(generation)의 개체군(population)들 중에서 그들끼리의 적합도(fitness : 우열의 법칙에서 살아남을 유전자들을 선택함)를 평가하고 가장 적합도가 높은 개체(individual)를 다음 세대의 개체군으로 만드는 알고리즘이다.

가지 모형을 결합한 결과, 제일 좋은 예측력을 보인 신경망 모형보다 전반적으로 예측력이 미미하기는 하지만 조금 더 우수함을 보여 모형 결합의 가능성을 보여 주었다.

한인구·조홍규(1996)는 국내기업을 대상으로 1991년~1993년 사이의 재무자료를 바탕으로 도산예측에 대한 MDA와 인공신경망의 예측정확성 비교 연구에서 인공신경망이 더 높은 예측력을 나타낸다는 결론을 얻었다.

한길석·이상로(1999)는 1992년부터 1998년까지 도산한 109개 기업과 비도산 기업 109개를 대상으로 신경망 모형과 로짓분석, 판별분석간의 도산예측능력을 비교한 결과 인공신경망의 도산예측율이 판별분석에 비해 높게 나타났지만 이는 미국기업의 도산연구에 비해 상대적으로 작은 수치이다. 우리나라의 도산예측 정확도가 미국의 것에 비해 낮은 이유는 도산결정에서 정부가 많은 역할을 수행하고 기업의 특성을 나타내는 자료 중 계량화되지 않은 질적자료가 상대적으로 많은 영향을 미치는 기업환경에 기인한다고 했다.

이상의 선행연구들의 분석기법과 예측을 결과를 대비해 보면 <표 1>에서 보는 바와 같다.

<표 1> 인공지능망을 이용한 도산예측에 관한 선행연구의 비교

선행연구	분석기법	예측을 결과
Odom and Sharda (1990)	다변량판별분석 인공신경망	인공신경망>다변량판별분석 (도산예측에 있어 신경망 응용의 가능성 선보임)
Liang, Chandler, Han and Roan(1991)	프로빗분석 ID3 인공신경망	인공신경망 > 프로빗 = ID3
Berry and Treigueiros (1991)	다변량판별분석 인공신경망	인공신경망>다변량판별분석
Tam and Kiang (1992)	선형판별분석 로짓모형 kNN ID3 인공신경망	인공신경망 모형이 가장 우수 (신경망은 예측정확성, 모형의 강건성, 데이터 추가의 적응성 높음)
이건창 · 김명종 · 김혁 (1994)	다변량판별분석 귀납적학습방법 인공신경망	인공신경망>귀납적학습방법>MDA (분석용 표본수의 증가에 따라 예측정확도 차이 증가)
Altman, Marco and Varetto (1994)	다변량판별분석 로짓모형 인공신경망	다변량판별분석>인공신경망>로짓 (신경망의 과도적합, 긴처리시간 등 한계점 지적)
Shin, Jo and Han (1996)	판별분석 로짓모형 인공신경망	인공신경망>로짓모형>MDA
한인구 · 조홍규 (1996)	다변량판별분석 인공신경망	인공신경망>다변량판별분석
한길석 · 이상로 (1999)	판별분석 로짓모형 인공신경망	인공신경망>로짓모형>MDA

IV. 연구방법론

1. 표본기업의 선정

1.1 분석대상기업 선정

선행연구에서 살펴본 것과 같이 도산예측에 관한 연구는 대부분 일반상장기업을 대상으로 하였으며 분석대상기업수가 한정되었다. 일반상장기업을 대상으로 도출된 부실예측모형을 기업환경과 제반 여건이 다른 벤처기업에 적용하는 데는 한계가 있을 것이므로 본 연구에서는 벤처기업을 대상으로 표본 추출하여 한국의 벤처기업 도산예측모형을 도출하고자 한다.

본 연구에서 표본기업은 벤처도산기업으로 여기서 말하는 벤처기업이란 KOSDAQ에 상장되어 벤처종목으로 지정된 기업체로 「벤처기업육성에 관한 특별 조치법 제2조」에 나와있는 중소기업을 말한다. 즉, 벤처금융이 10%이상 출자한 중소기업, 매출액 대비 연구개발비 비중이 5%이상인 중소기업, 특허권·실용신안권 등을 주된 부문으로 하여 사업화 하는 중소기업을 말한다.

본 연구에서 도산기업은 다음과 같은 도산의 개념적 정의에 해당하는 기업을 추출하였다. 부도가 발생하였거나 은행과의 거래가 정지된 기업, 회사정리절차를 개시한 기업, 영업활동을 정지하였거나 폐업한 기업, 자본잠식이 3년이상 계속된 기업, 코스닥증권시장에 등록취소를 신고하였거나 코스닥증권시장에 의하여 등록이 취소된

기업 등 5가지로 도산의 개념을 정의하였다.

위의 도산 개념에 의해 도산기업은 1999년, 2000년 중 도산기업으로 판정된 사실이 있는 업체들로 도산 직전 년도의 대차대조표와 손익계산서 등 재무제표의 자료수집이 가능한 업체를 선정했다. 1999년부터 2000년까지 도산한 제조업 벤처기업 34개를 도산기업으로 선정했으며, 이와 대응되는 정상기업은 표본기업에 대응하는 기업을 어떻게 선정하느냐에 따라 예측율이 달라질 수 있으므로 좀더 현실적인 가정 하에 선정하였다. 즉, 우량기업을 대응기업으로 선택하지 않고 짝짓기 표본추출방법을 이용하여 도산기업의 환경과 유사한 동종업종의 매출액을 고려하여 34개 기업을 임의 추출하였다.

최종 분석대상 기업의 수는 총 68개이며 이는 도산기업 34개, 정상기업 34개로 나누어진다. 정상기업의 표본은 도산기업의 표본과 업종, 매출액이 비슷한 기업으로 선정했으며 이러한 기준에 의해 정상기업의 표본으로 2001년말 여신불량 및 사고발생 사실이 없고, 1998년, 1999년도의 대차대조표와 손익계산서 등 재무제표의 자료수집이 가능한 업체를 대상으로 한다.

1.2 검증용 표본의 선정

모형의 구축 시 이용된 자료를 사용하여 계산된 자체예측력은 모형의 진정한 예측력을 과대 평가한다. 따라서, 객관적인 예측력에 대한 분석은 모형구축 시 이용되지 않은 검증용 표본을 이용하여 이루어져야 한다.

