

Robust covariance-based Mahalanobis distance를 적용한 사례기반추론: 기업부도 예측

Memory-Based Reasoning using a Robust covariance-based Mahalanobis distance model: Corporate bankruptcy prediction

안 효 준* Ahn, Hyo Jun

조 성 빈** Cho, Sungbin

Abstract

This study proposes a robust covariance-based Mahalanobis distance model as a memory-based reasoning method in order to predict corporate bankruptcy. Samples are drawn from the small and medium sized manufacturing companies. Variables are selected by the logistic regression method and the decision tree induction method after the single sample t-test as a preliminary selection process. For these two different variable groups, three MBR models are evaluated: Euclidean distance model, Euclidean distance after standardization model, and robust covariance-based Mahalanobis distance model. 25 nearest neighbors are picked for a reference group and then a simple voting rule is applied to solve bankruptcy problem. The analysis results indicate that compared to existing Euclidean distance model, the proposed model produces higher correct classification ratios.

Keywords: Robust covariance, Mahalanobis distance, bankruptcy prediction

* 서강대학교 일반대학원 경영학과

** 교신저자, 서강대학교 경영학과 교수, 이메일: sungbincho@sogang.ac.kr

이 연구는 2012년도 서강대학교 교내연구비 지원에 의한 연구임(201210030.01).

투고일 2016.11.14

수정일 2016.12.15

게재일 2016.12.31

1. 서 론

지난 반세기 동안 경영학 분야에서는 기업의 재무 건전성과 부도 위험을 설명하고 예측하기 위하여 많은 연구가 행하여져왔다(Altman, 1968; Beaver, 1966). 전통적 통계기법인 판별분석과 로지스틱 회귀분석의 적용에서 시작하여 기계학습이 등장하며 인공지능망, Support vector machine, 유전자 알고리즘, 베이지안 네트워크, 의사결정나무 추론 등 다양한 기법들이 개발되어왔다(Lee et al., 1994; Pendharkar, 2005; 조성빈 외, 2008).

인공지능 모델의 한 분야인 사례기반추론(Memory-based reasoning: MBR)은 1990년대부터 사용되어 왔는데, 목표 변수의 값을 설명하고 예측함에 있어 데이터를 구성하는 모든 리코드를 사용하는 대부분의 모델들과 달리, 예측할 리코드와 가장 유사한 또는 가장 비유사도가 적은 소수의 리코드만을 선별하여 사용한다는 점에서 특이한 구조를 갖는다고 말할 수 있다. 가장 인접한 이웃(nearest neighbors)이라 불리는 이 소수집단을 어떠한 기준에 의하여 선별하고 그 규모는 얼마나 할 것인가에 관한 다양한 논문들이 발표되어 왔다.

기업 재무상태가 나빠지면 통상 그 최종 결과로 기업부도로 이어지는데, 한 산업 분야 예를 들어 제조업의 경우, 데이터를 구성하는 모든 기업을 대상으로 하여 특정 기업의 부도/비부도 예측 모델에 적용하는 것은, 기업이 가진 수많은 특성을 고려할 때 변별력이 떨어질 수도 있다. 따라서 본 연구에서는 예측하고자 하는 특정 기업과 가장 유사하다고 판단되는 소수의 기업만을 데이터에서 추출하여 이 집단의 재무 상태를 참고하여 결과값을 도출하는 사례기반추론이 기업부도 예

측에 적절한 모형이 될 가능성에 주목하였다.

본 연구에서는 기존의 몇몇 연구에서 도입되었던 기업 간 유사도를 측정하는데 사용되었던 Mahalanobis distance를 채택함에 있어(Cho et al., 2010), 일반적인 변수 간의 공분산 구조를 쓰지 않고 robust covariance matrix라 불리는 개념을 적용한 Mahalanobis distance를 측정함으로써 이 개념을 적용한 사례기반추론의 기업부도 예측에의 유효성을 시험하고자 하였다.

