

SI 프로젝트의 효율성 평가를 위해 자료포괄분석과 기계학습을 결합한 하이브리드 분석

홍한국, 김종원, 서보라
동의대학교 경영정보학과

Hybrid approach combining Data Envelopment Analysis and Machine Learning to Evaluate the Efficiency of System Integration Projects

Han Kuk Hong, Jong Weon Kim, Bo Ra Seo
School of Management Information System, Donggeui Univ.

요 약

Data Envelopment Analysis (DEA), a non-parametric productivity analysis tool, has become an accepted approach for assessing efficiency in a wide range of fields. Despite of its extensive applications, some features of DEA remain bothersome. DEA offers no guidelines to where relatively inefficient DMU(Decision Making Unit) improve since a reference set of an inefficient DMU consists of several efficient DMUs and it doesn't provide a stepwise path for improving the efficiency of each inefficient DMU considering the difference of efficiency.

We aim to show that DEA can be used to evaluate the efficiency of System Integration Projects and suggest the methodology which overcomes the limitation of DEA through hybrid analysis utilizing DEA along with machine learning.

1. 서 론

다수의 투입물을 변환시켜 다수의 산출물을 제공하는 조직들의 상대적 효율성을 측정하는 문제는 조직의 현상을 측정하여 경쟁력 제고를 위한 방안을 마련한다는 차원에서 매우 중요한 사안이다. 조직의 상대적 효율성을 측정하는 방법으로서의 자료포괄분석(DEA : data envelopment analysis)은 Charnes, Cooper & Rhodes의 연구 [1978]에서 최초로 제시되어 최근 다양한 분야에 활발히 적용되고 있다.

DEA는 의사결정단위(DMU : decision making unit)라고 불리어지는 조직의 효율성을 유사한 투입·산출 과정을 갖는 다른 조직과 비교하여 상대적인 기술적 효율성 지수로서 평가하는데, 투입 요소(input factors)와 산출 요소(output factors)에 대한 가격 정보가 필요하지 않고, 사후적이며 경험적인 자료를 이용할 수 있으며, 모수 추정에 필요한 가정이 필요하지 않는 등 여러 가지 장점을 갖고 있다. 이러한 DEA방법론은 주로 다수의 투

입 요소(Multiple inputs)와 다수의 산출 요소(Multiple outputs)가 존재하며, 요소들의 가격 정보가 알려지지 않았거나 불완전한 조직, 예를 들어, 병원[Rutledge et al., 1995], 학교[Beasley, J.E., 1990 : Ahn 1987] 및 군대[Charnes et al., 1985] 등의 공공분야와 은행[Schaffnit, et al., 1997: Sherman 1995], 보험[민재형과 김진한, 1998] 등의 금융서비스 분야의 조직 효율성 평가에 적용되어 왔다.

이렇게 DEA방법론이 여러 분야에 적용되는 것은 여러 가지 이유가 있다. 첫째, 투입과 산출 요소에 대한 어떤 기본적인 가정이 필요하지 않다는 점이다. 둘째, 하나의 의사결정단위에 대해 다수의 투입 요소와 다수의 산출 요소를 동시에 고려할 수 있다는 점이다. 셋째, 효율적인 의사결정단위들과 비효율적인 의사결정단위들을 식별하는 절차를 제공하는 동시에 비효율적인 의사결정단위들 각각에 대해서 부족한 부분과 양을 제시해 준다는 점이 그 이유이다.

그러나, DEA방법론이 다양한 분야의 문제에 대

한 현실적 적용에 있어 몇 가지 단점이 있다. 첫째, DEA 방법론은 상대적으로 비효율적인 의사결정단위들이 개선해야 할 대상이 한 개 이상의 효율적인 의사결정단위들로 구성된 참조집합(Reference Set)으로 주어지기 때문에 구체적으로 어떤 의사결정단위를 타겟으로 하여 개선할지 가이드를 제공하지 않는다는 점이다. 둘째, 효율성의 크기에 따라 각각의 비효율적인 의사결정단위들의 효율성을 개선하기 위한 단계적인 개선 방향을 제공하지 못한다는 점이다.

우리는 앞에서 언급한 DEA의 특징들을 고려하여, DEA 방법론이 SI(system integration) 프로젝트의 효율성을 평가하는데 적용될 수 있다고 생각한다.

SI 프로젝트는 프로젝트를 요청한 고객의 요구에 따라 여러 종류의 정보시스템을 구축하고 유지하는데 필요한 모든 활동을 의미한다. SI 회사들은 보통 그들의 작업을 프로젝트 베이스로 수행한다. 고객으로부터 프로젝트 수주를 받으면, SI 회사는 프로젝트 팀을 구성한다. 프로젝트 수행 과정에서 예산이 초과되거나, 납기가 지연되거나, 고객의 요구수준을 만족시켜 주지 못하는 경우가 많이 있다. 따라서 SI 회사의 품질은 프로젝트의 효율성에 의해 결정되며, 또한 프로젝트 평가 결과는 프로젝트 멤버의 인센티브에 영향을 주기 때문에 정확한 프로젝트의 효율성 평가는 SI 회사에 있어 중요한 문제이다.

본 연구에서 우리는 DEA 방법론이 SI 프로젝트의 효율성 평가에 적용될 수 있다는 것을 보여주고 동시에 앞에서 언급한 DEA의 단점을 극복하는 방법론을 제시하는데 목적을 둔다.

본 연구의 구성은 다음과 같다. 서론 이후에는 DEA와 SI 프로젝트의 평가와 관련된 기존 연구들이 정리되고, 다음으로 우리가 제시하는 연구 방법론과 연구 결과가 제시된다. 마지막으로 연구의 결론과 한계가 기술된다.

II. 기존 연구

2.1 자료포괄분석 (DEA)

기존의 전통적인 계량 경제학적인 방법하에서는 조직의 생산성이 모수적인 방법에 의해 측정되었으며, 이 방법은 모수 추정을 위해 모집단에 대한 엄격한 가정을 필요로 한다. 또한 이 방법은 투입과 산출에 대한 비용과 가격 요소가 명확하

게 규명되기 어려운 조직에는 적용상의 많은 어려움이 따랐다. DEA는 이러한 어려움을 선형계획법을 이용함으로써 쉽게 극복할 수 있었으며 현재 다양한 분야에 활발히 응용되고 있다.

