

계량심리학의 방법론을 이용한 체계적인 전문가 지식구조분석 방법 : 비재무항목을 활용한 중소기업 신용평가전문가시스템 규칙개발에 적용*

이훈영** · 조옥래*** · 이시환****

A Psychometric Method for Structuring Expert Knowledge : Application to Developing Credit Analysis Expert Systems for Small-Medium Companies Using Nonfinancial Statement Information

Hoonyoung Lee** · Okrae Cho*** · Shihwan Lee****

Abstract

Translating expert knowledge into production rules has been the most difficult and time-consuming when building expert systems (Buchanan et al. 1983). Especially, building hierarchical structure, i.e. developing sequential or dominant relationship among production rules is one of the most important and difficult processes. Hierarchical relationship among rules has been typically determined in the course of interviewing human experts. Since this interviewing procedure is rather subjective, however, the hierarchically structured rules produced in terms of interviewing is widely exposed to the severe discussion about their validity (Nisbett and Wilson 1977; Ericsson and Simon 1980; Kellog 1982). We thus need an objective method to effectively translate human expert knowledge into structured rules. As such a method, this paper suggests the order analysis technique that has been studied in psychometrics (Cliff 1977; Reynolds 1981; Wise 1983).

In this paper, we briefly introduce the order analysis and explain how it can be applied to building

* 본 연구는 1996년도 학술진흥재단 자유공모과제 지원사업에서 지원을 받아 수행되었음.

** 경희대학교 정경대학 경영학부 조교수

*** 조흥은행 인력개발부 연수원 교수

**** 경희대학교 정경대학 회계학과 석사과정

hierarchical structure of production rules. We also illustrate how bankruptcy prediction rules of small-medium companies can be developed using this order analysis technique. Further, we validate the effectiveness of these rules developed by the order analysis, in comparison with those built by other methods. The rules developed by the proposed outperform those of the other traditional methods in effectively screening the bankrupted firms.

1. 서 론

전문가시스템(Expert Systems)에서 전문가란 전문지식을 기억하고 있고, 논리적 추론능력에 따라 결론을 도출하며, 설득력 있는 설명능력이 있는 사람으로 정의된다(이재규 1996). 우리나라에서도 전문가시스템(Expert Systems)에 관한 논의는 최근 학회와 업계의 많은 관심을 끌고 있으며, 여러 분야에서 그 효용성이 높게 인식되고 있다. 그러나 전문가시스템의 효용성에도 불구하고, 아직 전문가시스템의 개발 및 응용수준은 기초적인 단계에 있다. 전문가시스템의 발전이 정체되는 주요한 이유는 전문가의 지식을 효과적으로 추출하여 시스템 규칙으로 체계화하는 과정이 매우 어렵기 때문이다.

전문가의 지식을 추출해 전문가시스템의 생산법칙(production rule)으로 전환하기 위해 현재 사용하고 있는 방법들은 지식 공학자가 체계화되지 않은 전문가와의 인터뷰 내용과 전문가의 內觀에 주로 의존하고 있다. 이로 인해 비용과 시간이 많이 소요되는 것이 시스템의 개발에 큰 장애가 되고 있다(Buchanan et al. 1983). 일반적으로 기초지식이 부족한 전문가와의 인터뷰는 많은 시간과 인내를 요할 뿐 아니라, 지식 공학자들을 위한 효과적인 인터뷰 기술 개발 또한 상당한 교육과 연습을 필요로 한다. 더구나

전문가는 항상 정확하게 內觀할 수 없기 때문에 이러한 內觀에 의한 지식의 구축은 그 타당성에 대하여 많은 논란이 야기 되고 있다(Nisbett와 Wilson 1977; Ericsson과 Simon 1980; Kellog 1982). 그러나 계량심리적 도구를 사용하면 보다 체계적이고 효과적으로 전문가의 지식을 추출 체계화 할 수 있다. 또한 이러한 방식으로 추출된 규칙들은 통계적인 검증이 가능하므로 신뢰성이 높다.

계량심리학과 전문가시스템 개발은 지식 전달 및 판단, 그리고 의사결정과 같은 인간의 인지활동을 모형화 한다는 공통된 목적을 가지고 있다고 볼 수 있다. 그러나 최근 까지도 이들 두 분야간의 관계에 대한 공감대가 이루어 지지 않고 있었다. 즉 계량심리학자들은 인간의 심리적 과정을 수리적 모형으로 개발하는 데에 초점을 두고 연구해 왔고 (Coombs, Dawes 과 Tversky 1970), 컴퓨터 기술자와 경영정보 학자들은 인간인 전문가의 인지적 의사결정 과정을 대신할 수 있는 컴퓨터 시스템의 개발을 위한 지식베이스를 구축하는 데만 관심을 가지고 있었다. 따라서 발달된 각종 계량 심리적 도구를 사용하여 보다 체계적이고, 효과적이며, 신뢰성 있게 전문가의 지식을 추출하여 체계화 할 수 있는 방법에 대한 연구가 필요하다.

인간의 심리 상태 및 인식 과정을 수리적 모형으로 개발하는 데 초점을 둔 계량심리학이 이

러한 문제를 연구하기에 적합하다고 생각하여 적지 않은 연구가 수행되어 왔다. 예를 들어, Clancey(1984)는 계량심리적 방법이 전문가의 지식을 획득하는 과정에서 '문제의 구조적 분류' 및 '지식의 함축'에 매우 효과적이라고 주장하고 있다. 그는 전문가의 역할은 문제를 효과적으로 분할(decomposition), 분류(classification)하여 그에 맞는 적절한 방법을 제시하는 것으로 보고, 이러한 문제와 지식과의 효과적인 분류와 함축(abstraction)하는 전문가의 지식을 추출하는 계량심리학적 방법을 제시하였다. 그의 방법은 특정한 식사와 함께 하기에 적합한 포도주를 추천하는 전문가시스템인 'Wine Advisor'를 개발하는 데에 사용되었다(Teknowledge Inc).

또한 Butler와 Corter(1986)는 전문가의 지식 획득에 활용될 수 있는 계량심리학적 방법들을 체계적으로 소개하고, 그 중 additive tree와 extended tree에 의한 분석 방법의 효율성을 예시하였다.

전문가의 지식을 추출해 전문가 시스템의 생산법칙(production rule)으로 전환하기 위해 가장 중요한 부분중의 하나가 생산법칙 간의 계층적 구조이다. 즉 법칙간의 순서와 지배관계(order & dominance)이다. 예를 들어, 어느 한 규칙은 항상 다른 규칙에 선행되는 전제 조건이 되기도 한다. 이러한 경우, 여러 규칙간의 계층적 구조 분석은 대부분 소수 전문가와의 인터뷰 내용에 근거하여 수행된다. 그러나 이러한 정성적 분석방법은 분석자의 주관적인 판단에 많이 의존하여 객관성이 저하되는 단점이 있다. 따라서 가능한 한 주관성을 배제하고 객관성을 높이

기 위한 방법으로 계량심리학에서 연구되는 비대칭형 다차원척도법(DeSarbo and Manrai 1992; Holeman 1979)과 순서분석(Order Analysis) (Cliff 1977; Reynolds1981; Wise 1983)을 응용할 수 있다. 그러나 구체적으로 이러한 방법을 전문가의 지식추출에 적용한 연구 사례는 없었다.

본 연구에서는 이러한 계량심리적 도구를 개발하여, 실제로 중소기업평가 전문가시스템 개발에 필요한 비재무 항목간의 계층적 관계 존재 여부와 이런 계층적 관계의 지식체계화가 가능한지를 예시하고자 한다. 현재 우리나라 금융기관들이 운용하고 있는 신용평가제도는 대부분 재무적인 요인 뿐 아니라 비재무적인 요인도 포함하고 있다. 특히, 중소기업에 대한 신용평가에 있어서는 재무항목보다 비재무 항목에 더 많은 비중이 두어지고 있다.¹⁾ 이는 중소기업의 재무제표의 작성과정이 투명하지 않은 경우가 많아 신뢰도에 문제가 있기 때문이기도 하며, 또한 중소기업의 경우에는 경영자의 자질, 학벌 등 재무제표에 나타나지 않는 요소가 기업의 성공과 직접 연관되는 경우가 많기 때문이다. 따라서 실제 중소기업의 재무심사를 담당하는 심사자들은 종종 기업실사, 기업관계자와의 인터뷰 등을 통하여 비재무적인 요소를 파악하는데 많은 시간을 투자하고 있다. 그러나 비재무 항목에 대한 전문가의 지식은 문서화된 재무제표에 근거한 지식보다 체계적으로 정리하여 보존하기가 매우 어렵다. 지금까지 비재무 항목에 대한 심사와 분석은 거의 심사자의 경험과 주관에 의존해 온 것이 사실이다. 그러나 전문가의 지식

1) 96년도에 은행감독원이 개정 발표한 중소기업의 평가방법에는 비재무항목의 배점 비율이 종래 30%에서 최고 65%까지 높게 책정되었음.

