

채무불이행위험의 예측을 위한 CBR응용 (Applying CBR for Default Risk Forecasting)

김진백*

〈목 차〉

I. 서론	3.1 실험방법 및 설계
II. 채무불이행위험의 일반모형과 평가요소	3.2 규모별 예측결과의 비교
1. 채무불이행위험의 일반모형	3.3 업종별 예측결과의 비교
2. 채무불이행위험의 평가요소	3.4 실험결과의 평가
III. CBR을 이용한 채무불이행위험의 예측	IV. 결론
1. CBR의 개요	참고문헌
2. 사례베이스의 구조 및 평가사례의 표현	Abstract
3. 채무불이행위험의 예측	

I. 서론

IMF 사태 이후 기업의 자금흐름이 원활하지 못하여 각 기업들은 자금조달을 위해 다양한 수단을 이용하고 있으며, 그 중 대표적인 것이 신주, 회사채, 그리고 기업어음 등을 발행하는 것이다. 최근 직접금융 활성화정책에 따라 회사채의 발행규모가 급증하고 있다. 그 동안 우리나라에서 발행되는 회사채는 대부분 금융기관들이 원리금 지급을 보증하는 보증사채이기 때문에 채권투자자의 입장에서 채권평가와 채권분석은 큰 의미가 없었다. 그리고 회사채의 발행금리가 기업의 신용도나 원리금 상환능력과는 무관하게 시장금리보다 낮게 결정되어 온 것이 사실이다.

그러나 최근들어 무보증사채의 발행 증가와 은행의 부실로 인해 개인투자자 및 기관투자자의 입장에서 볼 때 정확한 채권평가가 점차 중요한 문제로 등장하고 있다. 투자자의 입장에서

* 부산여자대학 경영정보과 전임강사

채권을 평가하는 것은 미래의 원활한 현금흐름을 위해서이다. 따라서 미래에도 원활한 현금흐름을 계속 유지하기 위해서는 채권투자시 채권의 발행기업이 채무불이행을 할 위험의 정도를 평가해서 투자 여부를 결정해야만 한다.

채무불이행위험(default risk)을 예측하기 위해서는 기업 고유의 위험을 측정하여야 하나 이는 매우 복잡한 문제이다. 따라서 이러한 문제를 해결하기 위해서는 원인과 결과간의 명확한 이론적 근거를 필요로 하지 않는 CBR(case-based reasoning)이 채무불이행위험을 측정하는 유용한 도구가 될 수 있다. 본 논문에서는 채무불이행위험을 예측하기 위해 많이 이용되고 있는 다변량 판별분석기법과 인공지능분야의 CBR기법을 적용해서 채무불이행위험의 예측능력을 평가하기로 한다.

II. 채무불이행위험의 일반모형과 평가요소

1. 채무불이행위험의 일반모형

시장위험에 초점을 맞춘 자본자산가격결정모형(capital asset pricing model)이나 차익거래 가격결정모형(arbitrage pricing model) 등과 대조적으로 채무불이행위험모형은 기업고유의 채무불이행위험이 기대수익률에 미치는 영향을 분석하는 기법이다. 분산투자이론은 기업고유의 위험이 기대수익률 추정시 고려대상이 되지 않는 이유를 설명하고 있지만 이 논리는 현재와 같이 상향이익이 제한적이고 하향손실이 더욱 큰 상황에서는 적용되기 어렵다. 특히 채무불이행 비율은 경기침체에 크게 증가하는 것으로 알려져 있다. 따라서 회사채의 투자자들은 회사채에 대한 채무불이행위험을 통해 투자안의 가치를 평가하여야만 한다.

채무불이행위험은 현금흐름의 창출능력과 재무적 부담이라는 2가지 변수를 갖는 함수로서 다음과 같이 일반화할 수 있다(정한규 외 공동연구, 1998, p. 54).

$$DR = f(cf, fb)$$

- 여기서, DR : 채무불이행위험
 cf : 현금흐름의 창출능력
 fb : 재무적 부담

즉, 채무불이행위험은 재무적 부담에 비해 많은 현금흐름을 창출하면서 현금흐름이 안정적인 기업일수록 낮게 나타난다. 이때 재무적 부담은 법적으로 강제된 지급금인 이자 및 원금상환액만을 의미하며, 배당금이나 자본적 지출과 같이 법적 강제력이 없는 것은 포함되지 않는다. 따라서 기업이 현금흐름을 많이 창출해서 자산의 시장가치가 높은 기업일수록 채무불이행위험은 낮게 나타난다. 그리고 예측가능하고 안정적인 사업을 하는 기업은 경기순환적 혹은 변동성이 큰 사업을 하는 기업보다 채무불이행위험이 낮게 나타난다. 대부분의 채무불이행위험모형들은 재무적 부담에 대한 현금흐름비율을 측정하기 위해 재무비율은 반영하나 기업 고유의 현금흐름 변동성을 반영하기 위해 산업효과는 통제를 한다.

2. 채무불이행위험의 평가요소

현재 채무불이행위험의 척도로 가장 폭 넓게 이용되는 것은 채권등급이다. 채권등급은 개인적인 정보와 공개적인 정보를 바탕으로 평가될 수 있으나 평가기관들은 주로 공개적인 정보에 기초해서 등급을 결정한다. 회사채에 대한 등급평가기관 중 미국의 Standard and Poor's사와 Moody's사가 국제적으로 저명하며, 국내에는 (주)한국신용평가, (주)한국신용정보, (주)한국기업평가(주) 등이 있다.

채권등급은 주로 원리금지급능력과 현금흐름의 안정성 및 예측가능성 등을 나타내는 재무비율을 이용해서 결정되어지며 <표 1>은 채무불이행위험의 예측에 이용되는 주요 재무비율들을 나타낸 것이다.