DEA 방법론은 유사한 업무를 수행하는 조직들에 대해 다수의 투입 요소와 다수의 산출 요소의 가중된 크기를 비교하여 조직의 생산성을 측정하는 비모수적 모형이다. 즉, DEA 방법론은 모든 의사결정단위들이 효율적인 운영을 하기 위해 노력한다는 가정하에 경험적인 실제의 다수 투입량과 다수 산출량을 이용하여 조직의 생산성을 상대적으로 비교하는 것이다. 이 분석의 장점은 생산성 측정을 위해 기존에 사용되었던 비율 분석 [Bitran, 1984]이 요구하는 투입과 산출요소의 가중치를 결정할 필요가 없다는 것인데, DEA 방법론에서는 이러한 가중치들이 각 의사결정단위를 효율적인단위로 만들 수 있도록 자체적으로 설정된다. 즉, 모든 의사결정단위들의 투입에 대한 산출의 비율이 1.0을 초과해서는 안되며, 각 투입과 산출 요소의 가중치는 0보다 크다는 제약조건 하에 투입-산출 비율을 최대화 시킬 수 있는 요소의 가중치를 결정하는 모형으로 가공 투입량의 최소화(투입 기준) 또는 가공 산출량의 최대화(산출 기준) 형태로 재구성된 선형계획 모형이다.

DEA 방법론의 이론적 발전은 다양한 분야의 문제에 대한 현실적 적용을 보다 정확하게 하고자 하는 필요성에 따라 이루어졌다. 그 예로서, 의사결정단위 규모의 효과를 고려한 Banker, Charnes & Cooper의 연구[1984], 투입과 산출 요소에 정성적인 요소가 포함된 문제를 고려한 Banker & Morey의 연구[1986], 주관적 판단하에 가중치의 자율성을 통제하는 Roll & Golany의 연구[1989] 등을 들 수 있다.

또한 DEA와 다른 방법론과 결합하여 DEA의 한계를 극복하려는 시도가 많이 이루어지고 있다. DEA와 MOLP(Multiple Objective Linear Programming)를 결합하여 하나의 참조집합을 추정하는 Golany[1988]의 연구, DEA 방법론에 있어 가중치 제한 문제를 해결하기 위해 DEA와 MCDM(Multiple Criteria Decision Making)과의 하이브리드 방법론을 고려한 Stewart[1996]의 연구, 의사결정단위들의 효율성을 평가하기 위해 DEA와 인공지능망을 비교한 Antreas[1996]의 연구 등이 있다.

2.2 SI 프로젝트의 효율성 평가

SI 프로젝트의 효율성 평가는 프로젝트, 즉, 의사결정단위가 원하는 결과를 얻는데 프로젝트에 투입된 자원을 얼마나 잘 활용했는지에 대한 평가에 초점을 둔다. 따라서, 직관적으로 프로젝트 과정에서 투입되는 자원과 적합한 산출물을 정의하고, 투입 자원과 산출물을 잘 대변하는 평가 지표를 선정하는 것이 무엇보다 중요하다. 그러나, SI 프로젝트의 평가는 무엇을 평가하고, 또한 어떻게 평가할 것인지에 대해 당사자 간의 합의 도출이 쉽지는 않다.

Christopher et al. [1996]은 소프트웨어 개발 프로젝트의 산출 요소로, 고객만족도, 납기 준수율, 예산 준수율 및 제품 인도 후의 재작업 시간을 사용하였다.

고객만족도는 소프트웨어 시스템이 실제로 고객의 요구 수준에 어느 정도 만족했는지에 대해 고객이 직접 평가한다. 고객만족도의 지표 항목으로는 신뢰성, 가용성, 사용성, 설치 편의성, 유지용이성, 성능, 그리고 문서화 등이다.

납기 준수율과 예산 준수율은 계획된 기간 및 예산에 대해 어느 정도 준수했는지를 계획대비 실행 비율을 지표로 사용하였다. 납기가 지연되었거나, 예산을 초과한 프로젝트는 새로운 프로젝트를 지연시키고, 수익을 감소시킨다. 따라서, 초기에 일정과 견적을 계획할 때 현실적인 수준을 제시하는 것이 중요하다. 이것은 개발자와 발주자 간의 타협으로 해결할 문제이다. 발주자의 목표는 비용과 완성 일을 최소화하려는 반면에 개발자의 목표는 발주자와의 합의를 도출해 내는 것이다. 그러나 비현실적으로 낮은 가격과 촉박한 일정으로 설정되는 것이 일반적인 경향이다.

제품 인도 후의 재작업은 많은 SI 프로젝트에서 자주 발생하는 문제이다. 재작업은 요구사항의 잘못된 이해 또는 기술적인 설계의 실수에 의해 발생하는 결과이다. 이러한 재작업도 계획되지 않은 것으로 발생시, 일정과 수익에 악영향을 미친다.

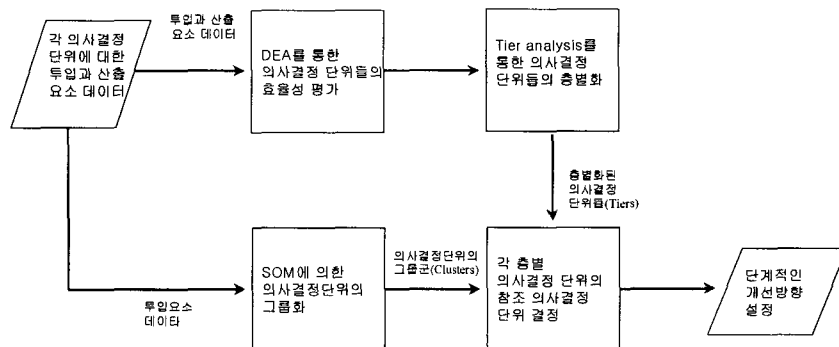
또한, Banker & Kemerer[1992]는 정보시스템 개발 프로젝트에 대한 산출 평가지표로 예산 퍼포먼스(Budget Performance), 납기 퍼포먼스(Schedule Performance), 사용자 만족도(User Satisfaction), 그리고 유지 난이도(Maintenance Complexity)를 사용하였다.

대부분의 정보시스템 평가에서 투입 요소로 프로젝트에 참여한 인력의 총 인시(man-hour), 하드웨어와 소프트웨어 등에 소요된 자재비가 주로 사용되었다[Mahmood et al, 1996].

III. 연구방법

본 장에서는 DEA방법론이 SI 프로젝트의 효율성 평가에 적용될 수 있으며, 동시에 DEA 방법의 단점을 극복할 수 있는 연구방법을 제시하고자 한다. <그림 1>은 우리가 제시하고자 하는 연구방법론을 도식화한 것이다. 이 방법론은 분석의 출발점에서부터 두 가지의 방향으로 구성되어 있다.

