韓國經營科學會誌
第33권 第2號
2008年 6月

도산예측을 위한 유전 알고리듬 기반 이진분류기법의 개발*

민재형**† · 정철우***

A GA-based Binary Classification Method for Bankruptcy Prediction*

Jae H. Min** · Chulwoo Jeong***

Abstract

The purpose of this paper is to propose a new binary classification method for predicting corporate failure based on genetic algorithm, and to validate its prediction power through empirical analysis. Establishing virtual companies representing bankrupt companies and non-bankrupt ones respectively, the proposed method measures the similarity between the virtual companies and the subject for prediction, and classifies the subject into either bankrupt or non-bankrupt one. The values of the classification variables of the virtual companies and the weights of the variables are determined by the proper model to maximize the hit ratio of training data set using genetic algorithm. In order to test the validity of the proposed method, we compare its prediction accuracy with ones of other existing methods such as multi-discriminant analysis, logistic regression, decision tree, and artificial neural network, and it is shown that the binary classification method we propose in this paper can serve as a promising alternative to the existing methods for bankruptcy prediction.

Keyword : Bankruptcy Prediction, Binary Classification, Genetic Algorithm

논문접수일 : 2007년 07월 20일 논문게재확정일 : 2008년 03월 24일
논문수정일(1차 : 2007년 11월 29일, 2차 : 2008년 02월 11일, 3차 : 2008년 3월 23일)
* 이 연구는 2007년도 서강대학교 교내연구비 지원에 의한 연구임(200701059.01).
** 서강대학교 경영학과
*** 서강대학교 경영학과
† 교신저자

1. 서 론

1997년 말 IMF 외환위기부터 2007년 국제결제은행(BIS)의 바젤 II 협약, 그리고 2009년 자본시장통합법 시행에 이르기까지 우리나라의 금융환경은 급속한 변화를 경험하고 있다. 특히, 바젤 II 협약의 시행은 금융기관의 자기자본 보유부담을 증가시켜 은행간 경쟁력 차이를 확대시킬 것으로 예상되며, 이에 따라 각 은행은 안전자산에 대한 선호 경향을 나타낼 것으로 보인다. 이러한 금융환경의 변화에 따라 우리나라 은행들의 위험관리에 대한 관심도 그 어느 때보다 증대되고 있는 실정인데, 특히, 위험관리 측면에서 기업도산예측은 은행의 여신심사에 있어 핵심적인 위치를 차지하고 있다. 기업도산예측이 위험관리 측면에서 중요한 이유는 정확한 기업도산예측을 통하여 은행은 대출의 부실화를 사전에 예방할 수 있고, 수익성을 제고할 수 있으며, 더 나아가서는 건전한 기업에 자금을 원활히 공급함으로써 금융중개기관으로서의 기능을 회복할 수 있기 때문이다.

그러나 오늘날 기업의 경영환경이 급변함에 따라 기업도산의 원인과 징후는 갈수록 복잡해지고 있으며, 이에 따라 기업도산예측을 위한 많은 기법들이 이미 나와 있음에도 불구하고 기업도산의 새로운 원인과 징후를 반영하기 위한 모형 개발의 필요성은 끊임없이 제기되고 있다.

기업도산예측에 관한 연구는 Beaver[6]가 처음으로 방법론적 토대를 마련한 이후 다변량 판별분석, 확률모형 등의 통계기반 모형을 거쳐 최근의 인공지능기법에 이르기까지 발전을 거듭해왔다. 도산예측 기법의 발달과정을 살펴보면 각각의 방법론들은 고유한 장점과 함께 단점도 가지고 있음을 보게 된다. 예를 들어, 다변량 판별분석은 모형이 비교적 단순하고 인지하기에도 쉬운 장점을 가진 반면, 정규성, 선형성 등의 엄격한 통계적 가정이 필요하기 때문에 수집된 자료가 그러한 가정을 위배하는 경우에는 예측력이 떨어진다는 단점을 갖고 있다. 반면, 인공지능기법의 대표적인 방법론인 인공신경망의 경우에는 통계적 가정에서 자유롭고 예측력도 우수하다는 장점이 있는 반면, 결과의 해석이 용이하지 않다는 단점을 갖고 있다. 결국, 도산예측기법의 발달과정은 기존 방법론이 갖는 단점을 보완함과 동시에 고유한 방법론적 장점을 갖는 기법을 개발하는 과정이었다고 할 수 있다. 따라서 기존에 많은 도산예측기법들이 개발되었음에도 불구하고 학계와 실무에서 새로운 방법론을 계속해서 연구하는 것은 상황과 목적에 따라 적절한 도산예측기법을 이용할 수 있도록 선택의 폭을 넓히는 의의를 갖는다.

본 연구에서는 인공지능기법의 일종인 유전 알고리듬(genetic algorithm : GA)을 기반으로 하여 기업도산예측을 위한 새로운 이진분류기법을 제안하고, 그 유용성에 관해 논하였다. 본 연구에서 제안하는 모형은 민재형과 정철우[2]에서 제안된 다집단 분류모형을 도산예측을 위한 이진분류에 적합하도록 변형, 발전시킨 것이다. 또한 본 연구에서는 실증분석을 통해 본 연구 모형의 예측성과를 기존의 다변량판별분석, 로지스틱 회귀분석, 의사결정나무, 인경신경망 모형과 비교하고, 본 연구에서 제안하는 기법이 기존 기법을 대체할 수 있는 유망한 대안 기법으로서 타당성을 가짐을 확인하였다.

2. 연구의 배경

기업도산예측을 위한 방법론적 발달과정은 크게 두 가지 줄기로 나누어 볼 수 있다. 하나는 Beaver[6]의 단일변량 판별분석 이후 시작된 통계기반 기법으로 이 기법은 Altman[3]의 다변량 판별분석을 거쳐, 로짓[24] 및 프로빗[34]과 같은 확률모형으로 발전하였다.

다른 하나는 1990년대부터 최근에 이르기까지 연구가 진행되고 있는 인공지능기법이다. 인공지능기법은 구체적으로 BPNN(back propagation trained neural network)[4, 7, 16, 18, 19, 26, 28, 29, 31], PNN(probabilistic neural networks)[32], SOM(self-organizing map)[13, 17, 18], Cascor(cascade correlation neural network)[15] 등과 같은 인공신경망의 다양한 방법들과 Fuzzy Set Theory[33], 의

사결정나무[10, 21]), 사례기반추론(Case-based Reasoning)[8, 12, 25], 유전 알고리듬[27, 30], Rough Sets Theory[9, 22, 23] 등과 같은 기법들을 포함하고 있다.

본 연구에서는 기업도산예측을 위한 방법론적 발달과정 중에서 인공지능기법의 일종인 유전 알고리듬을 응용하여 기업도산예측을 위한 새로운 이진분류기법을 제안한다.

2.1 유전 알고리듬

유전 알고리듬은 생태계의 선택(selection), 교차(cross over), 변이(mutation)와 같은 유전 메커니즘을 수리적으로 모형화한 것으로 최적해를 구하기 위해 방대하고 복잡한 공간을 확률적으로 탐색하는 특징을 갖고 있다. 선택, 교차, 변이 연산자를 모두 포함한 유전 알고리듬의 완성된 구조는 Holland[11]의 연구에서 처음으로 정리되었는데, 이러한 유전 알고리듬의 전형적인 구조를 도시하면 [그림 1]과 같다.

