

재무비율을 이용한 기업부도에측 모형의 예측력 비교 연구

The Comparative Study on Corporate Bankruptcy
Prediction Model Using Financial Ratios

2005 年

西江大學校 大學院
經營學科 財務管理專攻
裴 在 權

재무비율을 이용한 기업부도에측 모형의 예측력 비교 연구

The Comparative Study on Corporate Bankruptcy
Prediction Model Using Financial Ratios

지도교수 홍 광 현

이 논문을 경영학석사 학위논문으로 제출함

2006년 1월

서강대학교 대학원

경영학과 재무관리전공

배 재 권

논 문 인 준 서

배재권의 경영학석사 학위논문을 인준함

2006년 1월

주심 박 영 석 인

부심 홍 광 현 인

부심 김 진 화 인

감사의 글

2년 동안 재무관리를 전공하면서 내놓은 석사논문이 한없이 부족하기만 합니다. 아직 부족함이 많은 저에게 위 석사학위 논문은 박사과정에 있어서 집필해야 할 학회논문의 시발점이라 생각하고 좀더 정진하는 연구자가 되도록 노력하겠습니다.

본 학위논문이 나오기까지는 수많은 지인들의 도움이 있었기 때문에 가능했습니다.

먼저, 투자론에 대해 많은 가르침을 주셨던 지도교수님이신 홍광현 교수님, 기업가치평가 방법론에 대한 많은 가르침을 주셨으며 논문심사를 받게 해주신 박영석 교수님, 논문의 아이디어와 다양한 방법론에 대해 학습과 가르침을 주셨던 MIS 김진화 교수님께 감사의 말씀을 드립니다. 또한 재무관리의 기본적 이론에 대해 가르침을 주셨던 최운열 교수님, 파생상품과 옵션에 대해 가르침 주셨던 이정진 교수님, 그리고 국찬표 교수님께도 다시 한번 감사의 말씀을 올립니다.

또한 Support Vector Machine에 대한 이론과 궁금증에 대해 많은 이메일 학습과 가르침을 주셨던 이영찬 교수님, Bayesian Networks의 알고리즘에 대한 가르침 주셨던 MIS 이승훈 박사님께도 감사드립니다.

재무관리를 전공하면서 2년 동안 친동생처럼 보살펴 주신 우리 동기형님들과 누나들(친동생 같이 챙겨주셨던 민정누나와 기정이형, 배울 점이 너무 많았던 성용이형과 현식이형, 재무학 기본 지식이 부족한 저에게 많은 도움을 주셨던 창훈이형, 파트장 형님이신 재철이형, 많이 이야기를 못해 아쉬웠던 종우형, 절 항상 재미있게 해주셨고 친구처럼 대해주셨던 정순이

형, 항상 좋은 말씀만 많이 해주셨던 희준이형, 항상 학업에 열심히 하시고 대학원생으로서의 모범을 보여주셨던 주연이형과 재학이형, 누나였지만 얼마같이 느껴졌던 상현누나, 웃는 모습이 너무 보기 좋은 수경누나)에게 정말 감사의 말씀을 드립니다.

그리고 80클럽 리더인 소연이와 항상 밝게 생활해서 보기 좋은 영은이, 하나밖에 없었던 남자 동갑내기 지영이, 그리고 막내 귀여운 은서에게도 감사의 말씀을 드립니다. 또한 MIS와 재무학의 연결고리를 발견해주셨던 호진이형, 같이 졸업하지 못해 아쉬운 진영누나, 석사연구실에서 함께 공부하며 격려해주셨던 연승이형과 2학기 후배님들, 3학기 후배님들(항상 좋은 말씀과 조언 많이 해주셨던 순현이 형님과 승찬이 형님)에게도 감사의 말씀을 드립니다.

끝으로 저를 낳아주시고 길러주신 저에게 정말 소중한 아버지와 어머니께 감사의 말씀을 드립니다. 또한 하나밖에 없는 사랑스런 동생 재성이 그리고 서울에서 친아버지, 어머니처럼 저를 보살펴주신 이모부와 이모, 그리고 사랑스런 사촌동생 지영이, 준호와 함께 석사학위를 받는 기쁨을 나누고자 합니다.

마태오관 석사연구실에서

2005년 12월

배 재 권

< 목 차 >

제 1 장 서론	1
제 1 절 연구의 배경 및 목적.....	1
제 2 절 연구의 방법과 범위	4
제 2 장 부도예측모형에 관한 이론적 고찰	6
제 1 절 옵션가격결정모형과 예상도산확률모형... 6	
1. 옵션가격결정모형과 기업의 도산 가능성	
2. 예상도산확률모형을 통한 도산 가능성의 측정	
제 3 장 부도예측모형의 선행연구.....	19
제 1 절 주가정보를 이용한 부도예측모형	19
1. 이은주의 연구	
2. 이원흠, 이한득, 박상수의 연구	
3. 국찬표, 정완호의 연구	
제 2 절 재무비율을 이용한 부도예측모형	22
1. Altman(1968)의 다변량 판별분석	
2. Logit 분석에 의한 부도예측모형	
3. 인공신경망(Artificial Neural Networks)	

- 4. 규칙유도기법(Rule Induction)
- 5. 베이저안 망(Bayesian Networks)
- 6. SVM(Support Vector Machine)

제 4 장 연구 설계	41
제 1 절 자료 및 연구 설계	41
1. 표본기법 및 변수선정	
제 2 절 모형의 추정 및 분석	44
1. 다변량 판별분석	
2. 로지스틱 회귀분석	
3. 인공신경망	
4. 규칙유도기법	
5. 베이저안 망	
6. Support Vector Machine	
7. WP-ANN 통합방법론	
제 5 장 연구 결과	76
제 1 절 요약	76
제 6 장 결론 및 한계	78

제 1 절 연구의 결론.....	78
제 2 절 연구의 한계와 후속 연구과제.....	79
<참고문헌>.....	81

< 표 목 차 >

<표 3-1> 최종판별함수와 도산예측기준	23
<표 3-2> 신경망을 적용한 부도예측관련 연구(국외)	30
<표 3-3> 신경망을 적용한 부도예측관련 연구(국내)	31
<표 4-1> 선정된 재무비율 변수 목록	42
<표 4-2> 다변량 판별분석의 유의성 검증	45
<표 4-3> 함수의 집단중심점	46
<표 4-4> 판별분석의 결과	47
<표 4-5> 로지스틱 회귀모형 변수들의 유의성 검증.....	51
<표 4-6> 로지스틱 회귀분석의 결과	52
<표 4-7> 인공신경망의 모형	54
<표 4-8> 인공신경망의 결과	57
<표 4-9> 규칙유도기법의 결과.....	64
<표 4-10> Belief Network Power Constructor 1.0	66
<표 4-11> 베이지안 망의 결과.....	68
<표 4-12> 가우시안 RBF 함수의 SVM 결과.....	71
<표 4-13> SVM 분석 결과	72
<표 4-14> WP-ANN 통합방법론의 결과.....	75
<표 5-1> 전체결과	76

< 그림 목 차 >

<그림 1-1> 연구 분석의 흐름도	5
<그림 3-1> 다층신경망의 구조	27
<그림 4-1> 인공신경망의 모형	56
<그림 4-2> 규칙유도기법의 모형	59
<그림 4-3> Rule-set 모형	63
<그림 4-4> 베이지안 망 모형의 결과	67
<그림 4-5> WP-ANN 통합방법론	73
<그림 5-1> 모형별 평균예측력 비교	77

ABSTRACT

The Comparative Study on Corporate Bankruptcy Prediction Model Using Financial Ratios

Bae, Jae Kwon

Major in Finance

Dept. of Business Administration

Graduate School of Sogang University

This study proposes an integrated approach to predict corporate bankruptcy more accurately, which combines statistical methods with artificial intelligence techniques.

In order to develop more advanced approaches, this study attempts to combine statistical methods with AI techniques. This study selected two statistical methods - Multiple Discriminant Analysis and Logistic Regression and four AI techniques - Artificial Neural Networks, Rule Induction, Bayesian Networks, and Support Vector Machine. By Combining the above six methods, this study generates new approach, which are Voting with Performance & Weights from ANN.

Financial data about 1,888 corporations listed in Korea Credit Guarantee Fund are used in the analysis.

The purpose of this study is to evaluate the usefulness of integrated approach by applying the financial data and comparing the prediction accuracy.

The experiment results show that the Voting with Performance & Weights from ANN might be a promising approach to estimate the risk of corporate bankruptcy compared to the currently-in-use approaches. Also, this model works better than other models for all five data sets. The ranks are identified as Voting with Performance & Weights from ANN, Support Vector Machine, Logistic Regression, Artificial Neural Networks, Multiple Discriminant Analysis, Rule Induction, and Bayesian Networks.

Therefore, the Voting with Performance & Weights from ANN is recommended as a corporate bankruptcy prediction model in this study.

국문초록

재무비율을 이용한 기업부도예측 모형의 예측력 비교 연구

본 연구에서는 보다 효과적인 기업부도예측을 위하여, 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하였다. 이를 위하여 본 연구에서는 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석(Multiple Discriminant Analysis), 로지스틱 회귀분석(Logistic Regression)과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공신경망(Artificial Neural Networks), 규칙유도기법(Rule Induction), 베이지안 망(Bayesian Networks), Support Vector Machine(SVM) 등의 6가지 단일모형과 6가지 단일모형을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형을 제시하였다. 위 통합모형의 예측성가를 증명하기 위하여 기술신용보증기금이 보유하고 있는 1,888개 기업의 재무비율 자료를 기초로 분석하였고 기존의 단일모형 6가지와 통합모형 1가지, 총 7가지 방법론의 예측력 성과비교를 통하여 그 유용성을 검증해보고자 한다.

실험결과 본 연구에서 제안한 통합모형인 Voting with Performance & Weights from ANN 모형은 다변량 판별분석(Multiple Discriminant Analysis), 로지스틱 회귀분석(Logistic Regression), 인공신경망(Artificial Neural Networks), 규칙유도기법(Rule Induction), 베이지안 망(Bayesian Networks), Support Vector Machine(SVM) 등의 단일모형과 비교한 결과

가장 예측정확성이 우수한 것으로 나타났다. 따라서 본 연구를 통해 부도 예측에 있어서 통합모형인 Voting with Performance & Weights from ANN(WP-ANN) 방법론이 기존의 모형들에 비해 우수한 예측정확성을 나타낼 수 있었다.

제 1 장 서론

제 1 절 연구의 배경 및 목적

기업의 부도(bankruptcy)는 해당기업의 경영자, 종업원, 채권자, 투자자, 거래처, 고객 및 지역경제, 나아가 국가경제에 미치는 파급효과의 정도와 양상이 다양하게 나타난다. 또한 부도과정에서 제반 법적비용의 발생, 매출 감소, 영업활동의 마비 등 직·간접적인 피해가 엄청나게 크기 때문에 기업에 대한 부도예측은 다양한 이해관계자들의 중요한 실무적 관심사가 될 수밖에 없다. 1997년 이후 경기침체로 많은 기업들이 부실화되어 부도사태를 맞이하고 있다. 또한 시장개방의 가속화, 경기침체의 장기화, IMF 경제체제 등 기업을 둘러싼 경영환경이 급격히 변화되어 부도 위험도가 급증하고 있고 기업간 경쟁이 치열해져 우량기업조차 안정성장을 장담하기 어려운 실정이다. 특히 우리나라 국민경제의 근간을 이루고 있는 중소기업의 경영상태는 대기업의 공세적 경영에 밀려 갈수록 악화되고 있으며, 최근 중소기업의 잇따른 부도사태 및 부실화의 정도는 상식의 한계를 벗어나 심각한 수준에 와 있다.

우리나라의 중소기업은 사업체 수, 종업원 수에 있어서 모든 산업에서 막중한 비중을 차지하고 있을 뿐만 아니라 생산액, 부가가치 그리고 특히 수출에 있어서 중요한 위치를 차지하고 있어 국민경제 향상에 기여하는 바가 매우 크다. 과거 20-30년간 우리나라 중소기업은 대기업에 비해 경제적 보호 없이 발전을 해왔으며, IMF 금융위기를 겪으면서 대기업들의 부실과

맞물려 자금난을 겪고 있다.

대부분의 기업 부실화는 갑자기 발생하는 것이 아니라 경영 외적 요인과 경영 내적 요인이 복합적으로 작용하여 점진적으로 진행되는 현상이라고 할 수 있다. 따라서 이들 원인간의 관계를 체계적으로 분석하여 기업의 부도 가능성을 사전에 포착할 수 있다면 관련 금융기관은 물론 해당 기업의 경영자나 투자자, 그리고 채권자 등과 같은 이해관계자 모두가 조기에 기업 경영의 정상화를 위해 노력하거나 또는 정상적이고 합리적인 청산절차를 진행함으로써 사회·경제적으로 손실을 최소화할 수 있을 것이다¹⁾.

경영학, 경제학 및 회계학 분야에서 기업부도예측에 대한 연구가 광범위하게 수행되고 있으며, 과거 60년간 부도예측이 포함된 이론연구와 사례연구가 지속적으로 이루어졌다. 기존의 연구에서는 다변량 판별분석, 로지스틱 회귀분석 등의 전통적인 통계기법을 주로 사용하여 기업부도예측 문제를 해결하고자 하였지만, 통계적 기법이 갖고 있는 엄격한 가정이 만족되어야 한다는 방법론적인 한계를 갖고 있었다. 따라서 최근 인공지능 기법이 제반 경영학 문제에 보다 활발하게 적용되기 시작되면서 인공신경망(Artificial Neural Networks), 규칙유도기법(Rule Induction), 베이저안 망(Bayesian Networks), Support Vector Machine(SVM) 등과 같은 인공지능 관련 기법을 기업부도예측 문제에 적용하는 연구가 활발히 소개되었고 그 예측력 또한 기존의 전통적인 통계기법보다 더 높은 성과를 나타내었다²⁾.

본 연구에서는 보다 효과적인 재무비율을 이용한 기업부도예측을 위하여, 통계적 방법과 인공지능 방법의 예측정확성을 비교 분석하고자 한다.

1) 김진백, 이준석, “인공지능 기법을 이용한 중소기업 부도예측에 있어서 현금흐름변수의 유용성 검증,” 대한경영학회지, 제26권, 2000, pp.229-250.

2) 이건창, 김명중, 김혁, “기업도산예측을 위한 귀납적 학습지원 인공신경망 접근방법 : MDA, 귀납적 학습방법, 인공신경망 모형과의 성과비교,” 한국경영학회, 제23권 제2호, 1994, pp.109-144.

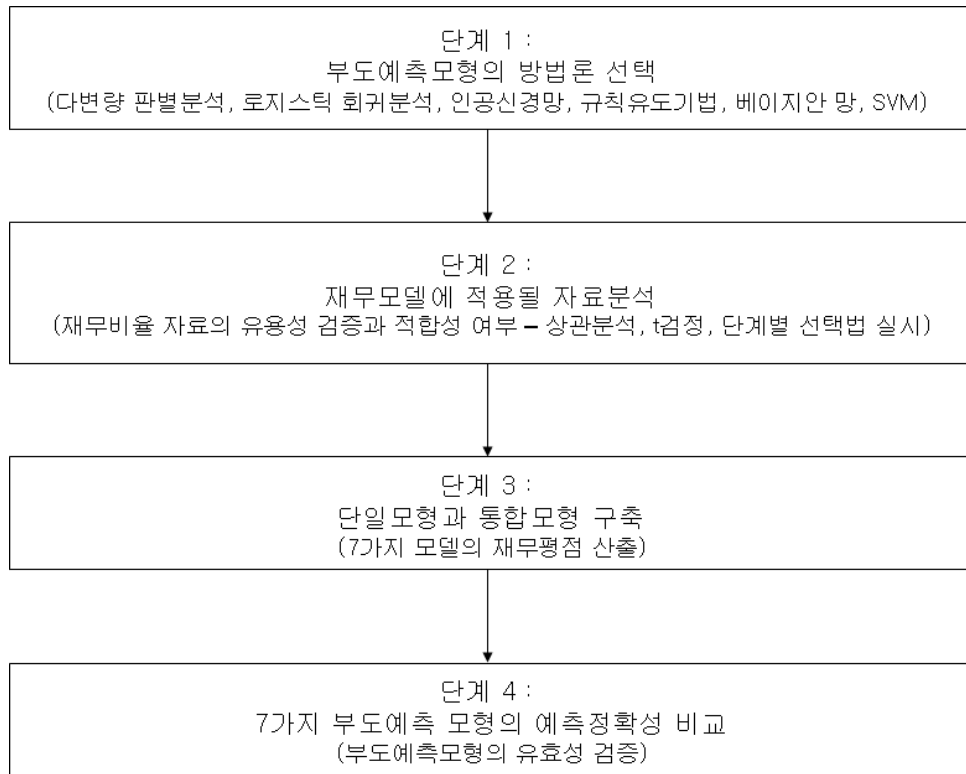
또한 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하여 기존의 모형과 예측정확성을 비교하여 그 유용성을 검증해보고자 하였다. 이를 위하여 본 연구에서는 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석(Multiple Discriminant Analysis), 로지스틱 회귀분석 (Logistic Regression Analysis)과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공신경망(Artificial Neural Networks), 규칙유도기법(Rule Induction), 베이지안 망(Bayesian Networks), SVM(Support Vector Machine) 등 총 6가지의 단일모형과 단일모형을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형의 예측정확성을 비교하여 분석하고자 한다. 위 통합모형의 예측성과를 증명하기 위하여 기술신용보증기금이 보유하고 있는 1,888개 기업의 재무비율 자료를 기초로 분석하였고 기존의 단일모형 6가지와 통합모형 1가지, 총 7가지 방법론의 예측력 성과 비교를 통하여 그 유용성을 검증해보고자 한다.

제 2 절 연구의 방법과 범위

본 연구에서는 보다 효과적인 재무비율을 이용한 기업부도예측을 위하여, 통계적 방법과 인공지능 방법의 예측정확성을 비교 분석하고자 한다. 또한 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하여 기존의 모형과 예측정확성을 비교하여 그 유용성을 검증해보고자 하였다.

제 1 장 서론에서는 연구의 배경 및 목적, 연구의 방법과 범위 등에 관해서 기술하였고, 제 2 장에서는 부도예측모형에 관한 이론적 고찰로서 옵션가격결정모형과 예상도산확률 모형에 대하여 기술하였다. 제 3 장에서는 부도예측모형의 선행연구로서 추가정보를 이용한 부도예측모형과 재무비율을 이용한 부도예측모형으로 분류하여 기술하였다. 제 4 장에서는 연구 설계 단계로 표본기업 및 변수선정 과정과 7가지 부도예측모형을 설계하고 분석하였다. 제 5 장에서는 연구모형에 대한 결과에 대해 해석하고 제 6 장에서는 본 연구의 결과를 요약하고, 연구의 한계점 및 향후 연구방향에 대하여 기술할 것이다.