<표 1> 채무불이행위험을 측정하는 데 이용되는 재무비율

재 무 비 율	관련 모형
순운전자본/총자산-순운전자본구성비율 이익잉여금/총자산-이익잉여금구성비율 영업이익/총자산-총자산영업이익률 자기자본의 시장가치/총부채-레버리지비율 매출액/총자산-총자산회전율	Z-Score 모형
log(총자산)-기업규모 log(매출액/총자산)-회전율 이익잉여금/총자산-누적 수익성 자기자본의 시장 혹은 장부가치/총부채-재무구조	K-Score 모형

그 외에도 한국은행이나 은행감독원에서는 자기자본비율, 유동비율, 고정장기적합률의 역수, 순운전자본/총자본, 자본금경상이익률, 총자본회전율, 경영자본회전율 등을 이용하고 있다(박정식과 신동령, 1997, p. 335; 남명수와 김대호, 1998, p. 340). 또한 Standard and Poor's사는 세전 이자보상비율, EBITDA 이자보상비율, (영업활동으로부터의 자금/총부채) 비율, (기업잉여현금흐름/총부채) 비율, 세전 총자본영업이익률, 매출액영업이익률, (장기부채/장기성자본) 비율, (총부채/총자본) 비율 등을 이용하고 있다(정한규 외 공동연구, 1998, p. 58).

채권등급이 자신의 재무비율보다 낮은 평가를 받은 기업이 있을 수 있다. 이는 채권등급의 최종 결정시 평가기관들의 주관적 판단이 개입될 수 있기 때문이다. 이는 현재의 재무비율상으로는 실적이 우수하나 다음 기간동안에는 실적이 떨어질 가능성이 높은 기업은 채권의 상환이 미래에 이루어지므로 낮은 평가를 받을 수 있기 때문이다. 그러나 대부분의 기업에 대해 재무비율은 채권등급을 추론하는데 합리적 근거로 이용되고 있다.

III. CBR을 이용한 채무불이행위험의 예측

1. CBR의 개요

사례란 경험을 나타내는 상황적 지식으로서(Kolodner, 1993, p. 13) 이를 이용해서 문제해결을 시도하는 것이 사례기반추론(Case-Based Reasoning : CBR)이다. 따라서 CBR이란 새로운 문제가 주어지면 이를 분석해서 과거의 경험 중 가장 유사한 경험을 검색한 후, 검색된 과거 경험을 직접 혹은 부분적인 수정을 통해 문제해결에 이용하는 과정을 의미한다.

Aamodt & Plaza(1994)는 CBR 과정을 검색, 재사용, 수정, 보관이라는 4단계로 나누었다. 검색(retrieve)단계는 문제가 주어지면 문제의 주요 속성을 추출한 다음 이들 속성을 이용해서 저장된 사례를 찾는 단계이다. 검색결과는 3가지로 요약될 수 있으며, 첫째는 둘 이상의 유사한 사례가 검색되는 경우, 둘째는 유일한 유사 사례가 검색되는 경우, 셋째는 유사한 사례가 검색되지 않는 경우이다(Mahapatra & Sen, 1994, p. 20).

재사용(reuse)단계는 검색결과로 주어진 유사사례를 이용해서 문제해결을 시도하는 단계이다. 검색결과로서 단일 사례가 검색되면 이를 통해 문제를 해결하게 되며 둘 이상의 유사한 사례가 검색되면 속성별 비교를 통해 유사성을 측정한 후 주어진 문제에 대해 가장 유사성 측정치가 높은 사례를 기저사례로 선택해서 문제의 해결을 시도한다.

수정(revise)단계는 선택된 기저사례가 문제와 상이한 차이가 있을 경우 이를 문제에 적합하게 변형시키는 단계이다. 이는 선택된 기저사례가 문제와 완전히 일치할 수도 있으나 경우에 따라서는 부분적인 차이가 발생할 수도 있기 때문이다. 따라서 이 경우에는 사용자와의 상호작용을 통해서 차이부분에 대한 부분적인 수정을 거쳐서 문제에 적합한 해를 생성하게 된다.

보관(retain)단계는 문제해결경험을 재사용하게 위해 저장하는 단계이다. 새로운 문제가 기존 사례의 문제기술부분과 완전히 일치하였을 경우 동일한 문제이므로 이에 대한 해결방법을 저장하면 중복저장이 된다. 따라서 이 경우에는 사용한 해결방법을 저장할 필요가 없다. 그러나 문제가 기존 사례의 문제기술부분과 차이가 있을 경우에는 사용한 해결안과 문제를 사례베이스에 저장해서 향후 의사결정에 이용할 수 있도록 한다. 따라서 CBR은 새로운 문제가 발생할 때마다 위의 과정을 반복함으로써 문제해결을 시도하게 된다.