이 시스템으로 체계화되어 있지 않기 때문에 심사자와 심사시기에 따른 편차가 크고 이는 심사의 보수성으로 연결되어 금융기관의 입장에서 보면 필요이상의 기회비용을 부담하고 있는 것이다. 따라서 심사의 전문성과 효율성을 높여 적극적으로 타 금융기관에 대한 경쟁력을 높이기 위해서는 회사 곳곳에 널려있는 전문가들의 지식을 체계화할 필요가 있다. 전문가 시스템의 개발을 통하여 심사에 관한 권한을 현장으로 대폭 이양할 수 있으며, 이는 경영의 효율성으로 이어질 수 있을 것이다. 따라서 중소기업 평가시 중요한 비재무 항목에 대한 심사전문가의 평가 지식을 체계화하여 컴퓨터 시스템(전문가시스템)으로 개발할 필요가 있다.

본 논문은 5장으로 구성되었다. 이후 2장에서는 순서분석(Order Analysis)의 개념과 ERGO 계수에 대하여 설명하고, 3장에서는 순서분석 방법을 이용하여 중소기업의 비재무 항목을 이용한 도산예측 규칙 개발을 예시 하였다. 4장에서는 3장에서 개발한 규칙의 신뢰성을 검증하기 위하여 상관관계가 높은 변수순서의 규칙과 무작위로 한 순서의 규칙, 그리고 전문가가 뽑은 순서의 규칙을 이용하여 기업도산을 가장 효과적으로 선별하는 정확도를 비교 논의하였으며, 끝으로 5장에서는 본 연구의 의의 및 한계점과 그리고 향후 연구방향에 대하여 언급하였다.

2. 순서분석(Order Analysis)

순서분석(order analysis)이란 어떤 속성간의 지배관계 행렬(item-by-item dominance matrices)를 통해 지배관계의 내부 일관성이 있다는

것을 어떤 값으로 나타내어 속성끼리의 종속적인 체인관계를 분석하는 것을 말한다. 이러한 속성간의 순서관계를 나타내는 계수를 Cliff계수 혹은 일관성계수(consistency index)라고 한다(Cliff 1977). 그리고 이러한 속성간의 사슬의 일관성을 최적화하는 방법으로 개발된 방법 중의 하나로 Guttman(1944)의 ERGO(Extracting Reliable Guttman Orders)알고리즘이 있다. 구체적으로 ERGO알고리즘은 다음과 같은 과정으로 사슬의 최적화를 이루고 있다.

예를 들어 표 2-1과 같은 S행렬이 있다고 하자. 행은 A, B, C, D 라는 Item을 나타내고 열은 가, 나, 다, 라 라는 Person을 나타내는 행렬에서 1은 Yes이고 0은No라고 하면 이것의 지배행렬 N은 $N = S' \times \bar{s}$ 식에 의해 도출된다.

〈표 2-1〉

	A	B	C	D
가	0	0	0	1
나	0	1	1	1
다	0	1	0	1
라	1	1	1	1

표 2-2과 같은 4x4행렬 (Item*Item)의 지배행렬을 다시 설명하면 i는 행, j는 열을 나타내고, i번째 행의 j번째열의 요소를 나타내는 N_{ij} 는 행의 변수 i는 On(혹은 'Yes'나 '1'로 표기)이고 열을 나타내는 변수 j가 Off(혹은 'No'나 '0'로 표기)하는 관측치의 개수를 의미한다. 예를 들어 〈표 2-2〉의 행렬에서 $N_{21}=2$ 은 전체 관측치 중에 변수 B는 On(1)이고 변수 A는 Off(0)인 사례 수가 2개임을 의미한다.

〈표 2-2〉의 행렬을 가능한 한 열의 합이 작은 변수를 좌측에 위치하도록 각 변수 열의 합의 오름차순으로 배열하면 〈표 2-3〉와 같이 된

다. 표 2-3의 행렬에서 좌측에 위치한 변수일수록 다른 변수가 1의 값을 가질 때, 0의 값을 갖게 되는 경우가 거의 없는 변수이고, 우측에 위치한 변수 일수록 다른 그 보다 좌측에 위치한 변수가 1일 때, 0의 값을 갖게 되는 경우가 많은 변수라고 볼 수 있다. 따라서 변수가 좌측에 위치할수록 대부분의 경우에서 1(On)의 값을 갖는 변수이고, 변수가 우측에 위치할수록 매우 드물게 1(On)의 값을 갖는다. 이러한 변수들 간의 상대적인 On-Off 관계를 나타내는 행렬에서 변수들 간의 다음과 같은 종속적 혹은 계층적인 관계를 추론하여 볼 수 있다. 즉 우측에 위치한 변수가 1(On)의 값을 가지면 그 보다 좌측에 위치한 변수도 1(On)의 값을 갖게 될 경우가 많다

<표 2-2>

	A	B	C	D
A	0	0	0	0
B	2	0	1	0
C	1	0	0	0
D	3	1	2	0

<표 2-3>

	D	B	C	A
D	0	1	2	3
B	0	0	1	2
C	0	0	0	1
A	0	0	0	0

위의 설명을 보다 쉽게 하기 위하여 다양한 난이성을 가진 수학시험 문제들간의 관계에 적용하여 설명하여 보자. 수학시험 문제 A, B, C, D가 있다고 할 때, 몇 학생들이 푼 결과는 S행렬로서 1은 답을 맞춘 경우이고 0은 틀린 경우를 의미한다고 하자. <표 2-2>은 수학문제의 지배행렬이고 이것을 다시 <표 2-3>의 형태로 정렬하여 보면 우리는 문제의 난이도가 A, C, B, D순 이라는 것을 짐작하여 볼 수 있다. 따라서 어려운 문제로부터 쉬운 문제에 이르는 A → C → B → D와 같은 계층적 관계사슬을 구상할 수 있다. 이 계층적 사슬의 의미는 다음과 같이

일반화하여 해석할 수 있다. 가장 어려운 A문제를 푼 학생은 그보다 조금 쉬운 C문제를 풀 수 있고, C문제를 푼 학생은 그보다 조금 쉬운 B를 풀 수 있다. 따라서 어려운 문제를 A를 푼 학생은 D문제를 풀었으리라 예상할 수 있다. 그러나 역으로 쉬운 문제 D를 푼 학생이 그 보다 어려운 문제 B, C, A 문제들을 풀었으리라 추론하기는 어렵다.

계층적 관계사슬에서의 변수의 순서와 더불어 중요한 점은 그 사슬이 얼마나 계층에 따른 일관성이 있는가를 측정하여 볼 필요가 있다. 이러한 사슬의 일관성정도를 측정하기 위하여 도입된 계수가 계층적 사슬의 일관성 계수이다. 계층적 관계사슬이 높은 일관성을 유지하기 위해서는, 기존의 사슬 A → C → B → D와 상반되는, 예를 들어 D → B나, D → A와 같은 역 방향의 요인이 없어야 한다. 이러한 역 방향의 관계가 많이 나타날수록 전체 문제의 난이도에 따른 계층적 관계사슬의 일관성이 떨어진다고 볼 수 있다.

변수의 계층적 관계행렬에서 일관성계수의 계산은 전체의 행렬을 계층적 관계의 순서에 맞게 배열했을 때의 행렬에서 상위행렬(upper matrix)에 위치한 값들은 전체변수의 계층적 관계에 일치하는 요인이고 그 외의 요인은 현재의 변수 계층적 관계에 일치하지 않는 요인이다. 따라서 전체 행렬의 합 S_m 과 올림차 순으로 정렬된 행렬의 상위행렬(upper matrix)의 합 S_u 는 다음과 같은 식으로 표현된다.

$$S_m = \sum_i \sum_j N_{ij}, \quad S_u = \sum_i \sum_{j>i} N_{ij}$$

변수 전환관계에 대한 인식이 없는 일반적인 행렬에서의 상위행렬의 기대값은 $1/2 S_m$ 이다.

