

# 기업도산예측을 위한 이진분류기법의 개발 Developing a Binary Classification Method for Bankruptcy Prediction

민재형<sup>a</sup> · 정철우<sup>b</sup>

<sup>a</sup> College of Business Administration, Sogang University, Seoul, Korea 121-742  
Tel: +82-2-705-8545, Fax: +82-2-715-8505, E-mail: jaemin@sogang.ac.kr

<sup>b</sup> College of Business Administration, Sogang University, Seoul, Korea 121-742  
Tel: +82-10-5605-0529, Fax: +82-2-715-8505, E-mail: manibyul@hanmail.net

## Abstract

본 연구는 유전 알고리즘에 기반한 새로운 도산예측기법을 개발하고 그 기법의 타당성 및 예측 우수성을 검증하는데 목적이 있다. 본 연구에서 제안하는 이진분류기법은 도산기업과 비도산기업을 대표할 수 있는 가상기업(virtual company)을 설정하고, 그 가상기업과 분류대상 기업 간의 유사도를 측정하여 도산여부를 분류하는 방법론으로, 가상기업의 변수 값과 각 변수의 가중치는 훈련용 자료의 분류정확도를 극대화할 수 있도록 유전 알고리즘을 이용하여 구하게 된다. 본 연구에서 제안하는 기법의 타당성을 검증하기 위해 기존의 도산예측기법과 예측성표를 실험을 통해 비교한 결과, 본 연구에서 개발한 기법의 예측력이 기존의 다변량판별분석, 로지스틱 회귀모형, 의사결정나무, 인공신경망 모형보다 높은 수준을 보이는 것을 확인하였다.

## 1. 서론

1997년 말 IMF 외환위기부터 2007년 국제결제은행(BIS)의 바젤 II 협약에 이르기까지의 급속한 금융환경의 변화로 인하여 우리나라 은행들의 위험관리에 대한 관심은 그 어느 때보다도 증대되고 있다. 이러한 상황에서 금융기관의 여신심사 기능 강화를 위한 논의가 활발히 진행되고 있는데, 특히 기업도산예측은 여신심사 기능의 핵심을 이룬다고 할 수 있다. 그 이유는 정확한 기업도산예측을 통하여 금융기관은 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 도산예측의 정확도 또한 떨어지고 있다. 따라서 기업도산예측을 위한 많은 기법들이 이미 나와 있음에도 불구하고 기존의 기법으로 반영하지 못했던 기업도산의 원인과 징후를 반영할 수 있는 새로운 모형을 개발하기 위한 연구는 끊임없이 이루어지고 있다.

본 연구에서는 인공지능기법의 일종인 유전 알고리즘(genetic algorithm)을 응용하여 기업도산예측을 위한 새로운 이진분류기법을 개발하였다. 그리고 실증분석을 통해 기존 기법과 예측력을 비교하고 그 우월성을 확인하였다.

## 2. 이론적 배경

기업도산예측을 위한 방법론적 발달과정은 크게 두 가지로 나누어 볼 수 있다. 하나는 Beaver(1966)의 단일변량 판별분석 이후 시작된 통계기반 기법으로 이 기법은 Altman(1968)의 다변량 판별분석을 거쳐, 로짓(Olshon, 1980) 및 프로빗(Zmijewski, 1984)과 같은 확률모형으로 발전하였다.

다른 하나는 1990년대부터 최근에 이르기까지 연구가 진행되고 있는 인공지능기법이다. 인공지능기법은 구체적으로 BPNN(back

propagation trained neural network)(Tam, 1991; Salchenberger, 1992; Wilson & Sharda, 1994; Leshno & Spector, 1996; Bell, 1997; Atiya, 2001; Swicegood & Clark, 2001; Lam, 2004; Lee et al., 2005), PNN(probabilistic neural networks)(Yang et al., 1999), SOM(self-organizing map)(Lee et al., 1996; Serrano-Cinca, 1996; Kaski et al., 2001; Lee et al., 2005), Cascor(cascade correlation neural network)(Lacher et al., 1995) 등과 같은 인공신경망의 다양한 방법들과 Fuzzy Set Theory(Zimmermann, 1996), 의사결정나무(Marais et al., 1984; Frydman et al., 1985), 사례기반추론(Case-based Reasoning)(Bryant, 1997; Jo et al., 1997; Park & Han, 2002), 유전 알고리즘(genetic algorithm: GA)(Varetto, 1998; Shin & Lee, 2002), SVM(Support Vector Machine)(Min & Lee, 2005), DEA(data envelopment analysis)(민재형과 정철우, 2006), Rough Sets Theory(Dimitras et al., 1999; McKee, 2000; McKee, 2003) 등과 같은 기법들을 포함하고 있다.

