

데이터마이닝을 이용한 창업보육센터의 평가모델 개발

이동엽¹ · 김진욱^{2*}

¹창원대학교 경영대학원 테크노MBA학과 / ²창원대학교 산업시스템공학과

Development of Evaluation Model in Business Incubator Using Data Mining Process

Dong Youb Lee¹ · Jin Wook Kim²

¹Dept. of Techno-MBA, Graduate School of Business Administration, Changwon National University, Gyeongnam 641-773

¹Dept. of Industrial and Systems Engineering, Changwon National University, Gyeongnam 641-773

Numerous countries promote business programs to revitalize local economy, increase employment, and nurture high-tech industries. Recently, a number of business incubators have been established and operated with aims to adapt to changing environment and increase economic competitiveness in Korea. To give satisfactory results of governmental policy, the requirement to develop the evaluation model to support effective operations of business incubators using the objective and rational criteria is growing. The purpose of this study is to develop evaluation model in Business Incubator using Data Mining Process. We suggested the evaluation model of business incubator, 'Score-5 RS' consists of making evaluation factor process using weighted sum and 5-grade classification and analyzing process by Decision Tree algorithm.

Keyword: Business Incubators, Data Mining, Decision Tree

1. 서론

한국 경제의 산업경쟁력 강화와 미래지향적인 산업구조로의 전환을 목적으로, 지난 1997년 외환위기를 거치면서 중소·벤처기업들의 창업 활성화를 위해 정부지원 하에 창업보육센터(BI, Business Incubator)가 집중적으로 설립되었다. 또한 졸업생 고용흡수와 이미지 개선, 외형 확장 등을 노린 각 대학의 이해와 맞물리면서 대학을 중심으로 경쟁적으로 건립되어(전체 창업보육센터의 70% 이상이 대학부설로 운영), 해를 거듭할수록 보육기능 면에서 실효성이 떨어지고 있으며 경제적인 부담을 떠안게 된 대학들의 재정난으로 BI 설립취지에 부응하는 역할을 기대하기 어렵게 되었다.

그동안 정부는 BI에 대해 신규건립과 확장건립비를 지원했는데 지난 2001년부터는 운영비를 차등지원 하였고, 중소기업

특별위원회에서는 2003년 12월 'BI 지원시책 조사 및 평가보고서'를 작성하는 등 문제점을 개선하기 위해 나섰지만 구체적인 행동으로 옮기지는 않고 있다.

하지만 2004년 들어 정부는 창업보육센터의 공익적인 기능에도 불구하고 운영비 지원을 중단하고 자립시키는 쪽으로 정책방향을 잡았고, 이에 대해 체계적인 계획(로드맵)이나 지원이 없을 경우 자칫 'BI 붕괴'로 이어질지도 모른다는 우려도 나타나고 있는 실정이다.

최근에는 더 이상 평가를 늦출 수 없다는 공감대가 형성되고 있으며, 투자를 효율화하기 위해서 먼저 옥석 가리기를 선행하고자 하고 있다.

중기청의 입장은 기본적으로 BI의 공실률이 1년 이상 연속해 50%를 초과하거나 사업계획서상 센터 설립예정기한을 2년 이상 초과한 경우 지정을 취소하겠다는 방침을 가지고 있으며,

*연락처 : 김진욱 교수, 641-773 경남 창원시 사림동 9번지 창원대학교 산업시스템공학과, Fax : 055-266-4464,

E-mail : jinwkim@sarim.changwon.ac.kr

2007년 07월 접수, 1회 수정 후 2007년 08월 게재확정.

또 지원받은 출연금의 용도를 위반해 사용하거나 지정 취소되는 BI에 대해서는 건물설립 등에 출연했던 정부의 지원금을 일부 회수하고자 하고 있다.

하지만 본 연구 결과에서도 볼 수 있듯이 공실률이 50%를 넘는 대학과 연구 기관은 전체의 3% 정도로 극히 소수에 불과하므로 보다 합리적인 평가모델이 없다면 객관적인 평가에 큰 어려움이 발생할 전망이다.

본 연구의 목적은 전국의 328개 창업보육센터(Yang *et al.*(2003)에 의하면 2003년 7월 기준, 7개 중앙부처에 의해 지정되어 운영 중인 보육센터 수는 총 361개로 조사됨)들에 대한 설문조사 결과자료들을 이용하여 데이터베이스를 구축한 후, 5등급 정도의 스코어와 데이터마이닝(Data Mining)기법 중 하나인 의사결정나무기법(decision tree)을 이용하여 창업보육센터의 평가모델을 개발하는데 있다. 그리고 본 평가모델에서 결과변수는 정성적 항목을 주로 평가하던 기존의 방법에서 탈피하여 정량적 항목만을 이용함으로써 보다 합리적인 모델을 제시하고자 하였다. 본 연구를 통해 개발된 모델의 명칭은 ‘Score-5RS(RuleSet)’으로 하였으며, 이를 이용하여 객관적인 평가에 도움이 되어 창업촉진을 통한 경제 활성화와 고용창출, 그리고 지역경제발전을 도모하고자하는 BI정책 본연의 목적에 부합하는 창업보육정책 실행에 도움이 되고자 하였다.

2. 기존 연구의 검토 및 평가모델 소개

2.1 기존 연구의 검토

창업보육사업에 대한 평가와 관련한 기존의 연구들을 연도순으로 정리하여 살펴보면, Mian, Sarfraz A.(1994)는 미국대학창업보육센터의 조직 및 관리, 정책, 성과차원의 평가를 6개 대학을 표본으로 하여 기존 연구결과를 토대로 13항목으로 구분하고 이에 대해서 탐색적 연구결과를 제시하였으며, 1997년에는 대학기술창업보육센터의 평가요인을 성과, 운영전략 및 실행, 지원서비스 및 효과 등의 3가지 차원에서 성과를 평가하기 위한 모델을 제시하였다(Mian, Sarfraz A.(1997)).