일관성 계수는 으뜸차순의 정렬된 변수 계층적 관계를 나타내는 행렬에서 전체 행렬의 계층적 관계의 강도를 나타내는 상위행렬의 합이 특별한 계층적 관계가 없을 때 나타날 수 있는 상위행렬의 합에 대한 비율 값에서 1을 뺀 값을 나타낸다. 1을 뺀 이유는 계수의 값이 0에서 1까지의 값을 가질 수 있도록 하기 위하여 빼었다. 따라서 일관성 계수(consistency index)는 다음과 같은 식으로 표현 할 수 있다.

$$C = \frac{S_u}{\frac{1}{2} S_m} - 1$$

따라서 전반적인 변수 전환관계에 어긋남이 없는 위의 예의 경우에 일관성계수 C는 1(10/5-1)이다. 그러나 일반적인 경우에는 일관성계수의 값이 0에서 1사이의 값을 갖는다. 이러한 일관성계수의 개념을 이용하여 각각의 속성 요소가 어느 계층적 관계사슬에 속하여야 가장 바람직한가를 평가하는 작업이 필요하다. 이 작업은 일종의 군집분석과 유사한 개념을 가지고 있다. 단 군집분석에서는 변수나 자극 점간의 거리로 변수의 군집이 결정되나 순서도 분석에 있어서는 각각의 변수에서부터 시작하여 새로운 변수가 각각의 시작된 변수를 포함하는 계층적 관계사슬에 들어 감에 따라 일관성 계수가 어떻게 변동하는가에 근거를 두고 분석한다. 일반적으로 변수전환 행렬에 적합한 일관성 계수의 임계치(cut-off value)를 정한다. 이 임계치에 따라 계층적 관계사슬의 형성을 구체화한다. 즉 각각의 새로운 변수가 기존의 사슬에 추가 됨으로써 작아지는 일관성계수가 임계치보다 클 경우에만 그 변수를 계층적 관계사슬에 추가 시키는 방식으로 사슬을 형성한다.

3. 비재무 항목을 이용한 도산 예측 규칙개발에의 적용

3.1 자료수집

본 연구를 위하여 C은행의 기업분석 부에서 신용조사 한 업체 중 1994년도부터 1997년 1/4분기까지 도산업체 79개, 정상업체 78개 중소기업체를 대상으로 자료를 수집하였다. 조사항목은 96년도 은행감독원이 발표한 중소기업 평가안 중의 비재무 항목과 시중은행이 중소기업의 평가 시 사용하는 비재무 항목, 외국계 은행의 도산업체 조기신호(early warning signal), 그리고 C은행의 기업분석 부에서 도산업체 사후관리 시 원인으로 규명된 항목들을 전부 종합 분석하여 선정하였다. 조사항목은 크게 내적인 요인과 외적인 요인으로 나눌 수 있다. 외적인 요인으로는 시장요인, 관계회사요인으로 중분류하고, 내적인 요인으로는 경영자요인, 그 회사의 사업성, 경영요인(경영형태), 기술력 요인, 인수 및 합병 등의 전략적인 요인, 재무요인, 판매요인, 매입과 생산요인, 노무 후생요인, 회계처리 및 금융신뢰도 요인과 사업의 기반이 되는 사업장과 관련된 요인으로 나누었으며, 총 100개의 비재무 항목으로 구성하였다. 그리고 이것을 C은행의 기업 분석가들에게 설문 의뢰하여 조사하기 곤란한 변수들과 주관성이 강하고, 판단의 개연성이 높은 항목을 삭제하여 최종적으로 67개 비재무 항목으로 구성하였으며, 이 변수를 독립변수로 하여 167개 중소기업을 조사하였다.

위의 방법으로 선정된 67개의 독립변수와 도산여부를 나타내는 종속변수와의 상관분석을 통하여 상관관계가 0.15미만인 변수들을 제거하여

40개의 변수를 추출하였다. 그리고 분석의 효율성을 위해 선정된 40개의 변수를 가지고 요인분석을 실시하여 고유치(eigenvalue)가 1이상인 요인을 선정, 총 11개의 요인을 추출 40개의 변수를 각 요인 별로 할당하였다. 추출된 11개의 요인과 각각의 요인으로 할당 된 변수를 요인에

대한 상관계수와 함께 <표 3-1>에 요약하였다. 또한 모두 이진변수로 구성된 각 요인의 신뢰도를 검증하였다. 따라서 신뢰도검증으로는, 등간 척도로 측정된 변수로 구성된 요인에 사용되는 Crombach's Alpha값 대신에, 이진변수로 구성된 변수의 신뢰도 검증에 사용되는 KR20(Kuder

<표 3-1> 각 요인별 주요 변수와 신뢰도 측정계수

요인 명	요인구성 변수	요인항목간의 상관계	KR 20
요인1 : 취약한 사업기반	감사의견이 비적정	0.769	0.814
	분식결산징후 있음	0.762	
	노조 혹은 노사 협의회 없음	0.563	
	기업이 외감 대상 이하법인	0.512	
	재해 보상제도 없음	0.400	
요인2 : OEM, 자기상표의 매출이 50%이하	자가상표매출 50%이하	0.834	0.875
	OEM업체	0.690	
	자기명외공사 매출 없음	0.740	
요인3 : 기술성 및 대외신뢰도 낮음	수상실적 없음	0.295	0.564
	대주주구성이 정부, 상장대기업 등 아님	0.423	
	우수 협력업체 아님	0.320	
	공인된 품질마크 없음	0.286	
요인4 : 원재료의 구매조건 전기 대비 악화	원재료 구매조건 악화	0.586	0.677
	원재료 구득난 있음	0.506	
	재고자산금액 평균치보다 과다	0.315	
요인5 : 거래처 부실화	부실채권 매출채권 5%이상	0.561	0.726
	도산한 거래처 있음	0.561	
요인6 : 업력 및 경력 평균 이하	회사의 동 업종 경력 평균이하	0.413	0.597
	업력 평균 이하	0.385	
	경영자 나이	0.370	
요인7 : 환경대응능력 미흡	금융거래불량	0.365	0.516
	노사분규, 설비노후화 등으로 생산장애 사실 있음	0.283	
	외부환경의 변화로 악영향	0.335	
요인8 : 판매조건 취약	매출채권기간 90일 이상	0.466	0.568
	안정적인 판매처 미확보	0.812	
	판매조건 전기보다 악화	0.401	
요인9 : 경영성과, 추정 평균이하	경영성과 3년 중 1회 이상 적자	0.245	0.403
	추정매출액증가율 20% 이하	0.245	
요인10 : 치열한 경쟁	동업종간 경쟁이 치열함	0.372	0.526
	산업동향 평균 이하	0.327	
	생산직 인력부족현상 있음	0.257	
요인11 : 무리한 사업확장	신규사업 투자 있음	1.0	-

-Richardson Formula 20) 점수²⁾를 산출하여 제시하였다. KR20값은 일반적으로 Crombach's Alpha값보다는 낮아 0.403에서 0.875사이의 값은 갖는 것으로 측정되었다. 신뢰도가 낮은 요인은 앞으로의 분석에 사용하지 않는 것이 원칙이나 본 논문의 목적이 자료의 분석에 있는 것이 아니라, 도산예측 규칙 개발의 예시에 있으므로 지금까지 추정된 모든 요인을 향후 분석에도 사용하기로 하였다.

따라서 <표 3-1>에서 첫번째 열은 요인의 이름을 두 번째 열은 요인을 이루고 있는 변수, 세 번째 열은 요인과 개별 항목간의 상관관계로 이는 요인분석의 경우에는 요인과 변수간의 요인 적재량(factor loading)에 해당하며, 마지막 열에는 KR20 값을 기재하였다. 끝으로 최종 결정된 요인 값은 그 요인을 이루고 있는 각 변수를 합하여 구하였으며, 요인의 이름은 도산의

방향성에 일치시켜 구성 변수들의 특성을 감안하여 전문가의 의견을 물어 아래와 같이 결정하였다. 예를 들면 공인품질마크가 있거나 우수 협력 업체인 경우에는 '기술성과 대외신뢰도가 높음' 으로 이름할 수 있으나, 도산을 방향으로 하는 경우에는 반대의 의미인 '기술성 및 대외신뢰도 낮음' 으로 요인의 이름을 정하였다.

3.2 도산예측 규칙 개발

도산예측의 규칙 개발을 위하여 전체 157개의 70%인 111개를 측정용 샘플데이터를 이용하여 규칙을 개발하고, 전체의 30%에 해당하는 46개의 검증용 샘플데이터로 개발된 규칙의 예측 정확도를 측정하는 방법으로 검증하였다. 측정용 샘플데이터의 요인 지배 행렬은 다음과 같은 과정으로 만들었다. 우선 각각의 중소기업

<표 3-2> 요인간의 비대칭적 계층적 관계 행렬 N

	요인 1	요인 2	요인 3	요인 4	요인 5	요인 6	요인 7	요인 8	요인 9	요인 10	요인 11	요인 12
요인 1	0	25	24	36	44	23	34	32	12	16	45	14
요인 2	11	0	14	28	34	18	31	29	11	7	33	14
요인 3	17	21	0	31	39	25	33	31	15	16	41	17
요인 4	9	15	11	0	22	14	17	11	8	11	19	8
요인 5	7	11	9	12	0	9	11	7	5	9	15	3
요인 6	13	22	22	31	36	0	29	25	10	18	36	14
요인 7	12	23	18	22	26	17	0	19	5	12	26	8
요인 8	11	22	17	17	23	14	20	0	5	12	25	5
요인 9	30	27	40	53	60	38	45	44	0	29	61	26
요인 10	22	7	29	44	52	34	40	39	17	0	51	24
요인 11	5	43	8	6	12	6	8	6	3	5	0	3
요인 12	13	27	23	34	39	23	29	25	7	17	42	0
Total	150	243	215	314	387	221	297	268	98	152	394	136

$$2) KR20 = [N(N-1)] [1 - \sum p_i q_i / \sigma_x^2]$$

위의 식에서는 N은 요인을 구성하는 항목의 수, P_i는 i항목에서 1의 비율, Q_i는 i항목에서 0의 비율, σ_x²는 요인의 분산을 뜻한다.