본 연구에서는 총 7가지의 부도예측모형을 제시하고 있으며 제 4장에서 시행될 모형설계 과정을 <그림 1-1>의 연구 분석 흐름도로 자세히 설명하고자 한다.



<그림 1-1> 연구 분석의 흐름도

연구 분석의 흐름도는 다음과 같이 구성되어 있다. 단계 1에서는 부도예측모형의 방법론을 선택한다. 본 논문에서 사용할 부도예측모형의 방법론은 다변량 판별분석, 로지스틱 회귀분석, 인공신경망, 규칙유도기법, 베이지안 망, SVM 등이다. 단계 2에서는 재무모델에 적용될 자료를 분석하는 단계로 상관분석, t검정, 단계별 선택법 등을 실시하여 재무비율 자료의 유용성 검증과 적합성 여부를 판단한다. 단계 3에서는 6가지 단일모형과 통합모형을 구축하며 각각의 모형에서 재무평점을 산출한다. 단계 4에서는 총 7가지 부도예측 모형의 예측정확성을 상호비교 분석하고 그 유용성을 검증해본다.

제 2 장 부도예측모형에 관한 이론적 고찰

제 1 절 옵션가격결정모형과 예상도산확률모형

1. 옵션가격결정모형과 기업의 도산 가능성

주가정보를 이용한 기업 도산 예측모형들은 주식이 갖는 조건부 청구권적인 성격을 근거하여 기업의 도산 가능성을 파악하고자 하고 있다. 이들 모형은 주가를 다른 모든 정보를 포함하고 있는 유일한 설명변수로 보고 있기 때문에 변수간의 다중공선성 문제가 제기될 수 없으며, 모형의 결과가 현시점의 절대적 주가수준보다는 주가의 시계열 변동 추세 또는 변동성에 더 큰 영향을 받게 된다는 특성을 갖고 있다.

Merton(1974)은 주식은 자산을 기초자산으로, 부채를 행사가격으로 하는 일종의 콜옵션임에 착안하여, 주식시장에서 관찰되는 시가총액을 이용해 자산 또는 부채의 시장가치를 산출할 수 있음을 주장하였다. 옵션가격결정모형을 이용하면 만기시점에 콜옵션이 행사되지 않을 확률을 구할 수 있는데, Merton(1974)에 의하면 이 확률은 만기시점의 자산가치가 부채가치보다 적을 확률을 의미한다. 특정시점의 자산가치가 부채가치보다 적을 경우 자산을 모두 매각한다 하더라도 부채를 모두 상환할 수 없기 때문에 기업은 도산상태에 이르게 될 것이다. 이러한 확률을 통해 실제 도산가능성을 산출하고자 하는 것이 예상도산확률모형의 기본적인 개념이다.

옵션가격결정모형에 의하면 주식의 가치는 식 (1)과 같이 표현될 수 있

다. 이 식을 살펴보면, 자기자본가치는 자산가치에서 부채가치를 차감한 것과 같다는 기본적인 회계원리를 반영하고 있다. 다만, 자산가치 변동성을 반영하고 있고 부채규모를 적정 할인율로 할인한 현재가치로 측정되고 있다는 점이 대차대조표와는 다른 특징이라 할 수 있다. 이러한 특징 때문에 이를 이용해 파악되는 기업의 도산 가능성은 현시점의 절대적 추가수준과 함께 그 변동성이나 변화추세에 의해서도 영향을 받게 된다.

$$V^E = V^A \cdot N(d_1) - e^{-r_T T} \cdot V^D \cdot N(d_2)$$

식 (1)

V^E 는 자기자본의 시장가치 = 시가총액

V^A 는 자산의 시장가치

V^D 는 부채의 장부가치

σ_A 는 자산가치의 변동성

T 는 부채의 만기

r_T 는 만기가 T인 무위험이자율

$N(\cdot)$ 는 표준누적정규분포 값

$$d_1 = \frac{\log\left(\frac{V^A}{V^D}\right) + \left(r + \frac{\sigma_A^2}{2}\right) \cdot T}{\sigma_A \cdot \sqrt{T}}$$

$$d_2 = \frac{\log\left(\frac{V^A}{V^D}\right) + \left(r - \frac{\sigma_A^2}{2}\right) \cdot T}{\sigma_A \cdot \sqrt{T}} = d_1 - \sigma_A \cdot \sqrt{T}$$

일반적으로 식 (1)과 같은 옵션가격결정모형의 $N(d_2)$ 는 만기시점 T 에 콜 옵션이 가치를 가질 확률을 의미한다. 따라서 식 (1)을 통해 구해지는 $N(-d_2)$ 또는 $1-N(d_2)$ 는 T 시점에 자산가치가 부채가치보다 작을 확률을 의미하고, 이는 기업의 도산 가능성을 나타내는 지표로 활용될 수 있을 것이다.

이원흠, 이한득, 박상수(2000)의 연구는 이러한 $N(-d_2)$ 가 기업의 도산 가능성을 파악하는데 유용한 변수임을 실증검증을 통해 밝히고 있다³⁾. 이들의 연구는 국내에서도 옵션가격결정모형을 이용해 기업의 도산 가능성을 파악할 수 있다는 가능성을 제시하였다는 점에서는 의의를 갖는다 할 수 있다. 그러나 위 연구의 결과에서 식 (1)의 $N(-d_2)$ 를 실제로 기업의 도산 가능성 측정에 활용하기에는 다음과 같은 한계점을 갖고 있다.

첫째, $N(-d_2)$ 는 콜옵션의 행사가격을 V^D , 즉 부채의 가치로 설정한 상태에서 산출된 값이라는 점이다. 이렇게 산출된 $N(-d_2)$ 는 T 시점에 자산가치가 부채가치 이하일 확률, 즉 자본잠식 가능성을 의미하게 된다. 현실적으로 자본잠식 상태에 있는 모든 기업이 도산하는 것은 아니기 때문에 이렇게 산출된 $N(-d_2)$ 를 그대로 활용하면 도산 가능성이 과대평가 될 수 있다.

둘째, $N(-d_2)$ 는 부채상환 이외의 목적을 위한 자산의 활용을 고려하지 않고 있다. 부채에 비해 자산가치가 클수록 $N(-d_2)$ 는 작아지지만, 자산가치가 높다 해서 반드시 부채 상환능력이 양호하다고 할 수는 없을 것이다. 따라서 $N(-d_2)$ 를 그대로 활용할 경우, 자산의 수익성이 낮거나 이자나 배당지급 또는 제3자에 대한 자산의 이전 등으로 인한 도산 가능성은 간과될 수 있다.

셋째, $N(-d_2)$ 는 몇 가지 제한된 가정⁴⁾과 무차익거래조건 하에서 산출되

3) 이원흠, 이한득, 박상수(2000)의 연구에서는 이러한 $N(-d_2)$ 를 산출하는 구체적인 방법과 유사한 검증 결과를 제시하고 있다.

4) 식 (1)이 성립되기 위해서는 자산가치가 Wiener process를 따른다는 등의 가정이 필요하다.

는 값으로서 실제 도산가능성과는 차이가 있을 수 있다는 점이다. 일반적으로 $N(-d_2)$ 는 투자자가 위험중립적 성향을 갖고 있다는 가정 하에서 산출되는 자본잠식 가능성으로서 d_2 의 수준에 따라 도산 가능성을 과대 또는 과소평가할 가능성이 있기 때문에, 현실적인 d_2 의 수준별 도산빈도를 파악할 필요가 있다.

국찬표, 정완호의 연구(2002)에서는 이러한 한계점을 극복하기 위해 도산 가능성으로 $N(-d_2)$ 대신 예상도산확률(EDF: Expected Default Frequency)을 활용하고 있는데, 이 값과 $N(-d_2)$ 의 차이는 다음과 같다. 첫째, 콜옵션의 행사가격으로 부채가치 대신 도산점(DP: Default Point)을 사용한다. 둘째, 자산가치를 부채상환에 사용될 수 있는 부분과 그렇지 않은 부분으로 구분하고 있다. 셋째, 도산 가능성의 지표로서 d_2 대신 도산거리(DD: Distance to Default)를 활용하고, 표준정규분포 대신 과거의 도산 빈도(Historical Default Frequency)를 이용해 예상도산확률을 계산 한다⁵⁾.

5) 도산거리를 측정함에 있어 Wiener Process 등의 제한적 가정이 불필요한 것은 아니다. 하지만 제한적 가정하에 산출된 현시점 자산가치와 그 변동성에 기초한 도산 가능성의 지표로서 d_2 대신 도산거리를 활용하며, 제한된 가정하에 산출된 도산거리에 기초하여 현실적 도산 가능성을 파악하기 위해 도산거리 수준별 과거 도산 빈도를 활용한다.

2. 예상도산확률모형을 통한 도산 가능성의 측정⁶⁾

식 (1)은 행사가격이 부채가치인 콜옵션의 가치를 나타내고 있기 때문에, T시점에 콜옵션의 가치가 없을 확률 $N(-d_2)$ 는 기업이 자본잠식 상태에 처할 가능성을 의미한다. 예상도산확률모형은 행사가격으로 부채규모 대신 T시점 이전에 기업이 도산하지 않기 위해 보유해야 할 최소한의 자산규모를 활용함으로써 자본잠식 가능성이 아닌 도산 가능성을 파악한다. 기업의 실제 자산규모가 행사가격으로 활용되는 자산규모 이하로 감소할 경우, 기업은 도산하게 된다는 의미에서 이러한 행사가격을 도산점(Default Point)이라고 한다.

도산점은 다음과 같이 파악될 수 있다. 먼저 자산규모가 T시점 이전에 상환해야 할 부채규모 이상이어야만 자산의 매각이나 자산을 담보로 한 자금조달을 통해 부채를 상환할 수 있기 때문에 도산점은 최소한 T시점 이전에 상환해야 할 부채규모 이상이어야 할 것이다. 그런데 자산 일부는 활용이 제한되어 있을 수 있고⁷⁾, 만기가 T시점 이후인 부채 일부는 기업이 재무적 곤경에 처할 경우 조기 상환이 요구될 수도 있다. 따라서 도산점은 T시점 이전에 상환해야 될 부채보다는 크고 기업의 총 부채보다는 작은 값을 갖게 된다.

한편, 기업이 재무적 곤경에 처할 경우 상환의 유보가 가능한 부채가 있다면 이는 T시점 이내의 도산 가능성에는 영향을 미치지 않기 때문에 도산점은 이러한 부채를 포함하지 않은 값을 갖게 된다. 반면, 다른 기업 부채에 대한 지급보증 등과 같이 현재는 부채로 인식되지 않은 우발채무가

6) 국찬표, 정완호, “기업 도산 예측에 관한 연구 : 추가정보를 이용하여”, 재무연구 제17권 제1호, 2002.

7) 예를 들어 T시점 이후에 상환해야 할 부채 조달을 위해 담보로 제공된 자산이나, 전세권 등이 설정되어 있는 자산은 T시점 이전의 부채 상환을 위해 활용될 수 없다.

있다면 도산점은 이를 고려하여 추정되어야 한다. 따라서, 도산점은 식 (2)를 통해 구해진다.

$$DP_{j,t,T} = SL_{j,t,T} - SL_{j,t,T}^D + x_{j,t} \cdot (LL_{j,t,T} - LL_{j,t,T}^D) + P_{j,t,T} \cdot CL_{j,t}$$

식 (2)

여기서, $DP_{j,t,T}$ 는 만기가 T인 경우의 j기업의 t시점 도산점

$SL_{j,t,T}$ 는 t시점 현재, 만기가 T보다 짧은 j기업의 부채규모

$SL_{j,t,T}^D$ 는 $SL_{j,t,T}$ 가운데 상환 유보가 가능한 부채규모

$LL_{j,t,T}$ 는 t시점 현재, 만기가 T보다 긴 j기업의 부채규모

$LL_{j,t,T}^D$ 는 $LL_{j,t,T}$ 가운데 상환 유보가 가능한 부채규모

$P_{j,t,T}$ 는 t시점 현재, T시점 이전에 j기업의 우발채무가 발생할 가능성

$CL_{j,t}$ 는 t시점의 j기업 우발채무규모

$$0 < x_{j,t} \leq 1$$

국찬표, 정완호의 연구에서는 T가 1년인 경우의 예상도산확률을 구하기 위해 도산점을 다음과 같이 산출하였다. 먼저 $SL_{j,t,T}$ 로는 t시점에 공시되어 있는 가장 최근 결산기 대차대조표상의 유동부채를, $LL_{j,t,T}$ 로는 같은 대차대조표의 부채총계에서 유동부채를 차감한 금액을 활용하였다. 단, 환율변동을 반영하기 위하여 외화부채의 경우에는 t시점의 환율을 이용해 원화금액을 다시 계산하였다.

또한, 모든 유동부채의 상환은 유보될 수 없음을 가정하여 $SL_{j,t,T}^D = 0$ 으로 계산하였으며, $LL_{j,t,T}^D$ 로는 대차대조표상의 퇴직급여충당금, 단체퇴직급여충당금, 퇴직보험충당금의 합계를 활용하였다. 이는 현실적으로 퇴직급여

의 경우 기업이 재무적 곤경에 처하면 일정 기간 지급 유보가 가능하며, 이의 지급여부는 도산 여부를 직접적으로 결정하지는 않기 때문이다. $CL_{j,t}$ 는 타 기업의 부채에 대한 지급보증 규모로 정의하였으며, $P_{j,t,T}$ 로는 피보증기업 예상도산확률이 산출 가능한 경우에는 예상도산확률을, 예상도산확률을 산출할 수 없는 경우에는 신용등급별로 산출해낸 과거 실제부도율을, 예상도산확률이나 신용등급이 모두 파악 불가능한 경우에는 보수적 추정을 위해 C등급의 실제부도율을 활용하였다⁸⁾.

$x_{j,t}$ 의 정확한 산출을 위해서는 각 시점별 부채규모나 담보 설정 내역이 파악되어야 할 것이다 그러나 이의 파악은 현실적으로 많은 제약이 따르기 때문에 위 연구에서는 $x_{j,t}$ 의 기업 및 시점별 차이는 반영하지 않았다. 하지만 산업별 자본집약도, 자본 조달 방법 등에 의해 영향을 받을 수 있음을 고려해 산업별로 다른 값을 적용하였다. I 산업에 속한 j 기업의 도산점을 계산하기 위한 $x_{j,t}$ 를 x_I 라고 하면, 이는 다음과 같은 과정을 통해 결정되었다.

먼저 $x_{j,t}$ 가 0.5임을 가정해 도산점을 계산한 다음 식 (3)과 같은 부채총계와 도산점 비율의 평균을 구하였다. 다음으로 특정산업내의 기업만을 대상으로 이 비율의 평균이 식 (3)을 통해 구한 전체 평균과 일치되도록 하는 산업별 x_I 를 구하였다.

$$\frac{\sum_t \sum_j \frac{DP_{j,t,T}|x_{j,t}=0.5}{L_{j,t}}}{N_j \cdot N_t} \quad \text{식 (3)}$$

8) 만기가 1년 이상 남은 보증채무도 $CL_{j,t}$ 에 포함시켰다. 피보증 기관이 1년 이내에 도산할 경우, 부채의 만기에 관계없이 도산 시점부터 보증기관의 자산 일부가 동결될 수 있기 때문이다.

여기서, $L_{j,t}$ 는 t시점의 j기업 부채총계

N_j 는 검증대상 기업의 수

N_t 는 기업별 도산점 산출 횟수

그런데 기업의 예상도산확률을 산출하기 위해서는 식 (1)에 부채가치 대신 도산점을 대입하는 것 외에 자산가치도 조정되어야 한다. 왜냐하면 t시점 현재 기업이 보유하고 있는 자산 모두가 부채의 상환을 위해 활용될 수 있는 것은 아니기 때문이다. 현재 보유하고 있는 자산 가운데 일부는 T시점 이전에 이자나 배당 지급을 위해 기업 외부로 유출될 수 있다. 따라서 예상도산확률을 산출하는데 필요한 기업의 자산가치와 그 변동을 파악하기 위해 식 (1)을 식 (4)와 같이 수정하여 활용할 수 있다.

$$V_{j,t}^E = (V_{j,t}^A - Q_{j,t,T}) \cdot N(d_1) - e^{-r_{t,T}T} \cdot DP_{j,t,T} \cdot N(d_2)$$

식 (4)

여기서, $Q_{j,t,T}$ 는 t~T시점 사이에 유출이 예상되는 j기업 자산규모

$$d_1 = \frac{\log\left(\frac{V_{j,t}^A - Q_{j,t,T}}{DP_{j,t,T}}\right) + \left(r_{t,T} + \frac{\sigma_{A,j,t}^2}{2}\right) \cdot T}{\sigma_{A,j,t} \cdot \sqrt{T}}$$

$$d_2 = d_1 - \sigma_A \cdot \sqrt{T}$$

위 연구에서는 $Q_{j,t,T}$ 로는 직전 결산기의 $[2 \times \text{이자비용} / (\text{기초 부채규모} + \text{기말 부채규모})] \times t$ 시점 부채규모와 직전 결산기 현금배당금의 합을 사용하였다. 또한 $r_{t,T}$ 로는 한국채권평가(주)에서 산출하고 있는 연속복리로 계산된 1년 만기 국고채의 spot rate을 이용하였다.