본 연구에서는 기업도산예측을 위한 방법론적 발달과정 중에서 인공지능기법의 일종인 유전 알고리즘을 응용하여 기업도산예측을 위한 새로운 이진분류기법을 제안한다. 본 연구의 모형은 민재형과 정철우(2007)의 모형을 기업도산예측에 맞추어 수정, 설계한 것으로, 이들은 군집분석의 기본 원리를 이용하였지만, 군집분석과는 달리 종속변수를 모형에 포함시킴으로써 실제군집에 대한 예측군집이 일치하는 비율을 최대화하는 새로운 그룹핑 기법을 제안하였다.

한편, 본 연구의 모형은 예측기업과 실제기업의 유사도를 기준으로 예측결과를 도출하는 방법론이라는 점에서 사례기반추론(CBR: case-based reasoning)과 공통점을 가진다. 그러나 본 연구의 모형은 다음과 같은 몇 가지 점에서 사례기반추론과 차별화된다.

첫째, 사례기반추론은 국지적(local) 학습방법인데 비해, 본 연구의 모형은 전체적(global) 학습방법이다. 즉, 사례기반추론은 각 평가대상 기업과 기존 기업 모두를 비교해서 가장 가까운 k개의 기업을 참조해 예측결과를 생성하는 국지적 학습방법을 이용한다. 반면, 본 연구에서 제시하는 모형은 유전 알고리즘을 이용하여 대표기업의 값을 구한 후, 대표기업과 평가대상 기업 간의 거리가 가장 가까운 대표기업의 도산여부에 따라 평가대상 기업을 분류하는 전체적 학습방법을 이용한다.

둘째, 본 연구의 모형은 사례기반추론에 비해 훨씬 효율적인 방법론이다. 사례기반추론은 속성변수별 가중치를 최적화하기 위해 훈련용 데이터를 두 세트로 나눌 필요가 있는데, 예를 들어, 1,000개의 데이터가 있다고 할 때 이를 6:4 비율로 나누어 600개는 입력사례로, 400개는 가중치를 최적화하기 위한 사례로 이용한다고 가정하자. 그러면 임의의 가중치에 대해서 계산해야 할 거리의 개수는  $600 \times 400 = 240,000$ 개가 된다. 또한 가중치가 최적화되었다고 할 때, 하나의 개별사례에 대해 계산해야 할 거리의 개수는 600개가 된다. 반면, 본 논문에서 제시하는 모형은 가중치를 계산하기 위해서 훈련용 데이터의 각 사례와 대표기업과의 거리만을 계산하면 되므로, 위의 예를 그대로 이용한다고 할 때 대표기업의 수를 8이라고 한다면 임의의 가중치에 대해서 계산해야 할 거리의 개수는  $8 \times 1,000 = 8,000$ 개로 감소한다. 또한 가중치가 최적화되었을 경우

에 계산해야 할 거리의 개수도 8개로 감소한다. 따라서 위의 예의 경우, 본 연구의 모형은 사례기반추론과 비교할 때 가중치를 찾기 위해 필요한 거리의 개수는 1/30, 최적화된 가중치를 이용하여 도산예측을 위해 계산해야 할 거리의 개수는 1/75에 불과한 매우 효율적인 방법이다. 그리고 이러한 차이는 가중치의 개수, 평가대상 기업의 개수, 기존 기업의 개수가 증가함에 따라 더욱 현격하게 벌어지게 된다. 또한 이러한 차이로 인해 도산예측에 걸리는 시간도 그만큼 차이가 나게 된다.