2. 선행 연구

건전한 재무구조를 가진 기업과 부도 판정이 난 기업 간의 차이를 설명하기 위한 노력은 오래 전에 시작되었는데, 단일변량 통계분석을 이용한 두 그룹 간의 차이를 분석했던 Beaver(1966)의 연구와 함께 Altman(1968)의 다변량판별 분석이 시도되었고, 여러 연구에서 판별분석모형을 채택하였다(Eisenbeis, 1978; Falbo, 1991). 재무 건전성에 관한 연구가 중요한 이슈로 부각되며 좀 더 다양한 통계적 모형들이 연구에 사용되기 시작하였고(Frydman et al., 1985; Ohlson, 1980; Srinivasan & Kim, 1998), 1980년대 소개된 인공지능 기법들도 부도 예측에 활용되었는데 상당 기간 동안 인공지능망 기법의 다양한 알고리즘이 개발되며 기존의 모델보다 약간 향상된 성과를 보여주었다(Barniv & Agrawal, 1997; Coates & Fant, 2000; Malhotra & Malhotra, 2002). 하지만, black box approach를 사용하는 인공지능망 모형은 관리적 시사점을 주기 어려워 실무에서의 활용성은 한계를 보이는 것으로 판단된다.

사례기반추론은 선명한 선행회귀식이나 복잡한

비선형모형과 같은 모델을 쓰기보다 기존 데이터에 저장된 지식과 정보에 근거하여 해답을 찾는 상대적으로 덜 복잡한 모형이라고 볼 수 있다. 인간의 의사결정 과정과 보다 흡사한 특징을 가지며 문제에 대한 광대한 지식과 경험이 없어도 기존 사례에 근거하여 답을 찾을 수 있는 장점을 가지고 있다. 데이터에 저장된 사례로부터 해결하고자 하는 문제와 가장 유사한 혹은 비유사성의 정도를 나타내는 거리를 정의하여 가까운 사례들을 추출하고(retrieve), 그 사례를 재사용하거나(reuse) 수정하여(revise) 해답을 찾는 방식을 이용한다. Bryant(1997)는 기업부도 예측에 사례기반추론 모형을 적용하였고, legal indexing과 information retrieval에 여러 가지 시도를 한 연구들(Elhadi & Vamos, 1999; Elhadi, 2000)과 Analytic Hierarchy Process를 접목시킨 연구(Park & Han, 2002) 등 많은 연구가 쏟아져 나오게 되었다.

사례기반추론에서 사용되는 거리의 개념은 다양한데, 일반적으로 유클리디안 거리(Euclidean distance)를 사용하여 왔다. 유클리디안 거리는 거리 계산에 사용되는 변수 간의 상관계수를 0이라고 가정했을 때의 거리 개념으로서 현재의 기하학에서 통상적으로 쓰이는 거리 개념이다. 하지만, 한 산업 군에 속한 여러 기업들의 재무비율 변수를 사용하여 거리를 측정할 때는 이 재무비율 변수들 간에 상관관계가 존재함을 쉽사리 확인할 수 있다. 이러한 기존 연구의 단점을 보완하기 위한 시도는 Cho et al.(2010)에 의하여 Mahalanobis distance 개념을 도입함으로써 이루어졌다. 이 연구에서는 거리 측정에 Mahalanobis distance와 변수에 가중치를 둔 경우에 여타의 비교 모델보다 우수한 예측력을 보여주었다. 사례기반추론 이외에도 Mahalanobis distance는 fuzzy

clustering이나 특정 public health problem의 패턴을 탐지하는데 적용되고 있다(Zhang et al., 2011; Wang et al., 2011).

3. 자료와 연구 모델

3.1 자료의 구성

이번 실험에는 한국신용보증재단에서 수집한 금융 데이터를 사용하였다. 비교적 타 산업에 비해 부도가 종종 일어나는 제조업 분야에서 자산 규모 10억 원에서 80억 원 대에 속한 기업을 대상으로 하였고 회계연도 1999년에서 2002년 동안 부도 판정이 난 500개의 기업과 정상인 500개의 기업을 분석대상으로 하였다.