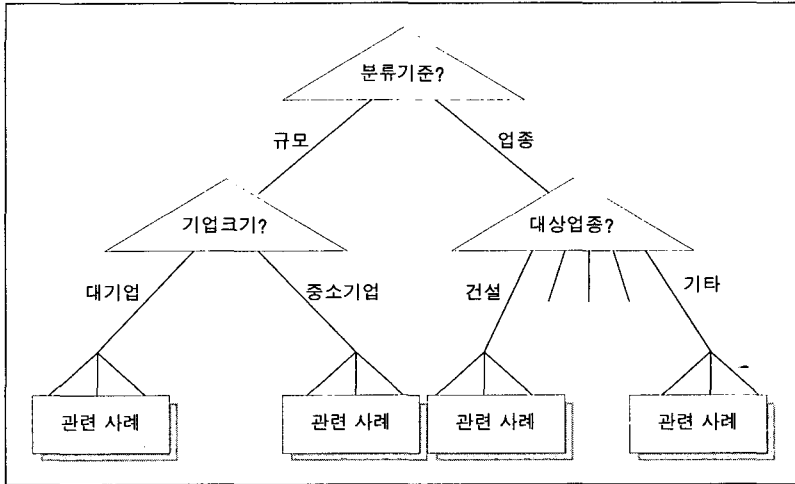
2. 사례베이스의 구조 및 평가사례의 표현

사례베이스의 구조는 사례의 연결형태를 의미하는 것으로 이는 검색시간의 주요 영향요인이 된다. 사례베이스를 구조화하는 방법에는 단순리스트나 배열을 이용한 평면구조(flat structure), 속성공유를 가능하게 하는 속성공유 네트워크구조(shared feature network), 그리고 질문을 노드로 표현하는 차별화 네트워크구조(discrimination network) 등이 있다. 채무불이행위험문제는 관련 사례들 간의 공통된 속성값이 많지 않다. 따라서 채무불이행위험을 평가하기 위한 사례베이스는 차별화 네트워크구조를 따르는 것이 좋다.

차별화 네트워크구조는 사례의 유사성측면보다는 차별화측면에 초점을 두고 사례들을 계층화한 구조이다(Kolodner, 1993, p. 300). 따라서 차별화 네트워크구조의 내부노드에는 해당 노드의 하위 항목들을 분류하기 위한 질문을 나타낸다. 차별화 네트워크구조에서는 사례들을 분류하기 위해 차원의 중요성 순서에 따라 주요 질문을 상위계층에 두게 된다. 차별화 네트워크구조는 속성공유 네트워크구조가 갖는 이점 이외에도 매 단계마다 단일 질문으로 검색이 이루어지므로 검색이 보다 효율적이며, 색인과 사례조직간의 연결이 직관적이다. 또한 속성과 속성값이 분리되어 있어 유용한 속성의 유지가 용이하다.

그러나 차별화 네트워크구조는 속성공유 네트워크구조의 단점 이외에도 결측정보(missing information)의 취급이 어려우며 단일 질문에 따라 구조화가 되었기 때문에 상위계층에서 잘못된 노드를 선택하였을 경우에는 원하는 결과를 얻지 못할 가능성이 속성공유 네트워크구조보다 크다는 문제점이 있다. 이러한 문제점을 해결하려면 중복 차별화 네트워크구조(redundant discrimination network)가 필요하다. 중복 차별화 네트워크구조는 여러 개의 차별화 네트워크

구조를 이용해서 각 노드를 조직화하는 방법이다. 여기서는 기업의 규모와 업종을 중심으로 2개의 차별화 네트워크구조를 이용하도록 한다. 따라서 채무불이행위험을 예측하기 위한 위험평가 사례베이스의 구조는 <그림 1>과 같다.



주) 삼각형 - 질문, 레이블 - 응답결과, 최종 노드 - 관련 사례

<그림 1> 위험평가 사례베이스의 차별화 네트워크구조

<그림 1>에서 첫 단계 노드는 기업의 규모 및 업종을 분류하기 위한 것으로 조건에 따라 규모와 업종에 따라 다른 차별화 네트워크를 이용하기 위해서이다. 만약 응답결과값이 “규모”라고 하면 좌측의 레이블값과 같으므로 좌측 경로를 따라 검색이 진행되며, 다음 질문은 기업의 규모에 관한 것이다. 이때 규모에 대한 응답결과값이 “대기업”이면 대기업의 사례들을 중심으로 검색을 하게 된다. 만약 첫 단계의 응답결과값이 “업종”이었다면 우측의 경로를 따라 업종의 세부분류를 위한 두 번째 질의가 주어진다. 두 번째 응답값에 따라 각각의 검색은 하위 네트워크구조를 대상으로 직접적인 사례검색이 이루어지게 된다.

일반적으로 사례는 문제기술 부분과 해결안 부분을 주요 구성요소로 갖는다(Mahapatra & Sen, 1994, p. 20). 문제기술 부분은 의사결정과 관련된 문제의 속성에 대한 정보를 가지며, 해결안 부분은 의사결정이나 문제해결에 필요한 정보를 가지고 있다. 따라서 각 사례들은 문제기술부분으로 개별기업들에 대한 재무비율을 가지며, 해결안부분으로 평가정보, 즉 채권등급에 대한 정보를 가진다. 이때 이들 각 재무비율들을 하나의 슬롯에 할당하여 프레임 형태로 사례들을 표현하면 <그림 2>와 같다.

문제기술부분	사례식별자	L-xx
	재무비율 1	xx.xx
	재무비율 2	xx.xx

해결안	채권등급	AAA

〈그림 2〉 프레임형태의 위험평가사례 표현 예

3. 채무불이행위험의 예측

기업부실(corporate financial distress)은 그 실태가 매우 다양하기 때문에 한마디로 정의하기는 어려우나 기업부실을 의미하는 용어로는 경제적 부실, 지급불능, 그리고 파산 등이 있다 (Altman, 1983, pp. 5~7). 기업의 부실은 순간적으로 일어나는 것이 아니다. 정상적인 기업이 부실기업이 되어 파산에 이르기까지는 점진적으로 수년에 걸쳐 진행이 된다. 따라서 기업의 부실로 인해 발생하는 채무불이행위험에 따른 손실을 최소화하기 위해 각 금융기관들은 자체적으로 여러 가지 정보를 이용해서 기업의 경영상태를 평가해서 대출업무를 수행하고 있다. 이때 채무불이행위험을 최소화함으로써 영업수익률을 제고시키기 위해서는 예측력이 뛰어난 모형을 이용해야만 할 것이다.