[그림 1]에 나타난 유전 알고리듬의 논리적 구조를 간단히 설명하면 다음과 같다. 우선, 알고리듬은 n개의 해를 임의로 생성하는데, 생성된 n개의 해로 이루어진 해의 집단을 population이라고 부른다. [그림 1]에서 population의 크기는 n이다. 이 population으로부터 $k(k \leq n)$개의 새로운 해를 만들어 내는데, 각각의 해는 선택, 교차, 변이의 단계를 거쳐 만들어진다. 이렇게 만들어진 k개의 해는 population 내의 k개의 해와 대치되어 새로운 population을 만들어낸다. 이러한 과정을 미리 설정한 정지조건(stopping condition)을 만족할 때까지 반복 수행한 후 population에 남은 k개의 해 중 목적함수에 가장 큰 기여를 한 해를 최적해로 도출하게 된다. 이러한 알고리듬의 구조적 특징으로 인해 유전 알고리듬은 다양한 제약조건을 포함한 문제 상황에서 목적함수를 최적화하는 모수를 추정하는데 널리 활용되어 왔다.

특히, 유전 알고리듬을 기업도산예측에 이용한 연구의 예로는 Varetto[30]와 Shin and Lee[27]를 들 수 있다. 이 중, Varetto[30]는 유전 알고리듬을 기반으로 두 가지 접근법에 따른 모형을 제시하였는데, 하나는 일종의 선형모형으로 유전 알고리듬을 이용하여 판별함수의 판별력을 최대화하도록 판별함수의 상수와 변수별 계수를 추정하도록 만들어졌다. 다른 하나는 규칙기반모형의 일종으로 유전 알고리듬을 이용하여 특정 개수의 if-then 규칙의 조합을 만들고, 이를 통해 GSR(genetic score by rules)이라는 판별점수를 산출하여 그 점수가 0보다 큰지의 여부에 따라 도산예측을 수행하도록 하였다. 한편, Shin and Lee[27]는 특정 변수의 조합에 대해 유전 알고리듬을 이용하여 예측력을 최대화하도

```
n개의 초기 염색체 생성 ;
repeat {
    for i = 1 to k {
        두 염색체 p1, p2 선택 ;
        offspringi = crossove(p1, p2) ;
        offspringi = mutation(offspringi) ;
       }
       offspring1, offspring2, …, offspringk를 population 내의 k개의 염색체와 대치 ;
    } until(정지 조건 만족) ;
남은 해 중 최상의 염색체를 return ;
```

자료원) 문병로, 「유전 알고리즘」, 1판, 다성출판사, 2001, p. 6.

[그림 1] 유전 알고리듬의 구조

록 각 변수의 임계값(cut-off value)과 조건, 그리고 도산여부를 추정함으로써 도출되는 규칙유도모형(rule inducing model)을 제시한 바 있다.

이와 같이 도산예측기법의 발달과정에서 유전 알고리듬은 그 자체가 하나의 기법으로서 역할을 하기보다는 기존의 통계기반 기법과 인공지능기법의 보완적 기법으로서의 역할을 담당해왔다. 본 연구에서도 유전 알고리듬은 본 연구에서 새로이 제시하는 이진분류기법을 구현하는데 있어 도산기업과 비도산기업을 대표할 수 있는 특정 개수의 대표기업 변수 값과 변수별 가중치를 추정하는 도구로서의 역할을 한다.

2.2 모형의 방법론적 장점

본 연구에서 제안하는 이진분류기법이 가지는 방법론적 장점을 좀 더 구체적으로 알아보기 위해 기존의 기법과 비교하여 설명하면 다음과 같다. 이러한 비교는 본 연구의 모형이 개발된 배경을 설명하기 위해서도 필요하다.

첫째, 본 연구의 모형은 관측값들을 몇 개의 군집으로 분류한다는 점에서 군집분석과 유사점을 가진다. 그러나 군집분석은 관측값이 속한 군집에 대한 정보 없이 독립변수만을 가지고 군집화 하는 방법인 데 반해, 본 연구의 모형은 관측값의 도산여부를 아는 상태에서 관측값의 도산여부와 관측값이 속한 군집을 대표하는 기업(대표기업)의 도산여부가 일치하는 비율을 최대화하도록 군집화하는 방법이라는 차이가 있다. 또한 군집분석은 군집화할 때 변수별 가중치 차이를 고려하지 않는 기법인데 반해, 본 연구의 모형은 예측정확도를 최대화하는데 있어 변수별 가중치를 반영한 모형이다. 아울러 본 연구의 모형은 유전 알고리듬을 이용함으로써 군집분석에서 초기값 설정에 따라 군집분석의 결과가 달라질 수 있는 문제점을 극복하도록 하였다.

둘째, 본 연구에서 제안하는 모형은 예측기업과 실제기업의 유사도를 기준으로 예측결과를 도출한다는 점에서 사례기반추론(CBR : case-based reasoning)과 공통점을 가진다. 그러나 본 연구의 모형은 전체적(global) 학습방법을 추구한다는 점에서 국지적(local) 학습방법을 이용하는 사례기반추론과 차별화된다. 즉, 사례기반추론은 각 평가대상 기업과 기존 기업 모두를 비교해서 가장 가까운 k개의 기업을 참조해 예측결과를 생성하는 국지적 학습방법을 이용한다. 이와 비교하여, 본 연구에서 제시하는 모형은 유전 알고리듬을 이용하여 대표기업의 변수 값을 구한 후, 대표기업과 평가대상 기업 간의 거리가 가장 가까운 대표기업의 도산여부에 따라 평가대상 기업을 분류하는 전체적 학습방법을 이용한다는 장점이 있다.

또한 본 연구의 모형은 사례기반추론보다 의미 있는 예측정보를 결과물로 제공해준다는 장점을 가진다. 즉, 사례기반추론은 결과물로서 평가대상 기업의 도산여부와 함께 평가대상 기업과 가장 유사한 사례기업을 보여준다. 이때 분석자는 사례기업의 어떠한 점이 평가대상 기업과 유사한지에 대한 정보를 얻기가 매우 어렵다. 이에 비해 본 연구의 모형은 결과물로서 평가대상 기업의 도산여부는 물론 평가대상 기업과 가장 유사한 대표기업을 보여준다. 분석자는 그 대표기업이 변수별로 다른 대표기업과 비교해서 어떠한 특징을 가지고 있는지 쉽게 파악할 수가 있으며, 이에 따라 그 대표기업과 같은 군집으로 분류된 평가대상 기업 역시 비슷한 특징을 가지고 있으리라는 논리적인 추측을 할 수 있는 장점이 있다.

3. 수리적 모형

본 연구에서 제안하는 이진분류 모형을 수리적으로 표현하면 다음과 같다. 우선, 대표기업과 관측값 사이의 동질성 또는 이질성을 나타내는 거리 개념을 표현하면 식 (1)과 같다.

$$d_{ki} = \sum_j (w_j / \sum_j w_j) |X_{kj} - X_{ij}| \qquad (1)$$

여기서, d_{ki} : 대표기업 k와 관측값 i 사이의 거리($i = 1, 2, \cdots, N$)

X_{kj} : 대표기업 k의 j번째 변수값($j = 1, 2, \cdots, M$)

X_{ij} : 관측값 i의 j번째 변수값

w_j : j번째 변수에 대한 가중치

식 (1)에서 d_{ki}는 관측값 i에 대해서 대표기업의 개수만큼 계산되어진다.[1] 그 중에서 가장 작은 값(d_{k^*i})에 상응하는 대표기업(k^*)의 도산여부에 따라 관측값 i를 분류한다. 즉, d_{ki}^*에 상응하는 대표기업(k^*)이 도산기업이면 관측값 i를 도산기업으로 분류하고, 반대의 경우이면 비도산기업으로 분류한다.

