

기업부도예측을 위한 통합알고리즘

배재권^a, 김진화^b

^a 서강대학교 경영학과

서울 마포구 신수동 1번지

Tel: +82-2-705-8860, Fax: +82-2-705-8519, E-mail: baejaekwon@sogang.ac.kr

^b 서강대학교 경영학과

서울 마포구 신수동 1번지

Tel: +82-2-705-8860, Fax: +82-2-705-8519, E-mail: jinhwakim@sogang.ac.kr

Abstract

본 연구에서는 보다 효과적인 기업부도예측을 위하여, 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하였다. 이를 위하여 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석, 로지스틱 회귀분석과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공신경망, 규칙유도기법, 베이저안 망의 5가지 방법론을 통합한 *Voting with Performance & Weights from ANN(WP-ANN)* 통합모형을 제시하였다. 실험결과, 본 연구에서 제안한 WP-ANN 통합모형은 다변량 판별분석, 로지스틱 회귀분석, 인공신경망, 규칙유도기법, 베이저안 망 등의 단일모형과 비교한 결과 가장 예측정확성이 우수한 것으로 나타났다. 따라서 본 연구를 통해 기업부도예측에 있어서 WP-ANN 통합모형이 기존의 모형들에 비해 우수한 예측정확성을 나타냄을 알 수 있었다.

Keywords:

기업부도예측; 통합모형; WP-ANN; 인공신경망

1. 서론

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, 매출감소, 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 부도위험도가 급증하고 있고 기업간 경쟁이 날로 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 해당기업의 경영자나 투자자, 그리고 채권자 등과 같은 이해관계자 모두가 조기에 기업 경영의 정상화를 위해 노력하거나 또는 정상적이고 합리적인 청산절차를 진행함으로써 경제적인 손실을 최소화할 수 있을 것이다[2].

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 인공지능망, 규칙유도기법, 베이저안 망 등과 같은 인공지능 관련 기법을 기업부도예측 문제에 적용하는 연구가 활발히 소개되었고 그 예측력 또한 기존의 전통적인 통계기법보다 더 높은 성과를 나타내었다[6].

본 연구에서는 보다 효과적인 기업부도예측을 위하여, 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하고자 한다. 이를 위하여 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석, 로지스틱 회귀분석과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공지능망, 규칙유도기법, 베이저안 망의 5가지 단일모형을 통합한 Voting with Performance & Weights from ANN(WP-ANN)모형을 제시한다. 이러한 통합모형의 성과를 증명하기 위해 기술신용보증기금이 보유하고 있는 1,888개 기업의 재무비율 자료를 기초로 분석하였고 그 결과를 통계적 방법과 인공지능 방법의 기존 단일모형과 성과비교를 통하여 그 유용성을 검증해보고자 한다.

본 연구는 다음과 같이 구성되어 있다. 제 2 장에서는 단일모형과 통합모형을 이용한 부도예측에 관련된 선행연구를 검토한다. 제 3 장에서는 본 연구에서 사용될 변수선정 작업과 분석절차 등을 설명한다. 제 4 장에서는 단일모형과 통합모형의 6가지 기업부도예측모형을 구축한다. 제 5 장에서는 검증데이터의 결과를 모아 단일모형과 통합모형의 예측력 성과비교를 통하여 그 유용성을 검증해본다. 마지막으로 제 6 장에서는 결론과 향후 연구방향에 대해 논의한다.

2. 이론적고찰

2.1 단일모형을 이용한 부도예측

어떻게 하면 기업부도를 좀더 정확하게 예측할 수 있을가에 대한 생각은 은행과 투자 기업의 주된 관심사중의 하나였다. 최근 몇 년 사이에 정교한 부도예측을 위해 많은 연구자들이 여러 가지 예측 모형을 제시하여 왔다.

1960년대에 다양한 분야의 연구자들이 본격적으로 기업부도예측모형을 연구하기 시작하였다. 전통적인 부도예측 모형에서 사용된 통계적 기법을 활용한 연구는 Beaver[12] 이후에 본격적으로 이루어졌는데, Beaver[12]는 단일 재무비율에 의하여 예측능력에 대한 분석을 시행하였다. 이후 Altman[10]은 단일 재무비율 분석이 아닌 다변량 판별분석을 이용하여 기존의 개별적으로 관찰되던 재무비율을 종합하고 단순화하여 부도예측을 시행하였다. 이후 다변량 판별분석을 통해 보다 향상된 부도예측 모형을

구축하고자 하는 많은 연구가 진행되었고, 이외에도 Ohlson[21]의 로짓 분석(Logit Analysis)과 프로빗 분석(Probit Analysis), 다중 회귀분석(Multiple Regression Analysis) 등의 다양한 통계방법론이 제시되었다.