그런데 식 (4)는 특정 시점의 자산가치 $V_{j,t}^A$ 가 Wiener process를 따른다는 등의 제한적 가정 하에 성립되는 것이며, 이러한 가정 하에서는 식 (5)의 관계가 성립된다. 식 (4)의 자산가치 $V_{j,t}^A$ 와 $\sigma_{A,j,T}$ 는 이러한 조건을 이용하여 다음과 같은 반복적 탐색과정(iterative search process)을 통해 구해진다.

$$\sigma_{E,j,t} = \frac{(V_{j,t}^A - Q_{j,t,T})}{V_{j,t}^E} \cdot N(d_1) \cdot \sigma_{A,j,t} \quad \text{식 (5)}$$

여기서 $\sigma_{E,j,t}$ 는 j기업의 t시점 시가총액 변동성

$\sigma_{A,j,T}$ 는 $(V_{j,t}^A - Q_{j,t,T})$ 의 변동성

먼저 식(4')의 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 에 적절한 초기값을 대입해 계산된 $V_{j,t}^E$ 와 $V_{j,t}^A$ 의 초기값을 식 (5)에 대입하여 $\sigma_{A,j,t}$ 를 계산할 수 있다. 이렇게 계산된 $\sigma_{A,j,t}$ 를 식 (4')에 대입하면 $V_{j,t}^A$ 를 다시 계산할 수 있다.

$$V_{j,t}^A = \frac{V_{j,t}^E + e^{-r_{t,T}T} \cdot DP_{j,t,T} \cdot N(d_2)}{N(d_1)} + Q_{j,t,T} \quad \text{식 (4')}$$

$V_{j,t}^{E'}$ 는 식 (4)의 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 에 특정한 값을 대입하여 계산된 $V_{j,t}^E$

식 (4')을 통해 계산된 $V_{j,t}^A$ 와 식 (5)를 이용해 계산된 $\sigma_{A,j,t}$ 를 초기값 대신 다시 식 (4)에 대입하면 $V_{j,t}^E$ 를 다시 계산할 수 있다. 이렇게 다시 계산된 $V_{j,t}^E$ 가 주식시장에서 관찰된 시가총액과 일치하지 않으면, 다시 계산된 $V_{j,t}^E$ 를 식 (5)에 대입하여 $\sigma_{A,j,t}$ 를 다시 구하고 이를 식 (4')에 대입해 $V_{j,t}^A$ 를 다시 구한 다음 새로 구한 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 를 다시 식 (4)에 대입해 구한 $V_{j,t}^E$ 가 시가총액과 일치하는가를 살펴본다. 식 (4)에서 구한 $V_{j,t}^E$ 와 시가총액이 일치할 때까지 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 를 구하는 과정을 반복해 최종적인 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 를 자산가치와 그 변동성으로 활용한다.

위 연구에서는 $V_{j,t}^E$ 로는 매월말 주식 종가를 이용해 계산한 시가총액을 활용하였다. 또한 주식수익률의 변동성이 크고 자기상관이 심하게 나타나는 국내 주식시장의 특성을 감안해 $\sigma_{A,j,T}$ 로는 식 (6)과 같은 회귀식을 이용해 추정하는 $\sigma_{E,j,t}^{yr}$ 을 활용하였는데, 검증기간 중 주식 수익률의 변동성이나 자기상관에 영향을 미칠 수 있는 제도변화가 빈번하였기 때문에 매달 회귀계수를 다시 추정하였다. 이때 현금배당에 대한 인위적인 배당락이 있었던 거래일의 시가총액은 변동성 계산에서 제외하였다.

$$\sigma_{E,j,t}^{yr} = \alpha_t + \sum_{n=0}^{23} \beta_{n,t} \cdot \sigma_{E,j,t-n}^{mo} + \epsilon_{j,t} \quad \text{식 (6)}$$

여기서 $\sigma_{E,j,t}^{yr}$ 은 t시점 이후 1년간의 j기업 일별 시가총액의 표준편차

$\sigma_{E,j,t}^{mo}$ 는 t시점 이전 1개월간의 j기업 일별 시가총액의 표준편차

$\alpha_t, \beta_{m,t}$ 는 회귀계수

반복적 탐색과정을 통해 자산의 가치와 그 변동성이 추정되면, 이 값을 이용해 식 (7)과 같은 도산거리(Distance to Default)⁹⁾를 산출한다. 도산거리는 예상도산확률을 결정하는 변수인데, d_2 를 이용해 도산 가능성을 파악하지 않고 이 값을 기준으로 예상도산확률을 파악한다는 것이 이원흙, 이한득, 박상수(2000) 등에 의해 제시된 옵션가격결정모형을 이용한 도산 예측모형과 구별되는 예상도산확률모형의 특징 가운데 하나라고 할 수 있다.

$$DD_{j,t,T} = \frac{\log \frac{(V_{j,t}^A - Q_{j,t,T})}{DP_{j,t,T}} + (\mu_{j,t,T} - \frac{\sigma_{A,j,t}^2}{2}) \cdot T}{\sigma_{A,j,t} \cdot \sqrt{T}} \quad \text{식 (7)}$$

여기서 $DD_{j,t,T}$ 는 만기가 T인 경우 j기업의 t시점 도산거리

$\mu_{j,t,T}$ 는 t~T시점 사이의 j기업 자산가치 성장률

식 (7)을 살펴보면 도산거리는 식 (4)의 d_2 에 무위험이자율 $r_{t,T}$ 대신 j기업의 자산가치 성장률 $\mu_{j,t,T}$ 을 대입하여 계산되는 값이다. d_2 가 현재의 자산가치가 무위험이자율로 증가할 때 T시점에 도산점 이하일 확률을 나타내는 지표라면, 도산거리는 현재의 자산가치가 $\mu_{j,t,T}$ 의 성장률로 증가할 때 T시점에 도산점 이하일 확률을 파악하기 위해 활용하는 지표라고 할 수 있다. 즉, 도산거리는 현재의 자산가치와 그 변동성이 $V_{j,t}^A$ 와 $\sigma_{A,j,t}$ 으로 주

9) 현재의 자산가치가 도산점과 얼마나 떨어져 있는가를 표준화하여 나타내고 있다는 의미에서 도산거리라고 한다.

어진 경우, $\mu_{j,t,T}$ 의 성장률로 증가하는 자산의 T시점 가치와 도산점 $DP_{j,t,T}$ 간의 괴리율을 자산변동성으로 표준화한 값이라 할 수 있다.

도산거리가 어떤 분포를 따르는지는 사전적으로 알 수 없기 때문에 예상도산확률은 식 (8)과 같이 과거의 도산거리별 실제 도산 빈도를 기준으로 산출된다. 또한 앞서의 과정을 통해 계산되는 자산가치나 그 변동성 또는 도산거리가 제한적 가정 하에 산출된 것이기 때문에 도산거리가 어느 정도 일 때 실제 도산 가능성은 얼마나 될지 사전적으로 파악하기 곤란하다. 따라서 도산거리를 통해 현실적 도산 가능성을 파악하기 위해 도산거리 수준별 과거 도산 빈도를 이용한다.

위 연구에서는 예상도산확률을 파악하기 위해 1985년 초부터 각 t시점 이전까지 매월달의 금융기관을 제외한 상장기업 도산거리(DD) 수준별 도산 빈도를 활용하였다. 즉, j기업의 t시점 예상도산확률은 1985년 1월부터 t-1시점(직전 월말)까지의 금융기관을 제외한 상장기업의 도산거리 가운데 j기업의 t시점 도산거리와 그 값이 유사한 경우의 수와 이러한 도산거리가 산출된 후 T기간(1년) 이내에 해당기업이 도산한 경우의 수의 비율로 계산된다.

$$EDF_{j,t,T} = \frac{\sum_{\tau=1}^{t-1} \sum_{i=1}^{N(\tau)} D_{i,\tau,T} \cdot S_{i,\tau,T}}{\sum_{\tau=1}^{t-1} \sum_{i=1}^{N(\tau)} S_{i,\tau,T}} \quad \text{식 (8)}$$

여기서, $EDF_{j,t,T}$ 는 j 기업이 t 시점으로부터 T 기간 이내에 도산할

가능성에 대한 예상도산확률

$N(\tau)$ 는 τ 시점에 도산거리가 산출된 기업의 수

τ 는 1985년 1월말을 1로 하고 매월마다 1씩 증가하는 수

$D_{i,\tau,T} = 1$, i 기업이 τ 시점 이후 T 기간 이내에 도산한 경우

$D_{i,\tau,T} = 0$, i 기업이 τ 시점 이후 T 기간 이내에 도산하지 않은 경우

$S_{i,\tau,T} = 1$, $DD_{j,t,T} - \Delta DD < DD_{i,t,T} < DD_{j,t,T} + \Delta DD$

$S_{i,\tau,T} = 0$, $DD_{i,\tau,T} \leq DD_{j,t,T} - \Delta DD$ 또는 $DD_{i,\tau,T} \geq DD_{j,t,T} + \Delta DD$

단, $D_{i,\tau,T}$ 는 I 기업의 τ 시점 도산거리

$\Delta DD = (1/n) \times DD_{j,t,T}$, 여기서 n 은 1보다 큰 임의의 수

n 은 $t > m \geq 0$ 인 모든 m 에 대해 $\frac{dEDF_t}{dDD_{t-m}} \leq 0$ 이 항상 성립될 수 있도록

록 결정되어야 하며 $DD_{i,\tau,T}$ 수준별로 다른 값을 가짐

제 3 장 부도예측모형의 선행연구

제 1 절 추가정보를 이용한 부도예측모형

1. 이은주의 연구¹⁰⁾

이은주의 논문은 KMV 모형을 중심으로 1994년에서 1998년까지를 연구 기간으로 정하였다. 우리나라의 금융위기가 있는 1997년을 기점으로 그 이전과 이후로 나누어 추가정보 이용 모델의 유의성에 대하여 분석하였다. 기업집단은 재무적 곤경 기업(실험집단)과 정상 운영 기업(통제집단)으로 각각 91개의 기업을 선정하여 쌍대비교(Paired Comparison)를 실시하였다. 이 연구에서는 Gauss로 모델을 구현하여 DD까지 산출하였다. DD의 쌍대비교를 통하여, 기업규모의 효과, 부채비율의 효과, 동종 산업 여부를 비교하였다.

연구 결과, IMF 이전의 DD보다 이후의 DD가 보다 유의하게 나왔으며, 기업규모가 작은 기업의 DD가 차이가 크게 나와 유의한 정보능력이 있는 것으로 해석하였다. IMF 이전에는 재무적 곤경기업과 정상기업간의 부채비율 차이가 클수록, 매칭된 두 기업이 동종 산업일수록 DD의 차이가 크게 나타나는 것으로 나타났다.

이 연구는 DD가 기업의 자본구조, 동종 산업, 기업의 자산 가치를 포함

10) 이은주, “신용위험평가 모형의 실증연구 : 한국 상장기업의 KMV모형 적용을 중심으로”, 연세대학교 석사학위 논문, 2000.

한 총체적 계수로서 EDF가 산업이나 국가의 경계를 넘어 사용될 수 있다는 KMV의 주장에 대한 실증을 분석한 논문이다.

2. 이원흠, 이한득, 박상수의 연구¹¹⁾

두 가지 상이한 이론적 배경을 가진 로짓분석기법과 Merton(1974)의 옵션모형에 의한 도산예측확률의 예측모델에 근거해서 실증분석을 시행한 결과를 보고하였다.

연구대상 기간은 1997년부터 1999년까지 3개년 동안으로 하였으며, 도산한 154개사를 중심으로 비금융 상장기업들을 위주로 실증분석 하였다.

현금흐름형 정보에 근거한 도산예측 모형으로 투하자산 회전율, 매출액영업이익률, 이자보상비율, 잉여현금흐름 상환능력 등을 이용하여 현금흐름과 현금창출, 이자지급, 원금상환을 변수로 구성하여 로짓 분석을 실시하였다. 각 변수간의 상관성과 각 연도별 변수의 설명력을 분석한 결과 투하자산회전율과 이자보상비율이 2년 시차를 사용한 도산예측모형의 판별력이 가장 좋은 것으로 나타났다.

이들의 연구에서는 Merton(1974)의 모형에 의한 부도의 예측 빈도(EDF)로 $N(-d_2)$ 를 사용하였다. 분석 방법으로는 대우그룹을 비롯한 1999년의 5대그룹을 벤치 마크하여 도산예상확률을 추정하였다. 검증결과 도산 발생 전 최소 2개월 내지 8개월 전부터 도산예상확률이 급증하는 것으로 나타났다. 또한 국가신용등급과 도산예상확률 간의 선후관계에 관한 Granger Causality Test를 실시하였다. 국가신용등급 자료로는 Fitch, Moody's, S&P의 등급자료를 이용하였다. F값과 유의도를 제시한 분석결과 EDF모

11) 이원흠, 이한득, 박상수, “현금흐름형 도산예측 모델과 옵션모델형 도산예상확률의 실증연구”, 증권학회지 제27권, 2000.

형의 예측력이 국가신용등급 정보에 비하여 선행성이 존재하는 것으로 나타났다.

3. 국찬표, 정완호의 연구¹²⁾

KMV의 EDF 모형이 실제 도산기업에 대한 사전적으로 도산 정보를 제공할 수 있는가를 국내 과거 시계열 자료를 이용하여 경험적으로 검증하였다. 이 논문에서는 다변량 판별분석, 프로빗모형, 생존분석, 분산분석, OPM 모형 등을 사용한 경우와의 예측결과 차이를 비교함으로써 국내에 유용한 모형을 파악해 보았다.

연구기간은 1998년 1월부터 2001년 5월까지로 하였으며, 비금융 상장사를 대상으로 도산거리(DD)와 예상도산확률(EDF)를 계산하였다.

예상도산확률을 파악하기 위해 1985년 초부터 각 시점 이전까지 매월말의 금융기관을 제외환 상장기업 도산거리 수준별 도산빈도를 활용하였다.

EDF를 0~3%, 3~20%, 20% 이상의 셋으로 구간화하여 우도비검증(Likelihood Ratio Test)를 실시하였다. 계급화된 구간별 도산의 경험적 실증으로 예상도산확률의 구간별 예측력을 분석한 결과 예상도산확률에 따른 도산 예측능력이 있는 것으로 나타났다.

또한, 본 모형의 기업도산에 대한 예측능력이 타모델보다 우월하게 나타남을 각 모형의 결과를 기준으로 도산 가능성의 순위를 정하고 이 순위가 높을수록 실제로 도산한 경우도 많은지를 비교하는 방법을 이용해 모형별 예측 능력을 비교하였다. 그 결과 예상도산확률 모형이 타 모형에 비하여 도산 가능성을 좀더 정확하게 예측할 수 있음을 파악하였다.

12) 국찬표, 정완호, “기업 도산 예측에 관한 연구 : 추가정보를 이용하여”, 재무연구 제17권 제1호, 2002.

제 2 절 재무비율을 이용한 기업부도예측

1. Altman(1968)의 다변량 판별분석

Altman의 연구에서는 단일 재무비율 분석이 아닌 다변량 판별분석(Multiple Discriminant Analysis)을 이용하여 기존의 개별적으로 관찰되던 재무비율을 종합하고 단순화하여 부도예측을 하였다. Altman은 기존의 단일변량분석 기법을 활용한 연구가 지니고 있던 한계를 다변량 판별분석을 통하여 해결하였으며, 기존의 단일변량분석에 비해 높은 예측력을 보여주었다. 그러나 이러한 분석방법은 다변량 분석에서 요구하고 있는 비교집단 간에 독립변수의 공분산행렬의 동일성 및 정규분포조건의 위배가능성을 내포하고 있다¹³⁾.

Altman의 Z-Score모형(1968)에서 표본선정은 기본적으로 쌍대표본추출 방식을 기본으로 하여 1946년부터 1965년 사이에 파산신청서를 제출한 33개의 제조기업을 부실기업으로 선정하였고, 독립변수는 기존 문헌에서 이용 빈도가 높은 변수 22개를 재무비율로 선정하였고, 이들 가운데 기업부실 예측력을 저하시키지 않는 범위 내에서 ① 독립변수의 상대적 공헌도를 결정하는 통계적 유의성, ② 변수간의 상관관계, ③ 예측정확성, ④ 분석자의 주관적 판단 등에 의해서 최종적으로 5개의 변수로 축소시켜, 판별함수를 도출하였다.

13) Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," The Journal of Finance, Vol. 23, No. 4, 1968, pp.589-609.

① Z-Score모형

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.0099X_5$$

X1 : 순운전자본/총자산 × 100

X2 : 이익잉여금/총자산 × 100

X3 : 영업이익(EBIT)/총자산 × 100

X4 : 자본의 시장가치/부채의 장부가치 × 100

X5 : 매출액/총자산 × 1 (총자산회전율)

Altman은 22개의 재무비율을 분석대상으로 하여 5개의 재무비율이 기업 도산을 예측하는데 가장 적합한 사실을 발견하였다.

Altman모델의 최종판별함수와 도산예측기준은 <표 3-1>과 같다.

등급	기준값	비고
A	$Z > 2.99$	모든 기업이 실패하지 않는 계층
B	$1.81 < Z < 2.99$	실패기업과 실패하지 않는 중간계층으로 질적 진단이 필요한 계층
C	$Z < 1.81$	모든 기업은 실패하는 계층

<표 3-1> 최종판별함수와 도산예측기준

Altman모델의 판별함수식에 의하여 산출된 Z값이 2.99보다 클 경우에는 건전기업으로 판별되고, Z값이 1.81이하일 경우에는 부도기업으로 판별되며, Z값이 1.81과 2.99 사이에 있는 기업은 중간계층으로 판별된다.

이렇게 나온 판별함수를 이용하여 추정용표본 33개를 대상으로 하여 예측력을 테스트해본 결과 부도 1년 전에는 95%, 2년 전에는 72%, 3년 전에는 48%, 4년 전에는 29%, 5년 전에는 36%의 정확성을 보였다. 예측력 테스트 결과 부도 2년 전까지는 높은 정확성을 보이다가 3년 전부터는 예측력이 크게 떨어지고 있다.

즉, 판별분석방법은 관측 대상의 특성 변수들의 값에 의해서 관측 대상을 다수의 그룹으로 분류하는 통계적 기법이다. 판별분석방법은 판별 변수들의 정규화분포 등의 제한적인 가설에 근거하고 있으므로 실제로 많은 문제점을 내포하고 있다. 실제로 특성변수들이 정규화분포를 따르지 않는 경우가 많으며, 이러한 경우에 판별능력에 문제가 생길 수 있다. 판별분석의 경우에 판별함수(discriminant function)가 정규분포의 평균(mean)과 공분산(covariance)에 의해 정해지기 때문에 예측 에러가 커질 수 있기 때문이다.

2. Logit 분석에 의한 부도예측모형

앞에서 언급한 다변량 판별분석이 갖는 한계점을 극복하기 위해서 Logit 분포와 Probit 분포를 이용한 부도예측모형이 개발되었다. Logit 분석은 판별분석에 비해 관련변수들이 정규분포이어야 한다는 가정이 전제될 필요가 없으며, 다만 선택확률이 logistic 함수를 취한다는 가정이 필요하다. Ohlson의 연구(1980)에서는 쌍대표본방식이 아닌 1970년부터 1976년 사이에 도산한 105개 기업을 부실기업표본으로, 건전기업표본으로는 2058개를 선정하였다. 독립변수는 9개를 임의로 선택하였는데 이들 변수는 다음과 같다.

SIZE : $\log(\text{총자산} / \text{GNP 물가지수})$

TLTA : 총부채 / 총자산

WCTA : 운전자본 / 총자산

CLCA : 유동부채 / 유동자산

OENEG : 유동부채 / 유동자산

NITA : 당기순이익 / 총자산

FUTL : 영업활동으로부터의 운전자본 / 총부채

INTWO : 과거 2년간 적자라면 1, 아니면 0

CHIN : $((NI_t - NI_{t-1}) / (NI_t + NI_{t-1}))$ 단, NI는 가장 최근의 순이익

판별점을 0.038로 할 때 가장 높은 예측력을 보여주었다. 이 때의 예측력은 부도의 경우 87.6%, 건전기업의 경우 82.6%이었다.