첫번째 방향에서는 DEA를 통한 의사결정단위들의 효율성을 평가하고, Tier analysis를 통해 의사결정단위들을 층별화한다. Tier analysis란 효율성의 크기별로 의사결정단위들을 층별화하는 방법이다. 즉, 첫번째 효율적인 의사결정단위들을 제외한 나머지 비효율적인 의사결정단위들에 대해 다시 DEA를 통한 효율성 평가를 실행하는 것

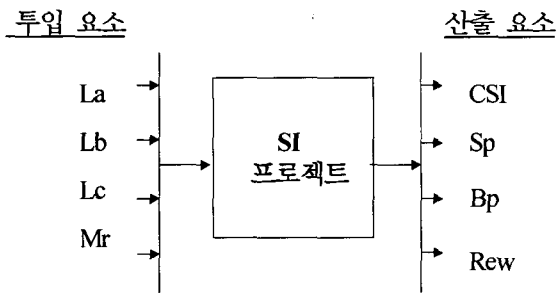


<그림 1> 연구 방법

으로, 남은 비효율적인 의사결정단위들이 충분히 적게 남을 때까지 DEA를 통해 반복해서 의사결정단위들을 증별화한다.

두번째 방향에서는 클러스터링 툴(clustering tool) 중에 하나인 SOM(Self-Organization Map)을 이용하여 비슷한 의사결정단위들끼리 몇 개의 그룹군으로 분류한다. 이러한 SOM에 의한 의사결정단위들의 그룹군과 DEA에 의한 Tiers를 통해 각 Tier에 있는 의사결정단위들이 벤치마킹해야 할 하나의 참조 의사결정단위가 결정되어진다. 우리는 이러한 참조 의사결정단위의 흐름을 "개선 방향(Improvement path)"이라고 부르며, 이는 비효율적인 의사결정단위들이 효율성을 개선하기 위해 단계적인 개선 방향과 구체적인 관리 포인트를 제시하여 준다.

3.1 SI 프로젝트의 효율성 평가 - 투입 요소와 산출 요소



<그림 2> SI 프로젝트의 평가 모델

본 연구에서 우리는 SI 프로젝트의 평가 모델을 <그림 2>에서와 같이 4개의 투입 요소와 4개의 산출 요소로 제한한다. 각 의사결정단위, 즉 SI 프로젝트에 사용된 기본적인 투입 요소는 소

프트웨어와 하드웨어와 같은 장비와 자재에 소요되는 자재비용, 그리고 참여한 인력의 총 투입 시간이다. 총 투입시간은 프로젝트 멤버의 경력에 따라 A급, B급, C급으로 분류되며, 단위는 인월(man-month)이다. 산출 요소는 고객만족도, 납기 준수율, 예산 준수율 및 인도 후 재작업 시간으로 하였다.

각 요소에 대한 측정 방법은 <표 1>과 같다.

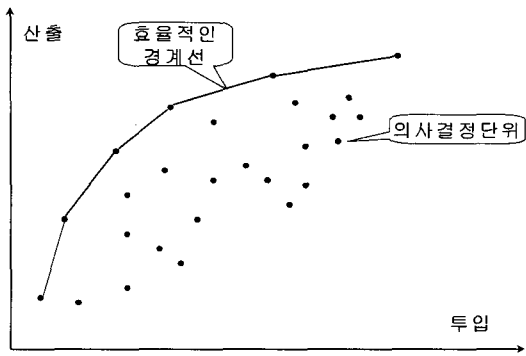
본 논문에서 포함되지 않은 SI 프로젝트의 퍼포먼스를 평가하는 다른 지표도 있다. 원시코드 라인수(SLOC : source lines of code)와 기능 점수(FPs : function points)와 같은 생산성 지표가 될 수 있다. 이들 지표는 그것을 측정하는데 있어서 신뢰성이 부족하기 때문에 본 연구에 포함되지 않았다.

3.2 DEA를 이용한 의사결정단위의 효율성 평가

DEA방법론을 통해 의사결정단위들을 효율적인 집단과 비효율적인 집단으로 구분한다. 이 방법론은 각 의사결정단위가 효율적인 경계선(Efficientfrontier)에 있거나, 또는 밑에 있다는 유일한 가정하에 하나의 의사결정단위에 대해 모든 의사결정단위와상대적으로 비교하여 최대의 효율값을 계산한다. <그림 3>에서와 같이 실선으로 묘사된 경계선에 표시되지 않은 비효율적인 의사결정단위는 효율적인 경계선에 있는 의사결정단위들 중에 λ 값이 양수인 의사결정단위들의 블록결합(convex combination)으로 효율값이 측정된다. 블록결합에 속한 의사결정단위들을 참조집합(reference set)이라고 하며, 효율값은 모두 1.0이다.

<표1> 변수설명

변수(Variable)		측정(Measurement)
투입 요소	A급 인력 투입 시간 (La)	경력 10년 이상의 인력이 투입된 총 인월 (man-months)
	B급 인력 투입 시간 (Lb)	경력 6~10년 인력이 투입된 총 인월 (man-months)
	C급 인력 투입 시간 (Lc)	경력 5년 이하의 인력이 투입된 총 인월 (man-months)
	자재 및 장비 비용 (Mr)	하드웨어, 소프트웨어 및 다른 기자재에 소요되는 총 비용
산출 요소	고객만족도 (CSI)	고객설문 조사
	납기 준수율 (Sp)	계획된 프로젝트 기간/실제프로젝트기간
	예산 준수율 (Bp)	실제 개발비용 / 계획된 예산
	인도 후 재작업 인월 (Rew)	5 인월(Buffer man-months) - 재작업 인월 (man-months)



<그림 3> DEA의 효율적인 경계선(Efficient frontier)

3.3 Tier Analysis를 통한 의사결정단위의 층별화

앞 절에서 우리는 SI 프로젝트의 효율성을 평가하기 위하여 DEA를 사용하였다. 의사결정단위들은 DEA를 통해 효율적인 그룹과 비효율적인 그룹으로 구분되어진다. DEA를 통한 의사결정단위들을 층별화하는 방법은 Thanassoulis[1995]에 의해 제시되었는데, 효율성의 크기가 아닌 투입 요소 배합의 특징에 따라 의사결정단위들을 층별화하였다.