통계적 기법도입 이후 인공지능망, 규칙유도기법, 베이저안 망 등 인공지능 분야의 연구 결과에 의한 다양한 소프트 컴퓨팅 기법(Soft Computing Technology)이 기업부도예측에 적용되어 전통적인 방법론보다 향상된 연구 결과를 가져오바 있다. 이러한 소프트 컴퓨팅 기법의 성공적인 적용은 통계적 모형이나 수리적 모형과 같은 전통적인 방법론이 동태적 복잡성을 띠고 있으며 급변하는 경영환경에의 적용에서 한계를 가지기 때문에, 이를 극복하기 위한 대안적 방법으로서의 유용성에 기인한다[7, 27].

2.2 통합모형을 이용한 부도예측

최근의 연구들은 전통적인 통계기법과 인공지능 기법의 성과를 비교, 분석하는 연구에서 나아가 통합방법론을 통해 모형의 예측력을 향상시키기 위한 방안들을 제시하고 있다. 이에 관련된 연구로 이진창 등[5, 6]의 연구에서는 인공지능망과 다변량 판별분석, 귀납적 학습방법, SOFM(self organizing feature map) 등의 기법을 통합한 귀납적 학습지원 인공지능망을 제시하였고 실증분석결과 향상된 예측성고를 나타내었다. 추취석 등[7]의 연구와 Shin and Lee[24]의 연구에서는 다수의 인공지능망 모형을 통합한 부도예측모형을 제시하였다. 다수의 신경망 모형의 결과에 따라 데이터를 분류하고 재분류된 데이터를 학습시켜 보다 나은 데이터의 패턴을 신경망에 적용하였다. 실험결과 다수의 신경망 모형을 통해 데이터를 분류 및 재학습시킨 결과 신경망 예측치와 실제 부도 사이의 일치 여부가 크게 개선되었다.

퍼지 기법은 경영 문제에도 두루 적용되어 왔으며 특히 금융공학 문제 및 경영 의사결정분야에 주로 활용되고 있다. 또한 퍼지 이론은 신용평가 문제에도 성공적으로 적용되고 있는데, 특히 사례기반추론 시스템과의 혼합적 적용에 의하여 기존의 연구에 비해 향상된 결과를 보여준 바 있다. 김경재와 한인구[1]의 연구에서는 퍼지신경망을 이용한 기업부실예측모형을 제안하였다. 이것은 기존 신경망에 퍼지집합의 개념을 적용하여 신경망 학습에 사용될 자료를 퍼지화하고 이를 신경망에 학습시키는 것이다. 퍼지신경망을 기업부도예측에 적용한 결과 기존의 신경망보다 우월한 예측성고를 나타내었다.

또한, 유전자 알고리즘이 기업부도예측에 적용된 연구가 많이 진행된 바 있는데, 이들 연구 중에서

기업부도예측 문제에 다양한 분류 기법들을 통합적으로 적용하기 위하여 유전자 알고리즘을 사용한다. Shin and Lee[25]에서는 유전자 알고리즘을 패턴인식과 학습이 뛰어난 인공신경망의 가중치를 훈련시키거나 아키텍처를 설정하는데 통합하여 사용하는 Neuro-Genetic 신경망 모형을 제시하였다. 외국의 연구로는 Anandarajan[11]이 유전자 알고리즘에 기반한 인공신경망 모형을 개발하였으며 이 모형을 역전파 인공신경망, 다변량 판별분석 모형과 예측성과를 비교하여 분석하였다. 또한 Chen[13]과 Abdelwashed[9]의 연구에서도 유전자 알고리즘에 기반한 인공신경망 모형을 제시하였다. Pendharkar[23]은 유전자 알고리즘에 기반한 새로운 Threshold-varying 인공신경망을 제시하였으며 이 모형을 역전파 인공신경망, 다변량 판별분석 모형 등과 예측성과를 상호 비교하여 분석하였다.