본 분석에서 모형의 검증을 위한 검증용 표본은 표본선정에 의해 분석용 표본으로 이용되지 않은 기업으로 하며, 표본선정 기준은 분

석용 표본과 동일하고 표본규모도 가급적 비슷한 수준으로 설계한다.

판별분석과 인공신경망 분석을 위해서는 추정표본과 검증표본이 필요하다. 기존의 연구들은 각 표본에서 도산기업과 정상기업의 구성비율을 50대 50으로 구성하였다. 그러나 현실적으로 도산기업은 전체 기업 중 소수부분만 차지하고 있다. 이러한 환경 하에서 도산기업과 정상기업의 비율을 일률적으로 50%씩 구성하면 Zmijewski(1984)가 지적한 것처럼 도산기업이 상대적으로 과대 추출되어 도산기업의 예측율이 과대 평가되며, 정상기업에 대한 예측율이 과소 평가될 수 있다. 즉, 각 표본에서 도산기업과 정상기업을 어떻게 구성하느냐에 따라 예측율이 달라질 수 있다. 하지만, 대부분의 연구결과 도산기업과 정상기업의 비율을 동일하게 했을 경우 가장 예측력이 높게 나타났다. 본 연구에서도 그 결과를 반영하여 동일비율의 도산기업과 정상기업을 표본으로 선정하였다.

특히, 동일한 표본선정 기준에 의해 모든 표본을 1차 추출한 후 이를 다시 무작위의 단계적 추출법에 의해 추정용과 검증용으로 분류하여 표본추출오류를 최소화하고자 하였으며, 추정용 표본과 검증용 표본의 표본 기업수는 <표 2>와 같다.

<표 2> 표본 기업수

구 분	정상기업	도산기업	소 계
추 정 용	17	17	34
검 증 용	17	17	34
합 계	34	34	68

2. 도산예측변수의 선정

본 연구에서는 기존연구에서의 변수선정의 주관성 배제 및 벤처기업의 중요한 변수들의 누락으로 인한 예측정확성 감소를 막기 위해 한국상장회사협의회의 기업정보 웨어하우스 TS2000에서 제공하는 139개 재무비율을 모두 도산예측변수로 사용하였다.

따라서 한국벤처기업의 도산예측모형을 구축하기 위하여 성장성비율 14개, 수익성비율 49개, 안정성비율 39개, 활동성비율 21개, 생산성비율 16개 총 139개 재무비율을 변수로 선정했다.

위의 재무비율에 대해 살펴보면, 성장성비율은 기업의 규모나 경영성과 등과 관련해서 전년대비, 동기대비, 추세대비 등을 비교하여 얼마만큼 성장 또는 증가하였는가, 아니면 마이너스 성장 또는 감소하였는가를 분석하는 것이다. 수익성비율은 투자자본에 대한 경영성과의 정도와 비용을 보전하고 이익을 낼 수 있는 능력에 관한 정보를 제공해 준다. 따라서 투자자, 경영자, 채권자 등 이해관계자들이 의사결정을 할 때에 가장 중요한 정보로 활용하고 있다. 안정성비율은 경기변동에 대한 기업의 장기적인 대응능력을 측정하고자 하는 비율이다. 일반적으로 고정자산에 많이 투자되어 있을수록 자본의 고정화가 많이 이루어진 것으로 볼 수 있으며, 따라서 경기변동에 대한 대응 능력이 적다고 할 수 있다. 활동성비율은 기업의 자원을 얼마만큼 효율적으로 사용하는지를 측정하는 척도가 되며 주로 매출액을 사용하여 비율을 구한다. 기업의 성장성을 나타내는 비율로 일반적으로 매출액이나 총자산의 증가율로써 기업의 성장성을 측정한다. 생산성비율은 기업이 경영활동을 하기 위하여 투자되는

여러 가지 생산요소가 얼마나 효율적으로 활용되었는가를 측정하는 비율로 생산량을 특정 생산요소투입량으로 나누어 구한다.

3. 실증분석 방법

분석대상 표본을 대상으로 도산기업과 정상기업을 잘 구분해 주는 재무비율변수를 알아보기 위하여 선정된 139개 변수에 대한 t-검정을 실시하고, 검정 결과 집단간 유의한 차이가 있는 것으로 나타난 변수를 선정한다.

선정된 변수를 이용하여 다변량판별분석 및 인공신경망의 도산예측모형을 도출한다.

다변량판별분석을 통하여 분석시 t-검정을 통해 나온 변수를 판별력이 있는 변수만 사용하여 판별함수를 결정하는 단계적 선택법과 모든 독립변수를 개별적인 판별력에 관계없이 동시에 고려하여 판별함수를 결정하는 동시적 방법의 2가지를 사용하여 실시했다. 또한, 한정된 분석기업수를 고려하여 표본수가 적을 때 모든 표본을 분석에 사용함으로써 예측의 정확성을 제고시키는 상호교차검증(Jackknife Method)을 사용하였다.

상호교차검증은 데이터를 임의로 상호 배타적인 K개의 집합으로 분할한다. 그리고 K-1개의 집합을 추정용 표본으로 사용하고, 나머지 1개의 집합을 검증용 표본으로 사용하는 방법으로 이 때 검증용 표본에 의해 계산된 오차는 불편추정량이 된다. 이러한 과정이 K번 반복되고, 검증용 표본의 오류는 K번의 검증용 표본에 대한 오차의 평균으로 계산된다. 이 방법을 사용하면 많은 표본을 학습에 사용할

수 있는 동시에 모든 표본을 검증 대상으로 삼을 수 있는 장점이 있을 뿐만 아니라, 각 모형의 일반화 오류율을 보다 정확히 측정할 수 있게 된다.

인공신경망을 이용한 도산예측모형 설계는 입력변수로 t-검정 결과 유의하게 나타난 변수를 모두 이용하는 방법과 판별분석의 단계적 선택법의 결과 선택된 변수를 이용하는 방법을 사용하였다. 또한, 판별분석의 Jackknife Method의 예측율을 비교하기 위해 상호교차검증 방법을 이용하여 예측율 분석을 실시하였다.