### 3. 수리적 모형

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

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

여기서,  $d_{ki}$ : 대표기업  $k$ 와 관측값  $i$  사이의 거리

( $i = 1, 2, \dots, N$ )

$X_{kj}$ : 대표기업  $k$ 의  $j$ 번째 변수값

( $j = 1, 2, \dots, M$ )

$X_{ij}$ : 관측값  $i$ 의  $j$ 번째 변수값

$w_j$ :  $j$ 번째 변수에 대한 가중치

식 (1)에서  $d_{ki}$ 는 관측값  $i$ 에 대해서 대표기업의 개수만큼 계산되어진다. 그 중에서 가장 작은 값( $d_{k^*i}$ )에 상응하는 대표기업( $k^*$ )의 도산여부에 따라 관측값  $i$ 를 분류한다. 즉,  $d_{k^*i}$ 에 상응하는 대표기업( $k^*$ )이 도산기업이면 관측값  $i$ 를 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 알고리즘을 이용하여 가중치( $w_j$ )와 대표기업의 변수( $X_{kj}$ )를 찾는다. 이러한 과정을 수식으로 나타내면 식 (2)와 같다.

$$\begin{aligned} \text{Max } H &= \frac{1}{N} \sum_{i=1}^N C_i \\ \text{subject to} \\ C_i &= 1 \quad \text{if } B(i) = B(k^*) \quad \forall i \in \{1, 2, \dots, N\}, \\ C_i &= 0 \quad \text{otherwise,} \\ d_{k^*i} &= \min(d_{ki}) \end{aligned} \quad (2)$$

여기서,  $H$ : 실제그룹과 예측그룹의 적중률

$N$ : 관측값 수

$C_i$ : 관측값  $i$ 의 실제그룹과 예측그룹의 일치여부

( $i = 1, 2, \dots, N$ )

$B(i)$ : 관측값  $i$ 의 실제 도산여부 ( $B(i) = 0, 1$ )

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

$k^*$ :  $d_{k^*i}$ 에 상응하는 대표기업

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

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

## 4. 실증분석

### 4.1 데이터 정리

본 연구에서는 실증분석을 위해 2001년부터 2004년까지의 우리나라 비외국 중소 제조업체를 대상으로 도산기업과 비도산기업에 대해 재무비율 자료를 수집하되, 도산기업에 대해서는 도산직전년도 자료를 수집하였다. 총 2,814개 기업의 27개 재무비율 자료에 대해 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 45개의 비도산기업을 무작위로 추출하여 제거함으로써 도산과 비도산기업의 수를 1,271개로 동일하게 맞추었다.

<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 변수에 대한 정의를 나타내고 있다. 변수선정을 위해 본 연구에서는 독립표본 t-검정, 판별분석, 요인분석 등 여러 가지 통계기법들을 다각적으로 이용하였다.

<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 로지스

<표 2> 변수 정의

변수	정의
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	순익분기점율
X25	인건비
X26	총자산순이익률
X27	EBIT 대 매출액

<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

틱 회귀분석에서 단계별선택법(stepwise selection method)의 진 입값은 F값의 유의확률 기준으로 0.01, 제거값은 0.05로 하였다. 의사결정나무에서 분리기준(splitting criterion)은 유의수준 0.02에 서 Chi-square 검정으로 하였고, 최대 뿌리 깊이는 6으로, 한 잎 에서 최소 관측값의 수는 5로 하였다.

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

<표 4> 재무비율 변수들의 회전성분행렬

변수	요인							
	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	0.050	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	0.855	-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)

이상의 분석 결과를 토대로 최종적인 변수를 선정하였는데, 변수 선정과정을 단계별로 요약하면 다음과 같다. 우선, 요인분석 결과 묶여진 각 요인의 변수들 중 판별분석의 단계별선택법, 로지스틱 회귀분석의 단계별선택법, 그리고 의사결정나무 분석 결과 도출된

<표 5> 변수선정단계

		요인							
		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	

규칙을 통해 선정된 변수들만을 선택하였다. 다음으로, 선택된 변수들 중에서 독립표본 t-검정 결과 유의한 변수로 선정되지 않은 변수는 제외하였다. 마지막으로, 각 요인에 속한 남은 변수들 가운데 의미상 중복되는 변수는 제외하였다. <표 5>는 이러한 변수선 정의 단계를 정리한 것이다.