민재형, 이영찬의 연구[3]에서는 자료포괄분석(data envelopment analysis:DEA)을 신용평점모형 개발에 도입하여 실용성이 높은 신용평점화 방법론을 제안하였다. 건전과 부도 여부에 관한 사전적인 정보가 요구되는 기존의 분석방법(다변량 판별분석, 로지스틱 회귀분석, 인공신경망)과는 달리 DEA를 이용한 신용평점모형은 고객기업의 사후적인 정보만으로도 신용평점을 산출할 수 있다는 장점이 있다. 그러나 신용평점모형은 아직까지 국내 금융기관에 실제 적용된 사례가 없는 탐색적 접근방법으로서, 기업의 사후적 정보만을 이용한 신용평점모형이기 때문에 다른 부도예측모형에 비해 판별력이 상대적으로 낮은 단점이 있다.

그러나 위 모형들은 전체적 통합이 아닌 부분적 통합 모형이라고 볼 수 있다. 따라서 보다 효과적인 기업부도예측을 위해 전체적인 통합형 기업부도예측 모형이 필요하다. 통합모형은 단일모형에 의해 추출된 결과들을 통합하여 하나의 결론을 도출할 수 있다는 점에서 부도예측 분야에 유용하게 쓰일 수 있을 것이다.

3. 연구방법

3.1 자료수집과 사전처리

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 비외국 제조중공업 기업의 2002년도 재무자료를 기준으로 하였고, 이에 대응하는 부도기업의 자료 역시 자산규모 10억 이상 70억 이하의 국내 비외국 제조중공업 기업의 자료를 기준으로 하였다. 그러나 일반적으로 부도기업은 건전기업보다 매년 발생하는

자료 건수가 적으므로 부도기업의 경우 1999년부터 2002년까지의 자료를 사용하였다.

본 연구에서 부도예측모형을 구축하기 위해 모형에 사용할 데이터로는 회계데이터로부터 파생된 재무비율을 사용하였다. 모형에 사용될 변수선정과 관련하여 먼저 총 83개의 재무비율을 전처리하였다. 전처리과정은 먼저 이상치 제거를 위하여 각 재무비율별 분포의 양측 1%의 데이터를 제거하고, 결측치는 각 비율의 평균값으로 대체하였다.

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 된다[22]. 분석에 사용할 재무비율의 선정은 두 단계로 이루어졌다. 먼저, 단일변량분석의 과정인 개별 독립 t검정을 거쳐 1차로 선정된 변수 54개의 건전 또는 부도기업의 분류에 유의한 재무비율들을 가려내었다($p < 0.01$). 단일변량분석을 통해 선택된 변수들에 대해 다시 다변량분석의 과정으로 단계별 로지스틱 회귀분석을 거쳐 2차로 선정된 변수 11개를 최종분석에 사용할 재무비율로 선정하였다($p < 0.05$). 선정된 변수 11개 항목과 산출식을 <표 1>로 나타내었다.

표 1 - 선정된 재무비율 변수 목록

변수명	산출식
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
자기자본경상이익율	= 경상이익/자기자본*100

모든 분석은 훈련용과 검증용의 두 가지 데이터 셋으로 구성되었으며 전체 데이터(1,888)의 60%는

훈련용으로 나머지 40%는 검증용으로 사용하였다. 보다 일반화된 연구결과를 얻기 위하여 본 연구에서는 상호검증방법을 사용하였다[26]. 따라서 총 5회에 걸친 상호검증방법을 실시하였다.

3.2 분석절차

본 연구는 <그림 1>과 같은 분석절차에 따라 진행된다.

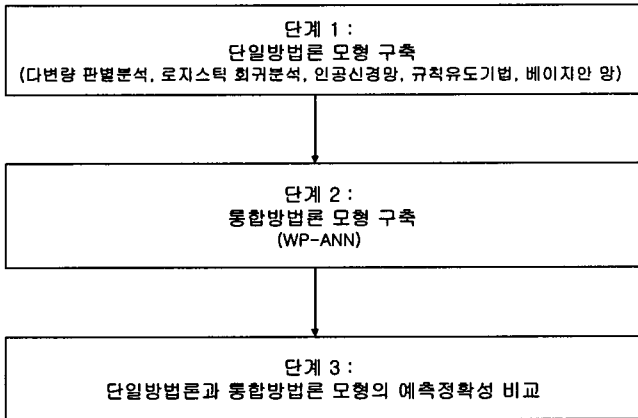


그림 1 - 분석절차

단계1에서는 다변량 판별분석, 로지스틱 회귀분석의 통계적 방법과 인공신경망, 규칙유도기법, 베이지안 망의 인공지능방법 등 5가지 단일모형을 구축한다. 단계2에서는 위 5가지 방법론을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형을 구축한다. 단계3에서는 단계1과 단계2에서 구축한 단일모형과 통합모형을 가지고 모형의 예측정확성을 비교한다.