이렇게 모든 관측값에 대하여 분류하였을 때 이 분류결과가 관측값들의 실제 도산여부와 일치하는 비율(H)을 계산한다. 이때 일치도가 최대화되도록 유전 알고리듬을 이용하여 가중치(w_j)와 대표기업의 변수(X_{kj}) 값을 찾는다. 이러한 과정을 수식으로 나타내면 식 (2)와 같다.

$$\text{Max } H = \frac{1}{N}\sum_{i=1}^{N} C_i$$

$$\text{subject to } C_i = 1 \quad \text{if } B(i) = B(k^*) \quad \forall i \in \{1, 2, \cdots, N\},$$

$$C_i = 0 \quad \text{otherwise,}$$

$$d_{k^*i} = \min(d_{ki}'s) \tag{2}$$

여기서, H : 실제 도산여부에 대한 예측 적중률

N : 관측값 수

C_i : 관측값 i의 실제 도산여부에 대한 예측의 일치여부($i = 1, 2, \cdots, N$)

$B(i)$: 관측값 i의 실제 도산여부($B(i) = 1, 0$)

$B(k^*)$: 대표기업 k^*의 도산여부($B(k^*) = 1, 0$)

k^* : d_{k^*i}에 상응하는 대표기업

d_{k^*i} : 대표기업 k와 관측값 i 사이의 거리 중 최소 거리

4. 실증분석

4.1 데이터 정리

본 연구에서는 실증분석을 위해 2001년부터 2004년까지의 우리나라 비외감 중소 제조업체를 대상으로 도산기업과 비도산기업에 대해 재무비율 자료를 수집하되, 도산기업에 대해서는 도산 직전 년도 자료를 수집하였다. 총 2,814개 기업의 27개 재무비율 자료를 변수별 평균과 표준편차를 이용하여 Z값으로 표준화한 후, 1단계 표본선정을 위해 Z값이 [-3, 3]의 범위를 벗어나는 관측값은 이상치(outlier)로 판단하여 표본에서 제외하였다. 그리고 2단계 표본선정을 위한 조치로 도산과 비도산기업의 수를 1,271개로 동일하게 맞추기 위해 45개의 비도산기업을 무작위로 추출하여 제거하였다. 이와 같은 과정을 거쳐 수집한 자료를 훈련용, 테스트용, 검증용 자료의 비율이 6 : 2 : 2가 되도록, 즉, 훈련용, 테스트용, 검증용 자료의 크기가 각각 1526개, 508개, 508개가 되도록 층화추출법에 의해 분류하였다. 본 연구에서 훈련용 자료는 모형 구축을 위해, 테스트용 자료는 최적의 대표기업 수를 찾기 위해, 검증용 자료는 최적 모형의 예측력을 확인하기 위한 용도로 사용하였다.[2]

1) 대표기업의 개수는 도산기업과 비도산기업에 대하여 분석자가 임의로 정한다. 본 연구에서는 대표기업의 개수를 도산기업과 비도산기업에 대해 각각 1개부터 5개까지 늘려가면서 실험을 진행하였다.

2) 여기서 데이터(data set)를 훈련용과 검증용으로 나누지 않고, 훈련용, 테스트용, 검증용 등 세 개의 그룹으로 나눈 이유는 본 연구 모형의 성과를 기존의 다른 기법의 성과와 공정하게 비교하기 위함이다. 만약 데이터를 훈련용과 검증용으로만 나누어서 실험을 진행하였다면 검증용 자료의 예측정확도는 최적 모형을 찾기 위해 쓰인 자료의 예측정확도이므로 당연히 높을 수밖에 없는데, 이를 기존 모형들의 예측정확도와 비교한다면 공정한 비교가 될 수 없다. 따라서 훈련용 자료와 테스트용 자료를 이용하여 모형 구축과 최적 모형을 찾고, 검증용 자료를 이용하여 그 예측정확도를 측정한 후, 이를 최적 모형의 예측정확도로 이용하는 것이 타당할 것이다.

〈표 1〉 표본 선정 단계별 표본 수

	비도산	도산	합계
원자료	1407	1407	2814
1단계 : 이상치 제거	1316	1271	2587
2단계 : 동일기업수 조정	1271	1271	2542

한편, <표 2>는 수집된 27개 재무비율 변수의 정의를 나타내고 있다. 변수선정을 위해 본 연구에서는 독립표본 t-검정, 판별분석, 로지스틱 회귀분석, 의사결정나무, 요인분석 등 여러 가지 통계기법을 다각적으로 이용하였다.

〈표 2〉 변수 정의

변 수	정 의
X1	부가가치율
X2	총자본투자효율
X3	총자산증가율
X4	매출액경상이익율
X5	매출액순이익율
X6	매출액영업이익율
X7	매출원가비율
X8	순금융비용대매출액
X9	총자본경상이익율
X10	총자본순이익율
X11	순운전자본비율
X12	유동부채대총자산
X13	자기자본비율
X14	차입금의존도
X15	총자산회전율
X16	경상수지비율
X17	운전자금회전기간
X18	회전기간
X19	총자산경상이익율
X20	감가상각비
X21	경영자산회전율
X22	금융비용대총비용비율
X23	순금융비용
X24	손익분기점율
X25	인건비
X26	총자산순이익률
X27	EBIT 대 매출액

<표 3>은 분석방법별로 선정된 변수들을 정리한 것이다. 독립표본 t-검정에서 유의수준은 0.01로 하여 기업의 도산 여부에 따라 통계적으로 차이를 보이는 변수를 선정하였다. 판별분석과 로지스틱 회귀분석에서 단계별선택법(stepwise selection method)의 진입값은 F값의 유의확률 기준으로 0.01, 제거값은 0.05로 하였다. 의사결정나무에서 분리기준(splitting criterion)은 유의수준 0.02에서 Chi-square 검정으로 하였고, 최대 뿌리 깊이는 6으로, 한 잎에서 최소 관측값의 수는 5로 하였다.

〈표 3〉 분석방법별 선정된 변수

	선정된 변수
독립표본 t-검정	X2, X3, X12, X13, X15, X20, X21, X23, X24, X25
판별분석	X2, X3, X9, X12, X15, X20, X21, X22, X24, X25
로지스틱 회귀분석	X2, X3, X7, X9, X12, X20, X21, X22, X25
의사결정나무	X2, X3, X4, X10, X15, X16, X20, X23, X24, X25

수집된 변수들에 대하여 요인분석을 실시한 결과, <표 4>와 같이 8개의 요인이 추출된 것을 확인할 수 있었다.

이상의 분석 결과를 토대로 최종적인 변수를 선정하였는데, 변수선정과정을 단계별로 요약하면 다음과 같다. 1단계로, 요인분석 결과 묶여진 각 요인의 변수들 중 판별분석의 단계별선택법, 로지스틱 회귀분석의 단계별선택법, 그리고 의사결정나무 분석 결과 도출된 규칙을 통해 선정된 변수들만을 선택하였다. 2단계로, 1단계에서 선택된 변수들 중에서 독립표본 t-검정 결과 유의한 변수로 선정되지 않은 변수는 제외하였다. 마지막 3단계로, 각 요인에 속한 남은 변수들 가운데 의미상 중복되는 변수는 제외하였다. <표 5>는 이러한 변수선정의 단계를 정리한 것이다.

예를 들어, <표 5>의 3단계에서 요인 5에 해당하는 X15(총자산회전율)와 X21(경영자산회전율) 중 X15만을 선정한 것은 두 변수의 의미가 유사하다