3.1 실험방법 및 설계

3.1.1 실험자료

채무불이행위험을 예측하기 위해 사용되는 자료는 기업의 신용평가를 주업무로 하는 (주)한국신용평가에서 발행하는 신용등급가이드(Ⅰ)과 『한국기업총람』의 상장법인편의 채권등급과 주요 재무비율들이다. 표본자료의 수는 총 111개이며, 이는 신용등급가이드(Ⅰ)에 제시된 회사채 등급평가부분에 제시된 기업을 중심으로 한 자료의 수이다. 신용등급가이드(Ⅰ)의 기업들을 중심으로 한 이유는 각 기법들의 예측결과를 전문가들이 예측한 결과와 일치하는 정도를 평가하기 위해서이다.

자료의 구성은 기업의 규모면에서 보면 대기업의 수는 86개이며, 중소기업의 수는 25개이다. 그리고 업종별로 보면, 일반제조업종이 38개, 통신관련업종이 13개, 석유화학업종이 21개, 건설업종이 14개, 서비스업종이 13개, 그리고 기타 업종(고무, 금속, 자동차, 어업 등)이 12개이다. 실험집단과 통제집단의 분류는 규모 및 업종별로 50%씩 나누었다.

3.1.2 예측모형에 대한 설명

기업의 부실은 원인과 진행과정이 다양해서 재무적 자료와 비재무적 자료를 통해서 다양하게 이루어지고 있다. 여기서는 재무적 자료를 통한 방법 중 특히 다변량예측모형을 중심으로 한 방법과 인공지능기법에 속하는 CBR기법을 중심으로 비교를 할 것이다. 다변량예측모형을 이용하는 이유는 단일변량예측모형의 경우 같은 기업이라도 어떤 재무비율을 채택하느냐에 따라 예측의 결과가 매우 상이할 수 있기 때문이다(박정식·신동령, 1997년, p. 325).

(1) Z-Score 모형

다변량예측모형 중 대표적인 모형은 판별분석기법을 이용한 것으로 Z-Score 모형이 효시로 되어 있다(Altman, 1968). Z-Score 모형은 초기표본으로 33개의 도산기업과 33개의 건전기업을 선정하였으며, 표본추출방법으로는 쌍표본추출방법(paired sampling technique)을 사용하였다. 판별분석에 일차적으로 사용된 재무비율은 기존의 관련 문헌에서 많이 이용되던 22개가 선정되었다. 이들 재무비율들을 유동성, 수익성, 레버리지, 지급능력, 그리고 활동성 등 5가지 범주로 나누어, 각 범주에서 대표적인 5개의 재무비율을 중심으로 Z 판별함수를 유도했다.

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5$$

여기서, X_1 = 순운전자본¹⁾/총자산,

X_2 = 유보이익/총자산,

X_3 = EBIT²⁾/총자산,

X_4 = 자기자본의 시장가치/부채의 장부가치,

X_5 = 매출액/총자산

위 공식에 각 기업의 재무비율을 대입해서 나온 Z값이 1.81보다 작으면 파산가능성이 높으며, 3.0보다 크면 파산가능성이 낮다고 판정된다. 그리고 Z값이 1.81과 3.0 사이에 있으면 해당 기업은 판정보류의 회색구간에 있다고 판정한다.

(2) K-Score 모형

K-Score 모형은 1990년에서 1993년까지 상장회사 중에서 34개의 부실기업과 61개의 건전기업을 표본으로 20개의 재무변수를 대상으로 한국기업의 부실을 예측할 수 있는 판별모형으

1) 유동자산 - 유동부채

2) 영업이익 = EBITDA-감가상각비

로 제시되었다(Altman, 1996). 최종모형은 상장기업에 적용할 수 있는 K1-Score 모형과 상장 및 비상장기업에 같이 이용할 수 있는 K2-Score 모형으로 선정된 재무변수는 각각 4개이다.

$$K1-Score = -17.862 + 1.472X_1 + 3.041X_2 + 14.839X_3 + 1.516X_4$$

$$K2-Score = -18.696 + 1.501X_1 + 2.706X_2 + 19.760X_3 + 1.146X_4$$

여기서, $X_1 = \log(\text{총자산})$,
 $X_2 = \log(\text{매출액}/\text{총자산})$,
 $X_3 = \text{유보이익}/\text{총자산}$,
 $X_4 = \text{자기자본의 장부가치}/\text{총부채}$,
 $X_4' = \text{자기자본의 시장가치}/\text{총부채}$

위의 공식에서 보는 것처럼 K1-Score 모형과 K2-Score 모형의 차이점은 자기자본의 장부가치를 사용하느냐 혹은 시장가치를 사용하느냐에 있다. K1-Score 모형과 K2-Score 모형의 판정기준은 <표 2>와 같다.

<표 2> K1-Score 모형과 K2-Score 모형의 판정기준

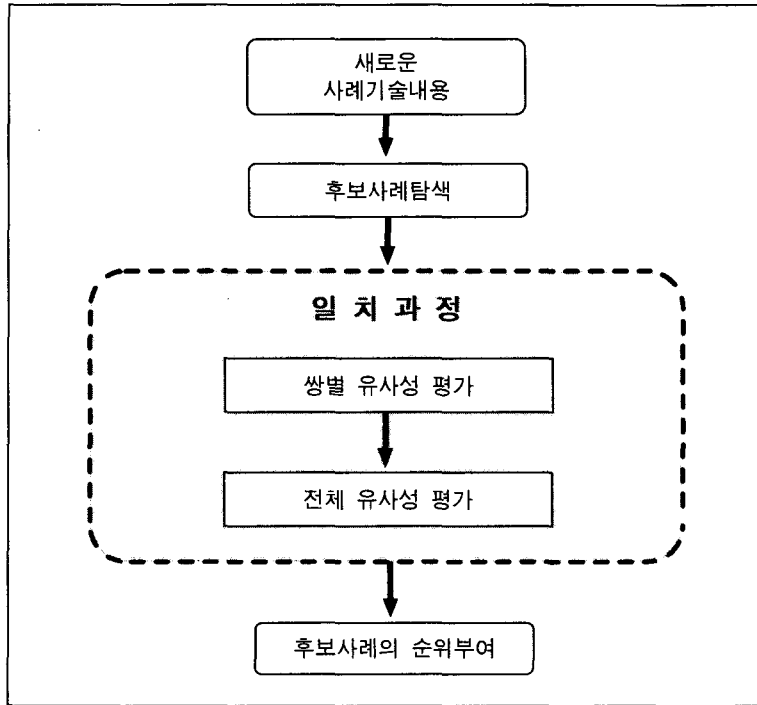
판 정	K1-Score 모형	K2-Score 모형
건 전	K1 - Score > 0.75	K2 - Score > 0.75
판정보류	-2.0 ≤ K1-Score ≤ 0.75	-2.30 ≤ K2-Score ≤ 0.75
부 실	K1 - Score < -2.0	K2 - Score < -2.30