4. 연구모형

4.1 다변량 판별분석 모형

다변량 판별분석은 사전에 정해진 집단(부도와 건전)을 판별해내는 선형판별함수를 도출하기 위한 통계적 기법이다. 선형판별함수는 집단내 분산대비 집단간 분산비율을 최대로 하는 통계적인 의사결정규칙을 생성하게 되는데, 이를 식으로 나타내면 식 (1)과 같은 형태이다[4].

$$Z = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (1)$$

여기서 Z값을 판별점수라 하고, w_i 는 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. 즉, 판별점수인 Z값이 일정한 판별점(Cut-off point)을 넘으면 건전기업으로, 판별점 이하인 경우에는 부도기업으로 판단한다.

다변량 판별분석은 독립변수가 다변량 정규분포를 따르고, 각 집단의 공분산행렬이 동일할 때는 유용하지만 부도기업의 경우 정규성에 대한 가정이 위배되는 경우가 많고, 집단별 공분산이 동일하다는 가정도 위배되는 경우가 많다. 특히, 독립변수간에 다중공선성이 존재할 경우 단계별 분석을 적용하게 되면 심각한 오류가 발생할 가능성이 높다[17].

본 연구에서는 다변량 판별분석을 위해 SPSS 12.0 프로그램을 사용하여 수행하였다.

4.2 로지스틱 회귀분석 모형

로지스틱 회귀분석은 비선형의 로지스틱 형태를 취하며 단지 2개의 값을 가지는 종속변수(건전, 부도)와 독립변수 사이의 인과관계를 밝히는 통계기법이다[21]. 로지스틱 회귀분석을 부도예측에 사용할 경우 기업의 설명변수의 관찰치벡터를 X_i 로 하고, 그 계수 β_i 를 추정한다면 기업의 부도확률은 로지스틱 함수에 의해 식 (2)와 같이 유도된다.

$$Y_i = \frac{1}{1 + \exp(-p)} \quad (2)$$

여기서, $p = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_iX_i$ 이다.

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

4.3 인공신경망 모형

인공지능의 한 분야인 인공신경망은 인간두뇌의 휴리스틱한 문제해결방법을 모형화한 것으로서 그 학습능력과 추론능력이 매우 뛰어난 것으로 알려져 있다. 인공신경망의 구조는 여러 가지 형태가 있으나, 가장 일반적으로 많이 쓰이는 형태는 관리학습(supervised learning)에 알맞은 다층 전향구조 인공신경망이다. 이는 입력층, 은닉층, 출력층의 삼층구조를 이루며 각 층마다 다수의 뉴런 또는 노드, 즉 처리단위를 소유하고 있다. 서로 다른 층에 존재하는 처리단위는 서로 연결되어 있으며 그러한 연결강도는 연결가중치(interconnection weight)로 계산된다.

일반적으로 신경망의 성능에 영향을 미치는 요인으로는 은닉층 수, 은닉노드 수, 학습 회수 등이 있는데, 어떤 값이 최적인지에 대한 일반적인 규칙은 없다. 다만, 분류 문제를 포함한 대부분의 문제에서 한 개의 은닉층으로도 만족할만한 결과를

얻을 수 있다는 선행 연구[19]를 토대로 본 연구에서도 은닉층이 하나인 3층 퍼셉트론을 사용하였다.

부도예측에 대한 인공지능망 구축과 평가를 위해 Clementine 8.1 프로그램을 사용하였으며 데이터 셋은 훈련용과 시험용, 그리고 검증용으로 구분하였다. 전체 데이터(1,888개)의 30%는 훈련용, 30%는 시험용, 40%는 검증용의 용도로 사용하여 분류 예측 정확도를 분석하였다. 인공지능망의 경우 학습과정에서 과대적합이 발생할 가능성이 크기 때문에 시험용 데이터를 사용하여 학습과정이 적절히 이루어졌는가를 확인하게 된다.