양현봉(2001)은 “창업보육센터 실태분석 및 정책효율화 방안”에서 대부분의 창업보육센터가 설립 후 2년이 경과되지 않아 보육센터의 운영성과를 평가하는 데에는 시기적으로 한계가 있다는 점을 감안하여, 창업보육센터의 발전 잠재력을 중점적으로 평가한다는 측면에서 센터의 운영능력 및 지원서비스 2부문으로 나누어 평가를 실시하였다.

한국창업보육센터협회(2001)의 평가내용을 보면 BI운영능력(400)과 BI지원시스템(300), 그리고 BI운영성과(300)의 세 개의 분야로 나누어 1000점 만점의 배점표를 만들어 평가를 시행하였는데, 창업보육센터의 여건과 평가의 형평성을 위하여 BI 지정년도별로 평가대상을 그룹화 하여 상대평가를 실시하였다.

이상석(2002)은 “AHP(Analytic Hierarchy Process)를 이용한 창업보육센터의 평가에 관한 연구”에서 창업보육센터의 평가요인을 목표 및 운영전략, 물적 및 인적자원, 보육지원서비스, 연계프로그램의 네 분야로 나누어 평가하였는데, Thomas L. Saaty(1980, 1982)에 의해서 개발된 계층분석과정(AHP)을 이용하였다.

안준모(2002)는 “소프트웨어지원센터 운영평가 및 공공 민간 협력모델 개발”에서 DEA(Data Envelopment Analysis ; 자료포괄분석)를 이용한 전국 SW지원센터의 운영효율성 평가를 행하였는데, 이때 투입요소로는 ‘총운영비’, ‘전담인력수’, ‘총입주실수’가 사용되었으며, 산출요소는 ‘입주기업수’, ‘졸업기업수’, ‘고용인력수’, ‘총수출액’ 이었다.

중소기업청의 2003년 보육센터 운영 평가 기준을 보면 보육시스템(25), 전문성(25), 지원서비스(20), 운영성과(30)의 네 개의 분야로 나누어 100점 만점의 배점표를 만들어 평가를 시행하였다.

양현봉(2003, 산업연구원)은 “창업보육정책의 평가 및 발전방안”에서 개별 창업보육센터를 평가함이 아니라 중앙정부가 추진하는 창업보육정책을 평가하는데 목적을 두고, 정책형성의 ‘적합성’과 정책집행의 ‘효율성’, 그리고 정책성과의 ‘효과성’에 대한 세부 평가항목을 만들어 배점을 한 다음 각각 가중치를 적용하고 이것들을 합산해서 해당 부처의 창업보육정책에 대한 종합평점을 산출하였다.

이상의 내용에서 보듯이 창업보육센터의 평가에 있어 대부분의 경우 정성적 항목을 주로 평가하고 있으므로 배점과 관련하여 논란이 발생할 가능성이 많음을 알 수 있었다. 하지만 최근 들어 개별 창업보육센터의 평가는 물론 중앙정부의 정책을 평가하고자 보다 합리적인 평가모델을 개발하기 위해 부단한 노력들이 이루어지고 있음을 알 수 있다.

2.2 기존 평가모델의 소개

기존연구들에서 창업보육센터를 평가하기 위해서 사용되어진 방법들은 크게 평가항목에 따른 합산점수 또는 가중합(Weighted Sum)방법과, AHP(Analytic Hierarchy Process; 계층분석과정) 방법, 그리고 DEA(Data Envelopment Analysis ; 자료포괄분석) 방법이 있다. 아래에 이들 3가지 방법의 설명 및 장단점에 대해 간략히 기술하였다.

2.2.1 평가항목에 따른 합산점수 또는 가중합(Weighted Sum) 방법

현재까지 주로 사용되어 왔던 평가방법으로 평가항목에 따라 사전에 가중치 또는 점수를 부여하여 피평가기관이 받은 항목별 점수를 가중합(Weighted Sum)이나, 단순합산점수로 평가하는 방법이다.

이러한 방법은 평가가 용이하기는 하나, 평가항목에 부여되는 가중치의 결정에 있어 주관적인 요소가 개입될 가능성이 크기 때문에 피평가기관의 특성을 적절하게 반영하기가 곤란하다.

2.2.2 AHP(Analytic Hierarchy Process; 계층분석과정) 방법

Thomas L. Saaty에 의해서 개발된 AHP 방법은 의사결정과정의 관련요소들을 분해한 후, 요소들을 이원비교(pairwise comparison)하여 상대적 중요도(weight)를 구한 다음, 산출된 가중치를 종합해서 각 대안들에 대한 복합가중치를 구해 순위평가를 한다.

AHP가 복잡한 문제를 계층구조화하여 관리 가능한 문제들로 분해하고, 정량적이고 정성적인 평가기준 모두를 고려할 수 있다는 장점은 있으나, 다음과 같은 단점도 있다(Go et al. (2001), Bana e Costa(2000), Belton. V.(1983), Dyer, J.(1990)). 첫째, 평가기준간의 가중치 산출과 평가기준별 대안들의 상대적 선호도 측정을 위해 수행되는 이원비교는 평가기준 및 비교대안의 수가 추가되거나 삭제될 때마다 다시 수행되어야 하는 번거로움이 있다. 둘째, 비교대상(n)의 수가 증가함에 따라 의사결정자가 판단해야 할 평가횟수(n(n-1)/2)는 급증하게 된다. 셋째, 이원비교행렬에 대한 일관성 비율의 해석과 타당성 여부가 명확히 검증되지 못하였다. 넷째, 9점 비율척도를 사용함으로써 선호의 기수적 일관성(cardinal consistency) 문제가 발생할 수 있다. 다섯째, 대안이 추가되거나 삭제될 경우 기존 대안의 순위역전현상이 발생할 수 있다.