신경망모형에 사용되는 데이터는 원천자료(raw data)가 아니고 정규화한(normalized) 값이다. 즉, 데이터를 신경망 모형에 입력하기 위해선 사전처리(preprocessing)과정을 필요로 한다. 데이터를 정규화하는 이유는 신경망 모형이 자료에 포함된 관련 특성에 주목하고 비관련 특성은 제외시키게 하기 위해서이다. 특히, 신경망모형은 원천자료 그대로와 같은 넓은 범위의 자료를 잘 처리하지 못한다. 신경망 모형에서는 원천자료를 정규화하여 주로 0에서 1사이, -1에서 1사이 또는 0.1에서 0.9사이의 범위값으로 스케일링(scaling)한다. 본 연구에서는 [-1, 1]사이의 범위값으로 정규화하였다. 이를 위해서 본 연구에서는 신경망프로그램인 NeuralSolution Version 4.18의 기능 중 'Normalized data file'을 이용하여 추정용표본을 정규화하였다.

V. 실증분석

1. T-검정 결과

139개의 재무비율 중 도산기업과 정상기업의 차이를 잘 나타내주는 재무비율을 선정하기 위하여 t-검정을 실시하였다. 그 결과 수익성비율 8개, 안정성비율 10개, 활동성비율 1개 등, 총 19개 변수가 유의한 결과를 나타냈다. 도산기업표본과 정상기업표본의 평균값과 두 집단간의 평균의 차이를 검정하는 t값은 <표 3>에서 보는 바와 같이 나타났다.

<표 3> t-검정 결과 유의한 변수들

변 수		내 용	
수익성 비율	X15 (매출액총이익율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	13.414 26.214 -2.78(0.0075)*
	X21 (총자본경상이익률)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	0.507 11.727 -3.01(0.0037)*
	X22 (총자본순이익률)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	-0.618 8.641 -2.93(0.0047)*
	X27 (자본금경상이익률)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	11.048 114.86 -3.37(0.0012)*
	X28 (자본금순이익률)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	-2.156 87.605 -3.43(0.0011)*
	X35 (매출원가 대 매출액비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	87.67 73.786 3.03(0.0038)*
	X44 (금융비용 대 총비용비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	4.779 0.728 4.18(0.0001)*

(계 속)

변 수		내 용	
안정성 비율	X48 (이자부담률)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	5.394 0.728 4.16(0.0001)*
	X69 (자기자본구성비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	27.182 58.552 -5.32(0.0001)*
	X70 (순자기자본비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	27.182 58.552 -5.32(0.0001)*
	X71 (타인자본구성비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	72.818 41.448 5.32(0.0001)*
	X74 (고정장기적합율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	98.725 56.767 4.04(0.0002)*
	X75 (유동비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	138.842 290.04 -3.24(0.0023)*
	X76 (당좌비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	100.75 237.72 -2.98(0.0047)*
	X81 (매출채권 대 매입채무비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	152.28 258.63 -2.94(0.0056)*
	X89 (순운전자본비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	6.499 32.198 -6.03(0.0001)*
	X90 (차입금의존도)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	44.062 20.301 -4.38(0.0001)*
	X94 (유보액대비율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	16.51 42.217 4.38(0.0001)*
활동성 비율	X105 (자기자본회전율)	도산기업표본 평균 정상기업표본 평균 t 값(P-value)	5.412 1.814 4.58(0.0001)*

*(P < 0.01)

t-검정결과 유의하게 나타난 19개의 변수들을 정의하면 <표 4>에 요약된 바와 같다. 이들 변수 중 매출원가 대 매출액비율, 이자부담률, 순자기자본비율, 타인자본구성비율, 유보액대비율 등 5개의 비율은 기존의 일반상장기업의 도산예측 연구에서는 언급이 되지 않았던 변수로 국내벤처기업의 도산과 관련된 고유한 변수라고 할 수 있겠다. 이들 변수들은 일반상장기업에 비해 벤처기업이 대부분의 자본을 타인자본에 의존하는 특성을 잘 나타내 주는 변수들이다. 매출액에 비해 비용항목인 매출원가가 높고, 타인자본에 대한 이자비용의 부담이 높고, 총자본에서 타인자본의 비율이 높으며, 유보액이 낮은 벤처기업은 도산할 확률이 높다고 할 수 있겠다.

<표 4> 유의한 도산예측변수의 정의

구 분	재 무 변 수
수익성 비율	X15: 매출액총이익률 $\{(매출액 - 매출원가) / 매출액\} \times 100$ X21: 총자본경상이익률 $(경상이익 / 총자본) \times 100$ X22: 총자본순이익률 $(당기순이익 / 총자본) \times 100$ X27: 자본금경상이익률 $(경상이익 / 자본금) \times 100$ X28: 자본금순이익률 $(당기순이익 / 자본금) \times 100$ X35: 매출원가 대 매출액비율 $(매출원가 / 매출액) \times 100$ X44: 금융비용 대 총비용비율 $(금융비용 / 총비용) \times 100$ X48: 이자부담률 $(이자비용 / 매출액) \times 100$
안정성 비율	X69: 자기자본구성비율 $(자기자본 / 총자본) \times 100$ X70: 순자기자본비율 $\{(자본금+자본준비금) / 총자본\} \times 100$ X71: 타인자본구성비율 $(타인자본 / 총자본) \times 100$ X74: 고정장기적합률 $\{고정자산 / (자기자본 + 고정부채)\} \times 100$ X75: 유동비율 $(유동자산 / 유동부채) \times 100$ X76: 당좌비율 $(당좌자산 / 유동부채) \times 100$ X81: 매출채권 대 매입채무비율 $(매출채권 / 매입채무) \times 100$ X89: 순운전자본비율 $(순운전자본 / 총자산) \times 100$ X90: 차입금의존도 $(차입금 / 총자산) \times 100$ X94: 유보액대비율 $\{(자기자본-자본금+자기주식) / 자본금\} \times 100$
활동성 비율	X105: 자기자본회전율 $(매출액 / 자기자본)$

2. 다변량판별분석

본 연구에서는 판별분석의 판별함수 도출방법으로 동시적 방법과 단계적 선택법을 모두를 사용하였다. 또한, 추정용표본 34개와 검증용표본 34개로 표본을 나누어 판별분석모형을 도출하고 예측율을 분석했으며, 표본수의 제한을 극복하는 방법으로 모든 표본을 분석에 이용함으로써 분석의 정확성을 제고하는 Jackknife method를 사용하여 예측율을 분석하였다.