평가사례를 표현하는 11개의 요인은 각 요인의 산술평균을 중심으로 이항변수로 바꾸었다. 즉 특정 사례의 요인 값이 그 요인의 산술평균보다 크면 1의 값을, 작으면 0의 값으로 변환하고 이에 각 사례에 도산사실(도산한 기업이면 1 건전 기업이면 0)을 추가하여 사례x요인(11x12)의 행렬 S를 만들었다. 행렬 S를 이용하여 요인간의 비대칭적인 계층적 관계 행렬N은 다음과 같이 표현할 수 있다.

$$N = S' \times \bar{S}$$

또한 행렬N의 구성 인자 n_{ij} 는 요인i는 1이고 요인j는 0인 전체 기업 사례의 수를 나타낸다. 예를 들어, $n_{12} = 37$ 이라는 것은 취약한 사업기반 (요인1)을 가졌으나 OEM, 자기상표매출 50%이하(요인2)가 아닌 기업의 수가 37개라는 것을 의미한다

<표 3-2>의 행렬을 각 요인의 열의 합이 작은 것을 좌측에 큰 것을 우측에 위치하도록 재배열하면 표 3-3과 같이 정렬된 요인간의 비대칭적 계층적 관계 행렬 N을 도출할 수 있다. 이

행렬의 좌측과 우측의 구분을 수학문제의 난이도와 비교하면, 좌측에 올수록 비교적 쉬운 문제이어서 적은 수의 학생들만이 풀지 못하였고, 우측으로 갈수록 난이도가 높아져 보다 많은 수의 학생이 풀지 못하였다고 유추 해석 할 수 있다.

<표 3-3>의 정렬된 행렬에서, 요인9(경영성과 나쁨)와 요인11(무리한 사업확장을 함)을 비교해보면, '경영성과'(요인9)가 나쁘나 '무리한 사업확장'(요인11)을 하지 않은 기업 수는 61개인 반면, '무리한 사업확장'(요인11)을 한 기업 중에 '경영성과'(요인9)가 나쁘지 않은 기업은 3개에 불과하다. 따라서 우리는 경영성과가 나쁘다고 무리한 사업확장을 하였다고는 볼 수 없으나, 무리한 사업확장을 하였으면 경영성과가 나쁘다고 할 수는 있다.

<표 3-3>의 행렬에서 요인의 순서는 각 열의 합을 올림차 순으로 배열한 것으로써 그 계층적 관계사슬의 강도는 역 방향의 전환요인이 적을 수록 높다. 이러한 계층적 관계의 강도에 대한 분석은 검증하고자 하는 사슬의 일관성 계수 값

<표 3-3> 정렬된 요인간의 비대칭적 계층적 관계 행렬 N

	요인 9	요인12	요인 1	요인10	요인 3	요인 6	요인 2	요인 8	요인 7	요인 4	요인 5	요인 11
요인 9	0	26	30	29	40	38	43	44	45	53	60	61
요인12	7	0	13	17	23	23	27	25	29	34	39	42
요인 1	12	14	0	16	24	23	25	32	34	36	44	45
요인10	17	24	22	0	29	34	27	39	40	44	52	51
요인 3	15	17	17	16	0	25	21	31	33	31	39	41
요인 6	10	14	13	18	22	0	22	25	29	31	36	36
요인 2	11	14	11	7	14	18	0	29	31	28	34	33
요인 8	5	5	11	12	17	14	22	0	20	17	23	25
요인 7	5	8	12	12	18	17	23	19	0	22	26	26
요인 4	8	8	9	11	11	14	15	11	17	0	22	19
요인 5	5	3	7	9	9	9	11	7	11	12	0	15
요인11	3	3	5	5	8	6	7	6	8	6	12	0
Total	98	136	150	152	215	221	243	268	297	314	387	394

에 의해 수행될 수 있다. 예를 들어 일관성 계수가 높으면 그 보다 작은 계수 값을 갖는 사슬보다 변수간의 계층적 관계가 강력한 것으로 생각할 수 있다. <표 3-4> 행렬의 행은 각각의 요인 하나 하나를 필수적으로 포함하는 계층적 관계사슬을 나타낸다. 각 행의 값들은 일관성계수 값으로서 일관성계수를 가장 높게 유지하는 방향으로 요인을 하나하나씩 늘려가면서 계층적 관계사슬을 만들 때, 점차 낮아지는 일관성 계수를 산출한 것이다.

예를 들어, 요인12를 필수적으로 포함하는 사슬(표 3-4의 둘째 행)을 생각하여 보자. 요인12만으로 구성된 사슬은 자기 하나뿐인 사슬로서 일관성 계수가 1이다. 다음으로 일관성계수를 가장 높게 유지하면서 추가될 수 있는 요인은 요인11로서 이 요인이 추가된 사슬(요인 11 → 요인 12)의 일관성계수는 1에서 0.867로 떨어지게 된다. 이와 같은 방식으로 하나하나 요인들이 추가적으로 연결됨에 따라 변하는 일관성계수 값이 <표 3-4>의 행렬이다. 따라서 행렬의

<표 3-4> 일관성 계수

	요인 9	요인12	요인 1	요인10	요인 3	요인 6	요인 2	요인 8	요인 7	요인 4	요인 5	요인 11
요인 9	1.000	0.817	0.632	0.519	0.465	0.587	0.490	0.770	0.547	0.667	0.731	0.906
요인12	0.817	1.000	0.632	0.519	0.465	0.587	0.490	0.770	0.547	0.667	0.731	0.867
요인 1	0.744	0.682	1.000	0.519	0.465	0.587	0.490	0.660	0.547	0.632	0.706	0.800
요인10	0.660	0.585	0.541	1.000	0.465	0.490	0.708	0.565	0.514	0.608	0.646	0.821
요인 3	0.690	0.650	0.574	0.512	1.000	0.465	0.488	0.628	0.538	0.599	0.662	0.673
요인 6	0.753	0.679	0.587	0.519	0.465	1.000	0.490	0.652	0.547	0.610	0.691	0.714
요인 2	0.734	0.675	0.596	0.541	0.465	0.490	1.000	0.626	0.514	0.574	0.667	0.650
요인 8	0.796	0.770	0.632	0.519	0.465	0.587	0.490	1.000	0.547	0.667	0.731	0.806
요인 7	0.800	0.741	0.615	0.519	0.465	0.547	0.490	0.651	1.000	0.583	0.700	0.784
요인 4	0.738	0.741	0.632	0.519	0.465	0.587	0.490	0.696	0.547	1.000	0.667	0.773
요인 5	0.786	0.857	0.632	0.519	0.465	0.587	0.490	0.731	0.547	0.667	1.000	0.761
요인11	0.906	0.817	0.632	0.519	0.465	0.587	0.490	0.770	0.547	0.667	0.731	1.000

<표 3-5> 각 사슬의 구성 요소 행렬

	요인 9	요인12	요인 1	요인10	요인 3	요인 6	요인 2	요인 8	요인 7	요인 4	요인 5	요인 11
사슬 1	1	1	0	0	0	0	0	1	0	1	1	1
사슬 2	1	1	0	0	0	0	0	1	0	1	1	1
사슬 3	1	1	1	0	0	0	0	1	0	0	1	1
사슬 4	1	0	0	1	0	0	1	0	0	0	0	1
사슬 5	1	0	0	0	1	0	0	0	0	0	1	1
사슬 6	1	1	0	0	0	1	0	1	0	0	1	1
사슬 7	1	1	0	0	0	0	1	0	0	0	1	0
사슬 8	1	1	0	0	0	0	0	1	0	1	1	1
사슬 9	1	1	0	0	0	0	0	1	1	0	1	1
사슬10	1	1	0	0	0	0	0	1	0	1	1	1
사슬11	1	1	0	0	0	0	0	1	0	1	1	1
사슬12	1	1	0	0	0	0	0	1	0	1	1	1

각행은 행렬의 대각선에 있는 요인을 필수적으로 포함하는 사슬의 일관성계수 값들이다

다음으로 일관성 계수의 임계치(cut-off value)를 정하는 것이 필요하다. 정해진 임계치에 따라 임계치보다 큰 일관성 계수를 유지하는 변수들만이 최종적인 계층적 관계사슬을 형성한다. 임계치를 정하는 데에는 일정한 규칙은 없으나 일반적인 신뢰도 검증에서와 같이 0.7이상 정도의 수준이 보통이다. 이 연구에서는 임계치를 0.85에서 0.65까지 0.05씩 낮추어 나가면서 어떤 요인이 추가되는가를 보고 이것에 따라 사슬을 구성하여 검증용 샘플데이터를 이용 규칙의 정확도를 알아보았다. <표 3-4>의 일관성계수의 행렬에서 임계치를 0.65로 할 경우 각 사슬의 구성 요인을 이항 변수로 표시한 행렬은 <표 3-5>와 같다.