4.4 규칙유도기법 모형

신경망 분석과 비교하여 규칙유도기법의 장점은 전체적인 과정에서 불필요한 요소들은 자동적으로 제거가 된다는 것이다. 이것은 사용자 및 분석자에게 데이터에 대해 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 의사결정나무 형식으로 되어 있어 예측 필드에 대해 영향력을 가지고 있는 필드들을 명확히 보여준다. 즉, 규칙이 어떻게 작용하는지 이해하기 위해 웹 분석이나 히스토그램과 같은 방법을 사용할 필요가 없다는 것이다. 규칙유도기법에서 예측에 대한 가장 중요한 필드는 나무 구조 첫 단계에서 분기되는 필드이다. 규칙유도기법시스템은 시스템과 사용자의 상호작용을 가능케 하는 자연어 처리 부분, 전문지식을 저장해 놓은 지식베이스, 지식베이스의 내용을 이용하여 사용자의 문제 해결을 도와주는 추론기관의 세 부분으로 구성되어 있다[8].

부도예측에 대한 규칙유도기법 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 모형을 구축한 후 검증용 데이터에 적용하여 분류 예측 정확도를 분석하였다.

4.5 베이지안 망 모형

베이지안 망은 특정 분야의 영역 지식을 확률적으로 표현하는 대표적인 수단으로서, 변수(특징)들 간의 확률적 의존관계(probabilistic dependency)를 나타내는 그래프와 각 변수별 조건부 확률들로 구성된다[18, 20]. 따라서 하나의 베이지안 망은 각 노드마다 하나의 조건부 확률표(conditional probability table, CPT)를 갖는 하나의 비순환 유향그래프(directed acyclic graph)로서 $G=<N, A>$,

$B=<N, A, \Theta>$ 으로 정의할 수 있다. 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[16].

부도예측에 대한 베이지안 망 구축과 평가를 위해 조건부 독립성 기반의 알고리즘인 Jie Cheng의 CBL 알고리즘과 그것을 구현한 Belief Network Power Constructor 1.0 프로그램을 이용하였다. Jie Cheng의 Belief Network Power Constructor 1.0 프로그램은 모두 3개의 세부 모듈로 구성되어 있다. 첫 번째 모듈에서는 분석 대상 데이터의 사전처리 준비를 하는 Data Preprocessor 단계로 데이터 분석의 전처리를 위해 특징을 축소하거나 필요한 경우 데이터의 정제, 변환, 이산화의 작업이 이루어진다. 두 번째 모듈인 Power Constructor에서는 분석 대상 데이터의 사전지식을 미리 입력하여 탐색공간을 줄일 수 있으며 사전지식을 이용하여 실험대상인 변수들간의 원인과 결과관계와 우선순위를 정할 수 있다[14, 15, 16]. 본 연구에서는 훈련용 데이터를 이용하여 구축된 베이지안 망 분류기의 학습 모형을 검증용 데이터에 적용하여 분류 예측 정확도를 분석하였다.

4.6 WP-ANN 모형

Voting with Performance & Weights from ANN(WP-ANN) 방법론은 5가지 단일모형의 예측결과를 입력변수로 한 인공지능망에서의 각 모형에 대한 가중치와 5가지 단일모형의 예측정확도를 가중치로 이용한 모형이다. 위 두 개의 가중치를 모두 곱한 후에 이 값을 부도와 건전의 예측결과 값에 대입하고 두 가지 예측결과 값 중에서 큰 값을 부도예측결과로 채택하는 방법론이 WP-ANN이다.

<그림 2>에서 보는 바와 같이 A기업을 부도기업으로 예측한 방법론은 다변량 판별분석, 로지스틱 회귀분석, 인공지능망이며 A기업을 건전기업을 예측한 방법론은 규칙유도기법, 베이지안 망이다. 인공지능망에서 나온 가중치와 예측정확도 가중치를 모두 곱한 후에 이 값을 부도와 건전의 예측결과 값에 대입한다. 부도기업으로 예상한 방법론의 식

(1)은 $(w_{MDA} \times Pw_{MDA}) + (w_{Logit} \times Pw_{Logit}) + (w_{ANN} \times Pw_{ANN})$ 이며 건전기업으로 예상한 방법론(Rule, Bayesian)의 식 (2)는 $(w_{Rule} \times Pw_{Rule}) + (w_{Baye} \times Pw_{Baye})$ 이다. 식 (1)과 식 (2)를 비교하여 큰 값을 부도예측결과로 채택한다. 위의 <그림 2>에서는 식 (1)의 부도예측값이 식 (2)의 건전예측값보다 크므로 WP-ANN의 결과는 식 (1)의 부도로 결정된다.