(3) CBR(Case-Based Reasoning)

CBR의 목적은 새로운 의사결정문제에 대해 가장 유용한 이전 사례를 추출하는 것이다. CBR의 검색절차는 <그림 2>와 같다(Gupta & Montazemi, 1997, p. 238). 첫 단계는 새로운 의사결정문제의 사례기술내용이 입력되면 후보사례들을 찾기 위해서 사례베이스를 탐색하는 것이다. 일반적으로 탐색은 잠재력이 있다고 판단되는 사례베이스의 일부 사례들만을 검토함으로써 검색의 효율성을 증가시킬 수 있다.

후보사례들에 대한 탐색이 이루어지고 나면 일치과정(matching process)이 수행된다. 일치과정에서는 후보사례와 새로운 문제간의 유사성을 평가한다. 유사성평가에는 쌍별 유사성평가(pair-wise similarity evaluation)와 전체 유사성평가(overall similarity evaluation)가 있다. 쌍별 유사성평가는 단일차원 일치원칙에 따라 새로운 의사결정문제와 후보사례들을 속성별로

비교하는 방법이다.



<그림 2> CBR의 검색절차

전체 유사성평가는 총합 일치원칙에 따라 쌍별 유사성평가의 총합을 토대로 각 사례와 새로운 문제간의 전반적인 유사성을 측정하는 방법이다. 전체 유사성평가를 위해 이용되는 인접일치함수(nearest-neighbor match function)는 다음과 같다.

$$NE(C_k) = \frac{\sum_{i=1}^n w_i * \text{sim}(f_i^I, f_i^R)}{\sum_{i=1}^n w_i}$$

- 여기서, C_k : k번째 검색된 사례
- w_i : 속성 i에 대한 가중치
- sim : 각 속성에 대한 유사성 함수
- f_i^I : 새로운 문제의 속성 i에 대한 값
- f_i^R : C_k 의 속성 i에 대한 값

위의 수식에서 $\text{sim}(f_i^l, f_i^r)$ 부분이 쌍별 유사성평가부분에 해당한다. 쌍별 유사성을 측정하기 위해 이용되는 $\text{sim}(f_i^l, f_i^r)$ 의 함수로는 유클리드 거리(Euclidian Distance) 계산방법이 많이 이용되고 있다. 따라서 유사성이 클수록 유클리드 거리값이 작으므로 $NE(C_k)$ 의 함수값이 가장 작은 사례를 기저사례로 선택한다. 본 논문에서는 각 속성에 대한 가중치는 모두 동일한 것으로 가정한다.

3.2 규모별 예측결과의 비교

표본데이터 111개를 대상으로 기업을 대기업(86개)과 중소기업(25개)으로 분류하여 Z-Score 모형, K1-Score 모형, K2-Score 모형을 이용해서 기업의 채무불이행 가능성을 평가하였다. 판정기준은 (주)한국신용평가, (주)한국신용정보, (주)한국기업평가에서 판정한 채권등급 중 가장 낮은 등급을 이용하였다.

채무불이행위험을 예측하기 위해 각 예측모형들의 결과와 채권등급의 일치정도를 대응할 경우, 투기적 요소가 적은 BBB 이상의 등급은 채무불이행위험이 낮은 것으로 판정하고, 그 이하의 등급은 채무불이행위험이 높은 것으로 판정하였다. 그리고 각 모형들의 계산값 중 판정보류영역의 값들은 채권등급과의 대응관계가 성립되기 어려워 제외시켰다.

〈표 3〉 규모에 따른 예측모형과 평가기관과의 예측일치도

업종		Z-Score 모형 (일치율)		K1-Score 모형 (일치율)		K2-Score 모형 (일치율)	
대기업	일치	39	0.494	47	0.723	49	0.758
	불일치	40		18		16	
중소기업	일치	14	0.667	9	0.692	12	0.706
	불일치	7		4		5	
총계	일치	53	0.530	56	0.718	61	0.744
	불일치	47		22		21	

주) 판정보류류 결과는 제외

〈표 3〉은 이상의 기준에 따라 대기업과 중소기업들의 예측결과를 평가기관들의 채권등급과 일치된 횟수와 이의 비율들을 나타내고 있다. 대기업의 경우에는 K2-Score 모형이 가장 높은 일치율(0.758)을 나타냈으며, Z-Score 모형이 가장 낮은 일치율(0.494)을 나타냈다. 중소기업

의 경우에도 K2-Score 모형이 가장 높은 일치율(0.706)을 나타냈으며, Z-Score 모형이 가장 낮은 일치율(0.667)을 나타냈다. 그러나 중소기업의 경우에는 대기업의 경우만큼 각 모형간의 일치율 차이가 크지 않은 것으로 나타났다. 평균적으로 보면 기업의 규모에 관계없이 채무불이행위험의 예측모형으로는 K2-Score 모형예측일치도가 높은 것으로 나타났다.