남주하(1998)의 연구에서 Logit 분석방법을 사용하여 부도예측모형을 개발하였다. 표본의 선정은 쌍대표본추출방식 기본으로 하여 97년 부도가 났거나 구제금융을 받은 경험이 있는 기업들을 대상으로 부실기업표본 47개를 선정하였고, 동일한 업종에서 자본금규모가 비슷한 건전기업을 47개를 선정하였다. 독립변수의 선정은 기업경영분석에 일반적으로 사용되고 있는 36개의 주요 재무비율을 선정하였다. 단계적 선택(stepwise)방식을 통해 최종모형 내에 포함된 변수들은 매출액영업이익률과 매출액경상이익률, 금융비용대매출액비율, 부채상환계수이고, 이들 변수로 구성된 모형의 전체의 예측력이 92%로 높은 예측결과를 보여주고 있다. 그 외에도 남주하, 김동수, 김명정(1995)에서도 Logit 분석방법을 이용하여 예측모형을 개발하였다. 유의성이 있는 3~4개의 변수를 이용하여 구성된 예측모형은 예측력이 99%를 넘었으며, 특히 부실기업에 대한 예측력은 100%이었다. 이렇게 높은 예측력을 보였던 이유는 모형 추정시 사용된 표본들을 이용하여 test한

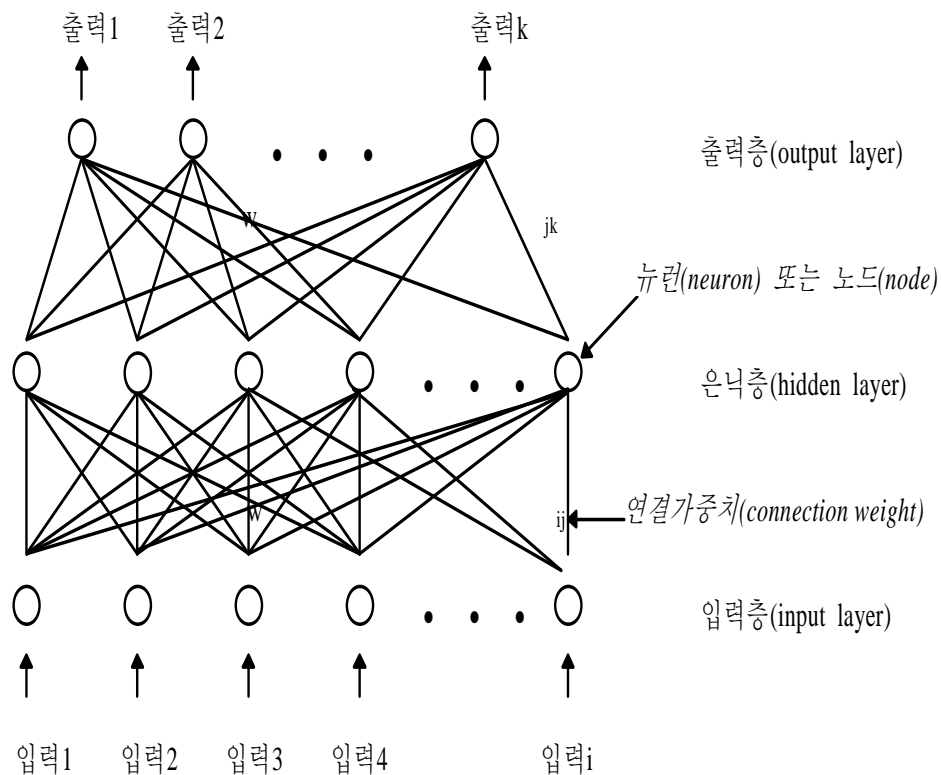
결과이기 때문에 예측력 자체에 큰 의미를 둔다는 것은 무리가 있다. 즉, 어떤 표본을 사용하여 test 하는가에 따라 예측력에 큰 차이가 생길 수 있어, 일반성이 떨어지는 경우가 발생할 가능성이 농후하다.

3. 인공신경망(Artificial Neural Networks)

신경망 분석은 인간의 두뇌가 문제를 해결하는 구조를 이용한 분석이다. 물론 인간의 두뇌만큼 다차원적으로 복잡한 문제를 해결할 수 있는 것이 아니라 한 부분에 대하여 단순히 인간의 신경구조의 작용원리를 응용한 것이라 할 수 있다.

전형적인 신경망은 망을 생성하는 층(layers)들이 배열된 뉴런(neurons)이란 것으로 구성되어 있다. 각각의 뉴런은 전체에서 어느 한 부분을 작용하는 과정 요소로 생각되어질 수 있다. 모든 뉴런을 연결하게 되면, 어떤 패턴(pattern)을 학습하고, 데이터간의 상호 관계를 학습하는 Network으로 발전하게 된다.

인공신경망의 구조는 여러 가지 형태가 있으나 가장 널리 사용되는 인공신경망의 구조는 다층신경망(Multi-Layer Neural Networks)이다. <그림 3-1>은 다층신경망의 구조를 간단히 나타낸 것이다.



<그림 3-1> 다층신경망의 구조

신경망을 이용하여 예측 모델을 구축할 때, 입력층(input layer)은 예측에 필요한 모든 필드들이 포함된다. 다음 출력층(output layer)에서는 목표필드(target filed)와 비교하여 나온 결과값들이 들어있다. 은닉층(hidden layer)은 입력층의 필드들을 결합하여 연결하는 뉴런들로 구성되어 있다. 일반적으로 은닉층은 그 수가 많지는 않지만 여러 개를 가질 수가 있다. 그리고 하나의 층에 있는 모든 뉴런들은 다음 층과 모두 연결이 되어져 있다.

신경망에서 입력층과 출력층 사이를 오가며 학습하는 것을 훈련

(training)이라고 말한다. 훈련을 하면서, 계속 새롭고 보지 못한 데이터들을 접하게 되고, 이런 경험을 기초로 하여 예측과 의사결정 모형을 만든다. 마치 인간의 학습과 같은 원리이다. 이러한 원리를 적용하여 다층신경망을 학습할 수 있는 역전파 알고리즘(Back-Propagation Algorithm)을 개발하였으며, 이로 인해서 인공신경망 연구가 활성화되는 계기가 되었다. 인공신경망을 응용할 수 있는 분야는 크게 세 분야로 나뉘어진다. 첫째, 패턴인식 분야로써 데이터 값의 특성에 따라 몇 개의 그룹으로 분류하는 방법이며, 본 논문의 기업부도예측이 이에 해당된다. 둘째, 시계열 데이터와 같이 시간에 따라서 변하는 데이터의 변화 패턴을 찾는 분야에 이용된다. 셋째, 인공신경망을 이용하여 임의의 비선형 함수를 추정하는 것으로 통계학의 regression 방법과 유사하다고 할 수 있으나 통계적 가설이 필요 없는 것이 특징이다.

인공신경망은 경영학 및 그 밖의 여러 분야에 응용되고 있다. <표 3-2>에서 보듯이 1980년대 후반부터 인공신경망 기법들이 부도예측 연구에 도입되어 많은 연구가 이루어졌다. 인공신경망 기법을 적용한 부도예측 모형에서는 주로 기업의 재무정보 등을 입력변수로, 기업의 부도여부를 출력변수로 설정하여 학습을 통해 이들의 관계를 추출하며 예측변수들간에 잠재적으로 숨겨진 상관관계를 탐색한다. 이러한 방법을 통해 인공신경망 기법이 부도예측과 관련하여 기존의 통계적 기법을 활용한 연구보다 더욱 향상된 예측 모형이 될 수 있음을 제시하고 있다. 이와 같은 인공신경망 모형을 적용한 부도예측 연구는 <표 3-3>에서와 마찬가지로 국내에서도 활발히 진행되고 있다.

Barniv et al.(1997)은 로짓 분석이나 NPDA보다 인공신경망이 데이터를 분류하는데 더 우수함을 보여주었으며, Zhang et al.(1999)은 로짓 분석과 인공신경망을 비교하여 2.28%의 성과를 개선시키는 결과를 보여주었다. 또

한, Davalos et al.(1999)은 인공신경망을 항공운송회사의 부도예측에 적용하였으며, Charalambous et al.(2000)은 인공신경망이 로짓 분석이나 역전파 알고리즘보다 우월한 결과를 보여준다는 결과를 도출하였다. 그리고 또한, Anandarajan et al.(2001)은 정성적 요인을 모형에 반영하기 위해 Zmijewski 점수를 사용하여 유전자 알고리즘에 기반한 인공신경망 모형이 다른 모형보다 판별오류 원가가 작다는 결과를 도출하였다.

연구자(국외)	연구기법	연구결과
Odom & Sharda (1990)	신경망과 판별분석 방법 비교	129개사(부도: 65, 건전: 64) 데이터로 적중률이 판별분석의 약 60~70%보다 약 80%로 우수함 증명
Tam & Kiang (1992)	ID3, 로짓 분석, 판별분석, 신경망 비교	118개사(부도 및 건전: 51) 데이터로 실험 결과 신경망이 예측정확성(predictive accuracy), 적응력(adaptability), 강건함 (robustness) 측면에서 부도예측 모형평 가에 좋은 방법이 될 수 있음을 보임
Bell (1997)	로짓 분석과 신경망 비교	1,059개사(부도: 131, 건전: 928) 데이터 로 실험결과 둘다 우수하나 신경망이 변 수간의 관계가 비선형일 경우 더 우수하 다는 가능성 제시
Jo et al. (1997)	신경망, 사례기반추론, 판별분석기법 비교	544개사(부도 및 건전: 272) 데이터로 판 별분석: 82.22%, 인공신경망: 83.79%, 사 례기반추론: 81.52% 결과 도출
Ran et al. (1997)	로짓 분석, 비모수다변량 판별분석(NPDA), 인공신경망	237개(폐합병: 49, 존속: 119, 청산: 69) 의 데이터를 이용하여 로짓 분석이나 NPDA보다 인공신경망이 더 우수한 분류 능력을 보여줌
Zhang et al. (1999)	인공신경망, 통계적 기법	220개(부도 및 건전 각각 110)의 자료를 바탕으로 인공신경망: 80.46%, 로짓 분 석: 78.18%의 결과 도출
Davalos et al. (1999)	다변량 판별분석, 인공신경망	신경망을 이용하여 테스트 집합인 26개 항공운송회사의 부도예측을 정확하게 함
Charralambous et al.(2000)	인공신경망, 로짓 분석, 역전파알고리즘	278개(건전 및 부도 각각 139)의 데이터 를 이용하여 로짓 분석이나 역전파알고리 즘보다 인공신경망이 우월한 결과를 보여 준다는 결과를 도출
Anandarajan et al. (2001)	인공신경망, 유전자알고리즘, 다변량 판별분석	유전자 알고리즘에 기반한 인공신경망 모 형이 다변량 판별분석 모형보다 판별오류 원가가 작다는 결과를 도출(정성적 요인 을 모형에 반영하기 위해 Zmijewski 점수 를 사용)

<표 3-2> 신경망을 적용한 부도예측관련 연구(국외)

연구자(국내)	연구기법	연구결과
이진창 (1993)	다변량판별분석 귀납적 학습방법, 인공신경망	166개(부도: 81, 건전: 85)의 데이터를 통해 각 연구기법들에 의해 예측력을 평가하여 인공 신경망, 귀납적 학습방법, MDA의 순으로 예측 력이 우수함을 보여줌
이진창 외 (1994)	인공신경망, 귀납적 학습방법, 다변량 판별분석, ILANN 통합모형	166개(부도: 81, 건전: 85)의 데이터 중 다변 량 판별분석이 70%, ILANN 모형이 92.5%로 인공신경망 기법이 우수함 증명
이진창 (1995)	HYNEN, 다변량 판별분석, ACLS, 퍼지 ARTMAP, 인공신경망	자료를 군집화하는 CNN(Clustering Neural Network)과 최종적인 출력을 제공하는 Output Neural Network)의 2단계로 구성 감독학습 인공신경망 모형과 비감독학습 인공 신경망 모형을 결합한 하이브리드 인공신경망 모형인 HYNEN 모형을 제안, HYNEN 모형성 과가 우수함
이제식, 차봉근 (1996)	인공신경망, 유전자 알고리즘	인공신경망을 적용함에 있어서 유전자 알고리 즘에 의하여 입력변수를 선정함
한인구 외 (1997)	인공신경망 통합모형, 판별분석, 로짓 분석	2,548개(부도 및 건전: 1274) 데이터의 세 가 지 모형 결과를 가중평균하고 GA를 통하여 가 중치를 구하여 78.7%에서 79.3%의 적중률을 79.1%에서 79.8%로 향상시킴
신경식 (2000)	인공신경망 통합모형, 판별분석, 유전자 알고리즘	2,088개사(부도 및 건전: 1,044) 데이터로 77.4%의 적중률을 75.8%로 향상시킴
홍승현, 신경식 (2003)	인공신경망, 유전자 알고리즘	528개사(부도 및 건전: 264) 데이터로 입력변 수 선정에 판별분석과 유전자 알고리즘을 각각 사용하여 적중률을 비교 분석함

<표 3-3> 신경망을 적용한 부도예측관련 연구(국내)

부도예측 분야에서 인공신경망을 적용하는 가장 중요한 연구 중의 하나
는 축적된 데이터를 이용하여 독립변수와 종속변수 간의 결합관계를 추출
해냄으로써 패턴인식, 분류, 예측 등의 기능을 수행하는 모형을 구축하는

것이다. 또한 인공신경망의 학습은 입력변수의 선정, 아키텍처, 데이터의 특성 등에 따라서 달라지고 예측치에도 차이가 있게 된다. 그러므로 최근의 연구들은 인공신경망과 기존 통계기법의 성과를 비교, 분석하는 연구에서 나아가 다양한 방법론을 혼합적으로 적용하여 예측력을 향상시키기 위한 다양한 연구가 진행되어 왔다¹⁴⁾.

이건창(1993)과 이진창 외(1994)는 다변량 판별분석, 귀납적 학습 방법, 인공신경망 등에 대하여 기업도산의 예측력과 성과를 비교하여 인공신경망 모형의 우수성을 보여주고 있다. 또한, 이진창(1995)은 비구조적인 의사결정 문제를 효과적으로 해결하기 위하여 감독학습 인공신경망 모형과 비지도학습 인공신경망 모형을 결합한 하이브리드 인공신경망 모형인 HYNEN(hybrid neural network) 모형을 제안하여 그 결과를 다변량 판별분석, ALCS(analog concept learning system), 퍼지 ARTMAP, 기존의 역전파 학습 인공신경망에 의한 실험결과와 비교하였다. 그리고 이진창, 한인구, 김명중(1996)은 보다 효과적인 기업신용평가를 위하여 통계적 방법인 MDA와 인공지능 방법인 인공신경망 모형을 휴리스틱한 방법으로 결합한 통합 모형을 제시하고 그 모형을 기존의 MDA와 인공신경망 방법에 의한 결과와 비교하였다.

이재식, 한재홍(1995)은 비재무정보를 입력변수에 반영함으로써 재무정보만을 이용한 모형에 비해 10% 정도 증가된 예측력을 보였다. 이재식, 차봉근(1996) 그리고 홍승현, 신경식(2003)은 부도예측을 위한 인공신경망 모형의 구축에 유전자 알고리즘을 사용하여 변수선정 과정에 유전자 알고리즘을 사용하였다. 저자들은 인공신경망 모형 구축에 있어서 유전자 알고리즘을 이용한 탐색을 변수 과정에 반영시킴으로써 최적의 입력변수군을 선정

14) 이재식, 한재홍, “인공신경망을 이용한 중소기업도산예측에 있어서의 비재무정보의 유용성 검증”, 한국지능정보시스템학회, Vol. 1, 1995, pp. 123-134.

하고 있다.

기존 연구의 대부분이 단일 인공신경망을 구축하여 예측력을 향상시키고자 하는 연구인데 반하여, 한인구 외 (1997), 신경식(2000)과 이견창 외 (1995, 1996)는 다수의 인공신경망 모형을 통합하여 상이한 계량 모형들을 통해서 얻어진 예측결과를 통합함으로써 예측력의 향상을 꾀하고 있다. 첫째로, 한인구 외 (1997)와 신경식(2000)은 다변량 판별분석 모형과 로짓 모형, 인공신경망 모형을 통해 추출된 결과값을 유전자 알고리즘을 사용하여 최적의 가중치로 가중평균함으로써 개별 모형보다 향상된 예측성과를 얻었다. 둘째로, 이견창(1995)은 첫 번째 연구와는 달리 구조적으로 비감독학습 인공신경망과 감독학습 인공신경망을 사용하는 하이브리드 방식의 의사결정 모형인 HYNEN(hybrid neural network) 모형을 제안하였다. 또한 이견창 외 (1996)는 통계적인 모형 중에서 가장 널리 활용되고 있는 MDA와 인공지능적인 방법으로서 널리 사용되고 있는 인공신경망 모형을 휴리스틱한 방법으로 결합하여 기존의 MDA 및 인공신경망 방법에 의한 결과와 비교하였다.

신경식(2000)은 통계적 분석 방법, 전문가의 의견, 유전자 알고리즘 등 다양한 변수군 선정 방법론을 활용하여 입력 변수군들을 선정하고, 각각의 변수군을 활용하여 모형을 구축하여 다음과 같은 방법으로 표현하였다.

$$OH_i = \text{Max}|0.5 - O_{ij}|$$
$$(i = 1, \dots, n, j = 1, \dots, m)$$

여기서 $O_{i,j}$ 는 어떤 기업(i)의 모형별 (j)출력값이며, OH_i 는 출력값($O_{i,j}$)과의 0.5와의 차이가 가장 큰 모형의 예측값이다. 이 모형을 통해 10%의

향상된 결과를 얻었다. 이러한 인공신경망 통합 모형 구축은 단일 모형을 사용하여 일반화를 시도함에 따라서 발생 가능한 한계를 극복하기 위한 목적을 가지고 있다. 이런 목적을 달성하기 위하여 세 가지 입력 변수군을 도입하여 이에 의하여 0과 1사이의 값을 각각 도출하고, 이들 중에 0혹은 1에 가장 가까운 결과를 보이는 예측결과를 선택함으로써 적중률을 높이고자 시도하였다¹⁵⁾.