2.1 동시적 방법

동시적 방법은 판별력이 큰 변수들이 차례로 도출되어지는 중간 결과에는 관심이 없고 가장 판별을 잘해 주는 변수에만 근거하여 독립변수 모두를 동시에 고려하여 판별함수를 구하는 것으로 독립변수들 각각의 판별력은 무시하고 독립변수 전체를 하나의 집합으로 보고 계산하는 방법이다. 판별분석에 사용한 변수는 t-검정결과 유의한 결과를 낸 19개변수 모두를 사용했다. 표본을 동비율의 추정용 표본과 검증용 표본으로 나누어 분석한 결과는 도산기업을 29.41%, 정상기업을 88.24%으로 예측하여 예측정확성(hit ratio)은 58.83%로 나타나 낮은 예측율을 보였다. (<표 5> 참조)

<표 5> 판별분석의 동시적 방법 결과

입력변수	표본구성		예 측 율		예측정확성
			도산	정상	
동시적방법	50:50	도산	5/17 (29.41%)	12/17* (70.59%)	58.83%
		정상	2/17** (11.76%)	15/17 (88.24%)	
	Jackknife	도산	22/34 (64.71%)	12/34* (35.29%)	66.18%
		정상	11/34** (32.35%)	23/34 (67.65%)	

(*: 1종오류, **: 2종오류)

또한, Jackknife method를 사용하여 예측율을 분석한 결과는 도산 기업을 64.71%, 정상기업을 67.65%로 예측하여 예측정확성(hit ratio)은 66.18%로 동비율의 추정용표본과 검증용표본으로 나누어 분석한 방법보다 더 나은 예측율을 나타냈다. (<표 5> 참조)

2.2 단계적선택법

단계적 방법은 각 독립변수의 크기에 의해서 한 번에 하나의 독립변수를 판별함수에 넣는 방법으로 탐색적으로 변수의 우선 순위를 결정하거나 집단을 분류하는 판별모형을 가설적으로 만들고자 할 때 매우 유용하다. 분류모형에 포함된 독립변수가 많으면 그 중에는 집단을 구분하는데 중요한 독립변수가 있는가 하면 그렇지 못한 변수가 포함될 수 있다. 이러한 경우에는 분류능력이 유의하지 않은 독립변수들을 제외시키는 것이 바람직할 것이다. 또한 독립변수의 수가 많아지면 변수들 사이에 다중공선성의 문제가 발생되기

때문에 개별 독립변수의 영향을 평가하는데 문제가 발생되게 된다. 이러한 점들을 방지하기 위해 대부분의 기존 연구에서도 이 방법을 사용하였다.

t-검정결과 유의하게 나타난 19개의 변수를 stepwise method를 사용하여 판별분석에 유의한 변수를 선별한 결과 수익성비율 중 X44(금융비용 대 총비용비율), 안정성비율 중 X89(순운전자본비율), X90(차입금의존도), 활동성비율 중 X105(자기자본회전률) 등, 총 4개의 변수가 유의하게 나타났다. 변수 선정기준으로 보통 Wilks' Lambda를 사용하는데 각각 0.6446, 0.5729, 0.5881, 0.5382로 나타났다. 4개의 변수를 통해 판별분석을 수행한 결과 (식 6)과 같은 판별함수가 도출되었다.

$$Z = 0.13358X_1 - 0.06346X_2 + 0.04371X_3 + 0.13945X_4 \quad (\text{식 6})$$

(X₁: 금융비용 대 총비용비율, X₂: 순운전자본비율, X₃: 차입금의존도, X₄: 자기자본회전율)

판별점 Z_c=1.28723으로 나와 Z≥1.28723이면 정상기업, Z<1.28723이면 도산기업으로 분류가 되었다. 추정용표본으로 판별함수를 도출을 하고 이 함수의 예측율을 검증하기 위해 검증용표본을 사용한 결과 예측율은 도산기업 70.59%, 정상기업 82.35%로 예측정확성(hit ratio)는 76.47%로 나타났다. 도산기업을 정상기업으로 분류하는 I종오류는 29.41%, 정상기업을 도산기업으로 분류하는 II종오류는 17.65%로 나타났다. (<표 6> 참조)

표본을 추정용표본, 검증용표본으로 나누지 않고 표본수의 제한을

극복하기 위해서 모든 표본을 분석에 이용해 정확성을 제고하는 Jackknife method를 사용하여 예측율을 측정한 결과는 도산기업 64%, 정상기업 88%로 예측정확성(hit ratio)은 76%로 나타나 동비율의 추정용, 검증용표본을 사용하는 방법과 예측율이 비슷했다.

(<표 6> 참조)

<표 6> 판별분석의 단계적 선택법 결과

입력변수	표본구성		예측율		예측정확성
			도산	정상	
단계적선택법	50:50	도산	12/17 (70.59%)	5/17* (29.41%)	76.47%
		정상	3/17** (17.65%)	14/17 (82.35%)	
	Jackknife	도산	22/34 (64.00%)	12/34* (36.00%)	77.94%
		정상	4/34** (12.00%)	30/34 (88.00%)	

(*: 1종오류, **: 2종오류)

3. 인공신경망

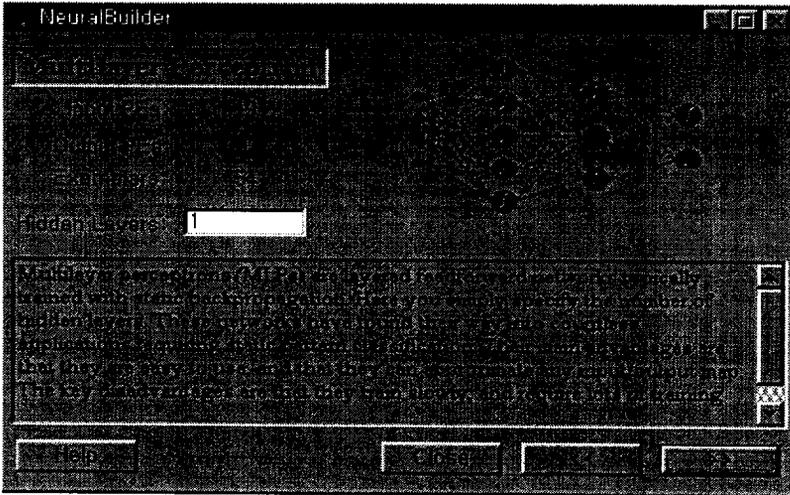
3.1 신경망 아키텍처의 선정

인공신경망 분석은 다층퍼셉트론(Multi-Layered Perceptron)모형을 선정하여 다층퍼셉트론에 주로 사용되는 학습알고리즘으로 역전파알고리즘(back-propagation algorithm)을 사용하였다.