〈표 4〉는 CBR기법을 통해 각 모형의 재무변수들을 문제기술부분의 변수로 활용하고, 해결안 부분의 값으로는 채권의 등급을 이용해서 추론한 결과를 나타내고 있다. 추론시 비교대상이 되는 사례베이스는 전체 표본의 50%를 임의적으로 선택해서 사례데이터로 선정하였기 때문에 대기업의 경우에는 43개 사례, 중소기업의 경우에는 13개의 사례를 가진다. 그리고 일치율의 계산시 채권등급이 정확히 일치하는 경우(○)와 근사하게 일치하는 경우(△)를 모두 일치하는 것으로 간주하였다.

〈표 4〉 규모별 추론결과

규 모		Z-Score 모형 변수 (일치율)		K1-Score 모형 변수 (일치율)		K2-Score 모형 변수 (일치율)		CB 자료수
대기업	○	16	0.674	15	0.651	17	0.744	43
	△	13		13		15		
	×	14		15		11		
중소기업	○	6	0.750	4	0.500	4	0.583	12
	△	3		2		3		
	×	3		6		5		
총 계		55	0.691	55	0.618	55	0.709	55

주) ○ - 엄밀일치, △ - 근사일치, × - 불일치

CB : 사례베이스

추론의 결과를 보면, 대기업의 경우에는 K2-Score 모형 변수를 활용한 방법이 보다 가장 높은 일치율(0.744)을 나타냈으며, K1-Score 모형이 가장 낮은 일치율(0.651)을 나타냈다. 그러나 중소기업의 경우에는 Z-Score 모형이 가장 높은 일치율(0.750)을 나타냈으며, K1-Score 모형이 가장 낮은 일치율(0.500)을 나타냈다. 중소기업의 경우에는 대기업의 경우보다 각 모형간의 일치율의 차이가 크게 나타났다. 기업의 규모에 관계없이 평균적으로 보면 K2-Score 모형이 Z-Score 모형보다 예측일치도가 높으나 이들 간의 차이가 크지 않았다.

3.3 업종별 예측결과의 비교

표본데이터를 6개의 업종으로 세분하여 각 업종별로 Z-Score 모형, K1-Score 모형, K2-Score 모형을 이용해서 기업의 채무불이행 가능성을 평가하였다. 대상업종별 데이터의 수는 건설업종이 14개, 통신관련업종이 13개, 서비스업종이 13개, 석유화학업종이 21개, 일반제조업종이 38개, 기타 업종이 12개이다. 그리고 비교시 규모별 예측에서와 같이 각 모형들의 계산 값 중 판정보류영역의 값들은 채권등급과의 대응관계가 성립되기 어려워 제외시켰다.

〈표 5〉 업종에 따른 예측모형과 평가기관과의 예측일치도

업종		Z-Score 모형 (일치율)		K1-Score 모형 (일치율)		K2-Score 모형 (일치율)	
건설	일치	10	0.714	4	0.400	4	0.444
	불일치	4		6		5	
통신 관련산업	일치	3	0.273	11	1.000	12	1.000
	불일치	8		0		0	
서비스업	일치	7	0.636	8	0.800	8	0.800
	불일치	4		2		2	
석유화학	일치	9	0.529	9	0.643	10	0.667
	불일치	8		5		5	
일반 제조업	일치	21	0.583	15	0.682	18	0.720
	불일치	15		7		7	
기타	일치	3	0.273	9	0.818	9	0.818
	불일치	8		2		2	
총계	일치	53	0.530	56	0.718	61	0.744
	불일치	47		22		21	

주) 판정보류 결과는 제외

〈표 5〉는 이상의 기준에 따라 재무기법을 통한 업종별 예측결과를 평가기관들의 채권등급과 일치된 횟수와 이의 비율들을 나타내고 있다. 건설업종의 경우에는 Z-Score 모형이 가장 높은 일치율(0.714)을 나타냈으며, K1-Score 모형이 가장 낮은 일치율(0.400)을 나타냈다. 통신관련업종의 경우에는 K1-Score 모형과 K2-Score 모형이 완전 일치율(1.000)을 나타냈으며, Z-Score 모형은 매우 낮은 일치율(0.273)을 나타냈다. 서비스업종의 경우에도 K1-Score 모형과 K2-Score 모형이 동일한 일치율(0.800)을 나타냈으며, Z-Score 모형은 상대적으로 낮은 일

치율(0.636)을 나타냈다. 석유화학업종의 경우에는 K2-Score 모형이 가장 높은 일치율(0.667)을 나타냈으며, Z-Score 모형은 상대적으로 낮은 일치율(0.529)을 나타냈다. 일반제조업종의 경우에도 K2-Score 모형이 가장 높은 일치율(0.720)을 나타냈으며, Z-Score 모형은 상대적으로 낮은 일치율(0.583)을 나타냈다. 그리고 기타 업종의 경우에는 K1-Score 모형과 K2-Score 모형이 동일한 일치율(0.818)을 나타냈으며, Z-Score 모형은 매우 낮은 일치율(0.273)을 나타냈다. 평균적으로 보면 업종에 관계없이 K2-Score 모형이 가장 일치율(0.744)이 높으며, Z-Score 모형이 상대적으로 낮은 일치율(0.530)을 나타냈다.