부도예측분야 이외의 선행연구¹⁶⁾에서는 Gallant(1988)는 인공신경망을 이용하여 대출 신청자들의 신용 및 기타 정보를 바탕으로 대출승인 여부를 결정하는 방법을 제시하였다. 또한 Collins(1990)는 인공신경망을 이용하여 기업의 채권할인율을 예측하는 연구를 발표하였다. 그 밖의 많은 연구에서 인공신경망을 이용한 방법은 채권 할인율 및 수익성 예측에 있어서 통계적 방법 및 수학적 방법보다 월등히 우수함이 입증되고 있다.

최근 인공신경망을 이용하여 주식시장을 효과적으로 분석하려는 연구가 활발히 진행되고 있으며, 주식매매의 시점을 포착하기 위한 연구도 진행되고 있다. 또한, 다양한 경제 지표를 이용하여 S&P 500 지수의 변동을 예측하는 인공신경망 모형이 Fishman(1991)에 의해 개발되었다. Fishman은 인공신경망 모형이 동일한 경제지표를 이용한 전문가 보다 정확한 예측성과가 있다고 주장하고 있다.

4. 규칙유도기법(Rule Induction)

신경망 분석과 비교하여 규칙유도기법(Rule Induction)의 장점은 전체적

15) 추휘석, 민지경, 이인호. “다수의 인공신경망 모형을 통한 기업데이터의 분류 및 부도예측에 관한 연구”. 연세경영연구. Vol. 41, No. 2, 2004.

16) 김락상, “인공신경망을 이용한 기업 도산 예측”, 한국동서경제연구, Vol. 16, 2004, pp. 65-80.

인 과정에서 불필요한 요소들은 자동적으로 제거가 된다는 것이다. 이것은 사용자 및 분석자에게 데이터에 대해 좀 더 많은 정보를 주게 되고, 사용자가 신경망 기법을 사용할 때 필드를 선택할 수 있는 기준을 마련하여 좀 더 효율적인 신경망 분석에 도움을 줄 수 있다. 또한 규칙유도기법은 의사결정 나무 형식으로 되어 있어 예측 필드에 대해 영향력을 가지고 있는 필드들을 명확히 보여준다. 즉, Rule이 어떻게 작용하는지 이해하기 위해 Web 분석이나 히스토그램과 같은 방법을 사용할 필요가 없다는 것이다. 또한 신경망과는 달리 모형에 대한 민감도 분석이 필요 없다. 규칙유도기법에서 예측에 대한 가장 중요한 필드는 나무 구조에서 첫 단계에서 분기되는 필드이다.

규칙유도기법 시스템은 시스템과 사용자의 상호작용(interface)을 가능케 하는 자연어 처리 부문, 전문지식을 저장해 놓은 지식 베이스(knowledge base), 지식 베이스의 내용을 이용하여 사용자의 문제 해결을 도와주는 추론기관(inference engine)의 세 부분으로 구성되어 있다.

규칙유도기법 모형을 설정하는 방법은 모형의 결과를 browse하여 해석하고, Rule의 형식을 의사결정 나무 구조에서 Rule-set으로 변환시키는 과정을 거친다.

5. 베이지안 망(Bayesian Networks)

베이지안 망(Bayesian network)은 특정 분야의 영역 지식(domain knowledge)을 확률적으로 표현하는 대표적인 수단으로서, 변수(특징)들 간의 확률적 의존 관계(probabilistic dependency)를 나타내는 그래프와 각 변수별 조건부 확률들로 구성된다. 따라서 하나의 베이지안 망은 각 노드마다 하나의 조건부 확률표(conditional probability table, CPT)를 갖는 하나

의 비순환 유허그래프(directed acyclic graph)로서 $G=\langle N, A \rangle$, $B=\langle N, A, \Theta \rangle$ 으로 정의할 수 있다¹⁷⁾. 이 때 각 노드 N 은 하나의 영역 변수를, 각 아크 A 는 두 변수간의 확률적 의존성을 나타내며, Θ 는 조건부 확률들의 집합을 나타낸다. 일반적으로, 하나의 베이지안 망은 다른 노드들에 배정된 값들을 기초로 특정 노드가 가질 값에 대한 조건부 확률을 계산하는데 이용할 수 있다. 따라서 하나의 베이지안 망은 한 개체의 다른 특징들의 값이 주어졌을 때 분류 클래스 노드(classification node)의 확률분포(probability distribution)를 구해 줌으로써 개체들에 대한 하나의 자동 분류기(classifier)로 이용될 수 있다. 즉 하나의 데이터 집합으로부터 베이지안 망을 학습할 때 베이지안 망의 각 노드는 데이터 집합의 각 특징을, 각 아크는 특징들 간의 의존성을 표현하게 되며, 이렇게 학습된 베이지안 망을 기초로 분류 클래스를 확률적으로 예측할 수 있다.

베이지안 망 분류기는 한 개체 j 가 클래스 C_i 에 속할 확률을 계산함으로써 그 개체를 분류하는 방법이다 그러한 확률은 식 (1)과 같이 표현되며, 이때 개체 j 는 $A_i = V_i$ 형태로서 특징과 값의 쌍으로 표현된다.

$$P(C_i \mid A_1 = V_{1j} \ \& \ ... \ \& \ A_N = V_{Nj}) \quad \text{식 (1)}$$

그러나 베이지안 망에서 변수의 수가 많아지면 각 변수들 간의 관계에 해당하는 값들도 많아져서 계산량이 급격히 요구된다. 그래서 이를 현실의 문제에 적용하기 위해서는 제약조건을 여러 가지로 두어 계산량을 줄이게 되는데 이런 여러 가지의 제약조건 형태가 베이지안 망의 유형이 된다. 베이지안 망의 유형은 학습성과를 결정하는데 매우 중요한 요소이다. 제약조

17) Cheng, J., Bell, A. and Liu, W., "Learning Belief Networks from Data: An Information Theory Based Approach" , Proceedings of ACM CIKM-97, 1997.

건과 관련하여 대표적인 유형들을 살펴보면, 각 변수들이 다른 변수와의 관계는 무시하고 분류 클래스와의 관계만을 갖는 제약조건을 인정한 나이브 베이저안 망(Naive Bayesian Network, NBN)이 있다. 그리고 트리 구조로 확장한 나이브 베이저안 망(Tree Augmented Nave-Bayes, TAN)과 베이저안 망으로 확장한 나이브 베이저안 망(Bayesian Network Augmented Nave-Bayes, BAN)이 있는데, 이는 나이브 베이저안 망에서 제한된 조건인 변수들 간의 의존성을 추가한 형태이다. 분류 클래스별로 망을 달리 가질 수 있는 베이저안 다중망(Bayesian Multi-Net)과 분류 클래스를 별도로 두지 않는 형태인 일반 베이저안 망(general Bayesian Network, GBN)을 베이저안 망의 대표적 유형으로 들 수 있다¹⁸⁾.

6. Support Vector Machine(SVM)

SVM은 Vapnik(1995)에 의해 개발된 분류기법으로, 입력공간과 관련된 비선형문제를 고차원의 특징공간에서의 선형문제로 대응시켜 나타내기 때문에 수학적으로 분석하는 것이 수월하다¹⁹⁾. 또한 SVM은 조정해야 할 파라미터의 수가 많지 않아 비교적 간단하게 학습에 영향을 미치는 요인들을 규명할 수 있다. 그리고 구조적 위험을 최소화함으로써 과대적합 문제에서 벗어날 수 있으며, 볼록함수를 최소화하는 학습을 진행하기 때문에 전역최적해(globally optimal solution)를 구할 수 있다는 점에서 인공신경망보다 우월한 기계학습기법으로 주목 받고 있다.

최근 몇 년간 SVM을 사용한 다양한 연구가 진행되었다. SVM은 문서분

18) Cheng, J., Bell and A., Liu, W., "An Algorithm for Bayesian Belief Network Construction from Data", Proceedings of AI & STAT-97, pp.83-90, Florida, 1997.

19) Hearst, M.A., Dumais, S.T., Osman, E., Platt, j. and Scholkopf, B., "Support vector machines." IEEE Intelligent Systems, Vol. 13, No. 4, 1998, pp.18-28.

류, 영상인식, 문자인식 등에서 뛰어난 일반화 성능을 보여주었다²⁰⁾. 또한, SVM을 재무분야에 적용한 연구도 있는데, 주로 시계열 예측 및 분류에 관한 것이다²¹⁾. 본 연구와 가장 유사한 연구로는 SVM을 사용하여 채권 신용등급을 예측한 연구를 들 수 있다²²⁾. 이 연구들에서 SVM은 일반화에 있어서 인공신경망이나 판별분석 등의 다른 분류기법들과 비교하여 비슷하거나 더 우수한 성능을 나타낸 것으로 보고 되었다. 본 연구에서는 이러한 연구 배경을 토대로 채권신용등급문제와는 다른 재무적 특성을 지닌 부도 예측에 SVM을 적용하여 보기로 한다. 이를 위해 SVM에 대하여 간단히 설명하고자 한다.

SVM에서는 모형구축용 데이터들을 서로 다른 두 개의 클래스로 분류할 때 분류의 기준이 되는 분리 경계면(hyperplane)을 학습 알고리즘을 이용하여 찾는다²³⁾. 따라서, SVM은 입력벡터 x 를 고차원의 특징공간(high dimensional feature space)으로 사상(mapping) 시킨 후 두 클래스 사이의 마진(margin)을 최대화시키는 분리 경계면을 찾는 것을 목적으로 한다. 이러한 최대마진 분리 경계면(maximum margin hyperplane)은 두 클래스 사이의 거리를 최대로 분리시킨다. 이때 최대마진 분리 경계면에 가장 근접한 모형구축용 데이터를 서포트 벡터(support vector)라고 부른다. 선형분리문제에서, 독립변수가 3개인 경우 분리 경계면은 식 (1)과 같다.

20) Joachims, T., "Test Categorization with Support Vector Machines," Proceedings of the European Conference on Machine Learning(ECML), 10th European Conference on Machine Learning, 1998, pp.137-142.

21) Tay, F.E.H. and Cao, L.J., "Application of Support Vector Machines in Financial Time Series Forecasting," Omega, Vol. 29, 2001, pp.309-317.

22) Huang, Z., Chen, H., Hsu, C-J., Chen, W-H. and Wu, S., "Credit Rating Analysis with Support Vector Machine and Neural Networks: A Market Comparative Study," Decision Support Systems, Vol. 37, No. 4, 2004, pp.543-558.

23) 이수용, 이일병, "Fuzzy 이론과 SVM을 이용한 KOSPI 200 지수 패턴분류기," 한국증권학회 제4차 정기학술발표회, 2002, pp.787-809.

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 \quad \text{식 (1)}$$

여기서 y 는 출력값이고, x_i 는 변수값, 그리고 4개의 w_i 는 학습 알고리즘에 의해 학습된 가중치이다. 상기 식에서 가중치 w_i 는 분리 경계면을 결정하는 파라미터이다. 이때 최대마진 분리경계면은 서포트 벡터를 사용해서 식 (2)와 같이 나타낼 수 있다.

$$y = b + \sum \alpha_i y_i x(i) \cdot x \quad \text{식 (2)}$$

여기서, y_i 는 모형구축용 데이터 $x(i)$ 의 분류값이고, \cdot 는 내적(dot product)이다. 벡터 x 는 모형검증용 데이터를 나타내고, 벡터 $x(i)$ 는 서포트 벡터(최대마진 분리 경계면에 가장 근접한 모형구축용 데이터)를 나타낸다. 이 식에서, b 와 α_i 는 분리 경계면을 결정하는 파라미터이다. 서포트 벡터를 찾아내고, 파라미터 b 와 α_i 를 결정하는 것은 선형적으로 제약된 이차계획문제(linearly constrained quadratic programming)를 푸는 것과 같다.

앞에서 언급한 바와 같이, SVM은 입력변수를 고차원의 특징 공간으로 이동시킴으로써 비선형 분류문제를 선형모형으로 근사시킨다. 비선형 분류 문제에서 사용될 식 (2)의 고차원 버전은 식 (3)과 같이 간단하게 나타낼 수 있다.

$$y = b + \sum \alpha_i y_i K(x(i), x) \quad \text{식 (3)}$$

상기 식에서 함수 $K(x(i), x)$ 는 커널함수라고 정의된다. 커널함수는 원래 데이터를 고차원 공간으로 사상시킴으로써 특징공간 내에 선형으로 분리가능한 입력 데이터셋을 만든다. 이때 사용될 수 있는 커널함수는 여러가지가 있으며 어떤 커널함수를 선택하는 것이 바람직한가는 문제에 따라 상이하고, 이는 SVM을 적용하는데 있어서 가장 중요한 요소 중의 하나이다. 일반적으로 많이 사용되는 커널함수로는 다항식커널(polyomial kernel)과 가우시안 RBF(Gaussian radial basis function)를 들 수 있다.

가우시안 RBF :

$$K(x, y) = \exp\left(-\frac{1}{\delta^2}(x - y)^2\right) \quad \text{식 (4)}$$

다항식 커널 :

$$K(x, y) = (xy + 1)^d \quad \text{식 (5)}$$

여기서 d 는 다항식 커널의 차수이고, δ^2 은 가우시안 RBF 커널의 대역폭이다.

분리가능한 문제에 있어서 상기 식의 계수 α_i 의 하한은 0이다. 분리가 불가능한 문제에서 SVM은 계수 α_i 의 하한 이외에 상한 C 를 추가함으로써 일반화된 결과를 얻을 수 있다.

제 4 장 연구 설계

본 연구에서는 우리나라 제조업 1,888개의 기업을 중심으로 하여 기업부실예측모형의 정립과 분석을 다변량 판별분석, 로지스틱 회귀분석, 인공신경망, 연관규칙기법, 베이지안 망, SVM(Support Vector Machine) 모형 등 총 6가지의 단일모형과 단일모형을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형 등의 총 7가지 방법론을 실행하고 각 결과를 비교 검토하도록 하겠다.

제 1 절 자료 및 연구 설계

1. 표본기업 및 변수선정

부도예측모형을 구축하기 위해 본 연구에서는 기술신용보증기금이 보유하고 있는 944개의 건전기업과 944개의 부도기업 등 총 1,888개 기업의 재무비율 자료를 수집하였다. 건전기업의 자료는 자산규모 10억 이상 70억 이하의 국내 비외감 제조중공업 기업의 2002년도 재무자료를 기준으로 하였고, 이에 대응하는 부도기업의 자료 역시 자산규모 10억 이상 70억 이하의 국내 비외감 제조중공업 기업의 자료를 기준으로 하였다. 그러나 일반적으로 부도기업은 건전기업보다 매년 발생하는 자료 건수가 적으므로 부도기업의 경우 1999년부터 2002년까지의 자료를 사용하였다.

분석에 사용할 재무비율은 모두 [0, 1] 사이의 값을 가지도록 단위를 조정하였다. 이와 같이 자료를 정규화(normalize)하게 되면 분석에 사용되는 모든 재무비율의 분산이 동일한 범위 내에 있게 되므로 측정 단위에 따른 예측오차를 줄일 수 있게 된다²⁴⁾. 분석에 사용될 변수의 선정으로는 처음에 고려된 변수 83개의 변수 중에서 개별 Independent sample t-test를 거쳐 1차로 선정된 변수 54개(alpha = 0.05)를 선정하였고 Stepwise logistic regression을 거쳐 2차로 선정된 변수 11개를 최종분석에 사용할 재무비율로 선정하였다. 선정된 변수 11개 항목과 산출식을 <표 4-1>로 나타내었다.

변 수 명	산 출 식
금융비용대매출액	= 이자비용/매출액*100
매출액순이익율	= 순이익/매출액*100
매출액영업이익율	= 영업이익/매출액*100
총자산경상이익율	= 경상이익/총자산*100
유동부채대총자산	= 유동부채/총자산*100
유형자산증가율	= (당기유형자산/전기유형자산-1)*100
경영자산회전율	= 매출액/[총자산-(건설중인자산+ 투자자산)]
순금융비용	= 이자비용 - 이자수익
순운전자본대총자산	= (유동자산-유동부채)/총자산*100
유동자산증가율	= 당기말 유동자산/전기말 유동자산*100-100
자기자본경상이익률	= 경상이익/자기자본*100

<표 4-1> 선정된 재무비율 변수 목록

24) Peel, M. J., D. A. Peel and P. F. Pope, "Predicting corporate failure-some results for the UK corporate sector," Omega, Vol. 14, No. 1, 1986, pp.5-12.

모든 분석은 훈련용과 검증용의 두 가지 데이터 셋으로 구성되었으며 전체 데이터의 60%(1132/1888)는 훈련용(training) 데이터 셋으로 사용하고, 나머지 40%(756/1888)는 검증용(test)으로 사용하였다. 또한 보다 일반화된 연구결과를 얻기 위하여 본 연구에서는 상호검증방법(cross-validation method)을 사용하였다²⁵⁾. 따라서 본 연구에서는 총 5회에 걸친 상호검증방법을 실시하였다. 즉, Data Set 1, Data Set 2, Data Set 3, Data Set 4, Data Set 5 등의 데이터 셋 변환을 통한 상호검증방법을 실시하였다.

25) Weiss, S. and Kulikowski, C., "Computer Systems That Learn," Morgan Kaufmann Publishers, Inc., 1991.

제 2 절 모형의 추정 및 분석

1. 다변량 판별분석(MDA: Multiple Discriminant Analysis)

다변량 판별분석은 로지스틱 회귀분석과 마찬가지로 등간척도나 비율척도로 측정된 독립변수와 명목척도인 종속변수를 이용한 분석기법으로 선형적으로 정의된 두 개 이상의 집단들을 가장 잘 판별할 수 있는 둘 이상의 독립변수의 선형조합을 찾아내는 과정을 포함한다. 다변량 판별함수는 식 (1)과 같은 형태이다²⁶⁾.

$$Z = W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nX_n$$

식 (1)

여기서, Z 는 판별점수를 나타내며 $W_i(i = 1, 2, \dots, n)$ 는 판별 가중치를 나타낸다. 그리고 $X_i(i = 1, 2, \dots, n)$ 는 독립변수인 재무비율을 나타낸다. 각 기업에 대해 계산된 판별점수를 절사점(cut-off)과 비교하여 어떤 집단에 속하는지를 결정하게 된다.

26) 박정민, 김정재, 한인구, "Support Vector Machine을 이용한 기업부도예측," 경영정보학연구, 제15권 제2호, 2005, pp.52-62.