2.2.3 DEA(Data Envelopment Analysis; 자료포괄분석) 방법

DEA 모형은 다수의 산출을 생산하기 위하여 여러 종류의 투입요소를 사용하는 일련의 경영체들에 대하여 상대적 효율성을 측정·평가하기 위한 선형계획법의 응용기법으로, 즉 일련의 경영체 집단으로부터 관측된 산출과 투입을 비교함으로써 효율성을 측정하고, 특정의 경영체를 다른 경영체들과 비교하여 상대적으로 효율적인가, 그리고, 효율성의 크기는 어느 정도나 되는가를 나타내준다. 그리하여 다수의 투입과 산출이 존재하는 경우, 이들을 적절한 방법으로 하나의 지수로 종합화하기 힘든 경우에 유용하게 사용될 수 있으며, 또한 투입자료를 통제 가능한 투입자료와 통제 불가능한 투입자료로 2분화하여 인구밀도, 경쟁상황 등과 같이 환경에 관련된 변수도 투입자료로 적용할 수 있다.

하지만, DEA 모형은 다음과 같은 한계점을 가진다. 첫째, 기술적 관계를 나타내지 못하므로, 투입자원을 보다 싼 가격에 구입하였는가에 대한 여부나, 동일 산출물이라도 보다 좋은 서비스를 제공하였는가를 알 수 없다. 둘째, 효율성 평가시 전체를 비교집합으로 사용하지 못하고 산출과 투입이 비슷한 규모를 갖는 경영체들을 비교대상으로 하기 때문에, 모든 집단을 대상으로 한 전체적인 평가가 곤란하다. 셋째, 경영체의 효율성을 평가하기 위하여 투입·산출 요소를 적절히 측정하고 선정하는 것 역시 DEA 모형 결과의 정확성을 결정하는 중요한 문제이다. 넷째, 평가 대상 경영체 집합의 크기가 충분하지 않은 경우, 투입·산출과 관련된 모든 요소를 모형에 전부 포함시키기 어렵다. 다른 여타의 모형과는 달리 DEA 모형 특성상, 투입 및 산출 요소 수가 많아지면 분석의 결과가 적절치 않아

질 가능성이 높다. 즉, 결정변수가 되는 투입·산출 요소의 수가 많아질수록 효율성이 높게 나타나는 경영체의 수가 많아지게 된다. 경우에 따라서는 모든 평가가 대상 경영체의 효율성이 1.0(100%)일 수도 있다. 이는 각 경영체가 나름대로 가장 유리한 가중치를 선택할 수 있도록 하기 때문이다.

2.3 데이터마이닝 소개

데이터마이닝은 대량의 데이터로부터 새롭고 의미 있는 정보를 찾아내는 과정을 일컫는 용어이다. 이러한 데이터마이닝의 알고리즘 및 기법에는 연관규칙(association rule), 유전자 알고리즘(genetic algorithm), 신경망(neural network), 의사결정나무(decision tree), 링크분석(link analysis), 클러스터링(clustering), 베이시안 네트워크(bayesian network), 메모리 기반 추론(memory-based reasoning), 귀납규칙(rule induction), 퍼지집합(fuzzy set) 등과 통계적인 접근 방법이 있다.

본 연구에서 사용한 알고리즘은 의사결정나무(decision tree) 기법 가운데 하나인 C5.0 알고리즘을 사용하였다. C5.0은 1993년에 J.Ross Quilan이 개발한 알고리즘으로, 정보 획득 값이 최댓값을 가지는 입력 필드를 분류 필드로 하여 나무모형을 만들게 되는 것이며, 일반적으로 이 알고리즘을 ID3(Iterative Dichomizer 3)라고 한다.

C5.0 분석은 목표(Target) 필드를 이용하여 다른 입력변수들의 분류를 통해 세분화 분석이나 목표 필드를 예측할 수 있는 분석인데, 엔트로피(Entropy)라는 데이터의 무질서 정도를 측정할 수 있는 지수(Index)를 이용하여 정보이득(Information Gains)라는 값을 도출해 내고, 이를 이용하여 분류를 하는 분석이다.

3. 데이터마이닝을 이용한 평가모델 개발

3.1 ‘Score-5 RS(RuleSet)’: 데이터마이닝을 이용한 5등급 척도 스코어 방법

전술한 기존의 평가모델들은 주로 정성적 항목을 평가에 이용하고 있으며, 정량적 항목을 이용하고 있더라도 전체에서 차지하는 비중은 너무나도 미약하므로 주관적인 요소가 많이 개입되었다고 할 수 있다. 그리하여 보다 합리적으로 창업보육센터를 평가하고자 개발된 ‘Score-5 RS’ 모델은 기존 모델들과 비교해 볼 때 다음과 같은 특성들을 지니고 있다.

첫째, 정량적 항목만을 사용하여 평가를 함으로써 주관적 요소를 배제하였다.

둘째, 정량적 항목들을 이용하여 설정한 결과변수에 대하여 창업보육센터 자체적으로 쉽게 확인할 수 있는 모든(일반, 정성적, 정량적) 항목들을 원인변수로 하는 데이터마이닝 분석을 통하여 Rule(RuleSet)을 얻을 수 있다.

셋째, 이 Rule(RuleSet)을 이용하여, 개발된 모델의 결과변수

와 연관성이 높은 일반항목 및 정성적·정량적 항목들을 알 수 있다.