<표 5>의 3단계에서 요인5에 해당하는 X15(총자산회전율)와 X21(경영자산회전율) 중 X15만을 선정한 것은 두 변수의 의미가 유사하다고 판단하였기 때문이다. 반면, 요인4에 해당하는 X12(유 동부채대총자산)와 X13(자기자본비율)은 기업의 도산여부에 서로 다르게 영향을 미치는 것으로 판단하여 둘 다 변수로 선정하였다. 또한 요인6에 해당하는 X20(감가상각비), X23(순금융비용), X25 (인건비) 역시 서로 다른 성격을 가진 비용으로 판단하여 모두 변수로 선정하였다. 그리하여 최종적으로 선정된 변수는 X2(총자본투 자효율), X3(총자산증가율), X12(유동부채대총자산), X13(자기자본 비율), X15(총자산회전율), X20(감가상각비), X23(순금융비용), X24(손익분기점율), X25(인건비) 등 9개 변수이다.

## 4.2 모형의 설계

본 연구에서는 도산·비도산 기업을 대표할 수 있는 대표기업의 수를 각각 1개부터 5개까지 늘려가면서 예측정확도의 추이를 살펴 최적의 대표기업 수를 구하였다. 대표기업의 수가 도산·비도산 기 업에 대해 각각 1개인 것을 모형1, 2개인 것을 모형2, 3개인 것을 모형3, 4개인 것을 모형4, 5개인 것을 모형5로 하여 실험을 진행하 였다. 이렇게 대표기업의 수를 변화시켜가며 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.

유전 알고리즘을 적용하는데 있어 본 연구모형이 가지는 염색체 (chromosome) 구조를 설명하면 다음과 같다. 각 염색체는 가중치 에 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 표 준정규분포를 따르는 확률변수를 발생시켜 이용하였다.

최적해를 찾기 위해 population 크기는 100으로 하였고, mutation rate는  $w_j$ 와  $X_{kj}$ 에 대해서 모두 0.1로, cross over rate 는  $w_j$ 와  $X_{kj}$ 에 대해서 모두 0.5로 하였다. Stopping condition은 '수동'으로 설정하였으며 최적해의 수렴 여부를 분석자가 판단하여 탐색을 멈추도록 하였다. 분석도구로는 유전 알고리즘 소프트웨어 인 Evolver 4.0을 사용하였다.

## 4.3 분석결과의 논의

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

<표 6> 모형별 최적해 탐색 횟수

모형	총시행횟수	최적해 탐색 시점의 시행횟수
1	3,955	1,898
2	21,150	14,088
3	24,368	22,366
4	21,497	12,829
5	21,383	15,470

<표 7>은 모형별 예측정확도를 정리한 것이다. <표 7>을 보면 훈련용 자료에 대해서는 대표기업의 수가 증가하면서 예측정확도는

<표 7> 모형별 예측정확도

모형	대표기업수	훈련용	검증용
1	2	75.8%	75.8%
2	4	76.6%	76.4%
3	6	77.1%	76.6%
4	8	77.5%	77.8%
5	10	78.9%	77.4%

계속해서 증가하는 추이를 볼 수 있다. 이에 반해 검증용 자료의 경우에는 대표기업의 수가 증가하면서 예측정확도는 증가하다가 대표기업 수 8을 정점으로 이후에는 오히려 예측정확도가 떨어지는 현상을 볼 수 있다. 이러한 현상은 대표기업의 수가 증가함에 따라 자료에 대한 모형의 적합도는 증가하지만, 일정 수 이상으로 대표기업의 수가 증가할 경우, 훈련용 자료에 대해서는 모형이 적합하지만 검증용 자료에 대해서는 오히려 예측정확도가 떨어지는 이른바 과적합 문제가 발생한 것으로 해석된다.

이제 예측정확도가 가장 뛰어난 모형4의 추가 분석결과를 기술하면 다음과 같다. <표 8>은 모형4의 변수별 표준 가중치를 나타낸 것으로 X13, X20, X23, X24, X25가 비교적 높은 수준을 보였고, X2, X3, X12, X15는 비교적 낮은 수준을 보였다. 이는 관측값들을 대표기업과의 거리에 따라 분류하는데 있어 상대적으로 큰 영향을 미친 변수들이 X13(자기자본비용), X20(감가상각비), X23(순금융비용), X24(손익분기점율), X25(인건비)임을 의미한다.