① 판별모형의 도출 (Data Set 1의 결과)

분류함수계수(discriminant function)

$$Z = -0.943 + 0.190(\text{금융비용대매출액}) - 0.124(\text{매출액순이익율}) + 0.099(\text{매출액영업이익율}) - 0.068(\text{총자본경상이익율}) + 0.009(\text{유동부채대총자산}) - 0.001(\text{유형자산증가율}) + 0.269(\text{경영자산회전율}) + 0.004(\text{순운전자본대총자산}) + 0.005(\text{유동자산증가율}) - 0.007(\text{자기자본경상이익률})$$

식 (2)

판별함수의 유의성 검증

Eigenvalue	정준상관계수	Wilk's Lambda	Chi-square	자유도	유의확률
0.484	0.571	0.674	444.452	10	0.000

<표 4-2> 다변량 판별분석의 유의성 검증

식 (2)에서 얻은 Z값은 상호검증방법을 통한 data set 1의 판별점수(discriminant score)이며, 변수 앞의 상수는 판별함수의 계수로서, 두 그룹을 가장 잘 구분할 수 있도록 판별분석 과정에서 추정되었다. 이러한 계수들의 크기와 부호는, 두 그룹으로 분류하는 과정에서, 측정변수들이 기여하고 있는 정도와 어떠한 방향을 가지고 있는지를 파악할 수 있게 만들어 준다.

다변량 판별분석의 유의성 검증은 <표 4-2>에서 보는 바와 같다.

고유값(Eigenvalue)이 0.484인데 이는 집단간 분산을 집단내 분산으로 나눈 값으로 값이 클수록 판별함수의 설명력이 증가한다. 정준상관계수는 0.571로 나타나고 있는데 이는 판별점수와 집단간의 관련성을 설명하는 것

으로 0~1의 값을 가지며 값이 클수록 판별력은 증가한다. 집단내 분산을 충분산으로 나눈 값인 Wilk's Lamda 값은 0.674, 이 값에 자유도를 고려한 chi-square는 444.452이며 유의확률은 5%의 유의수준에서 통계적으로 유의한 것으로 나타났다.

② 최적절사점 = 구분점(cutoff point)의 도출

판별함수를 추정 한 후에는 표본기업의 다변량 변수를 판별함수에 대입하여 표본기업의 Z평점을 산출한다. 그리고 분류정확도를 가장 높게 하는 최적절사점을 결정해야 한다. 판별함수가 아무리 잘 추정되었다 하더라도 최적절사점을 잘못 결정하면 분류오류가 커지게 되기 때문이다. 최적절사점(cutoff point)은 다음과 같다.

부도예측	함수
	1
건전	0.695
부도	- 0.695

<표 4-3> 함수의 집단중심점

<표 4-3>에서 보는 바와 같이 집단 중심점으로 건전집단은 0.695, 부도집단은 - 0.695이다.

이를 통해서 두 집단의 구분점은 식 (3)과 같이 계산할 수 있다.

$$\text{집단 구분점} = (N_2 \times GC_1) + (N_1 \times GC_2) / (N_1 + N_2) \quad \text{식(3)}$$

N_1 : 집단 1의 표본 수, N_2 : 집단 2의 표본수

GC_1 : 집단 1의 판별점수 집단평균

GC_2 : 집단 2의 판별점수 집단평균

$$\begin{aligned} \text{집단 구분점} &= (566 \times 0.695) + (566 \times -0.695) / (566 + 566) \\ &= 0 \end{aligned}$$

즉, 비부도집단과 부도집단의 구분점은 0가 된다.

각 case의 판별점수 값이 0보다 크면 건전집단, 작으면 부도집단에 속하는 것으로 판별한다.

③ 판별분석의 결과

판별분석의 결과에는 data set 1의 검증용 표본(756)을 이용하였다.

부도예측 분류결과		예측 소속집단		전체
		건전	부도	
빈도	건전	285	93	378
	부도	74	304	378
%	건전	75.4	24.6	100.0
	부도	19.6	80.4	100.0

<표 4-4> 판별분석의 결과

<표 4-4>에서 보는 바와 같이 검증용 표본을 이용한 결과가 원래의 집단 케이스 중 평균 77.90%가 올바르게 분류되었다.

실제 부도기업인 경우, 부도 기업을 맞춘 경우는 80.4%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우는 75.4% 이상의 적중률을 나타내었다. 그러므로 이 모형을 응용하고자 할 때에는 주로 부도기업을 추출해내고 예측할 때 사용할 경우 상당히 뛰어난 모형이 될 것이다. 다른 측면에서 보면, 만약 우리가 부도가 날 기업들을 알아내려는 의도가 있다면, 이 모형에서는 예측한 기업의 대략 80.4% 정도만이 올바르게 예측한 모형이 된다는 결론을 얻게 된다.

2. 로지스틱 회귀분석(Logistic Regression Analysis)

로지스틱 회귀분석은 비선형의 로지스틱 형태를 취하며 단지 2개의 값을 가지는 종속변수(건전, 부도)와 독립변수 사이의 인과관계를 밝히는 통계기법이다²⁷⁾. 로지스틱 회귀분석을 부도예측에 사용할 경우 기업의 설명변수의 관찰치벡터를 X_i 로 하고, 그 계수 β_i 를 추정한다며 기업의 부도확률은 로지스틱 함수에 의해 식 (4)와 같이 유도된다.

$$Y_i = \frac{1}{1 + \exp(-p)} \quad \text{식 (4)}$$

여기서, $P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$ 이다.

본 연구에서는 로지스틱 회귀분석 모형 구축을 위해 SPSS 12.0 프로그램을 사용하였고, Cutoff Point는 0.5를 기준으로 설정하였다.

27) Ohlson, J. A., "Financial ratios and probabilistic prediction of bankruptcy," Journal of Accounting Research, Vol. 18, No. 1, 1980, pp.109-131.

① 로지스틱 회귀모형의 도출 (Data Set 1의 결과)

모형에 추정된 계수 (B)

Logit 부도예측 = 1.146 - 0.370(금융비용대매출액) + 0.316(매출액순이익율) - 0.188(매출액영업이익율) + 0.075(총자산경상이익율) - 0.014(유동부채대총자산) + 0.002(유형자산증가율) - 0.347(경영자산회전율) + 0.005(순금융비용) - 0.006(순운전자본대총자산) - 0.007(유동자산증가율) + 0.010(자기자본경상이익률)

식 (5)

식 (5)의 로지스틱 회귀식은 data set 중에서 첫 번째(Data Set 1) 결과를 나타낸 것이다. 위에서 추정된 회귀식의 Logit점수를 구분점(Cutoff-point)과 비교하여 건전집단과 부도집단을 판별하게 되는 것이며 각 변수에 대한 계수들의 크기와 부호는 두 그룹으로 분류하는 과정에서, 측정변수들이 기여하고 있는 정도와 어떠한 방향을 가지고 있는지를 파악할 수 있게 만들어 준다.

② 로지스틱 회귀모형의 유의성 검증

변수명	계수(B)	S.E.	Wald	자유도	유의확률
금융비용대매출액	-0.370	0.051	52.340	1	0.000
매출액순이익율	0.316	0.056	31.512	1	0.000
매출액영업이익율	-0.188	0.034	29.958	1	0.000
총자산경상이익율	0.075	0.024	9.659	1	0.002
유동부채대총자산	-0.014	0.004	12.543	1	0.000
유형자산증가율	0.002	0.001	13.544	1	0.000
경영자산회전율	-0.347	0.108	10.386	1	0.001
순금융비용	0.005	0.001	12.690	1	0.000
순운전자본대총자산	-0.006	0.001	62.120	1	0.000
유동자산증가율	-0.007	0.001	37.789	1	0.000
자기자본경상이익률	0.010	0.004	6.312	1	0.012
Constant	1.146	0.310	13.652	1	0.000

<표 4-5> 로지스틱 회귀모형 변수들의 유의성 검증

<표 4-5>에서 보는 바와 같이 Cutoff값을 0.5로 설정하였으며 변수에 대한 Wald 통계량과 유의확률을 고려할 때 모든 계수의 추정치가 통계적으로 유의하였다.

즉, 부도예측과 11개의 재무비율 변수들과의 유의적인 관련이 있다는 가설을 검정한 결과 모든 변수가 부도예측모형과 유의적인 관련이 있는 것으로 나타났다.

③ 로지스틱 회귀분석의 결과

부도예측 분류결과		예측 소속집단		전체
		건전	부도	
빈도	건전	289	89	378
	부도	71	307	378
%	건전	76.5	23.5	100.0
	부도	18.8	81.2	100.0

<표 4-6> 로지스틱 회귀분석의 결과

<표 4-6>에서 보는 바와 같이 검증용 표본을 이용한 결과가 원래의 집단 케이스 중 평균 78.85%가 올바르게 분류되었다.

실제 부도기업인 경우, 부도 기업을 맞춘 경우는 81.2%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우는 76.5% 이상의 적중률을 나타내었다. 그러므로 이 모형을 응용하고자 할 때에는 주로 부도기업을 추출해내고 예측할 때 사용할 경우 상당히 뛰어난 모형이 될 것이다. 다른 측면에서 보면, 만약 우리가 부도가 날 기업들을 알아내려는 의도가 있다면, 이 모형에서는 예측한 기업의 대략 81.2% 정도만이 올바르게 예측한 모형이 된다는 결론을 얻게 된다.

3. 인공신경망(Artificial Neural Networks)

기업부도예측에 대한 인공신경망 구축과 평가를 위해 Clementine 8.1 프로그램을 사용하였으며 트레이닝용으로 분류된 데이터를 이용하여 신경망 모형을 구축한 후 테스트용 데이터에 적용하여 분류 예측정확도를 분석하였다. Clementine 8.1은 다층인식모형(Multi-Layer Perceptron : MLP)과 원형기준함수(Radial Basis Function : RBF)모형 등의 2가지 관리(supervised)학습하는 인공신경망 기법이 있는데 본 논문에서는 다층인식모형을 적용한다. 본 논문의 인공신경망의 구조는 역전파학습 알고리즘(Back-propagation algorithm)을 이용한 인공신경망을 사용한다.

① 신경망 부도 예측 구조 (Data Set1의 결과)

입력 레이어	11 뉴런
숨긴 레이어	5 뉴런
출력 레이어	1 뉴런
입력에 대한 상대적 중요도	
X127 (순운전자본/총자산)	0.72636
X13 (금융비용/매출액)	0.65534
X17 (매출액순이익율)	0.65452
X24 (총자산경상이익율)	0.60439
X18 (매출액영업이익율)	0.55372
X130 (유동자산증가율)	0.47374
X34 (유동부채/총자산)	0.28195
X103 (유형자산증가율)	0.27058
X111 (경영자산회전율)	0.26017
X135 (자기자본경상이익률)	0.24911
X126 (순금융비용)	0.21568

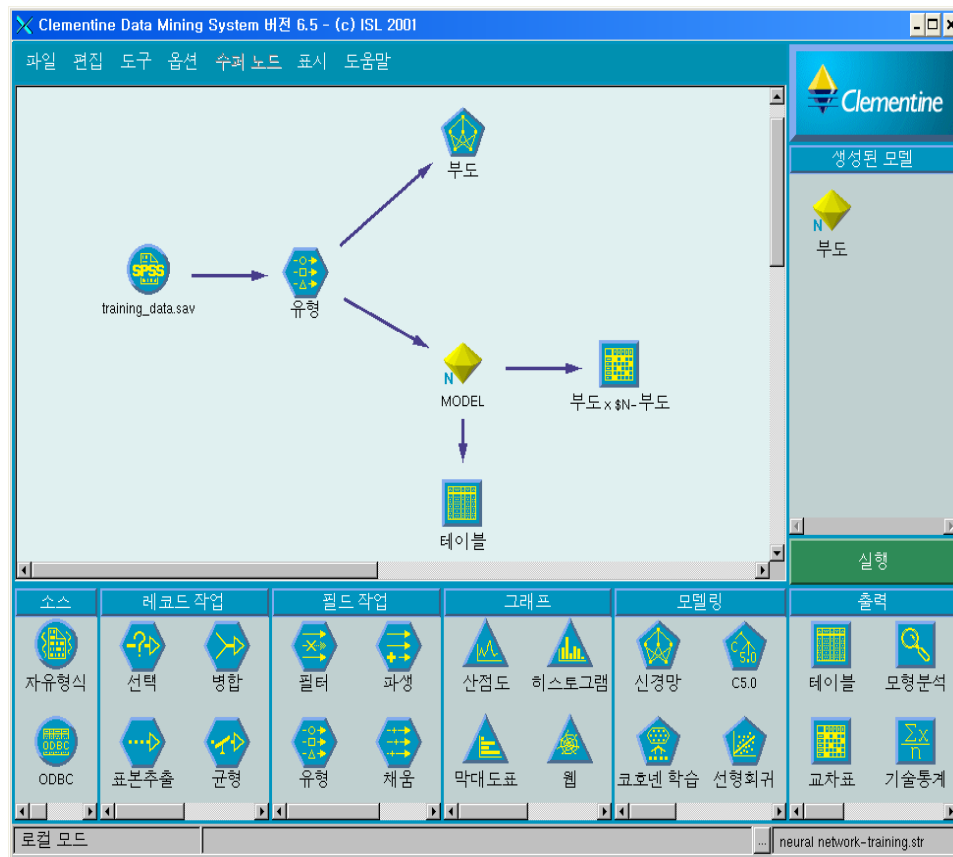
<표 4-7> 인공신경망의 모형

<표 4-7>에서 보는 바와 같이 인공신경망의 모형으로는 입력 레이어 11 뉴런, 숨긴 레이어 5뉴런, 출력 레이어 1뉴런으로 구성되어 있다. 입력 레이어는 본 논문에서 채택한 11개의 재무비율 변수를 말하며 출력 레이어 1 뉴런은 부도와 건전기업을 판별할 때 쓰이는 Output 뉴런이다. 마지막으로 숨긴 레이어는 인공신경망 모형 구축 시에 지정되는 것으로 1개의 숨긴 레

이어 층에 5개의 뉴런이 구성되어 있다는 것이다. 본 논문에서는 앞서 설명한 역전파학습 알고리즘을 이용한 인공신경망을 사용한다.

인공신경망 뉴런 설명 아래에는 입력 필드간의 중요도를 보여주는 항목들이 나타나는데 값이 큰 순서로 정렬이 되어져 있다. 이 값은 0.0과 1.0 사이에서 존재하며, 0.0 인 것은 별로 모형에 중요성을 가지지 못하는 입력 필드이고, 1.0 인 것은 가장 중요하여 전체를 설명할 수 있는 입력 필드이다. 따라서 본 논문에서는 인공신경망 모형을 구성하는데 가장 중요한 요소로는 X127(순운전자본대총자산), X13(금융비용대매출액), X17(매출액순이익율), X24(총자산경상이익율), X18(매출액영업이익율)순임을 알 수 있다.

② Clementine 8.1을 이용한 인공신경망 모형



<그림 4-1> 인공신경망의 모형

<그림 4-1>은 Clementine 8.1을 이용하여 인공신경망 모형 구축과정을 나타낸 것이다. 트레이닝용으로 분류된 데이터를 이용하여 신경망 모형을 구축한 후 테스트용 데이터에 적용하여 분류 예측정확도를 분석하였다.

③ 신경망의 결과

분류 예측 정확도 (%)

부도예측 분류결과		예측 소속집단		전체
		부도	건전	
%	부도	84.921	15.079	100.0
	건전	28.307	71.693	100.0

<표 4-8> 인공신경망의 결과

<표 4-8>에서 보는 바와 같이 검증용 표본을 이용한 결과가 원래의 집단 케이스 중 평균 78.31%가 올바르게 분류되었다.

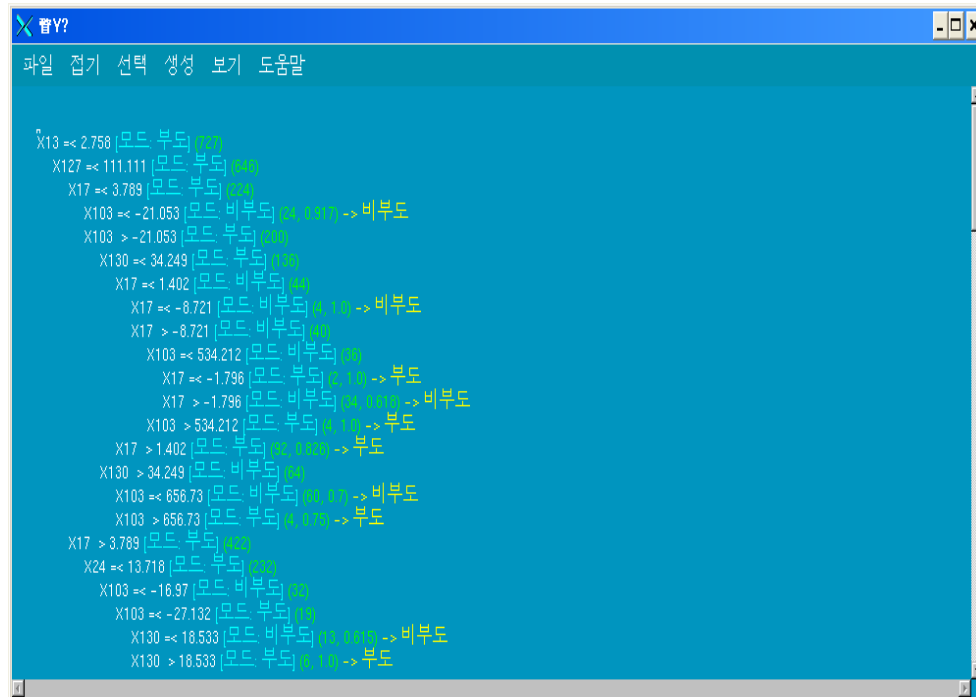
실제 부도기업인 경우, 부도 기업을 맞춘 경우는 85%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우는 72% 이상의 적중률을 나타내었다. 그러므로 이 모형을 응용하고자 할 때에는 주로 부도기업을 추출해내고 예측할 때 사용할 경우 상당히 뛰어난 모형이 될 것이다. 다른 측면에서 보면, 만약 우리가 부도가 날 기업들을 알아내려는 의도가 있다면, 이 모형에서는 예측한 기업의 대략 85% 정도만이 올바르게 예측한 모형이 된다는 결론을 얻게 된다.

4. 규칙유도기법(Rule Induction)

신경망 분석의 단점중의 하나는 블랙박스(black box)기법이라는 것이며 예측의 결과가 나오기까지의 과정과 그 구체적인 중간 모형을 알기가 어렵다는 것이다. 규칙유도기법(Rule Induction)은 바로 이런 신경망의 예측기능을 보완해 줄 수 있는 특징이 있다.

기업부도예측에 대한 규칙유도기법(Rule Induction) 구축과 평가를 위해 Clementine 8.1 프로그램을 사용하였다. Clementine 8.1은 C5.0 알고리즘과 Build Rule 알고리즘 등의 2가지의 규칙유도 알고리즘이 있는데 이 중에서 C5.0 알고리즘을 사용하였다. 훈련용으로 분류된 데이터를 이용하여 규칙유도기법(Rule Induction) 모형을 구축한 후 검증용 데이터에 적용하여 분류 예측정확도를 분석하였다.