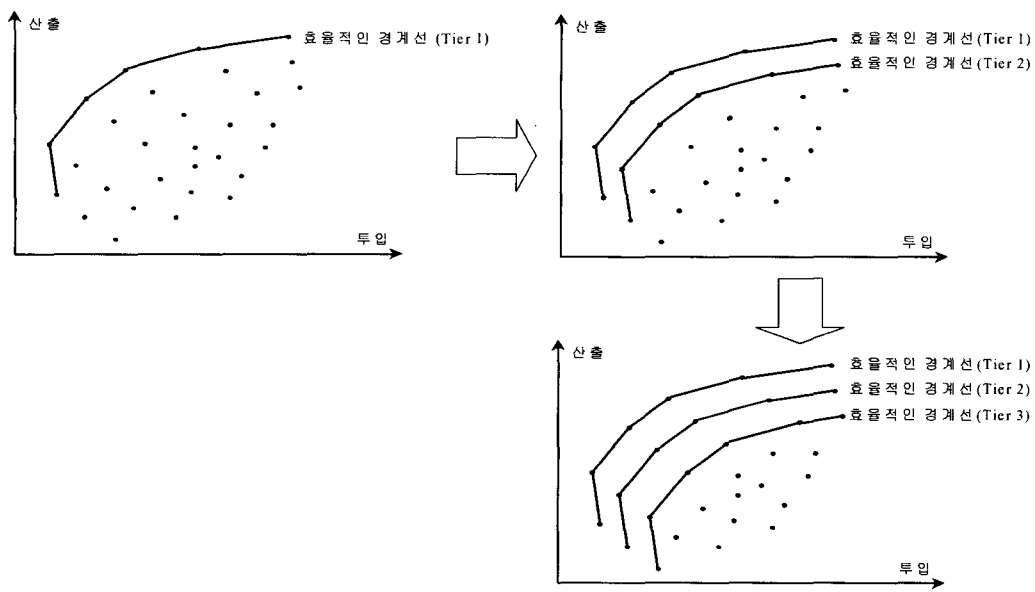
우리가 제안하는 Tier analysis는 효율성의 크기에 따라 의사결정단위들을 층별화하는 기법이다. 즉, Tier analysis의 첫번째 단계에서 DEA를 통해 모든 의사결정단위들의 상대적인 효율값을

계산한다. 그 결과, 효율값이 1.0인 효율적인 의사결정단위들이 나타난다. 우리는 이를 "Tier 1"이라고 부른다. 두번째 단계에서, Tier 1에 속하지 않은 비효율적인 의사결정단위들을 대상으로 다시 DEA를 실행한다. 두번째 단계에서 효율값이 1인 의사결정단위들을 "Tier 2"라고 부른다. 동일한 절차가 남아있는 의사결정단위의 수가 Banker et al. [1984] 이 제시한 것처럼 투입 요소와 산출 요소 수의 합($4 + 4 = 8$)의 3배($8 * 3 = 24$) 미만일 때까지 반복된다. 이것은 효율적인 의사결정단위들과 비효율적인 의사결정단위들을 구분하는 판별력의 기준이 된다. <그림 4>에서 보여주는 바와 같이 각 단계에 효율적인 그룹에 속하는 의사결정단위들이 각 단계에서 효율적인 경계선을 형성하기 때문에 우리는 이러한 절차를 "Tier analysis" 라고 부른다.

<그림 4>는 Tier 1에 있는 의사결정단위들의 효율값이 Tier 2보다 높고, Tier 2에 있는 의사결정단위들의 효율값이 Tier 3보다 높다. 우리는 하위 Tier에 있는 비효율적인 의사결정단위들의 단계적인 개선 방향을 제시하는데 사용한다.

3.4 SOM을 이용한 의사결정단위들의 그룹화

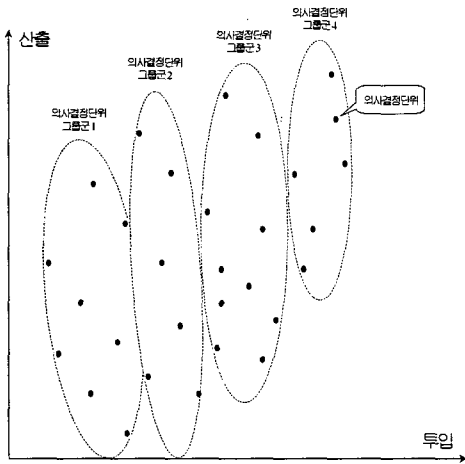
방법론의 두번째 방향에서 우리는 클러스터링 툴 중에 하나인 SOM을 사용한다. SOM은 앞 절에서 층별화된 의사결정단위의 Tier와 함께 비효율적인 의사결정단위들 각각의 벤치마킹 타겟과



<그림 4> Tier analysis의 절차

동시에 단계적인 개선 방향을 제시하기 위해 사용된다.

DEA는 비효율적인 의사결정단위들의 참조집합이 몇 개의 효율적인 의사결정단위들로 구성되어 있어 어느 의사결정단위를 개선타겟으로 할지 알 수가 없다. 따라서, 우리는 하나의 벤치마킹 타겟을 선택하기 위해 <그림 5>와 같이 SOM으로 하여금 4개의 투입 요소의 특징에 따라서 유사한 의사결정단위들끼리 그룹화 한다. 왜냐하면 투입 요소가 비슷하다는 것은 프로젝트에 투입한 등급 별(A, B, C) 인력, 자재 및 장비 비용의 규모가 비슷하거나, 투입된 배합 비율이 비슷하다는 의미로써, 가급적이면 참조집합 중에서 규모나 배합비율이 비슷한 의사결정단위를 개선 타겟으로 하는 것이 효율성을 개선하는데 용이하기 때문이다.



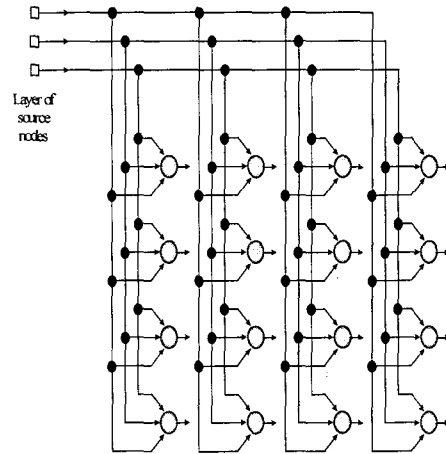
<그림 5> 의사결정단위들의 그룹군

SOM은 자기 조직화특성(self-organizing)을 가진 신경회로망의 일종이다. 신경회로망의 중요한 특징은 환경으로부터 얻어진 자료를 가지고 학습을 하고, 이를 통해 수행능력을 개선시키는 능력을 가지고 있다는 것이다.