넷째, Rule(RuleSet)을 통하여 모델의 개선을 위한 기본 자료를 제공 받을 수 있다. 즉, Rule(RuleSet)을 살펴봄으로써 결과변수 설정단계에서 미비한 점 및 개선방향을 찾아낼 수 있다. 그러므로 본 평가모델은 프로세스 내부에 수정 및 개선의 도구가 포함되어 있다고 할 수 있다.

3.2 'Score-5 RS' 모델의 프로세스

본 연구에서 개발된 'Score-5 RS' 모델의 프로세스는 크게 2개의 단계로 나눌 수 있는데, 1단계는 결과변수를 설정하는 단계이고, 2단계는 데이터마이닝 분석을 하는 단계이다. <표 1>에 'Score-5 RS' 모델의 프로세스를 간략히 나타내었다.

먼저 결과변수를 설정하는 1단계에 있어 ;

첫째로, 운영성과 중 가장 대표적이라 할 수 있는 '입주중인 업체 수'와 '입주기업들의 매출액', '입주기업들의 고용인수' 세 가지 항목들을 선택하였는데, '입주중인 업체 수'는 창업보육센터 고유기능인 입주기업의 수를 나타내는 것이고, 다음의 '입주기업들의 매출액'은 그 보육센터에 입주하고 있는 기업들이 벌어들이는 매출액으로서 입주기업의 성장성을 평가할 수 있으며, '입주기업들의 고용인수'는 지역사회에 기여하고 있는 정도를 알 수 있는 중요 항목이기 때문이었다.

결과변수에서 향후 창업보육센터의 자립도 등의 평가를 위해 필요한, 수익효율성과 관련된 임대료, 성공보수제 관련 수익, 장비 및 서비스사용료 등의 항목을 제외한 것은 성공보수제 등의 수익관련 다변화 제도들이 최근 들어 급증하고 있는 단계이기도 하거니와, 전국의 창업보육센터 중 92% 이상이 중앙정부로부터의 자금지원을 받고 있는 현실과, 수익 금액이 그 자금지원 및 투자 합계 금액에 견주어 매우 적은 규모이기 때문이었다.

둘째로는 선택된 3항목들에 대한 정량적 자료들을 순위별로 나열한 다음, 누적비율이 매 20% 증가할 때마다 1점을 부과하는 방식의 최저 1점, 최대 5점인 5등급 평가를 행하였다.

셋째로는 3항목들의 점수에 각각 가중치를 부여하여 합한

점수가 100점 만점이 되도록 환산하여 Scoring을 행한 다음, 이 결과를 순위별로 다시 누적비율 5등급 평가를 행하여 A, B, C, D, F의 5개 등급으로 구분하였다.

본 연구모델에서는 '입주중인 업체 수'에는 0.5의 가중치를, '입주기업들의 매출액'과 '입주기업들의 고용인수'에는 각각 0.25의 가중치를 부여하였는데, 이는 입주기업을 조사하는 과정에서 실제의 회사 및 공장이 창업보육센터 외부에 이미 존재하지만 공동개발 등의 필요에 의해 사무실 규모만을 입주한 사례들이 관찰되어 운영성과결과가 과장될 수 있기 때문에 '입주중인 업체 수'에 높은 가중치를 주었다.

다음으로, 데이터마이닝을 이용해 분석하는 2단계에서는 ;

총 133개의 항목으로 데이터베이스화된 설문항목들 중 구분을 위해 일련번호를 붙인 'a01'의 CODE와, 센터명 및 센터장의 이름을 나타내는 'a04', 'a05', 그리고 운영성과 항목인 'i01~i15', 창업보육사업 관련 정부정책 및 제도에 관한 설문항목인 'k01~k09'를 제외한 설문항목들을 모두 원인변수로 설정하여 의사결정나무기법 중의 하나인 C5.0 알고리즘을 이용하여 Rule (RuleSet)을 구하였다. 이 때 모델의 형성을 위해서는 표본을 50%씩 무작위 추출로 나누어 두 그룹으로 나눈 다음 한 그룹만을 사용하여 모델을 만들고, 나머지 한 그룹은 모델의 검증에 사용하였다.

설문항목들을 원인변수로 설정함에 있어서의 기준은 창업보육센터가 자체적으로 쉽게 확인할 수 있는 모든(일반, 정성적, 정량적) 항목들을 원인변수로 하였다. 그리고 원인변수에 운영성과 항목들을 포함하지 않은 이유는, 가급적 운영성과에 미치는 원인변수에 해당하는 항목들만을 이용하여 그 연관성을 유추하고자 하였기 때문이다. 본 연구에서 '입주중인 업체 수' 항목은 원인변수에 포함시켰는데, 이것은 창업보육센터의 평가에 사용되는 중요 운영성과 항목이기는 하지만 동시에 각 창업보육센터의 다른 운영성과들을 설명할 수 있는 기본항목이기 때문이었다.

3.3 평가대상

본 연구에서는 중앙정부가 1998~2003년도 기간 동안 지원

표 1. 'Score-5 RS' 모델의 프로세스

단계	내용	활동
1	결과변수의 설정	① 조사항목 중 결과변수 항목을 선택: (본 연구모델의 경우) 입주중인 업체수, 입주기업들의 매출액, 입주기업들의 고용인수
		② 선택된 각 항목들에 대한 정량적 자료들을 순위별로 나열한 다음, 누적비율이 매 20% 증가할 때마다 1점을 부과하는 방식으로 최저 1점, 최대 5점인 5등급 분류를 행함.
		③ 선택된 각 항목들의 점수에 각각 가중치를 부여하여 합한 점수가 100점 만점이 되도록 환산하여 Scoring을 행한 다음, 이 결과를 순위별로 누적비율 5등급 평가를 행하여 A, B, C, D, F의 5개 등급으로 구분함.
2	데이터마이닝 분석	④ 설문항목들을 원인변수로 설정하여 의사결정나무기법 중의 하나인 C5.0 알고리즘을 이용하여 Rule(Rule Set)을 구함.
		⑤ 모델 개발을 위해 무작위 추출해 사용하고 남은 나머지 50%의 표본을 사용하여 개발된 모델의 예측의 정확도를 확인.