다층퍼셉트론을 활용하고자 할 때에는, 입력층과 출력층은 물론

은닉층의 수, 은닉층에서의 계산단위의 수, 그리고 학습계수와 모멘텀(momentum)항의 계수 등과 같은 신경망 형태를 결정하여야 한다. 매우 복잡한 사상관계가 아니면 대부분의 경우에 있어 은닉층 하나면 충분하기 때문에, 본 연구에서는 [그림 4]와 같이 은닉층의 수가 하나인 모형으로 한정하였다. 다층퍼셉트론에서 은닉층의 노드 수를 결정하는 일은 어려운 일이며, 최적의 은닉 노드의 수는 학습 데이터의 수, 노이즈의 정도, 사상관계의 복잡성 등에 의해 결정된다. 이는 일반화 능력을 제고하는 방법마다 서로 다르다.

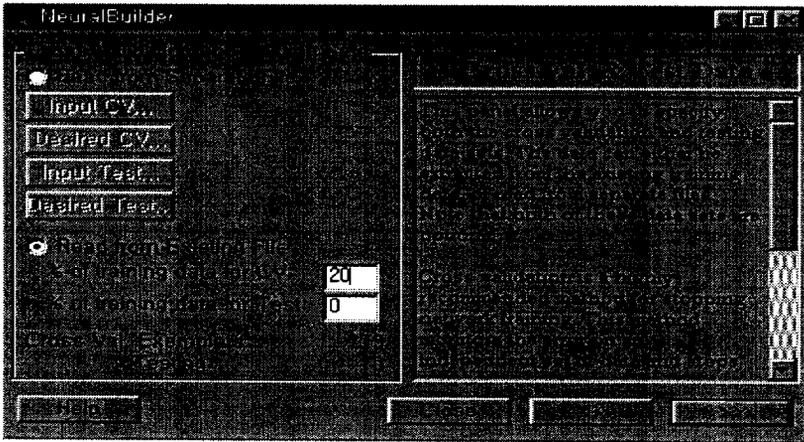
[그림 4] 네트워크의 은닉층의 수 설정



다층퍼셉트론의 경우 일반적으로 학습회수가 증가하면, 학습 데이터에 대한 오류율은 지속적으로 감소한다. 그러나 학습회수가 지나치게 많아지게 되면 예측능력은 오히려 떨어지게 된다. 이것은 과도적합(overfitting)¹³⁾ 문제 때문이다. 그러므로 언제까지 학습을 시켜

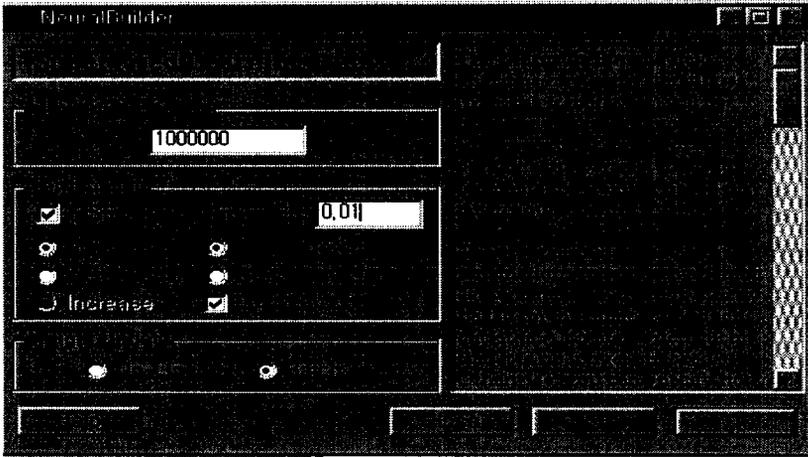
야 하는 지를 결정하는데 과도적합을 막기 위해서 조기정지법을 사용하였으며, 총 추정 표본의 20%를 시험 데이터로 사용하였다. ([그림 5] 참조) 조기정지법을 사용할 때 지역최소점을 피하기 위해서 여러 가지 노드의 수를 경험적으로 사용하였다. 본 연구에서는 최적의 노드 수를 결정하기 위해 여러 구조의 신경망을 실험한 뒤 그 결과를 통해 은닉층의 노드를 결정하였다. 학습횟수는 네트워크가 학습을 조기정지법에 의해 멈출 때까지 무한으로 설정하였다. ([그림 6] 참조)

[그림 5] 조기정지법에 사용할 시험표본(Cross-Validation Dataset)의 설정



- 13) 다층퍼셉트론의 오차를 최대한 낮추는 방향으로 학습을 지속했을 때, 데이터의 패턴을 학습하는 것이 아니라 데이터 자체를 암기함으로써 일반화 능력이 좋아지지 않고, 오히려 떨어지는 현상을 말한다.