〈표 4〉 재무비율 변수들의 회전성분행렬

변 수	요 인							
	1	2	3	4	5	6	7	8
X6	**0.988**	0.090	-0.064	0.052	0.009	0.003	-0.002	0.002
X4	**0.985**	0.092	0.117	0.060	0.023	-0.005	0.015	0.000
X27	**0.982**	0.094	0.145	0.055	0.006	-0.001	0.009	-0.004
X5	**0.980**	0.086	0.122	0.066	0.025	-0.007	0.032	0.003
X1	**0.968**	0.086	0.073	0.053	-0.034	0.063	-0.159	-0.031
X10	0.111	**0.940**	0.006	0.172	0.117	-0.057	-0.028	0.064
X26	0.101	**0.940**	0.010	0.193	0.122	-0.057	-0.005	-0.033
X19	0.105	**0.939**	0.009	0.182	0.147	-0.044	-0.004	-0.044
X9	0.114	**0.937**	0.003	0.169	0.136	-0.042	-0.027	0.066
X2	0.068	**0.521**	-0.014	0.039	0.311	0.255	-0.450	0.062
X17	0.066	0.003	**0.979**	0.031	-0.075	-0.011	-0.009	0.033
X18	0.066	0.003	**0.979**	0.031	-0.075	-0.011	-0.009	0.033
X8	-0.191	-0.010	**-0.907**	-0.023	-0.064	0.023	-0.016	0.014
X12	-0.089	-0.169	-0.003	**-0.856**	0.070	-0.020	0.072	0.014
X13	0.089	0.249	0.037	**0.834**	0.224	-0.020	-0.051	-0.080
X11	0.089	0.173	0.049	**0.806**	0.024	-0.121	0.149	0.097
X22	-0.047	-0.078	0.105	**-0.174**	-0.777	0.141	-0.111	-0.111
X15	0.015	0.402	-0.012	-0.103	**0.675**	-0.069	0.216	0.187
X21	-0.003	0.375	0.005	-0.125	**0.621**	-0.049	0.199	-0.138
X14	0.062	-0.090	-0.022	-0.428	**-0.537**	-0.034	0.123	0.123
X25	0.028	-0.060	-0.015	-0.034	0.024	**0.856**	-0.101	-0.020
X20	-0.005	-0.047	0.001	0.024	-0.047	**0.792**	-0.088	0.026
X23	0.024	0.008	-0.050	-0.167	-0.382	**0.642**	0.311	-0.004
X24	0.107	0.041	-0.028	-0.026	0.121	0.059	**0.760**	-0.057
X7	-0.198	-0.087	0.031	0.043	0.158	-0.114	**0.747**	0.074
X3	0.038	0.117	-0.041	-0.101	0.112	-0.011	-0.154	**0.751**
X16	0.049	0.052	-0.069	-0.067	0.049	-0.017	-0.117	**-0.652**

주) 요인추출방법 : 주성분분석.
회전방법 : Kaiser 정규화가 있는 배리맥스(varimax).

〈표 5〉 변수선정단계

	요 인							
	1	2	3	4	5	6	7	8
1단계	X4	X2, X9, X10		X11, X12, X13, X22	X15, X21	X20, X23, X25	X7, X24	X3, X16
2단계		X2		X12, X13	X15, X21	X20, X23, X25	X24	X3
3단계		X2		X12, X13	X15	X20, X23, X25	X24	X3

고 판단하였기 때문이다. 반면, 요인 4에 해당하는 X12(유동부채대총자산)와 X13(자기자본비율)은 기업의 도산여부에 서로 다르게 영향을 미치는 것으로 판단하여 둘 다 변수로 선정하였다. 또한 요인 6에 해당하는 X20(감가상각비), X23(순금융비용), X25(인건비) 역시 서로 다른 성격을 가진 비용으로 판단하여 모두 변수로 선정하였다. 그리하여 최종적으로 선정된 변수는 X2(총자본투자효율), X3(총자산증가율), X12(유동부채대총자산), X13(자기자본비율), X15(총자산회전율), X20(감가상각비), X23(순금융비용), X24(손익분기점율), X25(인건비) 등 9개 변수이다.

4.2 모형의 설계

본 연구에서는 도산 및 비도산 기업을 대표할 수 있는 대표기업의 수를 각각 1개부터 5개까지 늘려가면서 예측정확도의 추이를 살펴 최적의 대표기업 수를 구하였다. 대표기업의 수가 도산 및 비도산 기업에 대해 각각 1개인 것을 [모형 1], 2개인 것을 [모형 2], 3개인 것을 [모형 3], 4개인 것을 [모형 4], 5개인 것을 [모형 5]로 하여 실험을 진행하였다. 이렇게 대표기업의 수를 변화시키면서 실험을 진행한 이유는 대표기업의 수가 적으면 모형이 지나치게 단순화될 위험이 있고, 반면에 대표기업의 수가 너무 많으면 모형이 훈련용 자료에 과적합(overfit)될 위험이 있기 때문이다.

또한 도산 및 비도산 기업에 대하여 대표기업의 수를 1개씩 동일하게 증가시킨 이유는 최적 모형을 탐색하기 위한 시간을 절약하기 위해서이다. 즉, 도산 및 비도산 기업에 대하여 대표기업의 수를 다르게 설정하여(예를 들어, 도산기업의 대표기업 수는 1개, 비도산 기업의 대표기업 수는 2개 등) 최적 모형의 대표기업 수를 찾는다고 하면, 발생할 수 있는 경우의 수가 너무 많아져 실험시간이 매우 증가하게 된다. 따라서 우선 도산 및 비도산 기업에 대한 대표기업의 수를 1개씩 동일하게 증가시키면서 해당 모형의 예측정확도를 구하고, 이를 통해 최적 모형에 가까운 모형을 찾은 다음, 이에 근거하여 추가적으로 도산 및 비도산 기업에 대하여 대표기업의 수를 달리 설정한 모형에 대한 실험을 수행하는 것이 최적 모형의 탐색 시간을 줄일 수 있을 것이다. 이와 같은 추가적 모형에 대한 실험은 본 논문의 후반부에서 다루었다.

한편, 유전 알고리듬을 적용하는데 있어 본 연구 모형이 가지는 염색체(chromosome) 구조를 설명하면 다음과 같다. 각 염색체는 가중치에 대한 유전자 9개와 대표기업별 변수에 대하여 (대표기업의 수)×9개의 유전자로 이루어져 있다. 가중치 유전자는 [0, 1] 범위의 실수이고, 대표기업별 변수 유전자는 [-3, 3] 범위 내에서 실수값을 취하도록 설정하였다. 초기값은 가중치 유전자에 대해서는 [0, 1] 범위의 난수를 발생시켜 이용하였고, 대표기업별 변수 유전자는 표준정규분포를 따르는 확률변수를 발생시켜 이용하였다.

최적해를 찾기 위해 population 크기는 100으로 하였고, 변이율(mutation rate)은 w_j와 X_{kj}에 대해서 모두 0.1로, 교차율(cross over rate)은 w_j와 X_{kj}에 대해서 모두 0.5로 하였다. 정지 조건(stopping condition)은 총 시행 횟수가 20,000번이 되면 탐색을 멈추도록 하였다. 분석도구로는 유전 알고리듬 소프트웨어인 Evolver 4.0을 사용하였다.

4.3 분석결과의 논의

<표 6>은 예측정확도를 최대화하는 가중치와 대표기업 변수의 최적해를 탐색하기 위한 총시행횟수 및 최적해 탐색 시점의 시행횟수를 모형별로 정리한 것이다.

〈표 6〉 모형별 최적해 탐색 횟수

모형	총 시행 횟수	최적해 탐색 시점의 시행횟수
1	20,000	2,444
2	20,000	19,719
3	20,000	17,525
4	20,000	15,567
5	20,000	11,472

<표 7>은 모형별 예측정확도를 정리한 것이다. <표 7>을 보면 훈련용 자료에 대해서는 대표기업의 수가 증가함에 따라 예측정확도는 계속해서 증가하는 추이를 볼 수 있다. 이에 반해 테스트용 자료의 경우에는 대표기업의 수가 4일 때를 정점으로 가장 높았다가 대표기업의 수가 6일 때 낮아졌으며, 대표기업의 수가 8일 때 다시 높아졌다가 10이 되면 다시 낮아지는 모습을 보이고 있다. 이러한 현상은 대표기업의 수가 증가함에 따라 자료에 대한 모형의 적합도는 증가하지만, 일정 수 이상으로 대표기업의 수가 증가할 경우, 훈련용 자료에 대해서는 모형이 적합하지만 테스트용 자료에 대해서는 오히려 예측정확도가 떨어지는 이른바 과적합 문제가 발생한 것으로 해석된다. 예를 들어, [모형 4]의 경우 테스트용 자료의 예측정확도는 76.8%로서 높은 편이긴 하지만 훈련용 자료에 대한 예측정확도 77.9%에 비해 상대적으로 낮기 때문에 과적합 문제가 발생한 것으로 판단된다. 따라서 이 경우에는 과적합 문제가 발생하지 않으면서 테스트용 자료의 예측정확도가 가장 높은 [모형 2]를 최적의 모형으로 선택할 수 있다. 그리고 이러한 과정에 따라 선택한 [모형 2]의 검증용 자료에 대한 예측정확도는 78.4%로 가장 높게 나타났다.