일차적으로 single sample t-test를 사용하여 전체 83개의 재무비율 변수 중에서 유의한 56개의 변수를 사전 선별하였고, 이차적으로 로지스틱 회귀 분석(stepwise selection 사용)과 의사결정나무 추론(decision tree induction)을 적용하여 최종 8개의 변수를 선택하였다. 선택된 변수는 로지스틱 회귀분석의 경우 감가상각률, 매출액 경상 이익률, 경영자산 회전을, 순 금융 비율, 순 운전자본 대 총 자본, 유동자산 증가율, 인건비, 그리고 EBIDTA 대 매출액이 선정 되었다(<표 1> 참조). 의사결정나무 추론에서는 부가가치율, 금융비용 대 매출액, 매출액 경상 이익률, 영업 활동 후 총 부채, 현금영업이익+이자수익/차입금+이자비용, 고정비율, 인건비, 그리고 자본금 순 이익률이 선택되었다(<표 2> 참조).

<표 1> 로지스틱 회귀분석에서 선택된 변수

변수	정의
감가상각률	$\text{감가상각비} / [\text{전기상각대상자산} + \text{당기상각대상자산} + (\text{감가상각비} * 2)] * 200$ $= [\text{감가상각비} + \text{감가상각비}(\text{제조원가})] / [(\text{전기 건물} + \text{기계장치} + \text{차량운반구} + \text{기타유형자산}) + (\text{당기 건물} + \text{기계장치} + \text{차량운반구} + \text{기타유형자산}) + (\text{당기 감가상각비} + \text{감가상각비}(\text{제조원가})) * 2] * 200$
매출액 경상 이익율	경상이익/매출액*100
경영자산회전율	매출액/[총자산-(건설 중인자산+투자자산)]
순 금융 비용	이자비용-이자수익
순운전자본 대 총자본	(유동자산-유동부채)/총자본*100
유동자산증가율	당기말 유동자산/전기말 유동자산*100-100
인건비	급여+퇴직급여+급여(제조원가)+퇴직급여(제조원가)
EBITDA 대 매출액	(경상이익+이자비용+감가상각비+무형자산상각비)/매출액*100

<표 2> 의사결정나무 추론에서 선택된 변수

변수	정의
부가가치율	$[\text{법인세차감전순이익} + \text{급여} + \text{퇴직급여} + \text{급여}(\text{제조원가}) + \text{퇴직급여}(\text{제조원가}) + \text{임차료} + \text{임차료}(\text{제조원가}) + \text{감가상각비} + \text{감가상각비}(\text{제조원가}) + (\text{이자비용} - \text{이자수익})] / \text{매출액} * 100$
금융비용대매출액	이자비용/매출액*100
매출액 경상 이익율	경상이익/매출액*100
영업 활동 후 현금흐름/총부채	영업 활동 후 현금 흐름/총부채
현금영업이익+이자수익)/차입금+이자비용	현금영업이익+이자수익)/차입금+이자비용
고정비율	고정자산/자기자본*100
인건비	급여+퇴직급여+급여(제조원가)+퇴직급여(제조원가)
자본금순이익률	당기순이익/자본금*100