SOM에서는 뉴런이 N차원으로 구성된 격자의 노드에 위치하게 되며, 각각의 뉴런은 경쟁학습 과정에 있는 입력 패턴이나, 입력 패턴의 클래스에 따라 선택적으로 조정된다. 즉, 격자에 있는 뉴런의 공간의 위치가 입력 패턴의 고유의 특성과 일치하도록 입력 패턴의 지형적인 지도를 형성한다.

본 논문의 자료분석을 위해 사용된 SOM시스템은 여러 종류의 자기조직화 시스템 중에서 SOFM(Self-Organizing Feature Map)이라고 알

려진 신경회로망이다. 이 네트워크는 경쟁 학습(competitive learning)을 기초로 한다. 즉, 네트워크의 출력 뉴런을 활성화하기 위해서 서로 경쟁한 후 어떤 시점에서 단지 하나의 출력 뉴런만을 활성화시키는데, 경쟁에서 이긴 출력 뉴런을 승자 획득 뉴런(winner-takes-all neuron)이라고 한다. 즉, 0~1 사이의 값들로 정규화 된 Input vector 와 Weight vector들의 내적(inner product)이 가장 큰 뉴런이 승자가 되는 것이다.



<그림 6> 뉴런의 2차원 격자

$$\max_j (\hat{X}' \hat{W}_j)$$

\hat{X}' : Normalized input vector

\hat{W}_j : Normalized Weight vector

만약, Input vector와 가장 비슷한 뉴런을 찾기 위해서 Index function $i(x)$ 를 사용할 경우에는 다음의 조건식을 이용해 $i(x)$ 를 결정할 수 있다.

$$i(x) = \text{arg min}_j \|\hat{X}' - \hat{W}_j\|$$

네트워크이 스스로 자기조직화를 하기 위해서는 뉴런 j 의 Weight vector가 Input vector에 관련되어서 변화하여야 한다. Weight vector의 갱신 공식은Kohonen의 갱신 공식을 이용한다.

SOM의 알고리즘을 요약하면아래와 같다.

1. Initialization. Weight vector를 초기화 한다.
2. Similarity matching. 임의의 시점에서 Input vector와 내적값이 최소인 Winning neuron을 찾는다. 이때, 위에서 언급한

Index function을 사용할 수 있다.

$$i(n) = \underset{j}{\operatorname{argmin}} \|x(n) - w_j\|, \quad j=1, 2, \dots, N$$

3. *Updating.* Kohonen의 갱신공식을 이용해서 모든 뉴런의 Weight vector를 갱신한다.

$$w_j(n+1) = \begin{cases} w_j(n) + r(n)[x(n) - w_j(n)]\Lambda[i(n) - j, n] \\ w_j(n), & \text{otherwise} \end{cases}$$

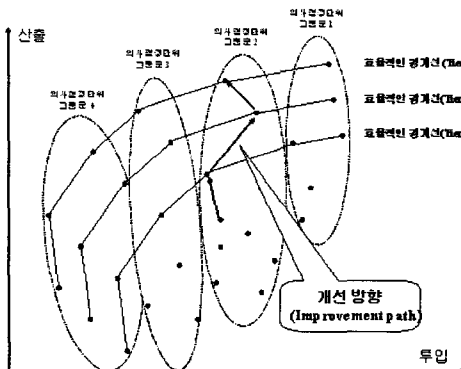
여기에서 $\Lambda[i(n) - j, n]$ 은 Neighborhood 함수로서 아래와 같은 값을 가진다.

$$\Lambda(i - j, n) = \begin{cases} 1 & \text{if } \|i - j\|_p \leq R(t) \\ 0 & \text{if } \|i - j\|_p > R(t) \end{cases}$$

4. *Continuation.* Stopping condition을 만족할 때까지 2번으로 되돌아 간다.

3.4 벤치마킹 타겟 및 개선 방향 결정

상위 Tier에 있는 효율적인 의사결정단위들은 하위 Tier에 있는 비효율적인 의사결정단위들의 참조 집합이 될 수 있다. 참조 집합에 있는 의사결정단위들 중에서 벤치마킹할 타겟은 어떻게 선택할 것인가? 우리는 상위 Tier에서 벤치마킹 타겟을 찾기위해 사전에 SOM을 이용하였다. 즉, <그림 7>과 같이 하위 Tier에 있는 의사결정단위와 유사한 투입 요소의 특징을 가진 의사결정단위를 타겟으로 선택하였다. 왜냐하면 투입의 규모나 인력의 여건이 비슷한 의사결정단위를 벤치마킹하는 것이 유리하기 때문이다.



<그림 7> Tier별 비효율적인 의사결정단위들의 벤치마킹 타겟과 개선 방향

DEA에 의한 Tier와 SOM에 의한 의사결정단위의 그룹군이 밝혀지면 우리는 <그림 7>에서와 같이 비효율적인 의사결정단위마다 효율성을 개선하기 위한 단계적인 방향을 찾아낼 수 있다.

IV. 방법론의 적용

4.1 DEA를 이용한 의사결정단위의 효율성 평가

우리가 사용한 데이터는 국내 S기업에서 1995년에서 1996년 사이에 수행된 50개의 SI 프로젝트의 결과이다. 이 프로젝트들은 제조, 금융서비스, 공공 사업 등을 대상으로 수행되었으며, Open System, Downsizing, Client/Server, Windows GUI 등의 기술을 응용해 고객의 특정한 정보시스템을 구축하는 일이었다. 참여한 인력의 인센티브 결정과 향후 프로젝트의 성공을 위한 자료의 축적을 목적으로 연말이면 그 해에 완료된 프로젝트들을 대상으로 평가를 한다. 따라서 무엇보다 평가의 객관성이 중요하다.

우리는 수행된 50개의 프로젝트들이 규모와 응용된 기술들은 다르지만 분석, 설계, 코딩 및 테스트 등 일련의 업무 프로세스는 동일하기 때문에 DEA방법론의 적용이 가능하다고 보았다. 우리는 DEA모형 중에 CCR 모형을 적용하였으며, <부록>에 간단히 모형에 대한 내용을 요약하였다. <표 2>는 CCR 모형을 이용하여 50개 SI프로젝트들의 효율성을 평가한 결과로서, 상대적으로 효율적인 SI 프로젝트와 비효율적인 SI 프로젝트를 구분하여 보여주고 있다.

여기서 효율값이 1.0인 효율적인 프로젝트가 프로젝트 P2와 P7을 포함해 총 13개이다. 프로젝트 P1은 효율값이 0.64인 비효율적인 프로젝트이며, 이는 실제 자원의 64%만 투입해도 현재의 산출 수준을 생산할 수 있다는 의미이다. 프로젝트 P3은 효율값이 0.89이며, 이 프로젝트도 11%의 자원을 초과해서 사용했다는 의미이다. 실제 50개의 프로젝트들 중에서 37개가 자원을 초과 투입하였다. 이러한 정보는 SI 회사가 실질적인 생산성 향상과 비용 감소가 필요함을 나타낸다.