〈표 7〉 모형별 예측정확도

모형	대표기업 수	훈련용	테스트용	검증용
1	2	76.4%	76.2%	75.8%
2	4	77.0%	77.0%	78.4%
3	6	77.1%	76.2%	77.8%
4	8	77.9%	76.8%	75.8%
5	10	78.2%	76.4%	70.5%

이제 예측정확도가 가장 뛰어난 [모형 2]의 추가 분석결과를 기술하면 다음과 같다. 우선, <표 8>은 [모형 2]의 변수별 표준 가중치를 나타낸 것으로 X3, X12, X13, X25의 가중치가 비교적 높은 수준을 보였고, X2, X15, X20, X23, X24의 가중치는 비교적 낮은 수준을 보였다. 이는 관측값들을 대표기업과의 거리에 따라 분류하는데 있어 상대적으로 큰 영향을 미친 변수들이 X3(총자산증가율), X12(유동부채대총자산), X13(자기자본비율), X25(인건비)임을 의미한다.

<표 9>는 [모형 2]의 대표기업별 변수에 대한 최적해를 나타내고 있다. 여기서 대표기업 1과 2는 비도산기업의 대표기업이고 대표기업 3과 4는 도산기업의 대표기업이다.

〈표 8〉 [모형 2]의 변수 가중치

가중치	X2	X3	X12	X13	X15	X20	X23	X24	X25
$w_j/\sum_j w_j$	0.0448	0.1776	0.1403	0.1395	0.0501	0.0909	0.0715	0.0717	0.2136
w_j	0.1694	0.6720	0.5310	0.5281	0.1896	0.3441	0.2707	0.2712	0.8084

〈표 9〉 [모형 2]의 대표기업별 변수 값

변수 / 대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1(비도산)	-2.4921	0.4066	1.5520	0.3040	1.7466	1.0213	1.1974	2.9719	0.8361
2(비도산)	-2.3118	-2.4352	2.7850	0.6320	0.2582	-2.3915	0.4086	0.7399	1.9994
3(도산)	-0.1427	-2.4112	0.7815	-0.7192	2.0606	2.4126	-1.5321	2.1588	-0.6153
4(도산)	2.0558	1.2644	1.8007	0.2825	2.3154	-2.0725	-0.8260	2.9237	-0.5044

이와 같이 구한 가중치와 변수 값을 가지고 관측값들을 분류한 결과는 <표 10>과 같다. <표 10>을 보면 [모형 2]의 4개 대표기업 중에서 3개 기업에 대해서만 관측값들이 분류되고, 대표기업 3에 대해서는 하나의 관측값도 분류되지 않았다. 이러한 결과는 대표기업 3이 모형구축에 있어 불필요할 뿐 아니라 추정시간의 지연 및 추정값의 편의(bias)와 같은 문제를 발생시킬 수 있는 요인이 될 수 있음을 암시한다. 따라서 대표기업의 수를 비도산기업에 대해서는 2개, 도산기업에 대해서는 1개만으로 하는 추가적인 모형을 만들어 분석을 시행하면 추정값의 편의 없이 [모형 2]와 비슷한 예측정확도를 가지면서 더 빠른 시간 내에 최적해를 구할 수 있을 것이라는 추측에 따라 추가적인 실험을 수행하였다.

총 20,000번의 시행을 통해 3,183번째 시행에서 최적해를 구하였는데, 이는 최적해를 더 빠른 시간 내에 얻어낼 수 있을 것이라는 추측과 일치하는 결과라고 할 수가 있다. 추가적 모형의 결과는 <표 11>부터 <표 14>와 같이 정리하였다.

<표 11>은 추가적인 실험 모형에 대한 변수별 최적 가중치이다. 이를 [모형 2]에 대한 결과(<표 8> 참조)와 비교하면, [모형 2]에서는 X3, X12, X13, X25가 비교적 높은 가중치 수준을 보이는 변수인데 반해, 추가적인 모형에서는 X3, X13, X20, X24, X25가 비교적 높은 가중치 수준을 보이는 변수로 작용한 것을 알 수 있다.

〈표 10〉 [모형 2]의 관측값 분류 결과

실제 도산여부 \ 대표기업		1	2	3	4	합계
훈련용	0	685	14	-	64	763
	1	273	67	-	423	763
	합계	958	81	-	487	1526
테스트용	0	231	3	-	20	254
	1	94	26	-	134	254
	합계	325	29	-	154	508
검증용	0	230	5	-	19	254
	1	86	26	-	142	254
	합계	316	31	-	161	508

〈표 11〉 추가적 모형의 변수 가중치

가중치	X2	X3	X12	X13	X15	X20	X23	X24	X25
$w_j/\sum_j w_j$	0.0319	0.2183	0.0378	0.1674	0.0196	0.1167	0.0276	0.1749	0.2057
w_j	0.1334	0.9127	0.1580	0.7000	0.0819	0.4877	0.1154	0.7313	0.8601

〈표 12〉 추가적 모형의 대표기업별 변수 값

변수 / 대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1(비도산)	2.2508	0.0855	1.3176	1.4516	2.2511	-2.9119	-0.6878	1.4422	2.6507
2(비도산)	-2.2057	0.1220	0.1508	2.9111	-0.1874	1.8721	0.7571	0.3975	0.0664
3(도산)	0.6379	0.2129	2.3068	1.7754	1.6116	-1.7152	0.2517	0.0458	-1.6165

한편, <표 12>는 추가적 모형의 대표기업별 변수 값을 나타내고 있다. 추가적 모형의 대표기업 1, 2, 3의 변수 값을 [모형 2]의 대표기업 1, 2, 4의 변수 값(<표 9> 참조)과 비교해 보면 다소 차이가 나는 것을 볼 수가 있다. 예를 들어, [모형 2]와 추가적 모형에서 모두 가중치가 높게 나타난 변수 X13에 대해 비교해 보면, [모형 2]에서는 대표기업 1과 2의 변수 값이 크고 대표기업 4의 변수 값이 작은데 반해, 추가적 모형에서는 대표기업 3의 변수 값이 대표기업 1과 대표기업 2의 변수 값 사이로 나타난 것을 볼 수 있다.

그 결과, 추가적 모형의 예측정확도 역시 [모형 2]의 예측정확도와 차이를 보였는데, 추가적 모형의 예측정확도는 훈련용 자료에 대해서는 77.3%, 테스트용 자료에 대해서는 78.0%로 나타나 [모형 2]의 예측정확도에 비해 높게 나타났으나, 검증용 자료에 대해서는 77.4%로 나타나 [모형 2]의 예측정확도에 비해 낮게 나타났다.

이러한 결과가 의미하는 바는 다음과 같다. [모형 2]는 과적합 문제를 발생시키지 않으면서 동시에 훈련용 자료의 분류정확도를 최대화하는 최적 모형을 탐색하는 과정에서 비도산 대표기업과 도산 대표기업의 수를 동일하게 두고 분석함으로써 대표기업 3에 대해서는 하나의 기업도 분류가 이루어지지 않았다. 반면에 추가적 모형에서는 비도산 대표기업의 수를 2, 도산 대표기업의 수를 1로 함으로써 탐색시간을 줄이는 동시에 훈련용 자료에 대해 더 높은 예측정확도를 보이는 변수 가중치와 대표기업별 변수 값이 탐색되었다. 그 결과, 훈련용 자료에 대해서는 [모형 2]보다 높은 예측정확도를 나타내었지만 검증용 자료에 대해서는 예측정확도가 [모형 2]보다 떨어지는 결과를 보이게 되었다. 그러나 이러한 결과를 과적합 문제가 발생한 것으로 볼 수는 없다. 왜냐하면 추가적 모형의 검증용 자료에 대한 예측정확도 77.4%는 [모형 2]의 78.4%에 비해 낮다고는 하지만 여전히 추가적 모형의 훈련용 자료에 대한 예측정확도 77.3%에 비하면 높게 나타났기 때문이다.