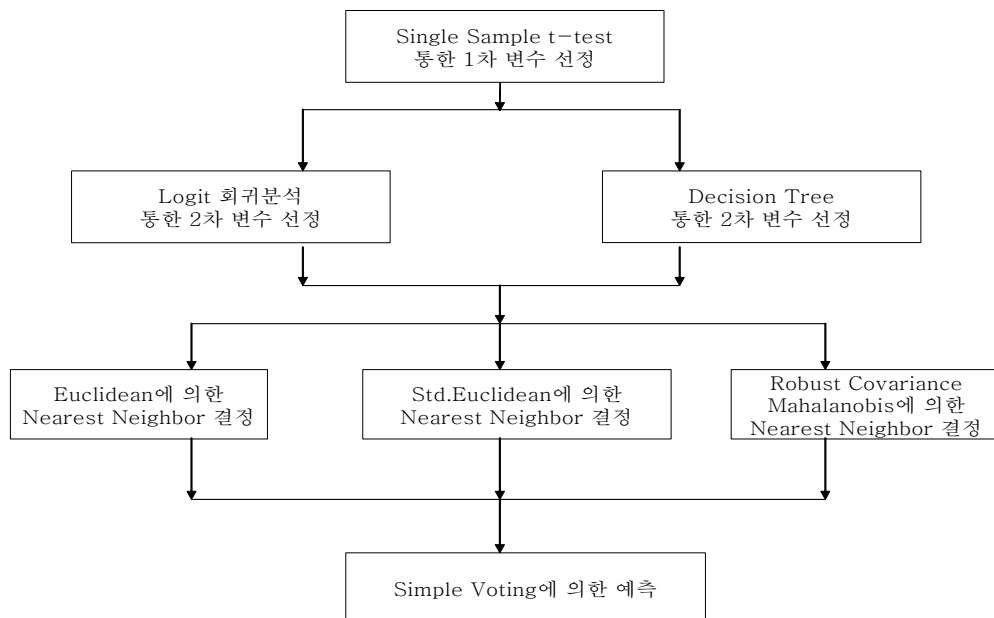
3.2 주요 알고리즘

<그림 1>은 본 연구의 주요 알고리즘을 묘사하고 있다. 분석에 사용되는 변수의 선정에는 일차

적으로 Single sample t-test를 사용하고 이차적으로 로지스틱 회귀분석과 의사결정나무 추론을 적용하여 최종적으로 8개의 재무비율 변수를 선정하였다. 훈련용 데이터(training data)와 테스트용 데이터(testing data)와의 거리는 세 가지 방법 즉, (1)

Euclidean distance, (2) Euclidean distance with standardized values, (3) robust covariance-based Mahalanobis distance를 사용하였다. 이렇게 테스트 용 데이터에 속한 개별 기업에 대하여 훈련용 데이터의 모든 기업과의 개별 거리를 계산하여 순위화 하였다.

현상이 생기는 원인은 거리 계산 시 평균과 공분산이 극단값에 민감하기 때문이다. 이러한 왜곡 현상을 감소시키기 위한 방안으로 robust한 추정량을 가진 공분산을 이용한 Mahalanobis 거리 개념을 도입하였다. 공분산 계산에는 FastMCD algorithm이 사용되었다(Hubert & Debruyne, 2010; Rousseeuw & Driessen, 1999).



<그림 1> 주요 알고리즘

개체 간 거리를 계산할 때 극단값(outlier)이 상당수 포함되면 masking 효과가 발생하여 결과를 왜곡시킬 수 있다. Masking 효과란 전통적 추정치(classical estimate)는 극단값에 의하여 큰 영향을 받아 tolerance ellipse는 이 극단값들을 포함하고 더 이상 극단값으로 인식하지 않게 되어 결과적으로 일반적 Mahalanobis 추정치는 극단값을 제대로 인식하지 못하는 현상을 말한다. 이러한 왜곡

일반적인 Mahalanobis distance 공식은 다음과 같다:

$$\sqrt{\sum (Y_{i,p} - Y_{j,p}) \Gamma^{-1} (Y_{i,p} - Y_{j,p})^T}$$

($Y_{i,p}$ = 기업 i 의 p 속성수준 벡터, $()^T$ = 벡터의 전치행렬,
 Γ^{-1} = 속성의 공분산 역행렬).

본 연구에서 쓰인 robust covariance-based Mahalanobis distance는 다음과 같다:

$$\sqrt{\sum (Y_{i,p} - Y_{j,p}) \Gamma_R^{-1} (Y_{i,p} - Y_{j,p})^T}$$

($Y_{i,p}$ = 기업 i 의 p 속성수준 벡터, $()^T$ = 벡터의 전치행렬,
 Γ_R^{-1} = 속성의 robust 공분산 역행렬).