<표 2> SI 프로젝트의 효율성 측정 결과

SI 프로젝트 (DMU)	효율값	참조 집합 (Reference set)	SI 프로젝트 (DMU)	효율값	참조 집합 (Reference set)
P1	0.64	P16 P29 P40 P41P45	P26	0.5	P7 P32
P2	1.0		P27	0.06	P8 P21 P40
P3	0.89	P40 P41 P45	P28	0.83	P7 P16 P29 P41
P4	0.36	P14 P16 P29 P47	P29	1.0	
P5	0.98	P16 P40 P46	P30	0.38	P40
P6	0.9	P21 P29 P40	P31	0.29	P8 P32 P45
P7	1.0		P32	1.0	
P8	1.0		P33	0.12	P8 P46
P9	0.44	P21 P40	P34	0.32	P21 P40
P10	0.02	P40 P45	P35	0.27	P16 P29 P45
P11	0.19	P7 P8 P16 P32	P36	0.43	P16 P36 P40 P46
P12	0.02	P16 P21 P32	P37	0.93	P8 P21 P32 P40
P13	0.25	P21 P29 P40	P38	0.75	P21 P46
P14	1.0		P39	0.76	P8 P16 P39
P15	0.15	P8 P21 P40 P45	P40	1.0	
P16	1.0		P41	1.0	
P17	0.27	P8 P40 P45	P42	0.07	P16 P40 P45
P18	0.57	P7 P8 P16 P32	P43	0.53	P21 P45
P19	0.39	P29 P40	P44	0.29	P8 P32 P40
P20	0.54	P16 P21 P32	P45	1.0	
P21	1.0	P8 P16 P40	P46	1.0	
P22	0.28		P47	1.0	
P23	0.09	P8 P40 P45	P48	0.63	P7 P32
P24	0.85	P7 P14 P16 P29	P49	0.14	P8 P16 P21 P32
P25	0.49	P8 P21 P40 P46	P50	0.44	P8 P21 P32 P40 45

4.2 Tier Analysis를 통한 의사결정단위의 층별화

우리는 50개의 SI 프로젝트를 Tier analysis를 통해 4개의 Tier로 층별화 하였다. 이 단계에서는 효율치 그 자체보다는 각 프로젝트가 어느 Tier에 속하는가가 중요하다.

1) 첫번째 Tier analysis에서 DEA에 의해 효율값이 1.0인 효율적인 의사결정단위들이 Tier 1이 되고 남아 있는 비효율적인 의사결정단위들이 두번째 DEA 적용시 후보가 된다. 첫번째 Tier analysis의 결과는 <표 3> 과 같다. P_nN 에서 n 은 SI 프로젝트가 몇 번째 Tier에 속하는지를 표시하며, N 은 SI 프로젝트의 번호를 의미한다.

<표 3> Tier analysis를 통한 SI 프로젝트의 층별화 - Tier 1

Tier	의사결정단위 (SI 프로젝트)	참조 집합 (Reference set)
1	P12, P17, P18, P14, P16, P21, P29, P32, P40, P41, P45, P46, P47	No reference set

2) 두번째 Tier analysis에서 우리는 Tier 1에 속하지 않는 비효율적인 의사결정단위들을 대상으로 DEA를 수행한다. 두번째 단계에서 효율값이 1.0인 의사결정단위들이 <표 4>와 같이 Tier 2가 된다. 동일한 절차가 투입과 산출 요소 수의 3배 미만일 때까지 반복된다.

<표 4> Tier analysis를 통한 SI 프로젝트의
층별화 - Tier 2

Tier	의사결정단위 (SI 프로젝트)	참조 집합 (Reference set) - Tier 1
2	P ₂₃	P ₁₄₀ P ₁₄₁ P ₁₄₅
	P ₂₅	P ₁₁₆ P ₁₄₁ P ₁₄₅
	P ₂₆	P ₁₂₁ P ₁₂₉ P ₁₄₀
	P ₂₂₄	P ₁₇ P ₁₁₄ P ₁₁₆ P ₁₂₉
	P ₂₂₈	P ₁₇ P ₁₁₆ P ₁₂₉ P ₁₄₁
	P ₂₃₇	P ₁₈ P ₁₂₁ P ₁₃₂ P ₁₄₀
	P ₂₃₈	P ₁₂₁ P ₁₄₆
	P ₂₃₉	P ₁₈ P ₁₁₆ P ₁₄₆
	P ₂₄₃	P ₁₂₁ P ₁₄₅
	P ₂₄₈	P ₁₇ P ₁₃₂

3) 세번째 Tier analysis의 결과는 <표 5>와 <표 6>과 같이 요약된다. 50개의 SI 프로젝트는 Tier analysis를 적용한 결과, 4개의 Tier로 층별화되었다.

<표 5> Tier analysis를 통한 SI 프로젝트의
층별화 - Tier 3

Tier	의사결정단위 (SI 프로젝트)	참조 집합 (Reference set) - Tier 1
3	P ₃₁	P ₂₆ P ₂₂₄
	P ₃₉	P ₂₅ P ₂₆ P ₂₃₇
	P ₃₁₇	P ₂₃ P ₂₆ P ₂₃₇ P ₂₄₃
	P ₃₁₈	P ₂₂₈ P ₂₃₇ P ₂₄₈
	P ₃₁₉	P ₂₆
	P ₃₂₀	P ₂₅ P ₂₂₄ P ₂₃₇
	P ₃₂₅	P ₂₅ P ₂₆ P ₂₃₇
	P ₃₂₆	P ₂₂₈ P ₂₃₇ P ₂₄₈
	P ₃₃₀	P ₂₆
	P ₃₃₆	P ₂₅ P ₂₆ P ₂₃₇ P ₂₄₃
	P ₃₅₀	P ₂₅ P ₂₆ P ₂₃₇

<표 6> Tier analysis를 통한 SI 프로젝트의
층별화 - Tier 4

Tier	의사결정단위 (SI 프로젝트)	참조 집합 (Reference set) - Tier 3
4	P ₄	P ₃₁ P ₃₁₈
	P ₄₁₀	P ₃₁₇ P ₃₅₀
	P ₄₁₁	P ₃₉ P ₃₂₀ P ₃₅₀
	P ₄₁₂	P ₃₁ P ₃₁₈ P ₃₂₅ P ₃₃₆