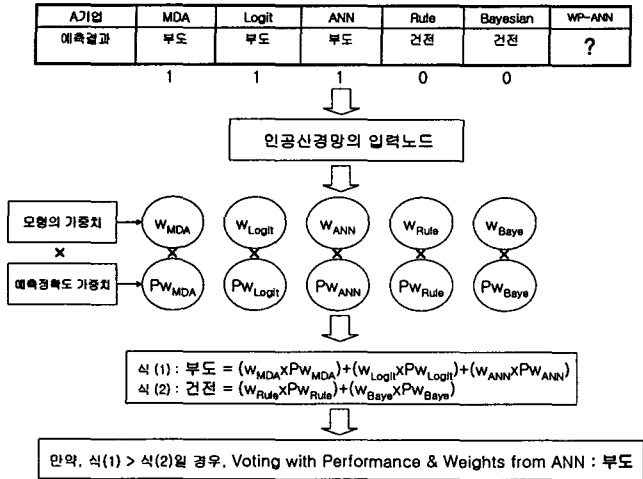


그림 2 - WP-ANN 모형

5. 연구결과

연구결과 분석을 위하여 검증데이터의 결과만을 모아 <표 2>를 구성하였고 <표 2>의 결과를 바탕으로 모형별 평균 예측력을 보여주는 <그림 3>을 구성하였다.

표 2 - 전체결과

구분	MDA	Logit	ANN	Rule	Baye	WP
Set 1	77.65	78.44	78.31	74.47	70.50	80.43
Set 2	78.57	78.84	79.23	71.16	72.62	80.29
Set 3	75.66	75.79	76.46	72.35	70.50	80.43
Set 4	76.46	77.65	76.32	72.49	69.84	80.02
Set 5	76.72	78.70	79.63	71.83	67.72	83.28
평균	77.01	77.88	77.99	72.46	70.24	80.89
Std	1.12	1.26	1.54	1.24	1.75	1.35

<표 2>와 <그림 3>의 결과에서 나타난 것과 같이 평균예측력은 통합모형인 WP-ANN 방법론이 전체 방법론과 비교하여 가장 예측력이 높았으며 데이터 셋 별로도 WP-ANN 통합모형이 해당 각 데이터 셋에서 예측력이 가장 높다는 것을 알 수 있다.

단일모형 중에서는 인공신경망 모형이 가장 예측력이 우수하였으며 위의 결과를 바탕으로 인공신경망과 예측정확도 가중치를 이용한 통합모형인 WP-ANN 방법론 또한 예측정확도가 높은 결과를 나타내었다고 해석할 수 있다.

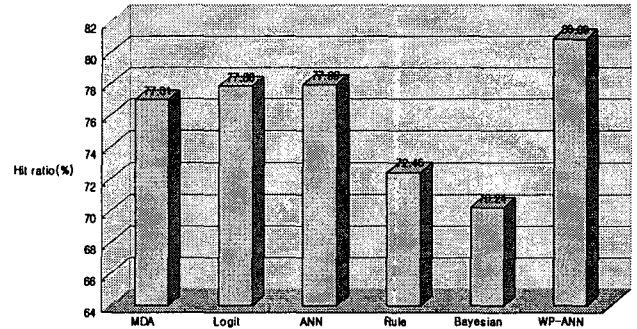


그림 3 - 모형별 평균예측력 비교

평균예측력을 모든 방법론과 비교하여 순서대로 나열하면 WP-ANN, 인공신경망, 로지스틱 회귀분석, 다변량 판별분석, 규칙유도기법, 베이저안 망의 순이다.

가장 예측력이 높은 WP-ANN 통합모형과 단일모형 간의 예측정확도 차이가 통계적으로 유의한지를 검증하기 위해 검증용 데이터를 대상으로 McNemar 검정을 실시하였다. McNemar 검정은 비모수 통계분석기법으로 이진값을 가지는 명목형 변수에 대해 관련이 있는 두 집단간의 차이를 검정할 때 사용된다. 특히, McNemar 검정은 동일한 대상에 대한 처리 전·후의 측정치 비교에 매우 유용한 것으로 알려져 있다.