① Generated Models 팔레트에 생성된 C5.0 Rule 노드 모형



<그림 4-2> 규칙유도기법의 모형

<그림 4-2>은 Clementine 8.1을 이용하여 규칙유도기법 결과를 의사결정나무 형식의 Tree 구조를 나타낸 것이다. 트레이닝용으로 분류된 데이터를 이용하여 규칙유도기법 모형을 구축한 후 테스트용 데이터에 적용하여 분류 예측정확도를 분석하였다.

② Rule-set 모형 (“IF.....THEN.....”형식)

Rule Induction 시스템에서 대부분의 지식은 가정-결과 형식(IF-THEN format)을 따르는 규칙의 형태로 저장되어 있다. Rule-set 모형 결과 규칙 대상이 부도인 경우 23개의 규칙, 규칙 대상이 건전인 경우 28개, 총 51개의 규칙이 도출되었으며 규칙이 부도인 경우 10가지 규칙을 <그림 3-4>로 나타내었다.

규칙 대상 부도:

규칙 #1 대상 부도:

조건식 $X_{13} \leq 2.758$
및 $X_{127} \leq 111.111$
및 $X_{17} > -8.721$
및 $X_{17} \leq -1.796$
및 $X_{103} > -21.053$
및 $X_{103} \leq 534.212$
및 $X_{130} \leq 34.249$
일치할 경우 \rightarrow 부도 (2, 1.0)

규칙 #2 대상 부도:

조건식 $X_{13} \leq 2.758$
및 $X_{127} \leq 111.111$
및 $X_{17} > -8.721$
및 $X_{17} \leq 1.402$
및 $X_{103} > 534.212$
및 $X_{130} \leq 34.249$
일치할 경우 \rightarrow 부도 (4, 1.0)

규칙 #3 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 > 1.402$

및 $X17 \leq 3.789$

및 $X103 > -21.053$

및 $X130 \leq 34.249$

일치할 경우 -> 부도 (92, 0.826)

규칙 #4 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 \leq 3.789$

및 $X103 > 656.73$

및 $X130 > 34.249$

일치할 경우 -> 부도 (4, 0.75)

규칙 #5 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 > 3.789$

및 $X24 \leq 13.718$

및 $X103 \leq -27.132$

및 $X130 > 18.533$

일치할 경우 -> 부도 (6, 1.0)

규칙 #6 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 > 3.789$

및 $X24 \leq 13.718$

및 $X103 > -16.97$

및 $X126 \leq 8$

일치할 경우 -> 부도 (105, 0.667)

규칙 #7 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 > 3.789$

및 $X24 \leq 13.718$

및 $X103 > -16.97$

및 $X126 > 8$

및 $X130 \leq 100$

및 $X111 \leq 1.731$

일치할 경우 -> 부도 (59, 1.0)

규칙 #8 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 16.021$

및 $X17 > 3.789$

및 $X24 \leq 10.458$

및 $X103 > -16.97$

및 $X126 > 8$

및 $X130 \leq 100$

및 $X111 > 1.731$

및 $X34 \leq 80.967$

일치할 경우 -> 부도 (4, 1.0)

규칙 #9 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 111.111$

및 $X17 > 3.789$

및 $X24 > 10.458$

및 $X24 \leq 13.718$

및 $X103 > -16.97$

및 $X126 > 8$

및 $X130 \leq 100$

및 $X111 > 1.731$

일치할 경우 -> 부도 (8, 1.0)

규칙 #10 대상 부도:

조건식 $X13 \leq 2.758$

및 $X127 \leq 35.593$

및 $X17 > 3.789$

및 $X24 \leq 13.718$

및 $X103 > -16.97$

및 $X126 > 8$

및 $X130 > 108.671$

일치할 경우 \rightarrow 부도 (9, 0.889)

<그림 4-3> Rule-set 모형

<그림 4-3>에서 나타난 Rule-set 모형 중 규칙 10을 설명하면 X13(금융비용대매출액)이 2.758보다 작거나 같고, X127(순운전자본대총자산)이 35.593보다 작거나 같으며, X17(매출액순이익율)이 3.789보다 크고, X24(총자산경상이익율)이 13.718보다 작거나 같고, X103(유형자산증가율)이 -16.97보다 크고, X126(순금융비용)이 8보다 크고, X130(유동자산증가율)이 108.671보다 크다는 등의 위 7가지 조건이 모두 만족한다면 이 기업들은 부도기업이라는 것이다. 부도결과 옆에 나타난 수치 중에서 9는 위 규칙을 만족시켰던 기업이 총 9개라는 것을 말해주고 있으며 0.889는 위 규칙의 신뢰도(confidence)를 의미한다. 즉, 이 규칙에 속한 사람이 맞게 분류된 비율을 나타내는 것이다.

기업부도예측에 가장 중요한 필드는 나무 구조에서 첫 단계에서 분기되는 필드이다. <그림 4-3> Rule-set 모형에서 나타난 것과 같이 위의 모형에서 기업부도예측에 가장 중요한 변수로는 X13(금융비용대매출액)으로 나타났다.

③ 규칙유도기법의 결과 - Data Set 1

부도예측 분류결과		예측 소속집단		전체
		부도	건전	
%	부도	72.636	28.042	100.0
	건전	23.016	76.984	100.0

<표 4-9> 규칙유도기법의 결과

<표 4-9>의 결과에서 나타난 것과 같이 Data Set 1의 규칙유도기법 결과 평균 예측정확도는 74.81%를 나타내었다. 즉, 실제 부도기업인 경우 부도 기업을 맞춘 경우는 약 73%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우도 77% 이상의 적중률을 나타내었다. 부도기업을 맞춘 경우에는 신경망 모형(85%의 적중률)에 비해서는 예측력이 떨어지나 건전기업을 맞춘 경우는 신경망 모형(72%의 적중률)에 비해서는 우수한 예측력을 보이고 있다. 그러나 총 분류정확도는 74.81%로 신경망에 비해 예측력이 떨어진다.

5. 베이지안 망(Bayesian Networks)

우리는 확률적 방법으로 학습이 가능한 베이지안 망을 재무분야, 그 중에서도 부도기업 예측과 관련된 특징들 간의 관련성을 표현하고 분석하는데 이용하고자 한다. 일반적으로 베이지안 망은 전산학적인 측면에서 여러 관심 대상들 간의 관련성에 대해 확률적으로 표현함으로써 그 구조 및 의존성을 잘 표현할 뿐만 아니라 수학적인 면에서도 견고성이 입증된 통계적인 도구라고 할 수 있다. 그리고 출력값에 대하여 단언적인 결과를 내는 타 알고리즘들과는 달리 확률을 기반으로 하여 입력값 및 출력값들을 조건적인 의존관계로 표현함으로써 누구나 쉽게 그 구조와 관계를 파악할 수 있는 장점이 있다. 또한 해당 영역 지식을 입력값으로 사용할 수 있어 이미 많은 연구에서 그 우수성을 입증한 도구라고 할 수 있다. 이러한 이유에서 베이지안 망은 우리가 적용하려는 부도예측 분야에 그 특징이 맞는 것으로 판단이 된다.

베이지안 망은 특정 분야의 영역 지식(domain knowledge)을 확률적으로 표현하는 대표적인 수단으로서, 변수(특징)들 간의 확률적 의존관계(probabilistic dependency)를 나타내는 그래프와 각 변수별 조건부 확률들로 구성된다²⁸⁾. 베이지안 망 학습을 위해서는 조건부 독립성 기반의 알고리즘인 Jie Cheng의 CBL 알고리즘과 그것을 구현한 Belief Network Power Constructor 1.0 프로그램을 이용하였다²⁹⁾.

28) Heckerman, D., Geiger, D., Chickering, D. M., "Learning Bayesian Networks: the Combination of Knowledge and Statistical Data," Technical Report MSR-TR-94, Microsoft Research, 1994.

29) Cheng, J., "Belief Network(BN) Power Constructor System," Available at <http://www.cs.ualberta.ca/~jcheng/bnpc.htm>, 1998.

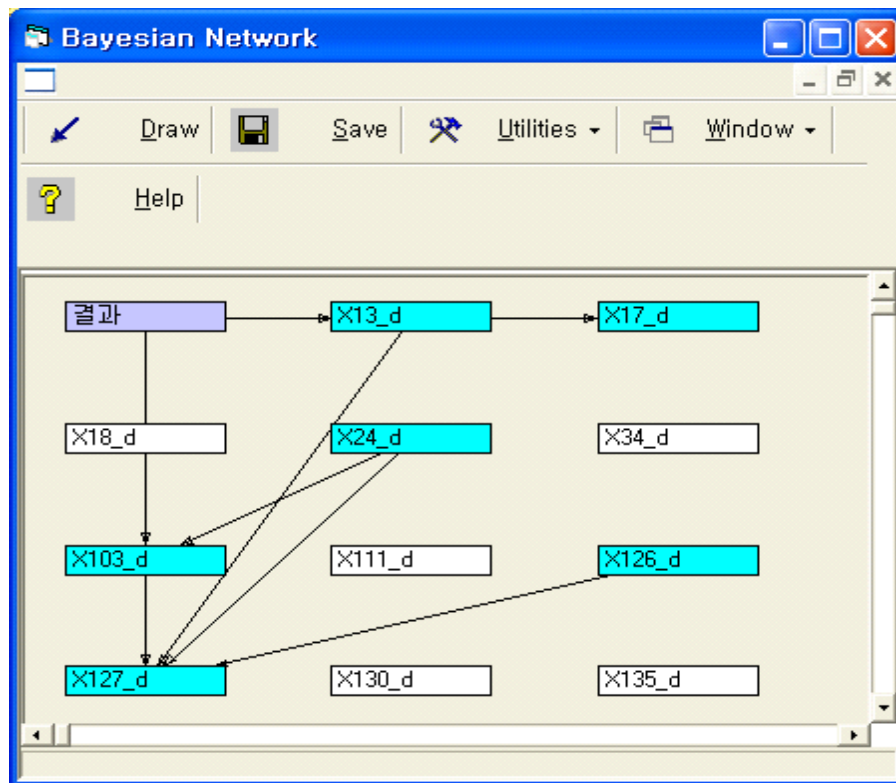
① 베이지안 망 모형

베이지안 망 모형 구축을 위해 사용할 Belief Network Power Predictor 프로그램은 <표 3-5>에서 보는 바와 같이 5가지의 규칙으로 구성되어 있다.

Belief Network Power Constructor 모형
(1) Complete ordering
(2) Partial ordering
(3) Cause & effects
(4) Forbidden links
(5) Root & leaf nodes

<표 4-10> Belief Network Power Constructor 1.0의 5가지 모형규칙

Belief Network Power Constructor 1.0의 5가지 모형규칙에서 분석결과 1번의 Complete ordering과 3번의 Cause & effects와의 관계에서만 규칙이 생성되었으며 그 결과는 <그림 4-4>와 같다.



<그림 4-4> 베이زي안 망 모형의 결과

베이زي안 망 모형의 결과는 <그림 4-4>에서 보는 바와 같이 X13(금융비용대매출액), X17(매출액순이익율), X18(매출액영업이익율), X24(총자산경상이익율), X103(유형자산증가율), X126(순금융비용), X127(순운전자본대총자산) 등 총 7개의 변수가 베이زي안 망 모형의 영향력 있는 변수로 선정되었다. 위의 선정된 변수에서 X127(순운전자본대총자산)변수는 X13(금융비용대매출액), X24(총자산경상이익율), X126(순금융비용), X103(유형자산증가율) 등의 변수들에 의해 영향을 받고 있다는 것을 알 수 있다.

② 베이지안 망의 결과 - Data Set 1

부도예측 분류결과		예측 소속집단		전체
		부도	건전	
빈도	부도	278	100	378
	건전	123	255	378
%	부도	73.5	26.5	100.0
	건전	32.5	67.5	100.0

<표 4-11> 베이지안 망의 결과

<표 4-11>에서 보는 바와 같이 검증용 표본을 이용한 결과가 원래의 집단 케이스 중 평균 70.50%가 올바르게 분류되었다.

실제 부도기업인 경우, 부도 기업을 맞춘 경우는 73.5%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우는 67.5% 이상의 적중률을 나타내었다. 그러므로 이 모형을 응용하고자 할 때에는 주로 부도기업을 추출해내고 예측할 때 사용할 경우 상당히 뛰어난 모형이 될 것이다. 다른 측면에서 보면, 만약 우리가 부도가 날 기업들을 알아내려는 의도가 있다면, 이 모형에서는 예측한 기업의 대략 74% 정도만이 올바르게 예측한 모형이 된다는 결론을 얻게 된다.

6. SVM(Support Vector Machine)

SVM 모형의 구축에 있어서 어떤 커널함수를 사용하느냐는 가장 중요한 문제 중의 하나이다. 본 연구에서는 SVM의 커널함수로서 가장 널리 사용되는 가우시안 RBF함수를 사용하였다. Tay and Cao(2001)에 의하면 SVM의 성능에 있어서 커널함수의 상한 C 와 커널 파라미터 γ 가 가장 중요한 역할을 한다고 보고 되었다. 따라서 적절한 상한 C 와 커널 파라미터 γ 를 선정하기 위해, 선행연구에서 SVM의 파라미터에 대해 제시된 일반적인 가이드를 따라 본 연구에서도 일정 범위 내에서 다양한 값을 대입하여 다양한 모형을 생성시켰다.

본 연구에서는 기업부도예측에 대한 Support Vector Machine(SVM) 구축과 평가를 위해 Libsvm-2.8 프로그램을 사용하였다.

① Support Vector Machine(SVM) 모형

SVM에서는 학습과정이 이루어지기 전에 사용자가 직접 파라미터 값을 결정해야 하며 파라미터 값에 따라 SVM의 성능은 달라진다. 이렇게 미리 결정해야 하는 파라미터는 학습과정에서 마진폭과 분류 오류 사이의 타협점(trade-off)을 찾아주는 오류 패널티(penalty) 변수 C 값과 비선형 SVM에 적용되는 커널함수의 파라미터이다. 본 연구에서는 기업부도예측 모형에 적합한 C 값과 커널함수의 파라미터 값을 결정하고, 이를 바탕으로 RBF 커널함수를 적용한 비선형 SVM의 분류 성능을 비교 분석하였다.

본 연구에서는 SVM 커널함수의 종류로 가우시안 RBF 함수를 사용하였으며 γ 의 범위는 0.1~1.0사이로 오류 패널티 변수 C 값은 10, 20, 40, 60,

80, 100 등으로 구분하여 최적의 모형을 찾고자 하였다.

γ	C	Set 1	Set 2	Set 3	Set 4	Set 5	평균
0.1	10	78.84	78.70	80.16	77.65	76.32	78.33
	20	78.17	78.57	80.29	77.25	76.46	78.15
	40	78.57	78.04	79.50	76.85	76.06	77.80
	60	79.10	77.78	79.76	77.25	76.59	78.10
	80	79.37	77.65	79.63	76.46	76.06	77.83
	100	79.37	77.25	79.63	76.32	75.66	77.65
0.2	10	78.84	77.12	79.76	76.46	76.98	77.83
	20	79.23	77.38	79.23	76.19	76.59	77.72
	40	79.10	77.51	81.08	76.06	76.32	78.01
	60	78.97	77.78	81.08	76.32	75.79	77.99
	80	78.31	77.51	80.82	76.59	76.59	77.96
	100	78.17	77.38	80.29	76.72	76.72	77.86
0.3	10	79.37	78.17	80.42	76.46	76.19	78.12
	20	79.10	78.04	81.98	77.69	76.59	78.68
	40	77.91	76.59	80.69	77.12	76.59	77.78
	60	77.12	76.72	80.55	76.98	76.46	77.57
	80	76.72	76.46	80.03	76.72	76.06	77.20
	100	76.59	76.72	80.29	77.38	75.66	77.33
0.4	10	79.23	77.65	81.22	76.46	76.46	78.20
	20	78.44	76.98	80.55	76.98	76.59	77.91
	40	76.85	76.19	81.16	77.12	75.79	77.42
	60	76.06	76.72	79.89	77.51	75.79	77.19
	80	75.93	76.59	78.97	77.12	75.53	76.83
	100	76.19	76.59	78.70	76.85	75	76.67
0.5	10	78.97	78.17	80.82	76.85	76.19	78.20
	20	77.38	76.19	80.03	77.12	76.59	77.46
	40	76.19	76.46	79.76	77.25	75.53	77.04
	60	76.46	76.59	79.36	76.98	74.47	76.77
	80	74.87	76.32	78.17	76.85	74.34	76.11
	100	74.60	76.06	77.51	76.46	73.94	75.71

0.6	10	78.17	76.98	80.42	77.38	77.12	78.01
	20	76.46	76.46	80.16	77.12	76.32	77.30
	40	76.19	76.32	78.97	77.51	74.87	76.77
	60	75.26	76.19	78.04	76.46	73.41	75.87
	80	74.74	76.19	76.85	76.72	73.94	75.69
	100	75.26	76.32	75.79	75.79	73.55	75.34
0.7	10	78.04	76.72	79.89	77.12	76.32	77.62
	20	76.59	76.32	79.50	77.78	75.40	77.12
	40	75.13	76.59	78.04	76.32	74.07	76.03
	60	74.34	76.72	76.59	75.93	73.94	75.50
	80	74.34	75.66	76.19	75.93	73.68	75.16
	100	73.28	75.93	75	75.4	72.88	74.50
0.8	10	76.72	76.46	79.76	77.38	75.93	77.25
	20	75.93	76.59	79.36	77.25	75	76.83
	40	73.94	76.98	76.98	76.32	74.21	75.69
	60	73.68	75.93	76.32	75.53	74.07	75.11
	80	73.15	75.53	75.13	75.53	73.02	74.47
	100	73.55	74.87	74.07	75.53	72.49	74.10
0.9	10	76.59	76.59	79.23	77.91	75.66	77.20
	20	74.74	76.59	78.04	77.12	74.74	76.25
	40	74.07	76.85	76.72	75.79	74.47	75.58
	60	73.02	75.40	75.40	75.79	73.15	74.55
	80	73.02	75.40	75.53	75.53	72.62	74.42
	100	72.75	74.74	74.60	75.13	72.49	73.94
1.0	10	76.46	76.72	79.50	77.65	75.66	77.20
	20	73.68	76.85	77.12	76.72	75	75.87
	40	73.28	75.93	75.93	75.40	73.81	74.87
	60	73.15	75.40	75.40	75.13	73.15	74.45
	80	72.75	74.34	74.47	75.40	71.83	73.76
	100	71.83	73.81	73.94	75.13	71.69	73.28