위의 행렬에서 도산여부, 즉 요인12를 처음으로 시작하는 계층적 관계사슬(<표 3-4>의 2행, 사슬 2)은 다음 <표3-6>과 같다. 이것은 앞의 수학문제 난이도에 관한 설명에서 본 것처럼 행렬의 맨 우측에서 시작하여 맨 좌측으로 이어지는 계층적 관계사슬로 나타낼 수 있다. 즉, '무리한 사업확장을 함 → 판매조건 취약 → 거래처 부실화 → 원재료 구매조건 전기 대비 악화 → 도산 → 경영성과 및 추정 평균이하'로 이어지는 계층적 관계사슬로 나타난다. 이것을 말로 풀어 쓰면 무리한 사업확장을 한 기업이나 거래처가 부실화 된 업체, 판매조건이 취약한 업체, 원재료 구매조건 전기 대비 악화 업체 중에는 도산업체가 많은 것으로 말할 수 있으나, 도산업체가 반드시 거래처가 부실하거나 판매조건이 취약한 업체 라고는 할 수는 없다. 이는 가장 쉬운 문제를 푼 학생에는 가장 어려운 문제를 푼 학생이 포함되어 있겠지만 이 학생이 가장

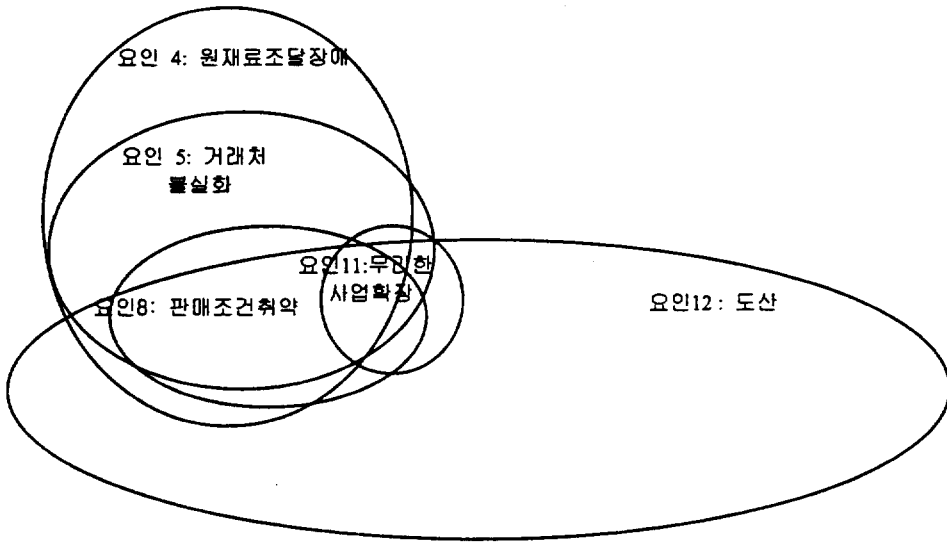
어려운 문제를 풀 수 있다고는 말할 수 없다는 것과 같다.

<표 3-6> 도산여부를 처음으로 시작하는 계층적 관계사슬

0.817 F9: 경영성과 및 추정 평균이하
1.000 F12: 부도
0.770 F8: 판매조건 취약
0.667 F4: 원재료 구매조건 전기 대비 악화
0.731 F5: 거래처 부실화
0.867 F11: 무리한 사업확장

위의 내용을 이해하기 쉽게 집합개념에 의한 다이어그램을 이용하여 설명하면 다음과 같다. [그림 3-1]에서 도산을 한 기업의 범위가 가장 크고, 다음으로 원재료 구매조건 전기 대비 악화 기업, 판매조건 취약한 기업, 무리한 사업확장을 한 기업의 순으로 그 범위가 크다. 무리한 사업확장을 한 기업들은 판매조건이 취약하며, 거래처가 부실하고 원재료 구매조건 전기 대비 악화로 대부분 도산하였다. 따라서 무리한 사업확장을 하면 도산한다고 추측할 수 있다. 그러나 역으로 도산을 한 기업은 모두 무리한 사업확장을 하였다고 추측하기는 어렵다.

따라서 이러한 계층적 관계사슬을 바탕으로 도산을 선별(screening)하는 규칙을 다음과 같이 생각하여 볼 수 있다. 일차적으로 무리한 사업확장을 한 기업은 도산으로 추정하고, 무리한 사업확장을 하지 않은 기업들은 사슬의 다음 단계인 '거래처의 부실 여부'로 도산을 추정하고, 거래처 또한 부실 하지 않은 기업은 '원재료의 구매조건 전기 대비 악화'로 도산을 추정하고, 원재료 구매조건 전기 대비 악화하지 않은 기업은 '판매조건취약 여부'로 도산을 선별하는 계층적 선별 규칙을 생각할 수 있다.



[그림 3-1] 집합개념에 의한 다이어그램

4. 규칙의 신뢰성 검증

본 장에서는 측정용 샘플데이터를 이용하여 순서분석에 의하여 개발된 규칙의 신뢰성을 검증할 위하여, 똑 같은 측정용 샘플데이터 (Estimation Sample Data)를 사용하여 순서분석 대신에 사용할 수 있는 여러 가지 방법에 의한 규칙을 개발하고, 개발된 규칙에다 검증용 샘플데이터(Holdout Sample Data)를 적용하여 그 효과를 비교 분석하였다. 즉 (1) 순서분석에 의하여 형성된 규칙, (2) 측정용 샘플데이터를 이용 사후측정 하여 각 단계별 가장 높은 예측율을 보이는 요인을 이용한 규칙의 조합, (3) 상관관계가 높은 요인의 순서에 의한 규칙, (4) 전문가의 의견에 의한 규칙과 (5) 무작위 추출에 의한 규칙을 검증용 샘플데이터에 적용시켜 그 결과를 비교 분석하였다.

4.1 순서분석에 의한 규칙

순서분석에 의하여 임계치(cut-off value)를 0.65로 할 때, '요인11 → 요인8 → 요인5 → 요인 4 → 요인12(도산)'에 이르는 사슬이 형성된다. 이러한 사슬의 계층적 구조로부터 불특정 다수의 업체 중에서 도산업체의 속성을 가지는 업체를 1차 스크린 하는 전문가시스템 규칙을 만들 수 있다.

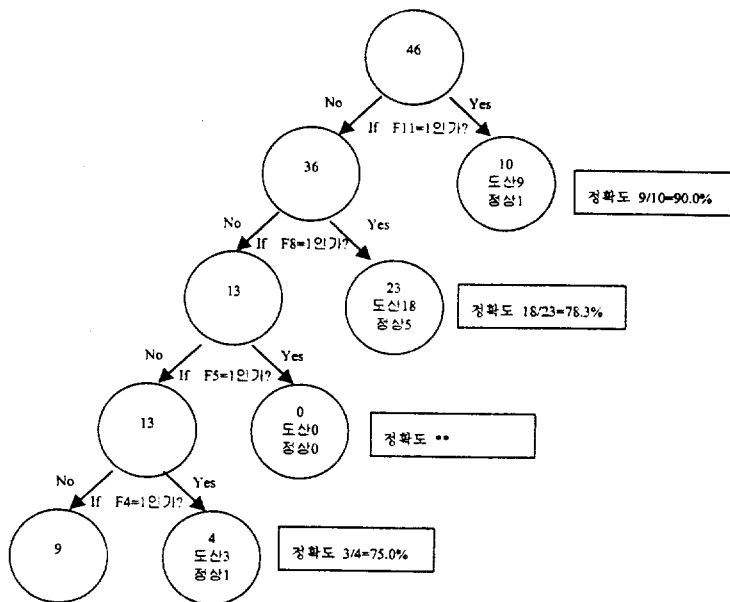
```

If 요인11=1 then 도산
Else If 요인8=1 then 도산
Else If 요인5=1 then 도산
Else If 요인4=1 then 도산
Else 세부적인 조사를 통하여 여신을 검토
Endif
    
```