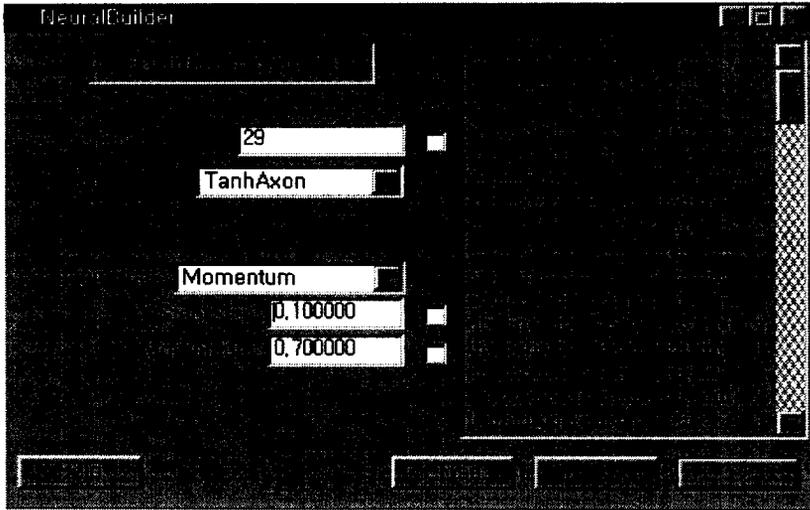
[그림 6] 학습회수 설정 및 조기정지법 설정



또한, 데이터의 표준화 부분으로 기존연구에서는 다층퍼셉트론에서 은닉층과 출력층에서는 입력값의 총합을 0과 1사이의 값으로 변환시키는 전이함수로 로지스틱 시그모이드(sigmoid) 함수를 주로 사용했으나, 본 연구에서는 [그림 7]에서 보는 것과 같이 NeuralSolutions 프로그램에서 권장하는 default 전이함수인 입력값의 총합을 -1과 1사이의 값으로 변환시키는 하이퍼볼릭탄젠트 시그모이드(Hyperbolic Tangent)함수, 일명 Tahn함수를 사용하였다. 양의 값들 및 음의 값들 모두를 산출하는 Tahn함수는 logistic함수와 같이 오직 양의 값들만을 산출하는 함수들보다 더 빠른 훈련을 산출하는 경향이 있으며, 이는 더 좋은 수치적 조건(numerical conditioning)때문이다.¹⁴⁾ 학습률은 0.1, 모멘텀 계수는 0.7로 NeuralSolutions 프로그램에서 권장하는 default값을 사용하였다.

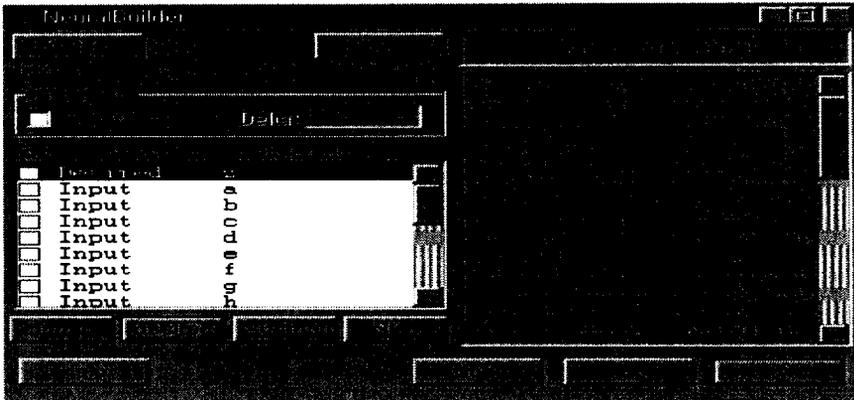
14) Jordan. M. I(1995), "Why the logistic function? A tutorial discussion on probabilities and neural networks", MIT Computational Cognitive Science Report 9503.

[그림 7] 은닉층 노드수 설정 및 전이함수 설정



입력변수의 선정에 있어서 기존 연구에서는 몇 가지 방법을 이용하고 있다. 일반적으로 통계적 방법론을 이용하여 신경망 입력변수를 결정하는 방법(이건창 1993; Han 등 1996, Shin & Han 1996), t-검정 등을 이용하여 입력변수를 결정하는 방법(Jo 등 1997) 등이 이용되고 있다. 본 연구에서는 판별분석의 Stepwise방법에 의해 도출한 변수를 입력변수로 이용하는 방법과 판별식의 동시적 방법, 즉 t-검정에서 유의한 결과를 나타낸 변수 19개를 모두 입력변수로 이용하는 방법 모두를 사용하였다. 또한, 판별분석의 제한된 표본수 극복방법인 Jackknife method에 의한 표본, 즉 67개의 추정용 표본으로 훈련시키고, 나머지 1개의 검증용 표본을 가지고 같은 방법으로 68번 신경망의 예측율을 검증하는 방법인 상호교차검증 방법을 사용하였다. ([그림 8] 참조)

[그림 8] 인공신경망 추정용 표본 입력



3.2 t-검정의 유의한 변수를 이용한 예측

t-검정을 통해 유의한 결과를 나타낸 19개의 입력변수를 이용하여 편의(bias)를 없애기 위해 임의로 추출한 추정용표본으로 네트워크를 훈련시키고 추정용표본과 동비율의 검증용표본으로 훈련된 네트워크의 예측력을 시험하였다. 은닉층은 1개로 했으며, 은닉층의 노드는 많은 실험을 통해 가장 예측율이 높게 나온 27개로 설정했으며 나머지 중요 설정은 프로그램에서 권장하는 default값으로 신경망의 아키텍처를 설정하였다. 신경망의 구조는 19 - 27 - 1의 형태로 형성되었다. 이 구조로 학습을 마친 네트워크의 예측율을 검증용표본을 통해 검증한 결과 도산기업은 82.35%, 정상기업은 88.24%로 예측정확성은 85.29%를 나타냈다. (<표 7> 참조)

<표 7> 인공지능경망의 t-검정 변수 사용 결과

입력변수	표본구성		예 측 율		예측정확성
			도산	정상	
t-검정 변수	50:50	도산	14/17 (82.35%)	3/17* (17.65%)	85.29%
		정상	2/17** (11.76%)	15/17 (88.24%)	
	Jackknife	도산	27/34 (79.41%)	7/34* (20.59%)	82.35%
		정상	5/34** (14.71%)	29/34 (85.29%)	

(*: 1종오류, **: 2종오류)

또한, 판별분석의 Jackknife method, 즉 상호교차검증 방법을 사용하여 추정용 표본 67개와 검증용 표본 1개를 가진 네트워크로 68번 예측을 실시하였다. 은닉층의 노드수는 많은 실험을 통해 예측율이 가장 높게 나온 최적의 노드수로 29가 결정되었다. 네트워크의 구조는 19 - 29 - 1 의 형태로 형성되었다. 이 구조로 68번 각각 다른 추정용 표본으로 학습을 한 후 검증용 표본으로 예측율을 검증한 결과 도산기업은 79.41%, 정상기업은 85.29%로 예측정확성은 82.35%로 나타났으며, 동비율의 추정용 표본과 검증용 표본을 사용한 방법보다 3%정도 낮게 나타났다. (<표 7> 참조)

3.3 선택변수를 이용한 예측

판별분석의 단계적 선택법(stepwise selection method)을 이용해 판별력이 있다고 나타난 4가지 변수를 이용하여 동비율의 추정용

표본과 검증용 표본을 사용하여 예측율을 측정하였다. 은닉층의 노드수는 경험을 통해서 2개가 최적의 은닉노드수로 결정됐다. 네트워크의 구조는 4 - 2 - 1의 형태를 갖게 되었다. 이 네트워크를 이용하여 추정용 표본으로 훈련시키고, 검증용 표본으로 예측율을 검증한 결과 도산기업은 76.47%, 정상기업은 82.35%로 예측정확성은 79.47%로 나타났다.