해 오고 있는 창업보육센터를 평가대상으로 설정한다는 전제 아래 조사한 결과를 바탕으로, 평가대상을 7개 중앙부처의 9개 창업보육사업으로 한정하였다(2003년 7월 기준, 총 361개로 조사됨). 9개 창업보육사업을 부처별로 살펴보면, 중소기업청 2개(일반창업보육사업, 여성창업보육사업), 정보통신부 2개(대학정보통신창업지원센터 지원사업, S/W지원센터 창업보육사업), 문화관광부 1개(문화산업지원센터 창업보육사업), 산업자원부 1개(테크노파크 창업보육사업), 농림부 1개(창업보육사업), 과학기술부 1개(신기술창업지원단 창업보육사업), 환경부 1개(환경신기술창업보육센터 지원사업)이다.

3.4 평가내용

본 연구에서 사용된 평가내용의 영역을 크게 살펴보면, 센터개요와 관련된 ‘일반사항’, 재원조달, 센터장 및 매니저, 운영 및 자문기구, 외부와의 네트워크, 운영정책에 관한 내용인 ‘창업보육센터의 운영능력’ 부분, 그리고 ‘창업보육센터의 지원서비스’ 및 ‘창업보육센터의 운영성과’ 부분이다. 덧붙여 평가내용과는 관련이 없지만 ‘창업보육사업 관련 정부 정책 및 제도’에 관한 내용도 조사하였다.

4. 평가모델의 적용결과

본 연구에 의해 개발되어진 ‘Score-5 RS’ 모델의 과정 중 1단계인 결과변수의 설정단계에서의 결과를 <표 2>에 나타내었다. 전국에 있는 창업보육센터들을 평가함에 있어 운영성과 중 가장 대표적이라 할 수 있는 정량적 항목인 ‘입주중인 업체수’와 ‘입주기업들의 매출액’, ‘입주기업들의 고용인수’를 사용하여 최종적으로 5등급 구분을 한 결과이다.

표 2. ‘Score-5 RS’ 과정 중 결과변수 설정단계에서의 결과

	등급	점수범위 (100점 만점)	빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	A	85~100	57	17.4	19.3	19.3
	B	70~80	49	14.9	16.6	35.9
	C	55~65	70	21.3	23.6	59.5
	D	40~50	59	18.0	19.9	79.4
	F	20~35	61	18.6	20.6	100.0
	합 계		296	90.2		
결 측			32	9.8		
합 계			328	100.0		

1단계 과정에서 설정된 결과변수에 대해 설문항목들 중 창업보육센터가 자체적으로 쉽게 확인할 수 있는 모든(일반, 정성적, 정량적) 항목들을 원인변수로 하여 2단계인 데이터마이닝 분석을 하였다. 데이터마이닝 분석은 설문조사 결과를 임의 추출하여 두 그룹으로 나눈 다음 한 그룹 즉 50% 만을 사용

하여 모델을 형성하였으며, 개발된 모델을 이용하여 미리 나누어 놓은 나머지 그룹에 적용함으로써 개발된 모델의 예측을 알아보았다. 이 결과를 <표 3>에 나타내었다. 각 평가등급에 대해 평균적으로 93% 정도의 높은 예측의 정확도를 나타내고 있음을 알 수 있었다.

표 3. ‘Score-5 RS’ 과정 중 데이터마이닝 분석단계에서의 예측율 결과

실제 예측	A	B	C	D	F
A	96.491	1.754	1.754	0	0
B	8.163	81.633	6.122	4.082	0
C	1.429	0	98.571	0	0
D	0	0	1.695	96.61	1.695
F	0	0	3.279	4.918	91.803

<표 4>에는 ‘Score-5 RS’ 모델의 Rule(RuleSet)에서 사용된 주요항목을 나타내었고, 각 등급의 평가 시 사용된 항목들에 표를 하였다. 처음에 원인변수로 사용하기 위해 조사되어진 항목수는 모두 106항목이었으나, 평가모델의 데이터마이닝 분석 결과 사용된 항목은 불과 37항목에 불과하며, 그 중 상위등급인 A, B 등급의 평가를 위해서는 20항목만이 사용되었다.

<그림 1>은 ‘Score-5 RS’ 모델의 Rule을 나타내며, 등급이 각 항목의 조건에 따라 나무구조로 구분되어져 있음을 알 수 있다.

Rule(RuleSet)에는 (○,○)와 같은 형태로 2개의 숫자가 적혀 있는데, 이것은 각각 생성된 각 가지의 레코드 수에 대한 정보를 나타내는 포함정도(coverage)와 결과의 신뢰 수준을 보여주는 신뢰도(confidence)를 나타낸다.

이처럼 정량적 항목으로만 설정된 결과변수가 설문항목의 다양한 일반, 정성적, 정량적 항목을 원인변수로 하여 해석될 수 있으며, Rule(RuleSet)에 나타난 원인변수 항목들이 여러 원인변수들 중에서도 특히 결과변수에 영향을 미치고 있다고 할 수 있다.