<표 9>는 모형4의 대표기업별 변수에 대한 최적해를 나타내고 있다. 여기서 대표기업 1부터 4까지는 비도산기업의 대표기업이고 대표기업 5부터 8까지는 도산기업의 대표기업이다.

이와 같이 구한 가중치와 변수값을 가지고 관측값들을 분류한 결

과는 <표 10>과 같다. <표 10>에서 훈련용 자료에 대한 결과를 보면 모형4의 8개 대표기업 중에서 6개에 대해서만 관측값들이 분류되었다. 그 중 대표기업 4에는 관측값이 2개, 대표기업 7과 8에는 관측값이 1개만이 동일 기업군으로 분류되고, 나머지 관측값들은 대표기업 1, 2, 6과 동일 기업군으로 분류된 것을 볼 수 있다. 검증용 자료에 대해서도 대부분의 관측값들이 대표기업 1, 2, 6과 동일 기업군으로 분류된 결과를 보였다.

이러한 결과를 근거로 대표기업의 수를 비도산기업에 대해 2개, 도산기업에 대해 1개만으로 하는 추가적 모형을 만들어 분석을 시행하면 모형4와 비슷한 예측정확도 및 최적해를 더 빠른 시간 내에 얻어낼 수 있을 것이라는 추측에 따라 추가적인 실험을 수행하였다. 총 21,468번의 시행을 통해 12,628개 시행에서 최적해를 구하였고, 그 결과는 <표 11> 및 <표 12>와 같다.

<표 11>은 추가적인 실험 모형에 대한 변수별 최적 가중치이다. 이를 모형4에 대한 결과(<표 8> 참조)와 비교했을 때, 모형4에서는 X13, X20, X23, X24, X25가 비교적 높은 수준을 보이는 변수

<표 10> 모형4의 관측값 분류 결과

실제 도산여부	대표기업							합계
	1	2	4	6	7	8		
훈련용	0	869	49	2	97	-	-	1017
	1	344	16	-	655	1	1	1017
	합계	1213	65	2	752	1	1	2034
검증용	0	208	18	-	27	-	1	254
	1	79	6	-	169	-	-	254
	합계	287	24	-	196	-	1	508

<표 8> 모형4의 변수 가중치

가중치	X2	X3	X12	X13	X15	X20	X23	X24	X25
$w_j / \sum_j w_j$	0.0520	0.0679	0.0189	0.1365	0.0078	0.2067	0.1582	0.1678	0.1842
$w_j$	0.1388	0.1815	0.0505	0.3646	0.0207	0.5523	0.4225	0.4482	0.4921

<표 9> 모형4의 대표기업별 변수 값

변수 대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1	-0.2170	-1.9132	0.0819	-0.2292	-1.5985	0.9406	1.2960	0.4204	1.2021
2	-1.4331	-2.3874	-0.9751	-0.0216	-1.1404	1.3888	1.3813	-1.5507	1.0174
3	-2.2204	1.6757	-2.3877	-2.3578	1.6868	-2.4887	0.6981	2.7685	1.6334
4	-1.0389	-1.2439	-1.6535	-2.7233	0.1774	-0.8597	-0.9763	2.7528	0.2030
5	1.9928	-0.7657	2.0836	-0.5866	1.5798	0.1230	0.1954	-0.7016	-2.5379
6	-1.4379	-0.1819	0.2775	0.9826	0.0838	2.8903	2.3139	1.0056	1.9763
7	0.6863	0.2016	-1.3127	0.2898	0.1606	-1.5441	0.8101	2.8140	1.5402
8	-2.5001	-1.7576	2.1868	1.2720	2.4976	-2.9720	-2.4994	-2.4209	2.2762

<표 11> 추가적 모형의 변수 가중치

가중치	X2	X3	X12	X13	X15	X20	X23	X24	X25
$w_j / \sum_j w_j$	0.0459	0.0712	0.0942	0.0508	0.0723	0.2159	0.1277	0.1124	0.2096
$w_j$	0.1830	0.2836	0.3753	0.2024	0.2879	0.8603	0.5086	0.4477	0.8352