P ₄₁₃	P ₃₁ P ₃₉ P ₃₃₀ P ₃₃₆ P ₃₅₀
P ₄₁₅	P ₂₅ P ₃₃₆ P ₃₅₀
P ₄₂₂	P ₂₀ P ₂₅ P ₃₅₀
P ₄₂₃	P ₃₁ P ₂₀ P ₂₅
P ₄₂₇	P ₃₉ P ₃₃₀ P ₃₃₆ P ₃₅₀
P ₄₃₁	P ₃₁ P ₃₅₀
P ₄₃₃	P ₂₅ P ₃₅₀
P ₄₃₄	P ₃₉ P ₂₅ P ₃₃₀ P ₃₃₆
P ₄₃₅	P ₃₁ P ₃₁₈
P ₄₄₂	P ₃₁ P ₂₅ P ₃₅₀
P ₄₄₄	P ₃₁ P ₂₅ P ₃₅₀
P ₄₄₉	P ₃₁ P ₃₁₈ P ₃₂₀ P ₃₂₅ P ₃₅₀

4.3 SOM을 이용한 의사결정단위들의 그룹화

본 연구에서 사용한 SOM은 3장에서 상기한 구조와 학습 알고리즘을 바탕으로 하였다. 구체적으로 3 3 노드를 가지는 SOM으로 입력 변수의 개수는 4개 투입요소 (La, Lb, Lc, Mr) 이고 학습 데이터의 개수는 50개의 SI 프로젝트들이다. 여기서 9개의 노드를 선택한 것은 이 정도 클러스터의 수가 관리가 용이한 수이기 때문이다. 학습 회수는 충분한 학습이 이루어질 수 있도록 20,000 epoch 동안 학습을 하는데, 학습 중 계산되어지는 Weight의 감소량이 0.01보다 작을 경우, 학습을 종료하는 것으로 했다.

<그림 8>은 Kohonen의 SOM을 이용하여 50개의 SI 프로젝트들을 4개의 그룹군으로 <그림 8>은 SOM의 결과를 보여준다. 각 그룹군 안에 있는 번호는 SI 프로젝트의 번호이다. 따라서, 그룹군 1은 프로젝트 P10 하나만 있으며, 반면에 그룹군 2는 22개의 투입 요소의 특징이 유사한 프로젝트들 끼리 모여 있는 것이다.

(1) 10	(2) 5, 6, 9, 13, 15, 17, 18, 19, 20, 23, 25, 27, 30, 31, 33, 34, 36, 38, 43, 46, 49, 50
(3) 1, 2, 3, 7, 8, 16, 21, 32, 37, 39, 40, 41, 44, 45, 47, 48	(4) 4, 11, 12, 14, 22, 24, 26, 28, 29, 35, 42

<그림 8> SOM에 의한 의사결정단위들의 그룹군

<표 7> 각 그룹군의 특징

그룹군	개수	La (평균)	Lb (평균)	Lc (평균)	Mr (평균)	CSI (평균)	Sp (평균)	Bp (평균)	Rew (평균)
1	1	56.0	43.3	51.0	61.2	64.8	0.51	0.91	0.11
2	22	3.9	8.1	4.5	5.7	71.0	0.93	1.03	2.0
3	16	1.6	4.8	3.9	5.2	90.1	1.1	1.02	2.4
4	11	10.5	18.3	11.8	19.3	81.5	1.12	1.11	2.3

<표 8> 단계적인 개선 방향의 세부적인 관리 포인트

그룹 (Tier)	의사결정단위 (SI 프로젝트)	투입 요소				산출 요소			
		La	Lb	Lc	Mr	CSI	Sp	Bp	Rew
1	P ₁₄₆	2.5	2.1	0.2	1.85	94.48	1.01	0.83	2.16
2	P ₂₅	0.9	2.6	0.8	1.9	78.88	0.96	0.9	1.95
3	P ₃₂₀	2.0	6.0	2.3	4.05	70.0	1.05	1.1	1.74
4	P ₄₂₃	15.0	14.0	9.1	15.99	68.62	0.75	1.11	1.64

<표 7>은 4개의 그룹군에 대해 각 그룹에 속한 프로젝트들의 개수, 투입과 산출요소의 평균값 등의 특징을 보여주고 있다. 특히, 그룹 1에는 프로젝트 P10 단 하나만 속해있다. P10은 정부를 대상으로 한 대형 프로젝트로서 다른 프로젝트보다 월등히 투입요소가 많이 소요된 반면, 산출요소는 크게 낮아, 다른 프로젝트와는 특징이 크게 다르기 때문이다.

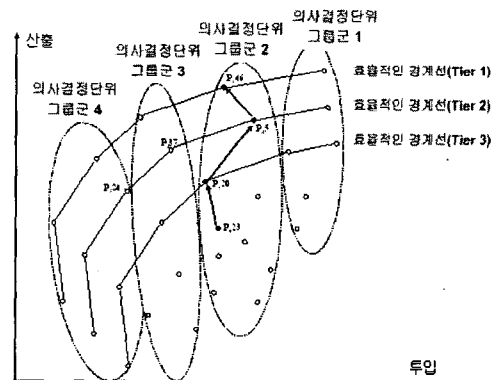
4.4 벤치마킹 타겟 및 개선 방향 결정

Tier analysis에 의해 50개의 SI 프로젝트들을 효율성의 크기별로 4개의 다른 Tier로 층별화하였다. 그리고 SOM에 의해 낮은 Tier에 있는 의사결정단위들이 효율성을 개선할 수 있는 방법을 알 수 있다. 그것이 어떻게 가능할까? 각 Tier에 있는 의사결정단위들은 바로 위의 상위 Tier에 있는 의사결정단위들 중에서 단지 하나의 참조 의사결정단위를 찾음으로써 그들의 효율성을 개선시킬 수 있다.

이것은 SOM에 의해 참조 집합 내에 있는 의사결정단위들 중에서 비슷한 특징을 가진 의사결정단위를 선택할 수 있기 때문이다. 예를 들어 Tier 3에 있는 프로젝트 P₃₂₀의 참조 집합은 상위 효율 경계선인 Tier 2에 있는 효율적인 프로젝트인 P₂₅, P₂₄와 P₂₇로 구성되어 있다. 이들 사이에서 우리는 프로젝트 P₂₅를 벤치마킹 타겟으로 결정한다. 이유는 SOM에 의해서 프로젝트 P₃₂₀과 같은 그룹군에 속하기 때문이다. 만약, 같은 그룹군에 속하는 프로젝트가 1개 이상일 경우 SOM의 결과 값이 가장 큰 프로젝트를 선택한다. 왜냐하면 비

효율적인 프로젝트와 가장 특징이 유사하기 때문에 그만큼 벤치마킹 타겟으로 하여 효율 개선이 용이하기 때문이다.