〈표 6〉 업종별 추론결과

업종		Z-Score 모형 변수 (일치율)	K1-Score 모형 변수 (일치율)	K2-Score 모형 변수 (일치율)	CB 자료수	
건설	○	1	0.571	0	0.000	7
	△	3		0		
	×	3		7		
통신 관련산업	○	1	0.500	1	0.500	7
	△	2		2		
	×	3		3		
서비스업	○	0	0.000	0	0.000	7
	△	0		0		
	×	6		6		
석유화학	○	0	0.000	1	0.100	11
	△	0		0		
	×	10		9		
일반 제조업	○	3	0.222	2	0.111	20
	△	1		0		
	×	14		16		
기타	○	0	0.333	0	0.333	6
	△	2		2		
	×	4		4		
총계		53		53		58

주) ○ - 엄밀일치, △ - 근사일치, × - 불일치

CB : 사례베이스

〈표 6〉은 업종별로 표본자료들을 분류해서 각 업종별로 50%씩을 무작위로 사례베이스의

자료로 선택해서 각 모형에서 이용되는 재무변수들을 이용해서 예측의 일치율을 비교한 결과이다. 건설업종의 경우에는 Z-Score 모형만이 약 반 정도의 일치율(0.571)을 나타냈다. 통신관련업종의 경우에는 모든 모형이 동일한 일치율(0.500)을 나타냈다. 서비스업종의 경우에도 모든 모형이 평가기관들의 예측결과와 완전히 다른 것으로 나타났다. 석유화학업종의 경우에는 K1-Score 모형과 K2-Score 모형만이 매우 낮은 일치율(0.100)을 나타냈다. 일반제조업종의 경우에도 K2-Score 모형과 Z-Score 모형은 일치율이 0.222로 매우 낮게 나타났으며, K1-Score 모형은 더욱 낮은 일치율(0.111)을 나타냈다. 그리고 기타 업종의 경우에는 모든 모형이 동일한 일치율(0.333)을 나타냈다. 평균적으로 보면 업종에 관계없이 모든 모형이 매우 낮은 일치율을 보였다.

3.4 실험결과의 평가

이상의 실험을 통해 인공지능분야의 CBR기법이 재무적 위험의 예측기법으로서 유효성을 알아보기 위해 채무불이행위험모형에서 이용되는 재무변수들을 중심으로 전문평가기관들의 채권등급결과를 통해 기존의 재무적 기법과의 예측력을 비교해 보았다. 먼저 규모별 예측결과를 요약하면 <표 7>과 같다. <표 7>의 내용을 토대로 재무적 기법간의 예측률 차이, CBR내에서 이용변수간의 예측률 차이, 그리고 재무적 기법과 CBR간의 예측률 차이 등 3가지 측면에서 분석을 하고자 한다.

<표 7> 재무적 기법과 CBR의 예측일치도

규 모	Z-Score 모형		K1-Score 모형		K2-Score 모형	
	재무적 기법	CBR	재무적 기법	CBR	재무적 기법	CBR
대 기업	0.494	0.674	0.723	0.651	0.758	0.744
중소기업	0.667	0.750	0.692	0.500	0.706	0.583
평 균	0.530	0.691	0.718	0.618	0.744	0.709

먼저 재무적 기법간의 차이를 비율검정을 통해 비교해 본 결과(<표 8> 참조), 대기업의 경우에는 K2-Score 모형이 Z-Score 모형과의 차이만이 유의한 것으로 나타났다($\alpha = 0.01$). 중소기업의 경우에는 어떤 재무적 기법간에도 유의한 차이가 없는 것으로 나타났다. 그러나 평균적으로 보면 여전히 K2-Score 모형과 Z-Score 모형간의 예측률 차이가 유의한 것으로 나타났다. 평균적으로는 K2-Score 모형이 Z-Score 모형보다 높은 예측률을 나타냈으나 중소기업의

실험 데이터 수가 상대적으로 부족하였기 때문에 K2-Score 모형이 Z-Score 모형보다 항상 우수한 예측기법이라고 단정하기는 어려울 것 같다.

〈표 8〉 재무적 기법간의 예측률 차이분석

$\alpha = 0.01$	Z-Score 모형	K2-Score 모형
대 기업	0.494	0.758**
중소기업	0.667	0.706
평균	0.530	0.744**

CBR의 경우, 대기업에 적용시 K2-Score 모형의 재무변수를 이용하였을 경우가 가장 우수한 것으로 나타났다(〈표 9〉 참조). 그러나 CBR을 중소기업에 적용하였을 경우에는 Z-Score 모형의 재무변수를 이용하는 경우가 가장 우수하였다. 평균적으로 보면, 대기업과 같이 K2-Score 모형의 재무변수를 이용하였을 경우가 가장 우수한 것으로 나타났다. 그러나 통계적 검정결과는 각 사용 변수간에 예측률 차이가 없는 것으로 나타났다($\alpha = 0.05$).

〈표 9〉 CBR의 예측률 차이분석

$\alpha = 0.05$	Z-Score 모형변수	K1-Score 모형변수	K2-Score 모형변수
대 기업	0.674	0.651	0.744
중소기업	0.750	0.500	0.583
평균	0.691	0.618	0.709

재무적 기법과 CBR간의 예측률 차이를 비교해 보기 위해, 실험데이터가 충분한 대기업과 전체 평균에 대해 가장 높은 예측률을 나타낸 K2-Score 모형을 이용해서 두 기법간의 예측률을 검정해 보았다(〈표 10〉 참조). 검정결과는 두 기법간의 차이가 유의하지 않은 것으로 나타났다($\alpha = 0.05$). 그러나 CBR이 많은 사례들을 필요로 하는 기법임에 비추어 볼 때 상대적으로 적은 사례의 수(55개)로 다변량판별모형과 근사한 예측률을 나타냈기 때문에 본 논문에서는 데이터의 제약으로 못하였지만 사례베이스를 정상적으로 구축하여 CBR을 적용한다면 보다 좋은 결과를 얻을 수 있을 것 같다.