<표 12> 추가적 모형의 대표기업별 변수 값

변수 대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1	-0.3764	-0.1279	1.2923	-2.8810	0.0643	0.7780	1.8540	0.4338	2.0366
2	-0.5120	-2.0394	1.6914	1.2092	-0.7629	2.3265	1.1731	-1.2252	1.3936
3	2.4773	0.9714	1.5091	-2.0950	2.1520	-0.2795	1.1757	-0.3802	-2.3241

<표 13> 대표기업별 가중 변수 값

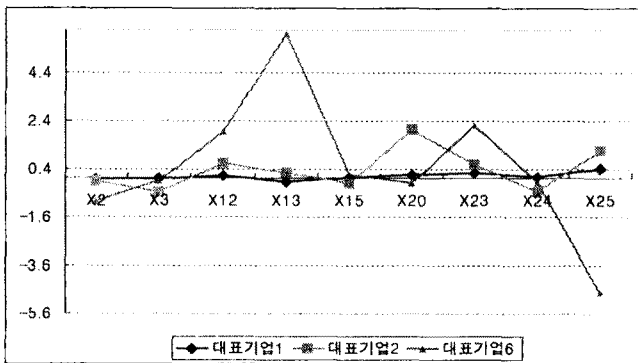
대표기업	X2	X3	X12	X13	X15	X20	X23	X24	X25
1	-0.0113	-0.1300	0.0015	-0.0313	-0.0124	0.1945	0.2050	0.0705	0.2215
2	-0.0745	-0.1622	-0.0184	-0.0029	-0.0089	0.2871	0.2185	-0.2602	0.1874
6	-0.0747	-0.0124	0.0052	0.1341	0.0007	0.5976	0.3660	0.1687	0.3641

였는데, 추가적인 모형에서는 이 중 X13이 빠져 있는 것을 알 수 있다.

한편, <표 12>는 추가적 모형의 대표기업별 변수 값을 나타내고 있다. 이를 모형4의 대표기업 1, 2, 6과 변수별로 비교해 보면 다소 차이가 나는 것을 볼 수가 있다. 예를 들어, X20의 값을 보면 모형4에서는 대표기업 1, 2의 값이 대표기업 3의 값에 비해 작은데, 추가적 모형에서는 그 반대의 현상을 보이는 것을 알 수 있다. 그 결과, 추가적 모형의 예측정확도 역시 모형4의 예측정확도와 차이를 보였는데, 추가적 모형의 예측정확도는 훈련용 자료에 대해서는 77.5%, 검증용 자료에 대해서는 76.2%로 나타나 모형4의 예측정확도에 비해 낮게 나타났다.

이러한 현상이 일어난 이유는 다음과 같이 설명할 수 있다. 모형4에서는 대표기업 1부터 4까지와 동일한 기업군에 해당하는 관측값은 비도산으로 분류되고, 대표기업 5부터 8까지와 동일한 기업군에 해당하는 관측값은 도산으로 분류되는데, 추가적 모형에서는 대표기업 1, 2와 동일한 기업군에 해당하는 관측값은 비도산으로 분류되고, 대표기업 3과 동일한 기업군에 해당하는 관측값은 도산으로 분류된다. 즉, 모형4에서는 비도산기업과 도산기업 각각에 대해서 4개씩의 대표기업 후보들이 예측정확도를 최대화하는데 기여하도록 경쟁을 하고, 추가적 모형은 비도산기업과 도산기업 각각에 대하여 대표기업의 수를 2개와 1개로 정한 상태에서 변수 값과 가중치를 찾도록 설계되었다. 따라서 모형4의 분석에서 도출된 대표기업 1, 2, 6의 값은 다른 후보들을 경쟁하여 이긴 값이고, 추가적 모형에서의 대표기업 값은 그러한 경쟁이 없이 나온 값이라는 차이가 있다. 결과적으로 보면 경쟁을 통해 도출된 값인 모형4의 해가 예측정확도의 최대화라는 측면에서 더 우수한 해임을 확인할 수 있었다. 다만, 대표기업의 수가 더 많은 모형5의 해가 모형4의 해보다 예측정확도 측면에서 낮은 수준을 보인 것은 전술한 바와 같이 모형5의 경우 검증용 자료에 대해 과적합 문제가 발생하였기 때문이다.

모형4의 분석결과 도출된 대표기업 1, 2, 6에 대하여 변수별 가중치와 변수 값을 곱하여 구한 가중 변수 값을 정리하면 <표 13> 및 <그림 1>과 같다.