최종적으로 nearest neighbors의 규모는 임의로 25개의 기업을 사용하였으며 테스트용 데이터의 기업 부도 여부를 결정하기 위하여 simple voting rule을 사용하였다. 순위가 높은 기업에 큰 가중치를 부여하는 weighted voting rule(Berry & Linoff, 2003)을 채택하지 않고 25개 기업에 동일한 가중치를 부여한 이유는 다음과 같다. 본 연구에서 거리측정을 위하여 선정한 8개의 재무비율 변수는 비유사도를 측정하는 완벽한 기준은 될 수 없기 때문에 25개 기업군에 포함 된 것이 중요하지 이 그룹 안에서 1위인지 25위인지는 상대적으로 덜 중요하다고 판단하였기 때문이다. 8개 변수의 선정에 다른 기준을 적용하였다면 25개 기업의 구성과 순위는 상당히 달라질 수도 있었을 것이다.

4. 실증 분석

4.1 실험 설계

전체 1,000개의 리코드(record)는 training data set: testing data set = 80: 20 으로 분할되었다. Training과 testing 데이터 각각 부도와 비부도 비율은 50: 50 으로 지정하였다. Training 데이터 전체에 대하여 testing 데이터의 개별 리코드와의 거리를 계산함에 있어, (1) Euclidean Distance; (2) Standardization을 시킨 데이터에 대한 Euclidean Distance; (3) Robust Covariance를 적용한 Mahalanobis Distance, 이렇게 세 가지 모델을 실험에 사용하였다.

기존의 Mahalanobis distance에 대비하여 robust covariance-based Mahalanobis distance를 쓰는 경우 변수 간의 공분산 매트릭스가 달라지는데, 예를 들면 <표 3>과 <표 4>는 validation set 1에 decision tree를 이용한 변수 선택의 경우이다. 두 표에서는 변수 간 상관성에 상당한 차이를 보임을 확인할 수 있다.

<표 3> 8개 변수에 대한 일반적인 공분산 매트릭스

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
X ₁	185.89	2.82	16.54	1.59	2.75	146.70	545.43	54.20
X ₂		6.57	-5.44	-0.45	-1.08	120.89	123.23	-29.50
X ₃			24.13	0.75	1.53	-99.04	-127.55	119.82
X ₄				0.59	0.64	-14.56	-36.59	3.81
X ₅					3.15	-46.02	-83.09	7.16
X ₆						37707.04	5137.12	2201.95
X ₇							100719.30	1444.25
X ₈								3030.49

<표 4> 8개 변수에 대한 robust 공분산 매트릭스

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
X ₁	190.46	5.56	9.88	1.00	1.10	222.42	500.01	-22.56
X ₂		6.75	-1.43	-0.15	-0.40	75.56	172.09	-16.70
X ₃			7.85	0.34	0.48	-44.73	-234.11	30.50
X ₄				0.18	0.20	0.68	-10.27	1.42
X ₅					0.50	-5.35	-20.45	3.15
X ₆						17570.63	7033.14	283.21
X ₇							84071.47	-83.81
X ₈								666.98

k nearest neighbors의 개수는 25개로 결정하고 simple voting rule을 적용하여 목표변수 값을 예측하였다. 테스트의 신뢰도를 높이기 위하여 five-fold cross validation 방법을 채택하였고 결과는 <표 5>에 요약되어 있다. Testing 데이터에 대한 정분류율(correct classification ratio)을 모델의 정확도를 나타내는 지표로 사용하였다.

나타났고, 표준화를 거친 후 Euclidean 거리 모형의 정분류율 평균은 0.665로 나타나 큰 차이를 보이지 않았다. Robust Covariance에 근거한 Mahalanobis 거리 모형의 정분류율 평균은 0.745로 앞의 두 모형보다 크게 높게 나타났다.

둘째, 의사결정나무 추론을 적용하여 최종 선택된 변수를 사용할 경우, 표준화 과정을 거치지 않

<표 5> 모델 분석 결과

validation set	Euclidean Distance		Standardized Euclidean Distance		Robust covariance Mahalanobis Distance	
	logit	decision tree	logit	decision tree	logit	decision tree
1	0.7	0.6	0.635	0.695	0.74	0.755
2	0.665	0.57	0.665	0.74	0.745	0.78
3	0.67	0.565	0.675	0.65	0.75	0.76
4	0.645	0.64	0.61	0.66	0.73	0.745
5	0.65	0.605	0.74	0.675	0.76	0.745
Average	0.666	0.596	0.665	0.684	0.745	0.757