이것은 기존의 평가모델들이 창업보육센터의 평가에 있어 정성적·정량적 항목을 사용하면서도 그 중요도 및 배점의 근거를 단순히 전문가의 판단에 의존하고 있던 부분을, 데이터마이닝 분석에 의해 합리적으로 설명할 수 있음을 보여주고 있다.

그러므로 결과변수의 설정과 데이터마이닝 분석의 2단계로 이루어진 ‘Score-5 RS’ 모델은 창업보육센터의 본연에 목적에 부합하는 정량적 항목만을 이용하여 결과변수를 설정하고 평가를 함으로써 보다 합리적이라 할 수 있으며, 또한 데이터마이닝 분석을 통하여 여러(일반, 정성적, 정량적) 항목들이 결과변수에 미치는 중요도를 알 수 있음으로써 기존의 평가모델에서의 배점선정시의 모호한 부분 역시 보다 명확히 설명할 수 있다는 장점이 있다. 뿐만 아니라 이 Rule(RuleSet)은 결과변수 설정단계에서의 미비한 점 및 개선 방향을 찾아낼 수 있는 기

본 자료가 된다.

표 4. 'Score-5 RS' 모델의 Rule에서 사용된 주요항목

No	설문 항목	내 용	설문항목 사용유무				
			A	B	C	D	F
1	a03	지역			*	*	*
2	a06	입주 후 개월			*	*	*
3	a07	센터규모(평)			*	*	
4	a08	입주면적(평)	*		*		
5	a09	입주면적/센터규모(%)				*	*
6	a10	건물취득형태				*	*
7	a11	전용성 여부	*	*			
8	a12	운영형태	*	*	*	*	*
9	a13	입주가능보육실수	*	*	*	*	*
10	a14	현입주업체수	*	*	*	*	*
11	a15	입주율	*	*	*	*	
12	a17	시제품제작-형태수			*	*	
13	a18	일반시관-형태수	*	*	*		
14	a20	벤처-특성수			*	*	*
15	a22	예비창업기업-특성수	*	*	*	*	
16	b01	건물-시설투자비 지원합계		*	*		
17	b03	운영비지원합계(-2003)				*	*
18	b08	기타 지원합계		*	*	*	
19	b09	총 지원합계			*	*	
20	c01	전담인원 지원수		*	*		
21	c03	센터장 직책겸임유무			*	*	*
22	c05	센터장의 필요자질-1	*	*	*		*
23	c07	센터장의 업무수행 보상방식	*	*			
24	c14	매니저 업무수행 보상방식			*		*
25	d07	전문가 풀 활용도			*	*	
26	d09	보육닥터 및 전문가 풀 제도 시행상황	*	*			
27	f02	입주기업 선정기준			*	*	*
28	f03	기본입주 개월			*	*	*
29	f04	입주연장 개월	*	*			
30	f06	면담개최	*		*		
31	g01	건물 및 설비 유지보수비 (02+03)		*	*	*	*
32	g02	전담인력 인건비(02+03)		*	*	*	
33	j02	섬유 의복 가족				*	
34	j03	화학 화공				*	
35	j08	정보통신 멀티미디어			*	*	
36	j15	기타 서비스업	*	*	*	*	
37	j16	입주기업 합계		*	*		

<그림 1>의 내용을 개략적으로 살펴보면 'A' 또는 'B'와 같은 높은 등급과 관련된 조건으로는 먼저 현입주업체수(A14)가 12를 초과하여야 한다. 다음으로 입주가능보육실수(A13)가 19를 초과할 경우에는 입주율(A15)이 57%를 초과하며, 입주면적(A08)이 550평을 초과하는 경우가 유리함을 알 수 있다. 그러나 입주가능보육실수(A13)가 19이하일 경우에는 기타 서비스업 업체수(J15)가 6보다 이하이면 예비창업기업 업체수(A22)가 0 이하 즉, 없어야만 높은 등급을 받을 수 있음을 알 수 있다. 또한 이 경우 기타 서비스업 업체수(J15)가 6을 초과하거나 예비

창업기업 업체수(A22)가 존재하는 경우에는 'A' 또는 'B'와 같은 높은 등급을 받는 것은 극히 어려움을 알 수 있다.

여기서 입주가능보육실수(A13)가 19이하일 경우에 있어 창업보육센터 설립의 취지와는 조금 거리가 멀게 느껴지는 기타 서비스업 업체수(J15)가 6을 초과하여 높은 등급을 받지 못함은 어느 정도 이해할 수 있지만, 예비창업기업 업체수(A22)가 없어야만 높은 등급을 받을 수 있다는 것은 처음엔 쉽게 이해가 되지 않았다. 그러나 본 연구에서 개발된 'Score-5 RS' 모델의 결과변수에 '입주기업들의 매출액'과 '입주기업들의 고용인수' 항목을 사용하였음을 생각해보면, 예비창업기업이 어느 정도의 규모와 체계를 갖춘 일반중소기업이나 벤처기업에 비교하여 불리하게 작용됨은 당연함을 알 수 있었다.

그러므로 'Score-5 RS' 모델의 Rule을 살펴봄으로써 향후의 개선방향으로 입주기업의 특성항목인 벤처기업, 일반중소기업, 예비창업기업 별로 가중치를 부여하여 결과변수를 선정한다면 보다 합리적인 평가가 될 수 있음을 알 수 있었다.