표 3 - McNemar 검정결과

	MDA'	Logit	ANN	Rule	Bayesian
WP-ANN	3.625 (0.057)*	2.186 ¹ (0.139) ²	2.008 (0.156)	14.810 (0.000)**	21.975 (0.000)**
MDA		0.092 (0.762)	0.126 (0.722)	3.781 (0.052)*	7.795 (0.005)**
Logit			0.000 (1.000)	10.864 (0.001)**	10.017 (0.002)**
ANN				5.575 (0.018)**	11.004 (0.001)**
Rule					0.699 (0.403)

1) McNemar 통계량 값 2) p-값

*) 유의수준 10%에서 통계적으로 유의함.

***) 유의수준 5%에서 통계적으로 유의함.

<표 3>에서 나타난 것과 같이 WP-ANN 통합모형은 다변량 판별분석, 규칙유도기법, 베이지안 망과는 통계적으로 유의한 차이를 나타내어 예측성고가 뛰어난 것을 확인할 수 있었으나 로지스틱 회귀분석과 인공신경망과는 그 차이가 통계적으로 유의하지 않았다. 한편, 다변량 판별분석은 규칙유도기법, 베이지안 망과 통계적으로 유의한 차이를 나타내었으나 로지스틱 회귀분석과 인공신경망과는 통계적으로 유의한 차이를 나타내지 않았다. 로지스틱 회귀분석은 규칙유도기법, 베이지안 망과 5% 수준에서 유의하였고 인공신경망과는 통계적으로 유의한 차이를 나타내지 않았다. 인공신경망은 규칙유도기법, 베이지안망과 5% 수준에서 유의하였고 규칙유도기법은 베이지안 망과 통계적으로 유의한 차이를 나타내지 않았다.

6. 결론

본 연구에서는 보다 효과적인 기업부도예측을 위하여, 통계적 방법과 인공지능 방법을 결합한 통합모형을 제시하였다. 이를 위하여 통계적인 모형 중에서 가장 널리 활용되고 있는 다변량 판별분석, 로지스틱 회귀분석과 인공지능적인 방법으로서 최근 널리 사용되고 있는 인공신경망, 규칙유도기법, 베이지안 망의 5가지 방법론을 통합한 Voting with Performance & Weights from ANN(WP-ANN) 통합모형을 제시하였다. 실험결과 본 연구에서 제안한 WP-ANN은 다변량 판별분석, 로지스틱 회귀분석, 인공신경망, 규칙유도기법, 베이지안 망의 단일모형과 비교한 결과 가장 예측정확성이 우수한 것으로 나타났다. 따라서 본 연구를 통해 부도예측에 있어서 WP-ANN 통합모형이 기존의 단일모형들에 비해 우수한 예측정확성을 나타냄을 알 수 있었다.

WP-ANN 모형은 단일모형을 통합하는데 있어 예측력 가중치를 부여하였고 단일모형에 의해 추출된 결과들을 신경망을 통하여 하나의 결론으로 도출하였다는 장점이 있으나 인공신경망에 기반한 통합모형이기 때문에 인공신경망과 관련된 단점, 즉 모형의 아키텍처 설정문제, 학습과정에서의 과도적합현상으로 인한 일반화 어려움이나 결과 해석의 어려움 등의 한계점을 가지고 있다.

지금까지 대부분의 부도예측모형은 발생주의에 기초한 재무비율(accrual-based financial ratios)에 근거하여 부도예측을 하였으나 재무비율들은 영업활동 중 경제적 자원의 취득과 사용에만 초점을 두기 때문에 현재 기업의 평가에 있어 중요성이 더해가고 있는 유동성을 나타내는 현금흐름을 정확히 나타내지는 못한다. 또한 재무비율과 같은 회계정보는 현재 가치가 아닌 과거의 역사적 가치를 반영하고 있다는 단점이 있으며 재무비율 상호간의 높은 상관관계에 의한 다중공선성의 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 단일모형을 DEA 기반으로 통합한 기업부도예측모형과 비재무정보를 포함한 통합부도예측모형을 개발한다면 보다 의미 있는 정보를 제공할 수 있을 것으로 기대된다.