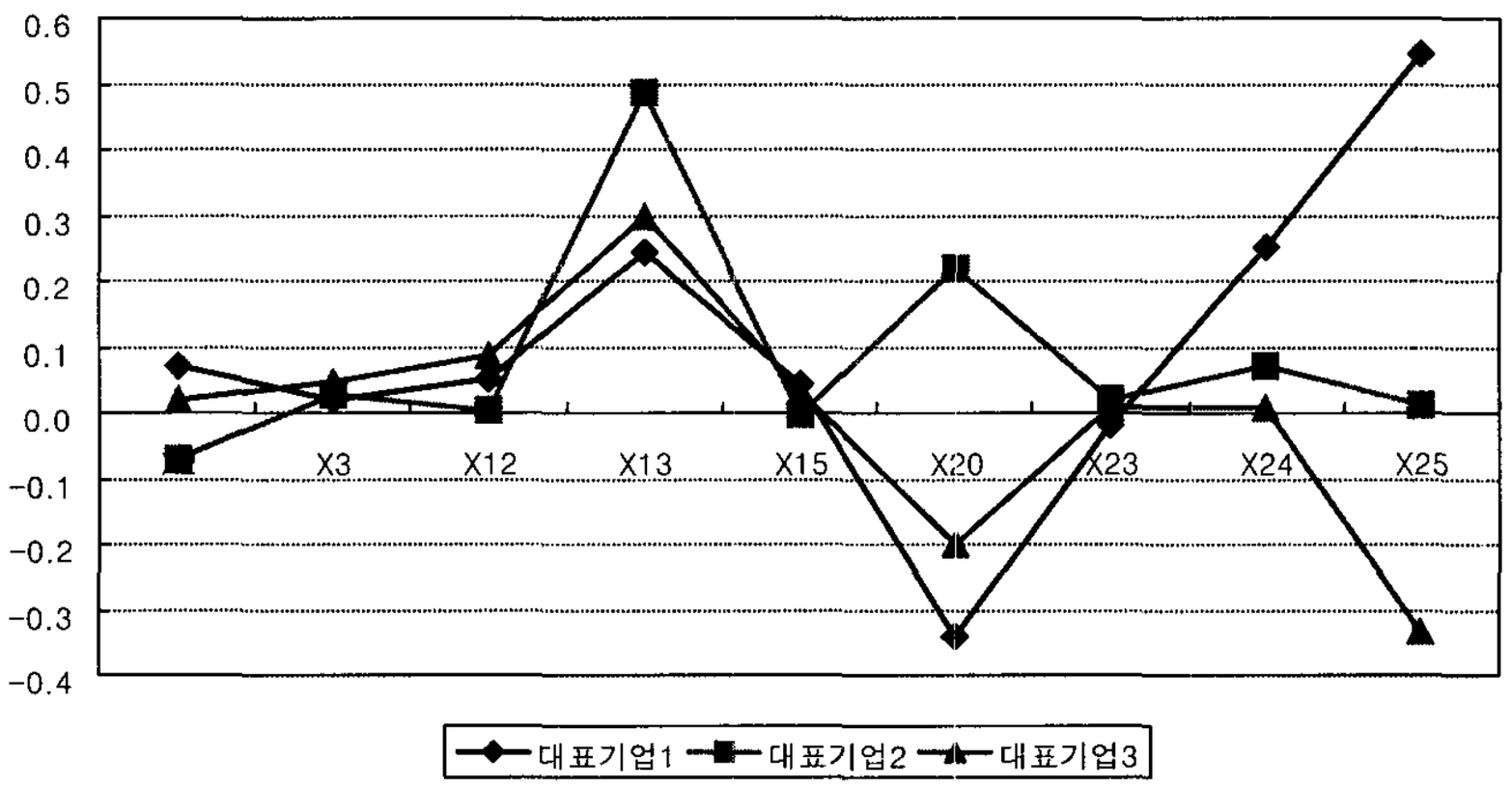
따라서 [모형 2]의 검증용 자료에 대한 예측정확도 78.4%가 추가적 모형의 예측정확도 77.4%에 비해 높게 나타났다고 해서 [모형 2]를 더 우수한 모형이라고 판단하는 것은 무리가 있다. 오히려 추가적 모형이 과적합 문제를 발생시키지 않으면서도 훈련용 자료, 테스트용 자료, 검증용 자료 모두에 대하여 안정적인 예측정확도를 보이는 해를 훨씬 빠른 시간 안에 탐색해 내었으므로 [모형 2]에 비해 더 우수한 모형이라고 평가할 수 있다. 추가적 모형의 관측값 분류 결과를 요약하면 <표 13>과 같다.

〈표 13〉 추가적 모형의 관측값 분류 결과

실제 도산여부 \ 대표기업		1	2	3	합계
훈련용	0	15	666	82	763
	1	3	261	499	763
	합계	18	927	581	1526
테스트용	0	5	228	21	254
	1	0	91	163	254
	합계	5	319	184	508
검증용	0	6	428	102	536
	1	1	160	519	680
	합계	7	588	621	1216

〈표 14〉 대표기업별 가중 변수 값

대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1(비도산)	0.0718	0.0187	0.0498	0.2431	0.0441	-0.3397	-0.0190	0.2523	0.5454
2(비도산)	-0.0704	0.0266	0.0057	0.4874	-0.0037	0.2184	0.0209	0.0695	0.0137
3(도산)	0.0204	0.0465	0.0872	0.2973	0.0316	-0.2001	0.0069	0.0080	-0.3326



[그림 2] 대표기업별 가중 변수 값

한편, 추가적 모형의 분석결과 도출된 변수별 가중치와 변수 값을 곱하여 구한 대표기업별 가중 변수 값을 정리하면 <표 14> 및 [그림 2]와 같다.[3)]

[그림 2]를 보면 도산기업에 속하는 대표기업 3은 비도산 대표기업인 1 및 2에 비해 X12(유동부채대총자산)의 값은 높은 반면, X24(손익분기점율)와 X25(인건비)의 값은 상대적으로 낮은 특징을 갖고 있음을 알 수 있다. 이러한 정보는 여신심사에 있어 유용하게 활용될 수 있을 것이다. 즉, 여신심사가 필요한 어떠한 기업이 유동부채대총자산이 동일 산업내의 다른 기업들에 비해 높고, 손익분기점율과 인건비가 상대적으로 낮은 기업이라면 일단 도산의 가능성을 의심할 필요가 있으며, 실제로 그 기업과 대표기업들과의 거리를 측정하여 대표기업 3과 동일한 기업군에 속하는지를 확인하고, 그 결과를 여신심사에 반영해야 할 것이다.