A14 =< 12 (151)
 A14 =< 9 (59)
 F03 =< 12 (23)
 C03 전입 (7)
 F02 사업계획 내용 및 질 (2, 0.5) => D
 F02 [기타 '사업형태 업종' 성장성 '센터시설과의 적합성', '센터에 대한 로열티나 지분소득잠재력' '신규 고용창출능력' '이전 사업 경험' '자금확보 재무능력', '지역특화산업 영위기업'] (5, 1.0) => C
 C03 겸입 (16)
 C14 [기타 '매월급여 지급' '부정기적 수당' '실적평가시 반영' 전허 없음] (13.867, 1.0) => F
 C14 정기적 수당 (2.133, 0.938) => C
 F03 > 12 (36)
 A20 =< 5 (34)
 F02 [기타 '사업형태 업종' 성장성 '센터시설과의 적합성', '센터에 대한 로열티나 지분소득잠재력' '신규 고용창출능력' '이전 사업 경험' '자금확보 재무능력'] (12, 1.0) => F
 F02 사업계획 내용 및 질 (21)
 A10 [기타 유희건물] (7)
 B03 =< 345 (4, 1.0) => D
 B03 > 345 (3, 1.0) => F
 A10 [신축 임대] (14, 1.0) => F
 F02 지역특화산업 영위기업 (1, 1.0) => D
 A20 > 5 (2, 1.0) => D
 A14 > 9 (92)
 A12 [기타 대학부설] (84)
 A06 =< 29 (18)
 G01 =< 29 (11.25, 0.856) => D
 G01 > 29 (6.75, 0.944) => F

A06 > 29 (66)
 A03 [강원 경기 대전 울산 제주] (14, 0.929) ⇒ D
 A03 경남 (5)
 A20 =< 0 (3, 1.0) ⇒ D
 A20 > 0 (2, 1.0) ⇒ C
 A03 경북 (5)
 A17 =< 1 (2, 1.0) ⇒ D
 A17 > 1 (3, 1.0) ⇒ C
 A03 광주 (7)
 D07 낮은편 (2, 1.0) ⇒ D
 D07 [높은편 매우높음 보통 아주낮음] (5, 0.8) ⇒ C
 A03 대구 (4)
 A15 =< 91 (2, 1.0) ⇒ D
 A15 > 91 (2, 1.0) ⇒ C
 A03 부산 (8)
 B09 =< 2316 (6, 1.0) ⇒ C
 B09 > 2316 (2, 1.0) ⇒ D
 A03 [서울 인천] (7, 0.857) ⇒ C
 A03 전남 (3, 0.667) ⇒ F
 A03 전북 (4)
 A09 =< 48 (2, 1.0) ⇒ D
 A09 > 48 (2, 1.0) ⇒ F
 A03 충남 (5)
 F03 =< 12 (2, 1.0) ⇒ C
 F03 > 12 (3, 1.0) ⇒ D
 A03 충북 (4)
 A07 =< 306 (2, 1.0) ⇒ D
 A07 > 306 (2, 0.5) ⇒ C
 A12 연구소부설 (8)
 C05 ['경영관리 능력' '기업의 애로상담능력' 기타 '네트워킹 능력' '법률관련 능력' '자금조달 능력' 협상능력] (5, 1.0) ⇒ C
 C05 ['기술지원 능력' '기업가적 능력'] (3, 1.0) ⇒ F
 A14 > 12 (145)
 A13 =< 19 (70.486)
 J15 =< 6 (66.429)
 A22 =< 0 (28)
 A18 =< 9 (18)
 J16 =< 54 (16)
 G01 =< 2 (2.133, 0.938) ⇒ C
 G01 > 2 (13.867, 1.0) ⇒ B
 J16 > 54 (2, 1.0) ⇒ C
 A18 > 9 (10)
 A11 전용 (6, 0.833) ⇒ B
 A11 공용 (4, 1.0) ⇒ A
 A22 > 0 (38.429)
 A22 =< 10 (34.942)
 B08 =< 18 (31.942)
 G02 =< 145 (27.675)
 C01 =< 1 (14.248, 0.964) ⇒ C
 C01 > 1 (13.428)
 B01 =< 800 (6.942, 0.856) ⇒ C
 B01 > 800 (6.485)
 J15 =< 0 (4, 1.0) ⇒ B
 J15 > 0 (2.485, 0.805) ⇒ C

G02 > 145 (4.267)
 J08 =< 1 (2.267, 0.559) ⇒ C
 J08 > 1 (2, 1.0) ⇒ D
 B08 > 18 (3, 1.0) ⇒ B
 A22 > 10 (3.486, 0.574) ⇒ D
 J15 > 6 (4.058, 0.986) ⇒ D
 A13 > 19 (74.514)
 C05 경영관리 능력 (31.514)
 A15 =< 57 (3.05, 0.656) ⇒ C
 A15 > 57 (28.464)
 F04 =< 12 (20)
 A12 [기타 대학부설] (17)
 D09 ['보육닥터 및 풀 제도 모두시행' '보육닥터제만 시행' '시행하지 않음'] (13.813, 0.855) ⇒ A
 D09 전문가풀 제도만 시행 (3.188, 0.941) ⇒ B
 A12 연구소부설 (3, 1.0) ⇒ B
 F04 > 12 (8.464, 1.0) ⇒ A
 C05 기술지원 능력 (12)
 F06 비정기적 (4)
 A08 =< 550 (2, 1.0) ⇒ C
 A08 > 550 (2, 0.5) ⇒ A
 F06 [개최않음 정기적] (8, 1.0) ⇒ A
 C05 ['기업가적 능력' 기타 '법률관련 능력' '자금조달 능력' 협상능력] (16, 1.0) ⇒ A
 C05 기업의 애로상담능력 (1, 1.0) ⇒ B
 C05 네트워킹 능력 (14)
 A11 전용 (12)
 C07 ['실적평가시 반영' 전혀없음] (5, 0.8) ⇒ A
 C07 [기타 '부정기적 수당' '정기적 수당'] (7, 1.0) ⇒ B
 A11 공용 (2, 0.5) ⇒ A

그림 1. 'Score-5 RS' 모델의 Rule

5. 결론 및 제언

5.1 결론

본 연구에서는 전국의 328개 창업보육센터의 설문조사 결과를 데이터베이스화 하여 창업보육센터의 평가모델인 'Score-5 RS (RuleSet)'를 개발 하였다.