[그림 1] 대표기업별 가중 변수 값

[그림 1]을 보면 도산기업에 속하는 대표기업 6은 X12(유동부채대총자산), X13(자기자본비율), X23(순금융비용)의 값은 높은 반면, X25(인건비)의 값은 매우 낮은 특징을 갖고 있음을 알 수 있다. 따라서 이러한 정보는 여신심사에 있어 유용하게 활용될 수 있을 것이다. 즉, 여신심사가 필요한 어떠한 기업이 유동부채대총자산, 자기자본비율, 순금융비용이 동일 산업내의 다른 기업들에 비해 높고,

인건비가 상대적으로 낮은 기업이라면 일단 도산의 가능성을 의심할 필요가 있으며, 실제로 그 기업과 대표기업들과의 거리를 측정하여 대표기업 6과 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 설계는 다음과 같은 방식으로 하였다. 모든 분석기법에 대해 변수는 본 연구와 동일하게 9개의 변수를 이용하였다. 자료는 본 연구의 모형을 만들기 위해 이용한 훈련용 자료를 8:2의 비율로 나누어 훈련용 자료와 테스트용 자료로 이용하였고, 검증용 자료는 그대로 검증용 자료로 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 14> 기법별 예측정확도

	훈련용	검증용
본 연구의 기법	77.5%	77.8%
판별분석	70.2%	69.1%
로지스틱 회귀분석	71.2%	70.7%
의사결정나무	76.5%	76.8%
인공신경망	78.1%	76.4%

<표 14>는 본 연구에서 개발한 기법과 기존 이진분류기법의 예측정확도를 비교한 것으로, 본 연구에서 제안한 기법이 검증용 자료를 기준으로 볼 때, 가장 높은 예측정확도를 보이는 것으로 나타났다.

#### 5. 결론

본 연구는 새로운 이진분류기법의 제안과 함께 실제 기업 자료를 이용하여 실증분석하고, 그 예측력을 기존의 기법과 비교함으로써 그 타당성을 확인하였다. 본 연구의 결과는 금융기관의 신용위험관리 측면과 새로운 방법론 개발이라는 학문적 측면에서 다음과 같은 기대효과를 갖는다.

첫째, 본 연구에서 개발한 이진분류기법은 금융기관의 신용위험 관리에 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 뿐

만 아니라 상품구매예측, 프로젝트 위험관리 등 다양한 분야에 확대 응용될 수 있을 것으로 기대한다.