참고문헌

- [1] 김경재, 한인구. (2001). "퍼지신경망을 이용한 기업부도예측", *한국지능정보시스템학회*, 제7권, 제1호, pp. 135-147.
- [2] 김진백, 이준석. (2000). "인공지능 기법을 이용한 중소기업 부도예측에 있어서 현금흐름변수의 유용성 검증", *대한경영학회지*, 제26권, pp. 229-250.
- [3] 민재형, 이영찬. (2004). "자료포괄분석(DEA)을 이용한 신용평점모형의 개발 - 정보통신업을 중심으로", *한국경영정보학회 춘계학술대회 발표논문집*, pp. 460-467.
- [4] 박정민, 김경재, 한인구. (2005). "Support Vector Machine을 이용한 기업부도예측", *경영정보학 연구*, 제15권, 제2호, pp. 52-62.
- [5] 이건창, 김명중, 김혁. (1994). "기업도산예측을 위한 귀납적 학습지원 인공신경망 접근방법 : MDA, 귀납적 학습방법, 인공신경망 모형과의 성과비교", *한국경영학회*, 제23권, 제2호, pp. 109-144.
- [6] 이건창, 한인구, 김명중. (1996). "통계적 모형과 인공지능 모형을 결합한 기업신용평가 모형에 관한 연구", *한국경영과학회지*, 제21권, 제1호, pp. 81-100.
- [7] 추휘석, 민지경, 이인호. (2004). "다수의 인공신경망 모형을 통한 기업데이터의 분류 및 부도예측에 관한 연구", *연세경영연구*, 제41권, 제2호, pp. 514-539.
- [8] 허준, 최병주, 정성원. (2001). *클레멘타인을 이용한 데이터마이닝*, 1판, SPSS아카데미.
- [9] Abdelwashed, T., and Amir, E.M. (2005). "New Evolutionary Bankruptcy Forecasting Model based on Genetic Algorithms and Neural Networks," *The 17th IEEE International Conference on Tools with Artificial Intelligence*, pp. 241-245.
- [10] Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, pp. 589-609.
- [11] Anandarajan, M., Lee, P., and Anandarajan, A. (2001).

- "Bankruptcy Prediction of Financially Stressed Firms: An Examination of the Predictive Accuracy of Artificially Neural Networks," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 10, pp. 69-81.
- [12] Beaver, W. (1966). "Financial ratios as predictors of failure," *Journal of Accounting Research*, Vol. 4, pp. 71-102.
- [13] Chen, M. C., and Huang, S.H. (2003). "Credit Scoring and Rejected Instances Reassigning through Evolutionary Computation Techniques," *Expert Systems with Applications*, Vol. 24, pp. 433-441.
- [14] Cheng, J. (1998). "Belief Network(BN) Power Constructor System," Available at <http://www.cs.ualberta.ca/~jcheng/bnpc.htm>.
- [15] Cheng, J., Bell, A. and Liu, W. (1997). "An Algorithm for Bayesian Belief Network Construction from Data," *Proceedings of AI & STAT-97*, pp. 83-90.
- [16] Cheng, J., Bell, A. and Liu, W. (1997). "Learning Belief Networks from Data : An Information Theory Based Approach," *International Conference on Information and Knowledge Management*, pp. 325-331.
- [17] Hair, J.F., R.E. Anderson, R.E. Tatham and W.C. Black. (1995). *Multivariate Data Analysis with Readings*, Prentice Hall.
- [18] Heckerman, D., Geiger, D., Chickering, D. M. (1994). "Learning Bayesian Networks: the Combination of Knowledge and Statistical Data," Technical Report MSR-TR-94, Microsoft Research.
- [19] Hornik, K. (1991). "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, Vol. 4, pp. 251-257.
- [20] Jensen, F. V. (1996). *An Introduction to Bayesian Network*, Springer-Verlag, New York.
- [21] Ohlson, J. A. (1980). "Financial ratios and probabilistic prediction of bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131.
- [22] Peel, M. J., D. A. Peel and P. F. Pope. (1986). "Predicting corporate failure-some results for the UK corporate sector," *Omega*, Vol. 14, No. 1, pp. 5-12.
- [23] Pendharkar, P.C. (2005). "A Threshold-varying Artificial Neural Network approach for Classification and its Application to Bankruptcy Prediction Problem," *Computers & Operations Research*, Vol. 32, pp. 2561-2582.
- [24] Shin, K. S., and Lee, K. J. (2004). "Bankruptcy Prediction Modeling Using Multiple Neural Network Models," *Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference*, KES.
- [25] Shin, K. S., and Lee, K. J. (2004). "Neuro-Genetic Approach for Bankruptcy Prediction Modeling," *Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference*, KES.
- [26] Weiss, S. and Kulikowski, C. (1991). *Computer Systems That Learn*, Morgan Kaufmann Publishers, Inc.
- [27] West, D. (2000). "Neural Network Credit Scoring Models," *Computers & Operations Research*, Vol. 27, pp. 1131-1152.