또한, 상호교차검증 방법으로 네트워크를 설계한 결과 최적의 은닉층의 노드수는 3으로 결정되어 사용할 네트워크의 구조는 4 - 3 - 1의 형태를 갖게 되었다. 이 구조로 68번 각각 다른 추정용표본으로 학습을 한 후 검증용표본으로 예측율을 검증한 결과 도산기업은 73.53%, 정상기업은 82.35%로 예측정확성은 77.94%로 나타났으며, t-검정의 유의한 변수를 이용한 예측과 마찬가지로 동비율의 추정용 표본과 검증용표본을 사용한 방법보다 약간 낮게 나타났다.

<표 8> 인공지능경망의 선택적 변수 사용 결과

입력변수	표본구성		예 측 율		예측정확성
			도산	정상	
선택적변수	50:50	도산	13/17 (76.47%)	4/17* (23.53%)	79.47%
		정상	3/17** (17.65%)	14/17 (82.35%)	
	Jackknife	도산	25/34 (73.53%)	9/34* (26.47%)	77.94%
		정상	6/34** (17.65%)	28/34 (82.35%)	

(*: 1종오류, **: 2종오류)

4. 종합 비교

지금까지 실험의 모든 결과는 <표 9>에 요약되어 있다.

<표 9> 전체결과

사용변수	도산예측모형	표본구성	예측정확성
t-검정 변수 (19개)	다변량판별분석	50:50	58.83%
		Jacknife	66.18%
	인공신경망	50:50	15.70%
		Jacknife	29.47%
선택적변수 (4개)	다변량판별분석	50:50	76.47%
		Jacknife	76.00%
	인공신경망	50:50	3%
		Jacknife	1.94%

좋은 결과를 낸 모형은 굵게 표시하였다. t-검정변수를 이용한 도산예측 실험에서는 표본구성비 50:50과 Jacknife method 모두 인공신경망이 각각 29.47%와 15.70%의 예측율이 다변량판별분석보다 좋게 나왔다. 또한 stepwise결과 도출된 선택적변수를 이용한 도산예측 실험에서도 인공신경망이 3%와 1.94%의 약간의 차이로 다변량판별분석보다 예측율이 좋게 나왔다. 전체적으로 본 연구의 모든 실험에서 인공신경망의 도산예측모형이 다변량판별분석의 도산예측모형보다 더 나은 예측정확성을 나타냈다.

다변량판별분석은 판별력이 있는 변수만을 판별함수도출에 이용했을 때의 판별분석모형이 변수들의 판별력에 관계없이 모든 변수를 사용한 판별분석모형보다 좋은 결과를 나타냈다.

인공신경망은 자체의 특성상 적은 수의 입력변수인 선택적변수를 사용했을 때보다 많은 수의 변수인 t-검정의 유의한 변수를 사용했을 때가 더 나은 예측율의 결과를 나타냈다.

VI. 결 론

본 연구는 한국벤처기업의 도산예측에 있어서 인공신경망모형과 다변량판별분석모형 간의 예측능력을 비교한 것이다. 이를 위해 1999년부터 2000년까지 도산한 34개 기업과 이에 대응하여 유사한 환경에 있는 정상기업 34개를 선정하여 139개의 재무비율로 분석을 수행하였다. 139개의 많은 재무비율을 t-검정을 실시하여 도산/정상기업의 유의한 평균 차이를 나타내는 19개의 변수를 도출하였다. 분석에서 사용한 변수는 선택적변수와 t-검정의 유의한 변수를 각각 사용해서 모형을 구축하였으며, 추정표본과 검증표본의 구성을 50:50과 예측의 정확성을 위해 Jackknife method를 사용하였다. 또한, 신경망의 과도적합(overfitting)의 문제를 해결하기위해 조기정지법을 사용하였다. 연구 결과를 요약하면 다음과 같다.

첫째, 실증분석의 결과는 실험한 전 모형에서 인공신경망이 다변량판별분석보다 더 좋은 예측력을 나타냈다. 특히, 인공신경망은 입력변수가 많았을 때가 입력변수가 적었을 때보다 다변량판별분석보다 예측율의 차가 더 크게 나타났다..

둘째, 본 연구를 통해 기존의 대기업 및 일반상장기업의 도산예측 모형변수에 비해 한국벤처기업 고유의 도산예측변수가 채택될 수 있음을 알 수 있었다. 매출원가대매출액비율, 이자부담률, 순자기자본구성비율, 타인자본구성비율, 유보액대비율 등 5개의 재무비율이 기존의 일반상장기업의 부실예측에 관한 연구에서는 선택되지 않았으나 본 연구에서는 도산예측변수로 선택되었다는 데 의의가 있다.

본 연구의 한계점은 다음과 같다.

첫째, 연구표본 수의 한계를 안고 있다. 검증용표본의 수가 적을 경우 몇 개 표본의 특이성이 검증용표본의 예측율을 상당히 크게 변화시킬 수 있으므로 제시한 검증용표본의 예측율이 향후 모형을 적용하여 얻게 될 실제 예측력과는 상당한 차이를 보일 위험이 존재하므로 더 많은 표본수의 확충이 필요하다.

둘째, 판별분석 모형 중 동시적방법 사용 시 예측율에 많은 영향을 줄 수 있는 변수들간의 다중공선성 문제가 발생할 수 있으므로 이에 대한 대응책 마련이 필요할 것이다.

셋째, 신경망의 학습알고리즘(learning algorithms)의 한계이다. 본 연구에서 적용된 백프로퍼게이션(back-propagation)은 본질적으로 전역최소값(global minimum)을 보장하지는 못한다는 것이다. 학습 과정을 통해 구해진 가중치는 물론 최상의 가중치이지만 지역최소값(local minimum)에 수렴할 가능성을 배제할 수 없다는 것이다. 따라서, 전역 최소값을 보장할 수 있는 다른 학습 알고리즘의 도입이 필요하다.