결과변수의 설정과 데이터마이닝 분석의 2단계로 이루어진 'Score-5 RS' 모델은 창업보육센터의 본연에 목적에 부합하는 정량적 항목만을 이용하여 결과변수를 설정하여 평가를 함으로써 보다 합리적이라 할 수 있으며, 또한 데이터마이닝 분석을 통하여 여러(일반, 정성적, 정량적) 항목들이 결과변수에 미치는 중요도를 알 수 있으므로 기존의 평가모델에서의 배점선정시의 모호한 부분 역시 보다 명확히 설명할 수 있다는 장점이 있다. 뿐만 아니라 Rule(RuleSet)은 결과변수 설정 단계에서의 미비한 점 및 개선 방향을 찾아낼 수 있는 기본 자료가 되었다.

‘Score-5 RS’ 특성들을 요약 정리하면 다음과 같다.

- 정량적 항목만을 이용하여 평가를 함으로써 주관적 요소를 배제하였다.
- 정량적 항목들을 이용하여 설정한 결과변수에 대하여 창업보육센터 자체적으로 쉽게 확인할 수 있는 모든(일반, 정성적, 정량적) 항목들을 원인변수로 하는 데이터마이닝 분석을 통하여 Rule(RuleSet)을 얻을 수 있다.
- 이 Rule(RuleSet)을 이용하여, 개발된 모델의 결과변수와 연관성이 높은 일반항목 및 정성적, 정량적 항목들을 알 수 있다.
- Rule(RuleSet)을 통하여 모델의 개선을 위한 기본 자료를 제공할 수 있다. 즉, Rule(RuleSet)을 살펴봄으로써 결과 변수 설정단계에서 미비한 점 및 개선방향을 찾아낼 수 있다. 그러므로 본 평가모델은 프로세스 내부에 수정 및 개선의 도구가 포함되어 있다고 할 수 있다.

5.2 제언

향후 평가를 수행함에 있어서 고려해야 할 사항으로 보다 공정한 평가를 위해서는, 결과변수에 사용되는 각 항목의 선택과 가중치의 설정에 대해 먼저 평가받는 센터들의 충분한 사전 동의와 합의가 필요하다고 생각된다.

그리고 본 연구 결과를 볼 때, 입주기업의 특성항목인 벤처기업, 일반중소기업, 예비창업기업 별로 가중치를 부여하여 결과변수를 선정한다면 보다 합리적인 평가가 될 수 있음을 알 수 있었다.

또한 최근 들어 임대료, 성공보수제, 장비 및 서비스사용료 등의 수익 관련 다변화 제도들이 급증하고 있으므로 향후에는 이와 관련한 항목들이 당연히 결과변수에 포함되어야 할 것이다.

마지막으로 본 연구에서 개발된 ‘Score-5 RS’ 모델은 정부 측면에서 창업보육사업 정책의 정비를 위하여 기존의 창업보육센터들을 평가를 하는 용도로 사용될 수도 있겠지만, 이 모델의 결과인 RuleSet을 기반으로 한 자가진단표를 만들어 참조용으로 창업보육센터들에 배포를 함으로써 창업보육센터의 운영에 있어 도움이 되리라 생각되어진다.

참고문헌

- Bana e Costa, C. and Vansnick, J. (2000), *A Fundamental Criticism to Saaty's use of the Eigenvalue Procedure to Derive Priorities*, *Cashier Du LAMSADE* 275.
- Belton, V. and Gear, T. (1983), *On a Short-Coming of Saaty's Method of Analytic Hierarchies*, *Omega*, 11(3), 228-230.
- Dyer, J. (1990), *Remarks on the Analytic Hierarchy Process*, *Management Science*, 36(3), 249-273.
- Go, G-G. and Lee, G-J. (2001), *Statistical Characteristics of Response consistency Parameters in Analytic Hierarchy Process*, *Korea Operations and management Science Society*, 26(4), 71-82.
- Lee, S. S. (2002), *A Study on the Evaluation of Business Incubator Using the Analytic Hierarchy Process*, *Korea Management Science Review*, 19(1).
- Mian, Sarfraz A. (1994), *US University-Sponsored Technology Incubator : An Overview of Management, Polices and Performance*, *Technovation*, 14(8), 515-529.
- Mian, Sarfraz A. (1997), *Assessing and Managing the University Technology Business Incubator: an integrative Framework*, *Journal of Business Venturing*, 12, 251-285.
- Saaty, T. L. (1980), *The Analytic Hierarchy Process*, MacGraw-Hill, New York.
- Saaty, T. L., Vargas, L. G. (1982), *The logic of Priorities*, Kluwer-Nijhoff Publishing.
- Yang, H-B., Song, H-Y., and Kim, H-S. (2003), *The Assessment and Policy Recommendations on the Korean Business Incubator Policy*, Korea Institute of Industry and Technology Information, Research Report.



이동엽

계명대학교 재료공학과 학사
계명대학교 재료공학과 석사
창원대학교 경영대학원 석사
예화테크 기술연구소장, T-toc 대표, 창원전
문대학 겸임교수 역임
현재: 프리랜서
관심분야: 데이터마이닝, 금융공학



김진욱

연세대학교 응용통계학과 학사
한국과학기술원 산업공학과 석사
한국과학기술원 산업공학과 박사
현재: 창원대학교 산업시스템공학과 교수
관심분야: 경제성공학, 금융공학