4.4 타 기법과의 성과비교

본 연구에서 개발한 이진분류기법의 판별력을 검증하기 위해 그 분류 성과를 기존의 다른 이진분류기법의 성과와 비교하였다. 분석기법별 모형의 설

3) <표 14>를 해석할 때는 각 대표기업의 변수 값을 전체적으로 볼 필요가 있다. 예를 들어, X2(총자본투자효율)의 값을 예로 들면, 비도산기업에 속하는 대표기업 2는 음의 값을 가지는 것으로 나타난 반면, 도산기업에 속하는 대표기업 3의 경우는 오히려 양의 값을 가진 것으로 나타나 일견 모순되게 여겨질 수도 있다. 그러나 변수 값에 대한 해석은 이렇듯 간단히 특정 변수만을 따로 떼어 개별적으로 파악하는 것이 아니라 다른 변수들과 함께 전체적으로 파악하는 것이다. 즉, 비도산기업에 속하는 대표기업2의 경우, X2(총자본투자효율)의 값은 음수이지만 X13(자기자본비율)의 값은 다른 대표기업에 비해 무척 높다. 대표기업 2의 X2 값이 다른 대표기업에 비해 상대적으로 낮은데도 불구하고 비도산기업이 된 이유는 X2의 값이 낮기 때문에 생기는 도산요인에 비해 X13의 값이 높은 데서 생기는 비도산요인의 크기가 훨씬 커 도산 요인을 상쇄하기 때문이다. 마찬가지 방법으로 대표기업 1과 대표기업 3의 도산여부에 대해서도 해석할 수 있다. 결국 <표 14>는 변수 전체적인 관점에서 비도산기업과 도산기업의 전형적인 모습을 세 가지로 제시한 것이다.

계는 다음과 같은 방식으로 하였다. 모든 분석기법에 대해 변수는 본 연구와 동일하게 9개의 변수를 이용하였다. 자료는 본 연구의 모형 구축 및 검증을 위해 이용한 훈련용, 테스트용, 검증용 자료를 그대로 이용하였다. 의사결정나무에서 분리기준(splitting criterion)은 유의수준 0.02에서 Chi-square 검정으로 하였고, 최대 뿌리 깊이는 6으로, 한 잎에서 최소 관측값의 수는 5로 하였다. 인공신경망 모형은 입력층과 출력층, 그리고 2개의 은닉층을 가진 4층 퍼셉트론(four layer perceptron)을 이용하였고, 각 은닉층의 노드 수는 입력변수의 개수와 동일한 9개씩으로 설정하였다.

〈표 15〉 기법별 예측정확도

	훈련용	검증용
본 연구의 기법	77.3%	77.4%
판별분석	70.2%	69.1%
로지스틱 회귀분석	71.2%	70.7%
의사결정나무	76.5%	76.8%
인공신경망	78.1%	76.4%

<표 15>는 본 연구의 자료를 이용하여 본 연구에서 개발한 기법과 기존 이진분류기법의 예측정확도를 비교한 것이다. <표 15>를 보면 본 연구에서 제안한 기법이 검증용 자료를 기준으로 볼 때, 가장 높은 예측정확도를 보이는 것으로 나타났다. 물론 이 결과는 특정 자료를 이용하여 도출한 표본 결과로 본 기법의 우월성을 일반화할 수 있는 통계적인 검정 결과는 아니지만 본 연구에서 제안하는 기법이 도산예측을 위한 기존의 기법을 대체할 수 있는 유망한 대안 기법임을 보여주는 수치 결과이다.

이러한 예측성과의 우수성과 함께 본 연구 모형은 실무적용 시 다음과 같은 장점을 갖는다.

첫째, 본 연구의 모형은 도산 여부의 분류와 함께 그 원인까지도 설명할 수 있는 장점을 갖는다. 본 모형은 부실기업으로 판정된 기업에 대해 그 원인이 무엇인지를 공정하게 파악하여 제시할 수 있다. 이러한 장점은 예측력은 뛰어나지만 결과의 해석이 모호함에 따라 현업에서 적용이 어려운 인공신경망과 비교할 때 더욱 크게 부각된다고 하겠다. 실제로 <표 15>를 보면, 본 연구의 기법이 기존의 이진분류기법들 중 예측성과가 우수하다고 알려진 인공신경망보다 나은 성과를 보이는 것으로 나타났는데, 이러한 결과는 본 연구의 기법이 부실로 판정된 기업의 경우, 그 원인까지 파악하여 제시할 수 있다는 장점과 결합하여 실무적으로 널리 활용될 수 있는 가능성을 보여주고 있다.

둘째, 본 연구의 모형은 선형성을 갖지 않는 재무비율 자료에 대해서도 적용이 가능하다. 오늘날 기업 도산의 원인과 징후는 매우 다양하게 나타나고 있다. 따라서 많은 도산예측모형에서 이용되고 있는 재무비율 자료의 경우 선형성이 지켜지지 않는 경우가 대부분이다. 본 연구의 모형은 이러한 재무비율의 비선형성도 수용하여 그와 일치하는 결과를 도출해 낼 수 있다. 실제로 본 연구의 실증분석 결과를 보면 어느 특정 변수에 대해 도산기업이 비도산기업에 비해 일률적으로 그 값이 높거나 낮은 식으로 나타나지 않음을 확인할 수 있다. 이는 재무비율 자료가 선형성을 따르지 않는데서 나타나는 자연스러운 현상이라고 보아야 할 것이다. 또한 이러한 현상은 현실 세계에서 나타나는 다양하고도 복잡한 도산의 징후를 반영하는 결과라고도 해석할 수 있다.

셋째, 본 연구의 모형은 유전 알고리듬에 기반하고 있기 때문에 모형 구축에 많은 시간이 소요되지만, 모형이 일단 구축되고 나면 새로운 기업에 대한 분류와 진단을 매우 짧은 시간에 간단히 수행할 수 있는 장점을 갖는다. 물론, 최근 급속한 발달을 보이고 있는 컴퓨팅 환경을 생각하면 이러한 장점은 크게 부각되지 않을 수도 있다. 그러나 기업에서 처리해야 되는 데이터의 용량도 컴퓨팅 환경이 발달하는 만큼이나 기하급수적으로 커지고 있음을 고려하면, 신속한 의사결정이 필요한 경영환경에서 본 연구 모형이 가지는 컴퓨팅 차원에서의 장점은 쉽게 간과할 수 없을 것이다.

5. 결 론

본 연구에서는 새로운 이진분류기법의 제안과 함께 실제 기업 자료를 이용하여 이를 실증분석하고, 그 예측력을 기존의 기법과 비교함으로써 도산예측을 위한 대안 기법으로서의 타당성을 확인하였다. 또한 본 연구에서 제안한 모형은 실무적 차원에서 기업의 도산·비도산 분류와 함께 다양하고도 복잡한 도산의 원인과 징후를 제시해 줄 수 있고, 재무비율 자료의 비선형성을 수용하며, 일단 모형이 구축하고 나면 모형의 적용은 매우 간편하게 이루어진다는 등 여러 가지 장점을 가지고 있음을 밝혔다.

본 연구의 결과는 금융기관의 신용위험관리 측면과 새로운 방법론 개발이라는 학문적 측면에서 다음과 같은 기대효과를 갖는다.

첫째, 본 연구에서 개발한 이진분류기법은 금융기관의 신용위험관리에 기여할 수 있다. 나날이 경쟁이 치열해지는 금융환경 하에서 정확한 기업도산예측은 금융기관의 수익성 제고 수단으로서 뿐만 아니라 기본적으로는 신용위험의 체계적 관리를 통한 금융기관의 생존 수단으로서의 역할을 수행할 수 있다. 또한 거시적으로는 기업의 재무건전성에 따라 자금을 효율적으로 배분함으로써 국가경제 발전에도 기여할 수 있을 것으로 기대한다.

둘째, 본 연구의 결과는 새로운 분류 방법론의 개발이라는 학문적인 기여와 함께, 개발한 기법의 응용범위 확대에도 기여를 한다. 본 연구의 목적은 단순히 기존의 기법을 기업도산예측에 적용하는데 그치지 않고 도산예측을 위한 새로운 분류기법을 개발하는데 있었다. 따라서 본 연구에서 개발한 이진분류기법은 기업도산예측 뿐만 아니라 상품구매예측, 프로젝트 위험관리 등 다양한 분야에 확대 응용될 수 있을 것으로 기대한다.

이러한 기여와 함께 본 연구는 다음과 같은 한계점을 가지며, 이러한 한계점은 향후 연구를 위한 방향을 제시한다.