넷째, 기업의 재무변수 외에 벤처기업의 특성을 잘 표현하는 비재무적인 질적 정보를 포함하여 재무정보와 비재무정보 간의 가중치도 함께 고려한 통합적인 연구가 필요할 것이다.

기업의 도산예측모형이 아무리 타당하고 객관성을 인정할 수 있는 모형이라 하더라도 미래를 정확히 예측한다는 것은 불가능할 것이다. 따라서 기업의 도산예측에 있어 계량적인 평가의 한계를 극복하기 위하여 주관적·전문적 평가방법을 병용하여야 할 것이다.

<참 고 문 헌>

1. 국내문헌

- 강종만 · 홍성희(1999), “부실예측모형의 적합성 분석”, 증권·금융연구, 제5권 제1호, pp.83-110.
- 김정호 · 김완표(1999), “벤처 붐의 조성과 NASDAQ 진출”, 삼성경제연구소 연구보고서.
- 김충련(1997), SAS라는 통계상자, 데이터플러스.
- 박순식 · 김병주(2000), “우리 나라 중소기업의 도산예측에 관한 연구”, Journal of Business Research, Vol. 15, No. 1, pp.27-53.
- 박용구 · 강신겸(1997), “벤처기업 육성을 위한 입지지원방안”, 삼성경제연구소 연구보고서.
- 이건창 · 김명종 · 김혁(1994), “기업도산예측을 위한 귀납적 학습지원 인공지능망 접근방법: MDA, 귀납적 학습방법, 인공지능망 모형과의 성과비교”, 경영학연구 제23권 제3호, pp.109-144.
- 임영규(1995), “기업부실화예측에 관한 실증적 연구”, 성균관대학교 대학원, 박사학위논문.
- 정현웅(1998), “기업도산예측에 관한 실증적 연구”, 재무관리논총, 제4권 제1호, pp.123-149.
- 최신재(1995), “기업도산예측을 위한 인공지능망과 귀납적 학습방법의 통합적 적용”, 성균관대학교 대학원, 박사학위논문.

- 한길석 · 이상로(1999), “인공신경망을 이용한 도산예측에 관한 연구”, 회계정보연구 제12권, pp.97-115.
- 한인구 · 조홍규(1996), “신경망을 이용한 도산예측”, 과학기술원, 춘계학술연구발표회 발표논문집, Vol.4, pp.135-146.
- Shin, K., Jo. H. and Han I.(1996), “Integration Methodology of Multiple Techniques Using Genetic Algorithms: A Case of Corporate Failure Prediction”, 경영과학 '96 추계학술대회.

2. 외국문헌

- Altman, Edward I(1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", The Journal of Finance, Vol. 23, Sep, 1968, No. 4, pp.589-609.
- _____, Robert G. Haldeman, P. Narayanan(1977), "ZETA ANALYSIS : A new model to identify bankruptcy risk of corporations", Journal of Banking and Finance, pp.29-54.
- _____, Marco, G. and Varetto, F(1994), "Corporate Distress Diagnosis Comparisons Using Linear Discriminant Analysis and Neural Networks", Journal of Banking and Finance, 18(3), pp.505-529.
- Beaver, W. H(1966), "Financial Ratios and Predications of Failure", Empirical Research in Accounting: Selected Studies, Supplement to Volume 5, Journal of Accounting Research, pp.71-111.
- Berry, R. and D. Treigueiros.(1991), "The Application of Neural Network Based Methods to the Extraction of Knowledge from Accounting Reports", IEEE International Joint Conference on Neural Networks, pp.136-146.
- Jordan. M. I(1995), "Why the Logistic Function? A tutorial discussion on probabilities and neural networks", MIT Computational Cognitive Science Report 9503.
- Liang, T. P, J. S. Chandler, Ingoo Han and J. Roan(1991), "An

Empirical Investigation of Some Data Effects on the Classification Accuracy of Profit, ID3 and Neural Networks", Working Paper, University of Illinois at Urbana-Champaign.

Lippmann, R(1987), "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, April, 1987, pp.4-22.

Nelson, M. C. and Illingworth, W. T(1991), A Practical Guide to Neural Nets, Addison-Wesley, Reading, MA.

Odom, M. D. and Sharda, R(1990), "A Neural Networks by Soft Weight-sharing", Neural Computation, 4(4), pp.473-493.

Tam, K and M. Kiang(1992), "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", Management Science, July, pp.926-947.

Vonk, E, L. C. Jain, R. P. Johnson(1997), Automatic Generation of Neural Network Architecture Using Evolutionary Computation, World Scientific Publishing Co. Pte. Ltd. p.25, p.58.

Walczak and Steven(1999), "Gaining Competitive Advantage for Trading in Emerging Capital Markets with Neural Networks", Journal of Management Information Systems, Fall, 1999, Vol.16 Issue 2, pp.178-194.

Wilson, Nicholas; Chong, Kwee S(1995), "Neural Network Simulation and the Prediction of Corporate Outcomes: Some Empirical Findings", International Journal of the Economics of Business, Vol.2, Issue 1, pp.31-51.

Woelfel and Joseph(1994), "Think Neural Networks", Marketing Tools, Apr/May, 1994, Issue N, pp.62-65.

Zmijeski, M. E(1994), "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", Studies on the Current Econometrics Issues in the Accounting Research, Journal of Accounting Research, pp.59-82.

A Study on the Prediction of Venture Firms' Failure Using Artificial Neural Network in Korea

Ji-Ho Cho

*Department of Business Administration, Graduate School,
Pukyong National University*

Abstract

The objectives of this study are ①to compare the prediction accuracy of neural networks model and multiple discriminant analysis model on the failure of venture firms in Korea, and ②to find the unique variables determining the prediction of venture firms' failure in Korea.

We selected 34 failed venture firms (1999-2000) and 34 existing venture firms with similar size in the same industry in KOSDAQ market as sample firms. And, we used 139 financial ratios in experiment. To reinforce the prediction accuracy, a stepwise method and simultaneous method were used in selecting variables. A 50:50 method and Jackknife method were also used in composing samples.

The result of the experiment is that neural networks model presented better prediction accuracy than multiple discriminant

analysis model. Also, the unique variables determining the prediction of venture firms' failure in Korea are as follows : cost of sales to net sales, interest & discount expenses to liabilities, stockholder's equity excluding cash dividends to total assets, debt to total assets, and reverses to total assets.

Key words : failure prediction, neural networks, multiple discriminant analysis, venture firm