〈표 10〉 재무적 기법과 CBR간의 예측률 차이분석

$\alpha = 0.05$	CBR	재무적 기법
대기업	0.744	0.758
평균	0.709	0.744

그리고 업종별 예측결과를 보면 CBR기법의 적용결과는 매우 낮은 예측률을 나타내었으며, 재무적 기법의 경우에는 여전히 높은 예측률을 나타내었다. CBR기법의 경우에는 상대적으로 Z-Score 모형변수를 해서 건설업종에 적용한 결과가 작은 사례의 수(7개)에 비해 매우 높은 예측정확률(0.571)을 나타냈으나 실험의 횟수가 적어 우연의 일치일 가능성이 높아 적합한 응용 도메인이란 결론을 내리기는 어렵다. 재무적 기법의 경우에는 통신관련업종의 경우 K1-Score 모형과 K2-Score 모형 모두가 완전히 전문평가기관과 일치하는 예측을 한 것으로 나타났으며, Z-Score 모형은 매우 낮은 일치율(0.273)을 나타냈다. 그러나 건설업종의 경우에는 Z-Score 모형이 0.714의 예측정확률을 나타냈으나, K1-Score 모형과 K2-Score 모형은 각각 0.400과 0.414로 낮게 나타났다. 따라서 재무적 기법은 적용업종에 따라 기법간의 예측률 차이가 있을 수 있음을 알 수 있다.

IV. 결 론

CBR은 반복적인 경험을 통해 문제해결방법을 습득하는 방법으로서 일종의 귀납적 추론방법이라 할 수 있다. 따라서 CBR의 대상영역은 경험이 풍부한 영역이 되는 것이 일반적이다. CBR기법은 진단, Helpdesk 업무, 식단구성, 의료, 계획수립, 예측 등 매우 다양한 분야에 적용될 수 있다. 본 논문에서는 예측분야를 대상으로 CBR기법을 적용해 보았다. 적용결과는 비교적 성공적이었다고 평가할 수 있었다. 특히 규모별로 기업들을 분류한 후 행한 실험에서는 적은 사례의 수에 비교해 볼 때 매우 높은 예측정확률을 나타내었다. 그러나 업종별로 데이터를 분류해서 적용한 경우에는 데이터의 부족으로 만족스러운 결과가 나오지는 못하였으나 충분한 데이터가 확보되면 재차 분석해 볼 필요성이 있는 것으로 판단된다.

그리고 본 논문에서 추구하지 못한 또다른 과제는 어떤 재무변수가 CBR기법의 적용시 보다 우수한 예측정확률을 나타내는가이다. 본 논문에서는 주요 다변량판별분석모형에서 이용하고 있는 재무변수들을 그대로 이용하였으나 이들 변수가 CBR기법에서도 효과적 변수들이란

결론을 내릴 수는 없을 것이다. 실험에서 나타난 바와 같이 모든 데이터들은 상장법인들의 데이터였으나 우수한 모형은 상장기업 전용 모형인 K1-Score 모형보다는 상장 및 비상장기업 모두에 적용가능한 K2-Score 모형인 것과 같이 이론과 실제의 적용결과는 다를 수 있기 때문이다. 따라서 재무적 기법들에 대한 문헌적 연구와 병행해서 CBR기법의 적용시 유효한 재무변수를 추출하는 것도 향후과제가 될 수 있을 것이다.

참 고 문 헌

1. 남명수 · 김대호, 신경영분석, 삼영사, 1998.
2. 박정식 · 신동령, 경영분석, 3판, 다산출판사, 1997.
3. 정한규 · 김철중 · 윤평식 역, 가치평가론, 경문사, 1998.
4. 한국신용평가주식회사, 신용등급가이드(I), 1998.1.
5. 한국신용평가주식회사, 한국기업총람, 상장법인편, 1998.
6. Aamodt, A. and E. Plaza, "Case Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches," *Artificial Intelligence Communications*, Vol. 7, No. 1, 1994, pp. 39~59.
7. Aha, D.W., "Tolerating noisy irrelevant and novel attributes in instance-based learning Algorithms," *International Journal of Man-Machine Studies*, Vol. 36, 1992, pp. 267~287.
8. Alterman, R., "Panel Discussion on Case Representation," *Proceedings : Case-Based Workshop*, May, 1989, p. 15.
9. Altman, E.I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, September, 1968, pp. 589~609.
10. Altman, E.I., *Corporate Financial Distress*, 1983, John Wiley & Sons, New York, pp. 5~7.
11. Altman, E.I., "Distress Classification of Korean Firms," 『은행의 신용위험관리』, 한국금융연구원, 1996, pp. 379~410.
12. Belkin, N.J. and W.B. Croft, "Information Filtering and Information Retrieval : Two Sides of the Same Coin," *Communications of the ACM*, Vol. 35, No.12, 1992, pp. 29~38.
13. Gupta, K.M. and A.R. Montazemi, "A Connectionist Approach for Similarity Assessment in Case-Based Reasoning Systems," *Decision Support Systems*, Vol. 19, 1997, pp. 237~253.
14. Kolodner, J. and W. Mark, "Case-Based Reasoning," *IEEE Expert*, 1992, pp. 5~6.
15. Kolodner, J., *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.
16. Mahapatra, R. and A. Sen, "Case Base Management Systems : Providing Database Support to Case-Based Reasoners," *Journal of Database Management*, Vol. 5, No. 2,

1994, pp. 19~29.

17. Plaza, E., "Cases as Terms : A Feature Term Approach to the Structured Representation of Cases," *Lecture Notes in AI*, 1010, Springer Verlag, Heidelberg, 1995, pp. 265~276.

Abstract

Applying CBR for Default Risk Forecasting

Kim, Jin-baek

Case-Based Reasoning(CBR) offers a new approach for developing knowledge based systems. In case-based approach the problem solving experience of the domain expert is encoded in the form of cases. CBR has successfully been applied to many kinds of problems such as design, planning, diagnosis and forecasting.

In this paper, CBR was applied for forecasting default risk. The applied result was successful in spite of the small casebase. Generally, CBR requires large casebase. So, if the number of data was large, the result was better. But in this paper, what financial variable was more forecastable was not tested. Next, this should be tested.