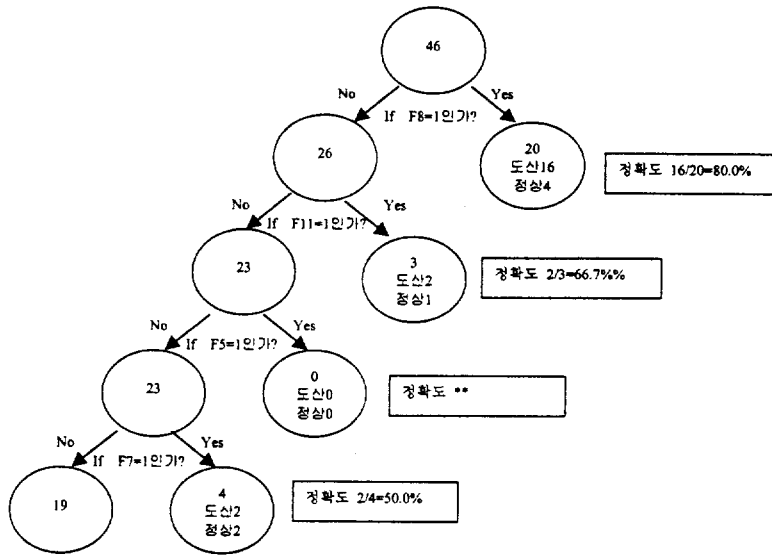
다음은 검증용 샘플데이터에 위의 규칙을 순서대로 적용하여 그 도산의 예측율을 추정하였다. 요인11을 제일 처음으로 적용한 결과 도산 기업으로 분류한 10개 업체 중 9개 업체를 도산으로 분류하여 도산 예측율이 90.0%, 두 번째 요인인 요인8을 앞의 정상업체로 분류한 36개 업체에 적용하니 총 23개 업체의 도산 업체로 분류한 것 중 18개 업체를 도산으로 분류하여 78.3%, 세 번째 요인인 요인 5로 13개 업체에 적용하니 도산으로 분류된 기업은 없었다. 마지막으로 요인4로 적용한 결과 도산으로 분류한 4개 업체중에서 3개를 예측하여 예측율은 75.0%를 보였다. 위에서의 4개의 단계적 규칙을 적용할 경우 전체적으로 77.8%의 도산 예측율을 보이고 있다

4.2 측정용 샘플데이터로 사후 평가하여 각 단계별 가장 높은 예측율을 보이는 요인의 순서대로 만든 규칙

측정용 샘플데이터를 이용하여 각 요인을 전부 개별적으로 적용하여 도산 예측율이 가장 높은 요인을 사용하는 방법으로 규칙을 만들어 이것을 검증용 샘플데이터에 적용하였다. 첫째로 가장 예측력이 높은 것은 요인8로서 80.0%의 적중률이 나타났다. 다음으로 요인8과 요인8을 사용한 규칙에서 걸러진 자료를 제외한 사례들을 나머지 요인11로 적용하여 본 결과 예측율이 66.7%로 나타났다. 같은 방법으로 하여 세 번째로 요인5를 적용한 결과 부도로 판별된 것은 없었고 다음으로 요인7을 적용한 결과 50.0%로 나타났다. 이러한 사후측정에 의한 규칙은 전체 74.1%로 나타났다. 측정용 샘플데이터를 이용한 사후측정 최고치는 모든 가능한 요인들을 전부 확인한 후에나 알 수 있어 요인의 수가 많은 경



[그림 4-1] 순서분석에 의한 규칙



[그림 4-2] 사후 평가하여 각 단계별 가장 높은 예측율을 보이는 요인으로 만든 규칙

우에는 어려움이 예상된다. 더욱이 규칙의 적용에 대한 이론적인 근거가 불충분하다는 단점이 있다.

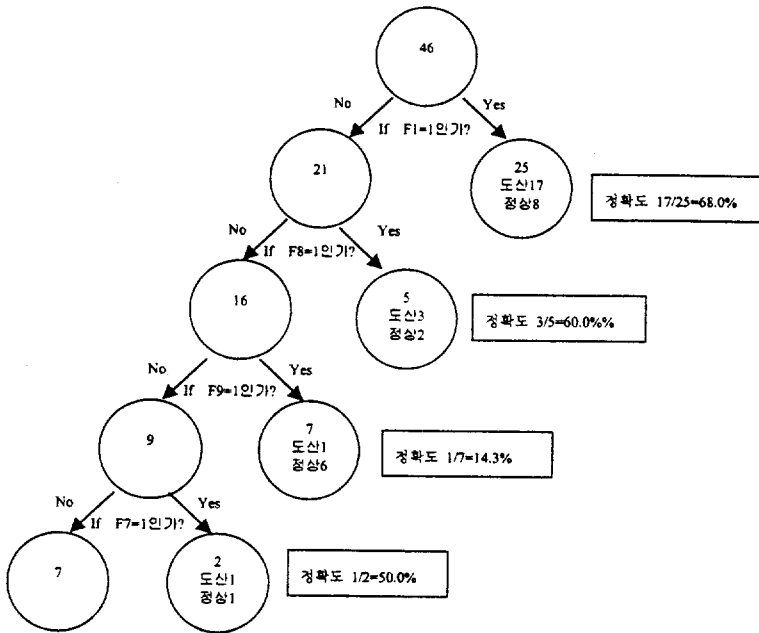
요인간의 상관관계를 이용한 도산 선별 규칙은 큰 효과가 없다고 할 수 있다.

4.3 상관관계가 높은 요인의 순서로 만든 규칙

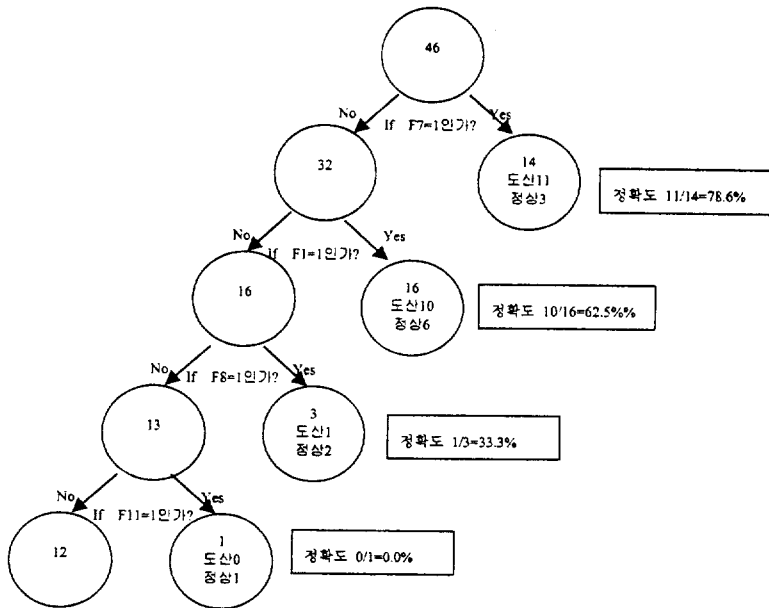
4.4 전문가의 의견에 따른 요인의 적용 순서에 의한 규칙

또 하나 가능한 방법으로 총 11개의 요인 중 상관관계가 가장 높은 순서에 따라 규칙을 만드는 방법이다. 상관관계는 가장 높은 4개의 요인은 요인1, 요인8, 요인9, 요인 7로서 상관관계가 높은 순으로 검증용 샘플데이터에 적용하여 그 예측율을 측정하였다. 그 결과 요인1을 적용하였을 때 68.0%, 요인8을 적용하였을 때 60.0%, 요인9을 적용하였을 때 14.3%, 요인 7을 적용하였을 때 50.0%로 나타났고 전체적으로는 56.4%로 상관관계가 높은 순서에 의한 규칙은 도산 예측율이 매우 낮은 것으로 나타났다. 따라서

다음으로 기업분석 전문가들 8명에게 가장 효과적으로 기업의 도산을 선별할 수 있는 요인 4개를 고르고 그 적용 순서에 대하여 의견을 물었다. 기업분석 전문가의 의견을 종합하면 요인 7 → 요인1 → 요인8 → 요인 11순으로 계층적 관계사슬이 형성되었다. 전문가들의 의견에 따르면 우선적으로 금융거래불량자이면 불량기업일 가능성이 가장 높고, 그 다음으로 분석결과와 비감사 대상 등의 기업의 취약한 기반 여부가, 그리고 판매조건이 취약성 여부, 마지막으로 무리한 사업확장을 보면 도산 기업을 읽을 수



[그림 4-3] 상관관계가 높은 순서



[그림 4-4] 전문가의 판단에 따른 요인의 적용 순서에 의한 규칙

있다고 말하고 있다. 이러한 순서 대로 규칙을 만들어 검증용 샘플데이터에 적용한 결과, 요인 7의 경우 78.6%, 요인1의 경우 62.5%, 요인8의 경우 33.3%, 요인 11의 경우 0.0%로 나타나 전체 도산 예측율은 64.7%이다.