## 참고문헌

- [1] 민재형, 정철우 (2006), "비모수기법을 이용한 기업도산예측 - DEA와 Negative DEA 교차증화기법의 개발", 「경영학연구」, 제35권, 제4호, pp.1157-1180.
- [2] 민재형, 정철우 (2007), "유전 알고리즘 기반 집단분류기법의 개발과 성과평가: 채권등급 평가를 중심으로", 「한국경영과학회지」, 제32권, 제1호, pp.61-76.
- [3] Altman, E.I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, Vol.23, No.4, pp.589-609.
- [4] Atiya, A.F. (2001), "Bankruptcy prediction for credit risk using neural networks: A survey and new results", *IEEE Transactions on Neural Networks*, Vol.12, No.4, pp.929-935.
- [5] Bandyopadhyay, S. and U. Maulik (2002), "Genetic clustering for automatic evolution of clusters and application to image classification", *Pattern Recognition*, Vol.35, No.6, pp.1197-1208.
- [6] Beaver, W.H. (1966), "Financial Ratios and Predictions of Failure", *Journal of Accounting Research*, Vol.4, Supplement, pp.71-111.
- [7] Bell, T.B. (1997), "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, pp.249-264.
- [8] Bryant, S.M. (1997), "A case-based reasoning approach to bankruptcy prediction modeling", *Intelligent Systems in Accounting, Finance and Management*, Vol.6, No.3, pp.195-214.
- [9] Dimitras, A.I., R. Slowinski, R. Susmaga, C. Zopounidis (1999), "Business failure prediction using rough sets", *European Journal of Operational Research*, Vol.114, No.2, pp.263-280.
- [10] Frydman, H., E.I. Altman, D. Kao (1985), "Introducing recursive partitioning for financial classification: The case of financial distress", *Journal of Finance*, Vol.40, No.1, pp.269-291.
- [11] Jo, H., I. Han, H. Lee (1997), "Bankruptcy prediction using case-based reasoning, neural network and discriminant analysis for bankruptcy prediction", *Expert Systems with Applications*, Vol.13, No.2, pp.97-108.
- [12] Kaski, S., J. Sinkkonen, J. Peltonen (2001), "Bankruptcy analysis with self-organizing maps in learning metrics", *IEEE Transaction on Neural Networks*, Vol.12, No.4, pp.936-947.
- [13] Kumar, P.R., V. Ravi (2007), "Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review", *European Journal of Operational Research*, Vol.180, No.1, pp.1-28.
- [14] Lacher, R.C., P.K. Coats, S.C. Sharma, L.F. Fantc (1995), "A neural network for classifying the financial health of a firm", *European Journal of Operational Research*, Vol.85, No.1, pp.53-65.
- [15] Lam, M. (2004), "Neural networks techniques for financial performance prediction: integrating fundamental and technical analysis", *Decision Support Systems*, Vol.34, No.4, pp.567-581.
- [16] Lee, K.C., I. Han, Y. Kwon (1996), "Hybrid neural network models for bankruptcy predictions", *Decision Support Systems*, Vol.18, No.1, pp.63-72.
- [17] Lee, K., D. Booth, P. Alam (2005), "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms", *Expert Systems with Applications*, Vol.29, No.1, pp.1-16.
- [18] Leshno, M., Y. Spector (1996), "Neural network prediction analysis: The bankruptcy case", *Neurocomputing*, Vol.10, No.2, pp.125-147.
- [19] Lin, H.J., F.W. Yang and Y.T. Kao (2005), "An Efficient GA-based Clustering Technique", *Tamkang Journal of Science and Engineering*, Vol.8, No.2, pp.113-122.
- [20] Marais, M.L., J. Patel, M. Wolfson (1984), "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, pp.87-114.
- [21] McKee, T.E. (2000), "Developing a bankruptcy prediction model via rough sets theory", *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol.9, No.3, pp.59-173.
- [22] McKee, T.E. (2003), "Rough sets bankruptcy prediction models versus auditor signaling rates", *Journal of Forecasting*, Vol.22, No.8, pp.569-589.
- [23] Min, J.H., Y.C. Lee (2005), "Bankruptcy prediction using support vector machine(SVM) with optimal choice of kernel function parameters", *Expert Systems with Applications*, Vol.28, No.4, pp.603-614.
- [24] Ohlson, J.A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy", *Journal of Accounting Research*, Vol.18, No.1, pp.109-131.
- [25] Park, C.S., I. Han (2002), "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction", *Expert Systems with Applications*, Vol.23, No.3, pp.255-264.
- [26] Salchenberger, L., C. Mine, N. Lash (1992), "Neural networks: A tool for predicting thrift failures", *Decision Sciences*, Vol.23, No.4, pp.899-916.
- [27] Shin, K.S., Y.J. Lee (2002), "A genetic algorithm application in bankruptcy prediction modeling", *Expert Systems with Applications*, Vol.23, No.3, pp.321-328.
- [28] Swicegood, P., J.A. Clark (2001), "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, pp.169-186.
- [29] Tam, K.Y. (1991), "Neural network models and the prediction of bank bankruptcy", *Omega*, Vol.19, No.5, pp.429-445.
- [30] Varetto, F. (1998), "Genetic algorithm applications in the analysis of insolvency risk", *Journal of Banking and Finance*, Vol.22, No.10-11, pp.1421-1439.
- [31] Wilson, R.L., R. Sharda (1994), "Bankruptcy prediction using neural networks", *Decision Support Systems*, Vol.11, No.5, pp.545-557.
- [32] Yang, Z.R., M.B. Platt, H.D Platt (1999), "Probability neural network in bankruptcy prediction", *Journal of Business Research*, Vol.44, No.2, pp.67-74.
- [33] Zimmermann, H.J. (1996), *Fuzzy set theory and its applications*, Kluwer Academic Publishers, London.
- [34] Zmijewski, M.E. (1984), "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", *Journal of Accounting Research*, Vol.22, Supplement, pp.59-82.