첫째, 본 연구에서 제안한 기법은 최적의 모형을 선택하는데 있어 분석자의 판단이 개입될 여지가 있다. 본 연구에서는 과적합 문제를 방지하기 위해 정확한 기준을 제시하기보다는 분석자의 판단에 따라 훈련용, 테스트용, 검증용 자료에 대하여 비교적 안정적인 예측력을 보이는 모형을 최적의 모형으로 선택하였다. 이러한 주관적인 판단을 보완하기 위한 연구가 향후 필요할 것이다.

둘째, 본 연구에서는 대표기업의 수를 결정하는데 있어 도산기업과 비도산기업에 대하여 대표기업의 수를 1개씩 증가시켜 이에 상응하는 모형들을 만들고, 이들을 서로 비교하여 최적 대표기업의 수를 결정함에 따라 많은 시간과 노력이 요구되었다. 따라서 향후 연구에서는 최적 대표기업의 수를 결정할 수 있는 방법을 수리적으로 모형에 반영하기 위한 연구가 필요할 것이다.

셋째, 본 연구에서 제안한 기법과 기존 기법들과의 성과 비교에 있어 통계적인 검정을 수행하지 않았다. 본 연구에서 기법들 간의 성과를 비교한 목적은 본 연구의 기법이 기존의 기법들에 비해 예측성과가 뛰어남을 강조하기 위한 것은 아니었다. 오히려 본 연구에서 제안한 기법이 가지는 방법론적, 실무적 장점에도 불구하고 그 예측력이 기존의 기법들에 비해 떨어진다면 그러한 장점들이 퇴색되기 때문에, 특정 자료를 이용한 비교 실험을 통하여 본 연구의 기법이 기존의 기법들에 비해 예측력이 떨어지지 않음을 보여주고자 한 것이었다. 그러나 본 연구에서는 기법 간의 성과 비교에 통계적인 검정 절차를 수반하지 않았으므로 그 결과를 일반화시키기에는 한계가 있다. 향후 연구에서는 실험계획을 통하여 본 연구의 기법이 다른 기법과 비교하여 어떠한 성과 차이를 보이는지, 그리고 그 차이는 통계적으로 유의한 지에 대한 검정이 필요할 것이다.

참 고 문 헌

[1] 문병로, 「유전 알고리즘」, 1판, 다성출판사, 2001.
[2] 민재형, 정철우, "유전 알고리듬 기반 집단분류

기법의 개발과 성과평가 : 채권등급 평가를 중심으로", 「한국경영과학회지」, 제32권, 제1호(2007), pp.61-76.

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

[4] Atiya, A.F., "Bankruptcy prediction for credit risk using neural networks : A survey and new results," *IEEE Transactions on Neural Networks*, Vol.12, No.4(2001), pp.929-935.

[5] Bandyopadhyay, S. and U. Maulik, "Genetic clustering for automatic evolution of clusters and applicaton to image classification," *Pattern Recognition*, Vol.35, No.6(2002), pp. 1197-1208.

[6] Beaver, W.H., "Financial Ratios and Predictions of Failure," *Journal of Accounting Research*, Vol.4, Supplement(1966), pp.71-111.

[7] Bell, T.B., "Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 6, No.3(1997), pp.249-264.

[8] Bryant, S.M., "A case-based reasoning approach to bankruptcy prediction modeling," Intelligent Systems in Accounting, *Finance and Management*, Vol.6, No.3(1997), pp.195-214.

[9] Dimitras, A.I., R. Slowinski, R. Susmaga, and C. Zopounidis, "Business failure prediction using rough sets," *European Journal of Operational Research*, Vol.114, No.2(1999), pp.263-280.

[10] Frydman, H., E.I. Altman, and D. Kao, "Introducing recursive partitioning for financial classification : The case of financial distress," *Journal of Finance*, Vol.40, No.1 (1985), pp.269-291.

[11] Holland, J.H., *Adaptation in natural and artificial systems*, The University of Michigan Press, Ann Arbor, MI, 1975.

[12] Jo, H., I. Han, and H. Lee, "Bankruptcy prediction using case-based reasoning, neural network and discriminant analysis for bankruptcy prediction," *Expert Systems with Applications*, Vol.13, No.2(1997), pp.97-108.

[13] Kaski, S., J. Sinkkonen, and J. Peltonen, "Bankruptcy analysis with self-organizing maps in learning metrics," *IEEE Transaction on Neural Networks*, Vol.12, No.4 (2001), pp.936-947.

[14] Kumar, P.R. and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review," *European Journal of Operational Research*, Vol.180, No.1(2007), pp.1-28.

[15] Lacher, R.C., P.K. Coats, S.C. Sharma, and L.F. Fantc, "A neural network for classifying the financial health of a firm," *European Journal of Operational Research*, Vol.85, No.1(1995), pp.53-65.

[16] Lam, M., "Neural networks techniques for financial performance prediction : integrating fundamental and technical analysis," *Decision Support Systems*, Vol.34, No.4 (2004), pp.567-581.

[17] Lee, K.C., I. Han, and Y. Kwon, "Hybrid neural network models for bankruptcy predictions," *Decision Support Systems*, Vol. 18, No.1(1996), pp.63-72.

[18] Lee, K, D. Booth, and P. Alam, "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of

Korean firms," *Expert Systems with Applications*, Vol.29, No.1(2005), pp.1-16.

[19] Leshno, M. and Y. Spector, "Neural network prediction analysis : The bankruptcy case," *Neurocomputing*, Vol.10, No.2(1996), pp.125-147.

[20] Lin, H.J., F.W. Yang and Y.T. Kao, "An Efficient GA-based Clustering Technique," *Tamkang Journal of Science and Engineering*, Vol.8, No.2(2005), pp.113-122.

[21] Marais, M.L., J. Patel, and M. Wolfson, "The experimental design of classification models : An application of recursive partitioning and bootstrapping to commercial bank loan classifications," *Journal of Accounting Research*, Vol.22, Supplement(1984), pp.87-114.

[22] McKee, T.E., "Developing a bankruptcy prediction model via rough sets theory," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol.9, No.3(2000), pp.59-173.

[23] McKee, T.E., "Rough sets bankruptcy prediction models versus auditor signaling rates," *Journal of Forecasting*, Vol.22, No.8(2003), pp.569-589.

[24] Ohlson, J.A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol.18, No.1 (1980), pp.109-131.

[25] Park, C.S. and I. Han, "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction," *Expert Systems with Applications*, Vol.23, No.3(2002), pp.255-264.

[26] Salchenberger, L., C. Mine, and N. Lash, "Neural networks : A tool for predicting thrift failures," *Decision Sciences*, Vol.23, No.4(1992), pp.899-916.

[27] Shin, K.S. and Y.J. Lee, "A genetic algorithm application in bankruptcy prediction modeling," Expert Systems with Applications, Vol.23, No.3(2002), pp.321-328.

[28] Swicegood, P. and J.A. Clark, "Off-site monitoring systems for predicting bank underperformance : A comparison of neural networks, discriminant analysis and professional human judgment," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol.10, No.3(2001), pp.169-186.

[29] Tam, K.Y., "Neural network models and the prediction of bank bankruptcy," *Omega*, Vol.19, No.5(1991), pp.429-445.

[30] Varetto, F., "Genetic algorithm applications in the analysis of insolvency risk," *Journal of Banking and Finance*, Vol.22, No.10-11 (1998), pp.1421-1439.

[31] Wilson, R.L. and R. Sharda, "Bankruptcy prediction using neural networks," Decision *Suppot Systems*, Vol.11, No.5(1994), pp.545-557.

[32] Yang, Z.R., M.B. Platt, and H.D Platt, "Probability neural network in bankruptcy prediction," *Journal of Business Research*, Vol.44, No.2(1999), pp.67-74.

[33] Zimmermann, H.J., *Fuzzy set theory and its applications*, Kluwer Academic Publishers, London, 1996.

[34] Zmijewski, M.E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, Vol.22, Supplement(1984), pp.59-82.