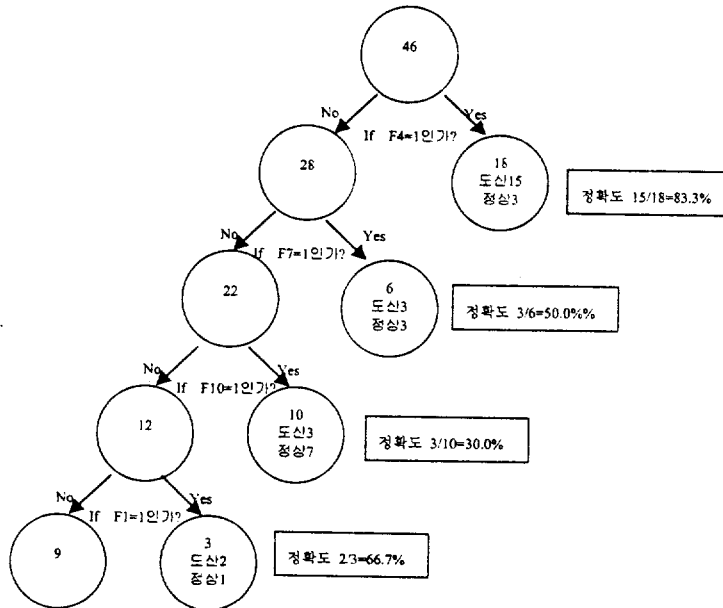
4.5 무작위 추출한 요인의 순서에 의한 규칙

마지막으로 무작위로 4개씩 요인을 조합하여 검증용 샘플데이터에 적용 그 예측율을 검증하였다. 10개의 조합 중 가장 예측력이 높은 조합은 요인4, 요인7, 요인10, 요인1으로서 그 결과는 첫번째 요인의 적용결과 18개를 도산으로 분류하였으나 그 중에 부도업체가 15개 있어 83.3%의 정확도를 나타내고, 두 번째 요인7의 적용함으로써 6개업체중 3개를 올바르게 판단하

여 50.0%, 세 번째 요인10의 적용으로 10개 업체 중 3개를 적중하여 30%, 마지막으로 요인1로 66.7%적중률을 보였다.

4.6 비교 분석 및 토의

이상으로 순서분석에 의한 방법을 비롯하여 상관관계가 높은 순서 등 5가지의 방법의 결과를 <표4-6>과 <표4-7>에 요약하였다. 표4-6은 측정용 샘플데이터의 결과를 <표4-7>을 검증용 샘플데이터의 결과를 요약하였다. 표4-6을 보면, 측정용 샘플데이터를 이용한 평가에서는 5가지 방법 중에 당연히 사후측정 최고치에 의한 방법이 가장 높은 78.7%의 정확도를 나타냈고, 다음으로 순서분석에 의한 방법이 75%로 가장 높은 선별력을 보여 주고 있다. 상관관계의 크기에



[그림 4-5] 무작위 요인의 순서에 의한 규칙

의한 방법은 전문가의 의견에 의한 방법과 유사하여 전문가의 판단 기준은 주로 경험에 의한 상관관계에 많은 비중을 두고 있음을 알 수 있다. 그러나 이러한 단순 상관관계는 우리가 추구하고 있는 기업 도산을 조기 스크린하기 위한 시스템 규칙을 만드는 데에는 큰 도움이 되지 않음을 실험 결과에서 보여 주고 있다. 또한 무작위 추출 방법을 10번 시도하여 그 중에 가장 높은 전체 예측율이 60.6%에 그친 것을 보면,

아무런 판단 기준이 없이 무작위로 만들어진 규칙에는 한계가 있음을 알 수 있다.

그러나 <표4-7>에서 보듯이, 검증용 자료를 이용한 평가에서는 순서분석에 의한 방법이 77.8%로 가장 높은 선별력을 나타냈고, 그 다음으로 사후추정 최고치에 의한 방법 74.1%의 정확도를 기록했다. 따라서 검증용 샘플데이터와 측정용 샘플데이터에 의한 정확도 분석결과 순서분석에 의한 방법이 규칙개발에 가장 바람직

<표 4-6> 측정용 샘플데이터(Estimation Sample Data)의 각 방법별 규칙의 정확도 비교 표

구 분	첫 번째 적용요인		두 번째 적용요인		세 번째 적용요인		네 번째 적용요인		종합 예측율
	요인	정확도	요인	정확도	요인	정확도	요인	정확도	
순서 분석에 의한 방법	요인 11	14/17 82.4%	요인 8	22/25 88.0%	요인 5	5/7 71.4%	요인 4	1/7 14.3%	42/56 75.0%
사후추정 최고치에 의한 방법	요인 8	31/36 86.1%	요인 11	5/6 83.3%	요인 5	5/7 71.4%	요인 7	7/12 58.3%	48/61 78.7%
상관관계 크기에 의한 방법	요인 1	33/57 75.4%	요인 8	7/11 63.6%	요인 9	5/20 25.0%	요인 7	0/3 0.0%	55/94 58.5%
전문가의 판단에 의한 방법	요인 7	27/35 77.1%	요인 1	22/34 64.7%	요인 8	5/8 62.5%	요인 11	1/1 100.0%	55/78 70.5%
무작위 추출에 의한 방법	요인 4	22/30 73.3%	요인 7	16/22 72.7%	요인 10	12/30 40.0%	요인 1	3/6 50.0%	54/85 60.2%

정확도 % = 옳게 판별한 개수/전체 판별한 개수

<표 4-7> 검증용 샘플데이터(Holdout Sample Data)의 각 방법별 규칙의 예측 정확도 비교 표

구 분	첫 번째 적용요인		두 번째 적용요인		세 번째 적용요인		네 번째 적용요인		종합 예측율
	요인	정확도	요인	정확도	요인	정확도	요인	정확도	
순서 분석에 의한 방법	요인 11	9/10 90.0%	요인 8	9/13 69.2%	요인 5	**	요인 4	3/4 75.0%	21/27 77.8%
사후추정 최고치에 의한 방법	요인 8	16/20 80.0%	요인 11	2/3 66.7%	요인 5	**	요인 7	2/4 50.0%	20/27 74.1%
상관관계 크기에 의한 방법	요인 1	17/25 68.0%	요인 8	3/5 60.0%	요인 9	1/7 14.3%	요인 7	1/2 50.0%	22/39 56.4%
전문가의 판단에 의한 방법	요인 7	11/14 78.6%	요인 1	10/16 62.5%	요인 8	1/3 33.3%	요인 11	0/1 0.0%	22/34 64.7%
무작위 추출에 의한 방법	요인 4	15/18 83.3%	요인 7	3/6 50.0%	요인 10	3/10 30.0%	요인 1	2/3 66.7%	23/37 62.2%

정확도 % = 옳게 판별한 개수/전체 판별한 개수

**는 부도로 판별한 기업이 없음.

한 방법임을 알 수 있다.

다음으로 임계치의 변화에 따라 추가되는 요인과 예측정확도의 변화를 표 4-8에 요약하여 보았다. 임계치 0.85에서 요인11이 선택되어서 정확도 90.0%로서 가장 높고, 임계치 0.75에서 요인 11과 요인8이 선택되어서 정확도78.3%로서 사후측정 최고치에 의한 방법과 함께 가장 높다. 또한 임계치 0.7에서는 요인 5가 추가되었으나 검증용 자료에서는 더 이상 부도로 분류된 것이 없어서 정확도는 78.3%로 같게 되었다. 마지막으로 임계치 0.65에서는 요인11, 요인8, 요인5, 요인 4가 선택되어서 전체 정확도는 77.8%로 5가지 방법 중에 가장 높았다. 임계치 0.65미만의 경우에는 대부분의 부도 기업이 선별된 다음이기 때문에 의미가 없어 분석은 생략하였다. 결과를 보면 일반적으로 임계치가 낮아질수록 선택되는 요인의 수가 증가하고 전체 정확도는 하락한다. 그러나 순서분석에 의한 스크리닝 규칙은 임계치의 변동에 무관하게 가장 높은 정확도를 나타내고 있다. 따라서 본 실험의 결과 순서분석에 의한 규칙 설정은 비교적 신뢰성이 높은 방법으로 생각 할 수 있다.

실험결과, 순서분석에 의한 규칙은 비재무 항목에 의한 기업 도산을 일차적으로 스크리닝(Screening)하기 위한 계층적 규칙을 설정하는데 유용한 방법으로 사용 될 수 있음을 알 수 있다. 더 나아가 순서분석은 전문가 시스템 개발 시 필요한 규칙의 계층적 구조를 체계화 하기 위한 하나의 효과적인 방법으로 응용될 수 있을 것이다.