<표 4-12> 가우시안 RBF 함수의 SVM 결과

커널	C	γ	예측정확도
RBF	20	0.3	78.68%

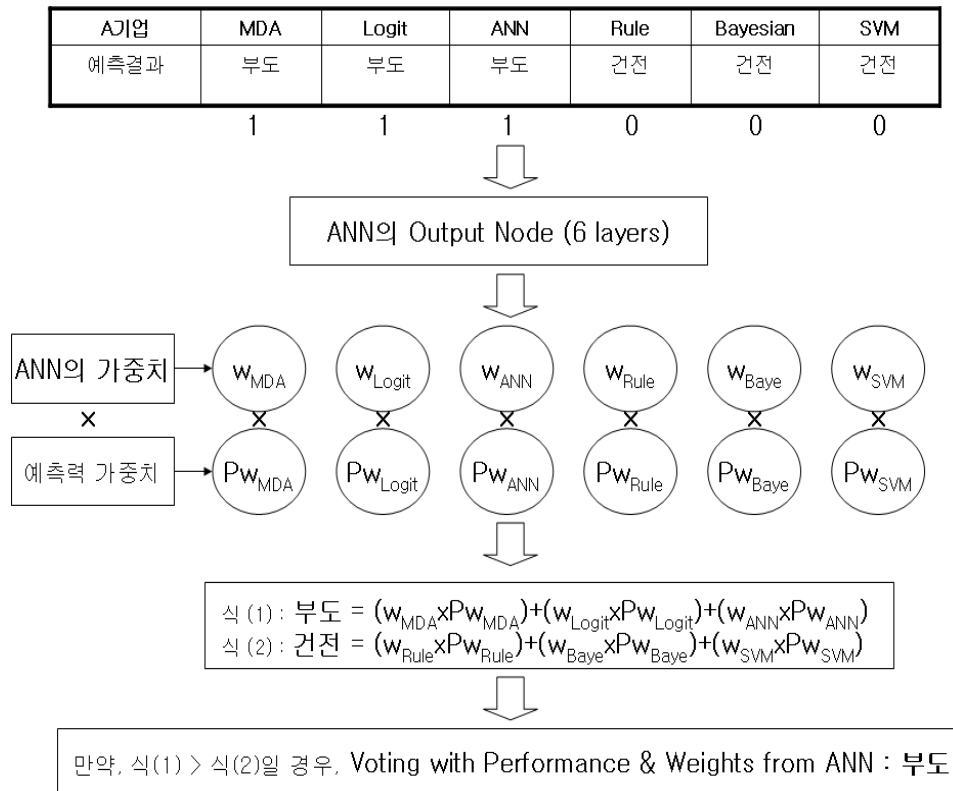
<표 4-13> SVM의 분석 결과

<표 4-12>에서 보는 바와 같이 교차타당성을 이용한 격자탐색 기법을 도입하여 최적의 C 와 γ 의 값을 도출하고, 이를 이용하여 SVM 분석을 수행하였다. 11개 재무비율 변수로 구성된 훈련용 데이터를 대상으로 격자탐색을 수행한 결과, C 와 γ 의 값이 각각 20, 0.3일 때 교차타당성의 예측성 결과가 가장 우수한 것으로 나타났으며, 이때의 예측성과는 78.68%이다.

7. Voting with Performance & Weights from ANN (WP-ANN)

WP-ANN 방법론은 앞에서 구축한 단일모형 6가지, 즉 통계적 방법과 인공지능 방법을 결합한 통합모형을 말한다.

WP-ANN 모형은 6가지 연구모형에서의 인공신경망 가중치(w)와 예측 정확도 가중치(Pw)를 모두 곱한 후에 이 값을 부도와 건전의 두 가지 예측결과 값에 대입한다. 두 가지 예측결과 값 중에서 큰 값을 부도예측결과로 채택하는 방법론이 WP-ANN 통합방법론이다.



<그림 4-5> WP-ANN 통합방법론

<그림 4-5>에서 보는 바와 같이 A기업을 부도기업으로 예측한 방법론은 다변량 판별분석(MDA), 로지스틱 회귀분석(Logit), 인공신경망(ANN)이며, A기업을 건전기업으로 예측한 방법론은 규칙유도기법(Rule), 베이저안망(Bayesian), Support Vector Machine(SVM)이다. 각 방법론에 해당하는 인공신경망 가중치(w)와 예측정확도 가중치(Pw)를 모두 곱한 후에 이 값을 부도와 건전의 두 가지 예측결과 값에 대입한다.

부도기업으로 예상한 방법론(MDA, Logit, ANN)의 식 (1)은 $(w_{MDA} \times Pw_{MDA}) + (w_{Logit} \times Pw_{Logit}) + (w_{ANN} \times Pw_{ANN})$ 이다.

건전기업으로 예상한 방법론(Rule, Bayesian, SVM)의 식 (2)는 $(w_{Rule} \times Pw_{Rule}) + (w_{Bayesian} \times Pw_{Bayesian}) + (w_{SVM} \times Pw_{SVM})$ 이다.

따라서, 식 (1)과 식 (2)를 비교하여 큰 값을 부도예측결과로 채택한다.

위의 <그림 4-5>에서는 식 (1)의 부도예측값이 식 (2)의 건전예측값보다 크므로 Voting with Performance & Weights from ANN(WP-ANN)의 결과는 식 (1)의 부도로 결정된다.

① WP-ANN 통합방법론의 결과

부도예측 분류결과		예측 소속집단		전체
		건전	부도	
빈도	건전	282	89	378
	부도	71	325	378
%	건전	74.6	23.5	100.0
	부도	18.8	86.0	100.0

<표 4-14> WP-ANN 통합방법론의 결과

<표 4-14>에서 보는 바와 같이 검증용 표본을 이용한 결과가 원래의 집단 케이스 중 평균 80.30%가 올바르게 분류되었다.

실제 부도기업인 경우, 부도 기업을 맞춘 경우는 86%의 적중률을 나타내었고, 실제 건전기업인 경우 건전기업을 맞춘 경우는 74.6% 이상의 적중률을 나타내었다. 그러므로 이 모형을 응용하고자 할 때에는 주로 부도기업을 추출해내고 예측할 때 사용할 경우 상당히 뛰어난 모형이 될 것이다. 다른 측면에서 보면, 만약 우리가 부도가 날 기업들을 알아내려는 의도가 있다면, 이 모형에서는 예측한 기업의 대략 86% 정도만이 올바르게 예측한 모형이 된다는 결론을 얻게 된다.

제 5 장 연구 결과

제 1 절 요약

1. 예측 정확도의 비교

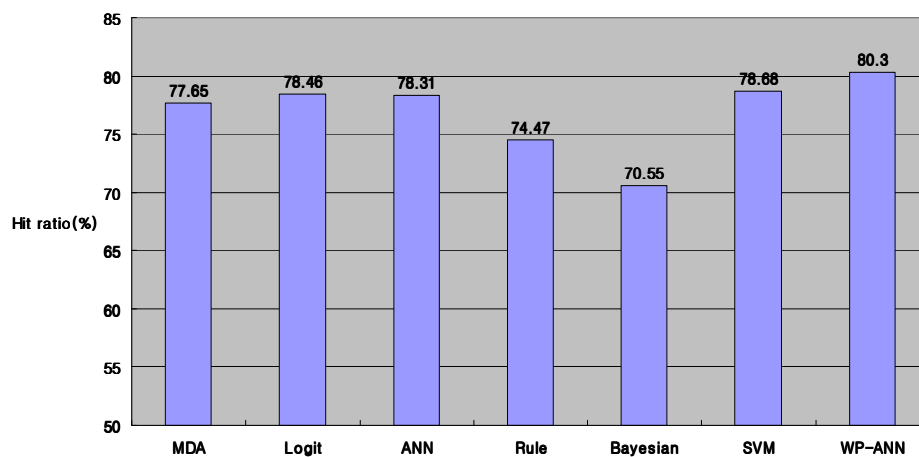
종 류	Set 1	Set 2	Set 3	Set 4	Set 5	Average	Standard deviation
MDA	77.90	77.25	80.95	76.65	75.50	77.65	2.05
Logit	78.85	78.10	81.45	77.70	76.20	78.46	1.93
ANN	78.31	77.88	81.59	77.75	76.00	78.31	2.04
Rule	74.81	74.58	77.23	73.69	72.03	74.47	1.89
Bayesian	70.50	69.86	73.82	70.09	68.51	70.55	1.97
SVM	79.10	78.04	81.98	77.69	76.59	78.68	2.05
WP-ANN	80.30	80.02	83.41	79.90	77.82	80.30	2.00

<표 5-1> 전체결과

<표 5-1>과 <그림 5-1>의 결과에서 나타난 것과 같이 평균예측력은 통합모형인 Voting with Performance & Weights from ANN(WP-ANN) 방

법론이 전체 방법론과 비교하여 가장 예측력이 높았으며 5가지 데이터 셋 별 결과로도 항상 WP-ANN 방법론의 결과가 가장 우수하다는 것을 알 수 있다.

단일모형 중에서는 Support Vector Machine(SVM) 방법론의 평균예측력이 가장 우수하였으며 베이지안 망(Bayesian) 방법론은 가장 낮은 평균예측력을 나타내었다.



<그림 5-1> 모형별 평균예측력 비교

평균예측력을 모든 방법론과 비교하여 순서대로 나열하면 Voting with Performance & Weights from ANN(WP-ANN), Support Vector Machine(SVM), 로지스틱 회귀분석(Logit), 인공신경망(ANN), 다변량 관별 분석(MDA), 규칙유도기법(Rule), 베이지안 망(Bayesian) 등의 순으로 분류 예측 정확도를 보이고 있다.

결론적으로, 본 연구에서는 Voting with Performance & Weights from ANN(WP-ANN) 방법론을 기업부도예측 통합모형에 추천하고자 한다.

제 6 장 결론 및 한계

제 1 절 연구의 결론

본 연구에서는 보다 효과적인 기업부도예측을 위하여, 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하였다. 이를 위하여 본 연구에서는 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석(Multiple Discriminant Analysis), 로지스틱 회귀분석(Logistic Regression)과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공신경망(Artificial Neural Networks), 규칙유도기법(Rule Induction), 베이저안 망(Bayesian Networks), Support Vector Machine(SVM) 등의 6가지 단일모형과 6가지 단일모형을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형을 제시하였다.

실험결과 본 연구에서 제안한 통합모형인 Voting with Performance & Weights from ANN(WP-ANN)모형은 다변량 판별분석, 로지스틱 회귀분석, 인공신경망, 규칙유도기법, 베이저안 망, SVM 등의 단일모형과 비교한 결과 가장 예측정확성이 우수한 것으로 나타났다. 따라서 본 연구를 통해 부도예측에 있어서 통합모형인 Voting with Performance & Weights from ANN(WP-ANN)방법론이 기존의 모형들에 비해 우수한 예측정확성을 나타냄을 알 수 있었다.

제 2 절 연구의 한계와 후속 연구과제

재무비율 등의 재무정보를 이용하는 기업부도예측모형의 경우, 각 모형이 활용하는 통계적 기법이나 사용하는 변수와 관계없이 공통적으로 재무정보의 적시성과 투명성이 제한될 수밖에 없다는 한계점을 갖고 있다. 즉, 재무제표를 통해 파악되는 재무정보는 분식결산 등을 통해 왜곡되어 있을 가능성이 상존하며, 현재 시점의 정보가 아닌 과거 특정시점의 정보이고, 이 또한 수시로 파악하기는 곤란하다는 단점이 있다는 것이다. 따라서 이를 이용하여 미래의 부도 가능성을 예측하기 위해서는 모든 정보가 정확하고 과거의 추세가 미래에도 안정적으로 지속된다는 가정이 필요하다는 것이다.

이러한 근본적인 한계 외에도 몇 가지 기술적인 문제가 있을 수 있다.

첫째, 독립변수로 활용되는 재무비율을 산출할 수 없는 기업이 있다는 점이다. 예를 들어 자본잠식 기업의 경우 부채비율 등의 재무비율은 산출하기가 곤란하다. 따라서 이러한 기업들의 경우 동일한 기준에서 도산 가능성을 예측하기 곤란하다.

둘째, 산업별로 재무비율의 의미가 상이할 수 있다는 점이다. 예를 들어 일반 제조업체의 재고자산 회전율과 도·소매업 또는 건설업의 재고자산 회전율은 동일한 기준에서 상대비교가 곤란하다.

셋째, 기업별로 또는 시점별로 회계기준이 상이할 수 있다는 점이다. 예를 들어 국내기업의 경우 감가상각시 정액법과 정률법을 모두 인정하고 있는데 적용하는 방법에 따라 상당한 금액 차이를 초래할 수 있다.

지금까지 대부분의 부도예측모형은 발생주의에 기초한 재무비율(accrual-based financial ratios)에 근거하여 부도예측을 하였으나 이 비율

들은 영업활동 중 경제적 자원의 취득과 사용에만 초점을 두기 때문에 현재 기업의 평가에 있어 중요성이 더해가고 있는 유동성을 나타내는 현금흐름을 정확히 나타내지는 못한다. 또한 재무비율 상호간의 높은 상관관계에 의한 다중공선성(multicollinearity)의 가능성과 상이한 회계처리방식에 의한 재무제표의 다양성, 비상장기업들의 경우 재무제표의 신뢰성문제 등으로 인해 부도예측모형의 예측력을 떨어뜨리고 있다. 특히 본 연구에서 분석대상으로 하고 있는 중소기업의 경우 재무정보보다 비재무정보가 기업의 신용위험을 파악하는데 중요한 역할을 할 수 있으므로 향후 연구에서는 비재무정보를 포함한 통합부도예측모형의 개발이 필요할 것으로 생각된다.

참 고 문 헌

<국내문헌>

- 김락상, “인공신경망을 이용한 기업 도산 예측,” *한국동서경제연구*, Vol. 16, 2004, pp. 65-80.
- 김진백, 이준석, “인공지능 기법을 이용한 중소기업 부도예측에 있어서 현금흐름변수의 유용성 검증,” *대한경영학회지*, 제26권, 2000, pp. 229-250.
- 국찬표, 정완호, “기업 도산 예측에 관한 연구 : 추가정보를 이용하여,” *재무연구* 제17권 제1호, 2002.
- 박정민, 김경재, 한인구, “Support Vector Machine을 이용한 기업부도예측,” *경영정보학연구*, 제15권 제2호, 2005, pp. 52-62.
- 이건창, 김명중, 김혁, “기업도산예측을 위한 귀납적 학습지원 인공신경망 접근방법 : MDA, 귀납적 학습방법, 인공신경망 모형과의 성과비교,” *한국경영학회*, 제23권 제2호, 1994, pp. 109-144.
- 이건창, 한인구, 김명중, “통계적 모형과 인공지능 모형을 결합한 기업신용평가 모형에 관한 연구,” *한국경영과학회지*, 제21권 제1호, 1996, pp. 81-100.
- 이수용, 이일병, “Fuzzy 이론과 SVM을 이용한 KOSPI 200 지수 패턴분류기,” *한국증권학회 제4차 정기학술발표회*, 2002, pp. 787-809.
- 이은주, “신용위험평가 모형의 실증연구 : 한국 상장기업의 KMV모형 적용을 중심으로,” 연세대학교 석사학위 논문, 2000.
- 이원흠, 이한득, 박상수, “현금흐름형 도산예측 모형과 옵션모델형 도산예상확률의 실증연구,” *증권학회지* 제27권, 2000.

이재식, 한재홍, “인공신경망을 이용한 중소기업도산예측에 있어서의 비재무정보의 유용성 검증,” *한국지능정보시스템학회*, Vol. 1, 1995, pp. 123-134.

추휘석, 민지경, 이인호. “다수의 인공신경망 모형을 통한 기업데이터의 분류 및 부도예측에 관한 연구,” *연세경영연구*. Vol. 41, No. 2, 2004.

<국외 문헌>

- Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, 1968, pp. 589-609.
- Beaver, W., "Financial ratios as predictors of failure," *Journal of Accounting Research*, Vol. 4, 1966, pp. 71-102.
- Collins, E., S. Ghosh, and C. Scofield, "An Application of a Multiple Neural Network Learning System to Emulation of Mortgage Underwriting Judgements," Working Paper. Nestor, Inc. 1989.
- Cheng, J., "Belief Network(BN) Power Constructor System," Available at <http://www.cs.ualberta.ca/~jcheng/bnpc.htm>, 1998.
- Cheng, J., Bell, A. and Liu, W., "Learning Belief Networks from Data : An Information Theory Based Approach," *Proceedings of ACM CIKM-97*, 1997.
- Cheng, J., Bell, A. and Liu, W., "An Algorithm for Bayesian Belief Network Construction from Data," *Proceedings of AI & STAT-97*, pp. 83-90, Florida, 1997.
- Fishman, M., D. Barr, and W. Loick, "Using Neural Networks in Market Analysis," *Technical Analysis of Stock and Commodities*, 1991, pp. 18-25.
- Gallant, S.I., "Connectionist Expert Systems," *Communications of the ACM* : pp. 152-169.
- Hearst, M.A., Dumais, S.T., Osman, E., Platt, j. and Scholkopf, B., "Support vector machines." *IEEE Intelligent Systems*, Vol. 13, No. 4,

- 1998, pp. 18-28.
- Heckerman, D., Geiger, D., Chickering, D. M., "Learning Bayesian Networks: the Combination of Knowledge and Statistical Data," *Technical Report MSR-TR-94*, Microsoft Research, 1994.
- Huang, Z., Chen, H., Hsu, C-J., Chen, W-H. and Wu, S., "Credit Rating Analysis with Support Vector Machine and Neural Networks: A Market Comparative Study," *Decision Support Systems*, Vol. 37, No. 4, 2004, pp. 543-558.
- Jensen, F. V., *An Introduction to Bayesian Network*, New York, Springer-Verlag, 1996.
- Joachims, T., "Test Categorization with Support Vector Machines," *Proceedings of the European Conference on Machine Learning(ECML), 10th European Conference on Machine Learning*, 1998, pp. 137-142.
- Ohlson, J. A., "Financial ratios and probabilistic prediction of bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, 1980, pp. 109-131.
- Peel, M. J., D. A. Peel and P. F. Pope, "Predicting corporate failure-some results for the UK corporate sector," *Omega*, Vol. 14, No. 1, 1986, pp. 5-12.
- Tay, F.E.H. and Cao, L.J., "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, Vol. 29, 2001, pp. 309-317.
- West, D., "Neural Network Credit Scoring Models," *Computers & Operations Research*, Vol. 27, 2000, pp. 1131-1152.
- Weiss, S. and Kulikowski, C., "Computer Systems That Learn," *Morgan Kaufmann Publishers, Inc.*, 1991.