4.2 모델 결과 분석

분석을 통하여 다음과 같은 결과를 얻었다. 첫째, 로지스틱 회귀분석을 적용하여 변수를 최종 선택한 자료에 표준화 과정을 거치지 않은 경우 Euclidean 거리 모형의 정분류율 평균은 0.666으로

고 Euclidean 거리 모형을 사용하면 정분류율 평균은 0.596로 가장 낮은 성과를 나타났다. 표준화 과정을 거치면 Euclidean 거리 모형의 정분류율 평균은 0.684로 상승하였다. Robust Covariance Mahalanobis 거리 모형의 정분류율 평균은 0.757로 나타나 가장 높은 예측률을 보였다.

Robust Covariance에 근거한 Mahalanobis 거리 모형의 두 가지 변수 선택 방법에서 대략 0.75 정도의 정분류율을 나타냄으로써 두 유형의 Euclidean 거리 모형보다 약 0.08 정도 높은 예측률을 보였다. 그 외에도, 변수 선택에서 의사결정나무 추론을 적용한 후 표준화 되지 않은 Euclidean 거리 모형을 적용할 경우 성과가 가장 나쁘게 나타났는데, 이는 의사결정나무 추론의 타 모형에 대한 상대적 효과가 우수하게 나타날 수 있는 조건이 변수 간의 교호 효과(interaction effect)가 있을 경우임을 고려하면 어느 정도 납득이 될 수 있을 것이다. 다시 말하여, 변수 간 상관관계를 무시하는 Euclidean 거리를 변수 간 상관관계를 고려하여 선택한 변수에 적용하면 성과가 좋지 않게 나타날 가능성이 커질 수 있을 것이다.

5. 결론 및 시사점

기업부도 예측을 다룬 다양한 계량적 모형 중에서 본 연구는 사례기반추론 모형을 선택하여 새로운 거리 개념의 적용가능성을 확인하고자 시도하였다. 사례기반추론의 성과는 목표하는 대상과 가장 유사한 대상들을 어떠한 기준에 의하여 선정하는가에 따라 여타의 계량 모형보다 낮은 성과를 산출하기도 하고 높은 결과를 산출하기도 한다. 변수 간 공분산 구조를 감안하여 변수 간 거리를 측정하는 Mahalanobis 거리 개념에 극단값의 부작용을 감소시킬 수 있는 robust covariance를 접목시킴으로써 예측정확도가 향상될 수 있는 여지가 있는지 실험하였고 실험결과는 긍정적으로 나타난 것으로 보인다.

로지스틱 회귀분석에 비하여 의사결정나무 추론을 이용하여 변수를 선정할 경우 표준화 후 Euclidean 거리 모형과 Mahalanobis 거리 모형의

성과가 우수하게 나타났다. 이는 많은 변수를 포함하는 데이터에서 일정한 수의 변수를 가려내는 예비 작업을 필요로 하는 일반 연구에 시사하는 바가 있을 것으로 사료된다. 연구의 본 모형이 교호 효과를 인지하여 결과를 내는 모형일 경우 의사결정나무 추론을 통한 변수 선택은 좋은 예비 작업이 될 수 있을 것이다.

본 연구의 성과와 시사점이 일반화되기 위해서는 기업부도 예측 뿐 아니라 다양한 형태를 갖는 여러 산업 분야에 대한 테스트가 필요해 보인다. 본 연구에서는 변수에 가중치를 주거나 또 다른 voting rule을 적용하는 것과 같은 시도는 하지 않고, 탐색적 차원에서 robust covariance를 적용한 Mahalanobis 거리 모형의 사례기반추론 분야에서 적용 가능성을 타진해보았다. 향후 연구에서는 이러한 다양한 옵션들을 포함하는 시도가 필요할 것으로 보인다. 또한 k nearest neighbors를 결정할 때 훈련용 데이터를 training data와 validation data 그룹으로 나누어 validation data의 정분류율을 최대화하는 k 를 선정한다면 모형의 정확도를 좀 더 향상시킬 수 있을 것이다.