5. 결 론

전문가시스템(Expert Systems)은 최근 학회와 업계의 많은 관심을 끌고 있다. 그러나 전문가의 지식을 효과적으로 추출하여 시스템 규칙으로 체계화하는 과정이 매우 어렵기 때문에, 전문가시스템의 개발 및 응용은 아직 기초적인 단계에 있다.

전문가의 지식을 추출해 전문가 시스템의 생산법칙(production rule)으로 전환하기 위한 방법들은 대부분 전문가와의 인터뷰 내용과 전문

<표 4-8> 임계치의 변화에 따른 순서 분석적용과 다른 방법의 도산분류 정확도 비교 표

구 분	0.85		0.75		0.70		0.65	
	요인	정확도	요인	정확도	요인	정확도	요인	정확도
순서 분석에 의한 방법	요인 11	9/10 90.0%	요인 8	18/23 78.3%	요인 5	18/23 78.3%	요인 4	21/27 77.8%
사후측정 최고치에 의한 방법	요인 8	16/20 80.0%	요인 11	18/23 78.3%	요인 5	18/23 78.3%	요인 7	20/27 74.1%
상관관계 크기에 의한 방법	요인 1	17/25 68.0%	요인 8	20/30 66.7%	요인 9	21/37 56.8%	요인 7	22/39 56.4%
전문가의 판단에 의한 방법	요인 7	11/14 78.6%	요인 1	20/30 66.7%	요인 8	21/33 63.6%	요인 11	22/34 64.7%
무작위 추출에 의한 방법	요인 4	15/18 83.3%	요인 7	18/24 75.0%	요인 10	21/34 61.8%	요인 1	23/37 62.2%

가의 內觀에 의존하고 있어 효과적인 전문가시스템의 개발에 큰 장애가 되고 있다. 또한 단순히 內觀에 의한 지식의 구축은 그 타당성에 대하여 많은 논란의 여지가 있다.

전문가의 지식을 추출해 전문가 시스템의 생산법칙(production rule)으로 전환하기 위해 가장 중요한 부분 중의 하나가 생산법칙 간의 순서와 지배관계(order & dominance)를 나타내는 계층적 구조이다. 본 연구에서는 순서분석에 의하여 시스템의 규칙에 사용할 요인의 추출은 물론 그들 간의 계층적 구조를 발견하는 방법을 활용하여, 실제로 중소기업의 도산 기업 스크리닝 규칙 개발에 적용, 그 효용성을 검증하였다.

본 연구의 결과는 다음 세 가지 측면에서 의미가 있다. 첫째, 계량심리학적 도구를 활용한 새로운 전문가 지식추출 체계화 방법의 개발이다. 전문가시스템(Expert Systems)개발에 있어 전문가로부터 지식을 추출하는 과정은 비용과 시간이 가장 많이 소요될 뿐 아니라, 추출된 지식 구조(규칙)의 타당성에 대해서도 많은 논란의 여지가 있다. 왜냐하면 지식의 체계화 과정이 너무나 임의적이고 주관적인 요소가 많기 때문이다. 본 연구에서 시도한 방법은 보다 효율적이면서 객관적이어서 타당성과 신뢰성이 높은 새로운 방법이다.

둘째, 인간의 지식 전달과 판단이나 의사결정과 같은 인간의 인지활동을 모형화 한다는 의미에서 공통된 목적을 지향하고 있는 계량심리학과 경영정보학의 두 분야간의 관계에 대한 공감대 형성하여 향후 두 분야간의 효과적인 학제간 연구의 가능성을 제시한다.

셋째, 비재무 항목을 활용한 효과적인 중소기업의 신용평가 도구의 개발이다. 중소기업에 대한 신용평가에 있어서는 재무항목보다 비재무

항목이 더 중요하다. 그러나 비재무 항목에 대한 전문가의 지식을 체계적으로 정리하여 보존하기가 매우 어렵기 때문에 지금까지 비재무 항목에 대한 심사와 분석은 전적으로 심사자의 경험과 주관에 의존해왔다. 그러나 전문가의 지식이 시스템으로 체계화되어 있지 않기 때문에 심사자와 심사시기에 따른 편차가 크고 이는 심사의 보수성과 심사자의 자의성 개입으로 연결되어 금융기관의 입장에서 보면 필요이상의 기회비용을 부담하고 있는 실정이다. 따라서 중소기업 평가 시 중요한 비재무 항목에 대한 심사전문가의 평가 지식을 체계화하여 기업 심사의 전문성과 효율성 향상을 위한 전문가시스템 개발의 기초(Prototype의 개발)를 제공 할 수 있다는 데에도 본 연구의 의미가 있다.

그러나 본 연구는 여러 제한 점과 그에 따른 향후 연구과제를 제시하고 있다. 그 중에 가장 대표적인 것은 사슬 형성을 위한 일관성 계수의 임계치(cut-off value)를 효과적으로 정하는 규칙이 없어 임의성이 있다는 것이다. 단순히 신뢰도 검증의 경우와 같이 0.7정도의 수준을 유지하는 것이 바람직하다는 휴리스틱(heuristic)에 의존하는 것이 고작이었다. 따라서 일관성계수 수준의 변경에 따른 규칙의 종합적 예측력을 검증하려고 하였다. 따라서 임계치 결정에 관한 보다 정교한 추가적인 연구가 필요하다.

또한 이렇게 순서분석을 통하여 만들어진 규칙은 본 연구에서 처음 시도되었다. 따라서 보다 많은 영역에 순서분석에 의한 전문가 규칙 개발 가능성에 대하여 추가적인 연구가 필요하다.

참고 문헌

- [1] 이재규 외5인, "전문가시스템 원리와 개발", 법영사, 1996, p18.
- [2] Buchanan, B. G., Barstow, D., Bechtal, R., Bennet, J., Clancy, W., Kulikowski, C., Mitchell, T., and Waterman, D. A., "Constructing an Expert System," *Building Expert Systems*, Hayes-Roth, F., Waterman, D. A., and Lenat, D. B. (eds.), Addison-Wesley, 1983
- [3] Butler, A. K. and Corter, E. J., "Use of Psychometric Tools for Knowledge Acquisition: A Case Study," *Artificial Intelligence and Statistics*, Gale, William (eds.), Addison-Wesley Publishing Co., Reading, Massachusetts, 1986.
- [4] Carmines, G. Edward and Zeller, A. R., "Reliability and Validity Assessment," *Sage Publication*, 1977, p48.
- [5] Clancy, W. J. , "Classification Problem Solving," AAA-84: *Proceedings of the National Conference on Artificial Intelligence*, Los Angeles, AAAI, 1984.
- [6] Cliff, N., "A Basic Theory Generalizable to Tailored Testing," *Psychometrika*, Vol. 2(1977), pp.327-400.
- [7] Coombs, C. H., Dawes, R. M., and Tversky, A., "Mathematical Psychology: An Elementary Introduction," Englewood Cliffs, N.N.(eds), *Prentice-Hall*, 1970.
- [8] Corter, J. and Tversky A., "Extended Similarity Trees," *Psychometrika*, Vol. 51, No. 3(1986), pp.319-45.
- [9] DeSarbo, W., and Manrai, A., "A New Multidimensional Scaling Methodology for the Analysis of Asymmetric Proximity Data in Marketing Research," *Marketing Science*, Vol. 11, No. 2(1992), pp.1-20.
- [10] Ericsson, K. A., and Simon, H. A., "Verbal Reports as Data," *Psychological Review*, Vol. 87(1980), pp.215-251.
- [11] Guttman, L.A., "A Basis for Scaling Qualitative Data," *American Sociological Review*, Vol. 9(1944), pp.139-150.
- [12] Holeman E.W., "Monotonic Models for Asymmetric Proximities," *Journal of Mathematical Psychology*, Vol. 20(1979), pp.1-15.
- [13] Kellog, R. T. (1982), "When Can We Introspect Accurately About Mental Processes?," *Memory and Cognition*, Vol. 10(1982), pp.141-144.
- [14] Nisbett, R. E., and Wilson, T. D., "Telling More Than We Can Know: Verbal Reports of Mental Processes," *Psychological Review*, Vol. 84(1977), pp.231-259.
- [15] Reynolds, T. J., "ERGO: A New Approach to Multidimensional Item Analysis" *Educational and Psychological Measurement*, Vol. 41(1981), pp.643-659.
- [16] "SAS Language Reference Version 6 First Edition," SAS Institute Inc.
- [17] Sattath, S. and Tversky, A., "Additive Similarity Trees," *Psychometrika*, No. 42(1977), pp.319-45.

[18] Wise, S., "Comparisons of Order Analysis and Factor Analysis in Assessing the Dimensionality of Binary Data," *Applied*

Psychological Measurement, Vol. 7, No. 3(1983), Summer, pp.311-321.