참고문헌

- [1] 조성빈, 김진화, 장성우(2008), “유전자 다중 에이전트 규칙유도를 이용한 기업 부도 예측 방안”, 한국경영공학회지, 13(1), 1-11.
- [2] Altman, E.(1968), “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, Journal of Finance, 23(3), 589-609.
- [3] Barniv, R., Agrawal, A.(1997), “Predicting the outcome following bankruptcy filing: a three-state classification using neural networks”,

- International Journal of Intelligent Systems in Accounting, Finance, and Management, 6(3), 177-194.
- [4] Beaver, W.(1966), "Financial ratios as predictors of failure", Journal of Accounting Research, 4(1), 71-127.
- [5] Berry, M.A., Linoff, C.S.(2003), Data Mining Techniques, Wiley.
- [6] Bryant, S.M.(1997), "A case-based reasoning approach to bankruptcy prediction modeling," Intelligent Systems in Accounting, Finance and Management, 6, 195-214.
- [7] Cho, S., Hong, H., Ha, B-C.(2010), "A hybrid approach based on the combination of variable selection and using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction", Expert Systems with Applications, 37(4), 3482-3488.
- [8] Coates, P., Fant, L.(2000), "Recognizing financial distress patterns using a neural network tools", Financial Management, 9(2), 119-144.
- [9] Eisenbeis, R.A.(1978), "Problems in applying discriminant analysis in credit scoring methods", Journal of Banking and Finance, 2, 205-219.
- [10] Elhadi, M.T., Vamos, T.(1999), "An IR-CBR approach to legal indexing and retrieval in bankruptcy law," Tenth proceedings in Database and Expert Systems Applications, 769-774.
- [11] Elhadi, M.T.(2000), "Bankruptcy support system: taking advantage of information retrieval and case-based reasoning," Expert Systems with Applications, 18(3), 215-219.
- [12] Falbo, P.(1991), "Credit scoring by enlarged discriminant models", Omega, 19(4), 275-289.
- [13] Frydman, H.E., Altman, E.I., and Kao, D.(1985), "Introducing recursive partitioning for financial classification: the case of financial distress", Journal of Finance, 40(1), 269-291.
- [14] Hubert, M., Debruyne, M.(2010), "Minimum covariance determinant", WIREs Computational Statistics, 2(1), 36-43.
- [15] Lee, K.C., Kim, M.J., Kim, H.(1994), "An inductive learning - assisted neural network approach to bankruptcy prediction: comparison with MDA, inductive learning, and neural network models", Journal of Management Research, 23(3), 109-114.
- [16] Malhotra, R., Malhotra, D.K.(2002), "Differentiating between good credits and bad credits using neuro-fuzzy systems", European Journal of Operation Research, 136(2), 190-211.
- [17] Ohlson, J.A.(1980), "Financial ratios and probabilistic prediction of bankruptcy", Journal of Accounting Research, 18(1), 109-131.
- [18] Park, C., Han, I.(2002), "A case-based reasoning with the feature weights derived by Analytic Hierarchy Process for bankruptcy prediction," Expert Systems with Applications, 23(3), 255-264.
- [19] Pendharkar, P.C.(2005), "A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem", Computers & Operations Research, 32, 2561-2582.
- [20] Rousseeuw, P.J., Driessen, K.V.(1999), "A fast algorithm for the minimum covariance determinant estimator, Technometrics, 41(3), 212-223.

- [21] Srinivasan, V., Kim, Y.H.(1998), "Designing expert financial systems: a case study of corporate financial management", Financial Management, 5, 32-43.
- [22] Wang, P.C., Su, C.T., Chen, K.H., and Chen, N.H.,(2011), "The application fo rough set and Mahalanobis distance to enhance the quality of OSA diagnosis, Expert Systems with Applications, 38(6), 7828-7836.
- [23] Zhang, Y., Huang, D., Ji, M., and Xie, F.,(2011), "Image segmentation using PSO and PCM with Mahalanobis distance", Expert Systems with Applications, 38(7), 9036-9040.