Tier별로 각 의사결정단위의 벤치마킹 타겟을 식별할 수 있는 앞의 결과를 토대로 우리는 Tier 1을 제외한 모든 Tier에서 의사결정단위들의 단계적인 개선방향(stepwise improvement path)을 찾을 수 있다. 예를 들어, <그림 9>에서와 같이 우리는 프로젝트 P₄₂₃의 개선방향이 P₄₂₃ -> P₃₂₀ -> P₂₅ -> P₁₄₆임을 알 수 있다.



<그림9> Tier 4에 있는 의사결정단위 P₄₂₃의 개선방향

<표 8>은 프로젝트 P₄₂₃의 단계적인 개선방향에 대해 구체적으로 어떤 투입 요소를 줄이고 어떤 산출 요소를 높여야 할지 관리 포인트를 보여준다. 프로젝트 P₃₂₀은 프로젝트 P₄₂₃의 최종 목표인 프로젝트 P₁₄₆으로 개선하기 위한 첫번째 벤치마킹 타겟이다. 이 프로젝트는 P₄₂₃보다 남기

준수율이 매우 높다. 그리고 두번째 벤치마킹 대상인 프로젝트 P₂₅는 프로젝트 P₃₂₀과 비교하여 더 적은 자원을 갖고도 비슷한 수준의 산출 수준을 생산하고 있다. 마지막 벤치마킹 타겟인 P₁₄₆은 프로젝트 P₂₅보다 고객만족도(CSI)가 매우 높다. 따라서 프로젝트 P₄₂₃의 단계적인 개선방향에 있어 관심을 가져야 할 관리 포인트는 먼저, 납기 준수율을 향상시키고, 다음에 자원 전반에 대해 효율성을 기하고, 마지막으로 고객만족도를 향상시킨다.

V. 결론

DEA는 의사결정단위(DMU : decision making unit)라고 불리어지는 조직의 효율성을 유사한 투입·산출 과정을 갖는 다른 조직과 비교하여 상대적인 기술적 효율성 지수로써 평가하는데, 투입·산출 요소에 대한 가격 정보가 필요치 않고, 사후적이며 경험적인 자료를 이용할 수 있으며, 모수 추정에 필요한 가정이 필요하지 않는 등 여러 가지 장점을 갖고 있다.

그러나, DEA방법론은 다양한 분야에 적용함에 있어, 몇가지 단점이 있다. 첫째, DEA방법론은 상대적으로 비효율적인 의사결정단위들이 개선해야 할 대상, 즉, 참조 집합(reference set)이 한 개 이상의 효율적인 의사결정단위들로 주어지기 때문에, 구체적으로 어떤 의사결정단위를 타겟으로 하여 개선할지 가이드를 제공하지 않는다는 점이다. 둘째, 효율성의 크기에 따라 각각의 비효율적인 의사결정단위의 효율성을 개선하기 위한 단계적인 개선방향을 제공하지 못한다는 점이다.

이러한 DEA의 한계를 극복하기 위해 우리가 제안한 방법론은 DEA와 기계 학습(machine learning)을 이용한 하이브리드 분석(hybrid analysis)이다. 이 방법론은 다음과 같이 요약된다.

방법론은 크게 두 가지의 방향으로 구성되어 있다. 첫번째 방향에서는 DEA를 통한 의사결정단위의 효율성을 평가하고, Tier analysis를 통해 의사결정단위들을 총별화한다. 두번째 방향에서는 클러스터링 툴(clustering tool) 중에 하나인 SOM(Self-Organization Map)을 이용하여 투입 요소의 특징이 유사한 의사결정단위들끼리 몇 개의 그룹군으로 분류한다. 이러한 SOM에 의한 의사결정단위들의 그룹군들과 DEA에 의한 Tier들을 통해 각 Tier에 있는 의사결정단위들이 하나의

벤치마킹 타겟을 결정할 수 있으며, 비효율적인 의사결정단위가 효율성을 높이기 위한 단계적인 개선방향과 구체적인 관리상의 포인트도 제시해 준다.

그러나 제시된 방법론에는 몇 가지의 한계를 내포하고 있다. 이러한 한계는 향후 연구 과제이다. 첫째, 프로젝트 복잡성, 사용한 하드웨어와 소프트웨어 툴들의 품질과 같은 환경적 요소(environmental factors)가 SI 프로젝트의 효율성에 크게 영향을 미친다. 불행히도 데이터의 불충분으로 환경적인 요소는 본 논문에서는 포함할 수 없었다. 향후 DEA분석에서는 이들 요소를 정성적인 변수(categorical variables)로 DEA모형에 포함시킬 수 있다.

둘째, 제시한 모델은 회사 내부의 프로젝트 절차 준수율, 설계검토 실시율 등 프로젝트 진행 중에 나타나는 프로젝트 중간관리 지표를 포함하지 않았다. 경험이 풍부한 SI 프로젝트 관리자는 이러한 변수들이 SI 프로젝트를 관리하는데 중요하다는 것을 알지만, 프로젝트의 효율성과 어떤 관계에 있는지 규명하지는 못했다. 향후 연구에 우리는 이러한 중간관리 지표가 제한된 자원 내에서 프로젝트의 퍼포먼스에 어떻게 영향을 미치는가를 파악하기 위해 DEA를 적용할 계획이다.

참고 문헌

- 민재형, 김진한, "DEA를 이용한 손해보험회사의 효율성 측정에 관한 연구" 한국경영과학회지, Vol 23, No 2, 1998, pp. 201-217.
- A. Charnes, C.T. Clark, W.W. Cooper and B. Golany (1985), "A development study of Data Envelopment Analysis in measuring the efficiency of maintenance units in the U.S. Air Force", *Annals of Operations Research* 2, pp.95-112.
- Ahn, T.S. (1987), *Efficiency and related issues in higher education: A data envelopment analysis approach*, Ph.D. Dissertation (The University of Texas at Austin)
- Antreas D., Stephen P.(1996), "A Comparison of Data Envelopment Analysis and Artificial Neural Networks as Tools for Assessing the Efficiency of Decision Making Units", *Journal of the Operational Research Society* 47, pp 1000-1016.

- Banker, R. D., A. Charnes and W. W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, 30(1984), pp.1078-1092.
- Banker, R. D. and Morey, R. C., "The use of Categorical Variables in Data Envelopment Analysis," *Management Science*, 32(1986), pp.1613-1627.
- Beasley, J.E. (1990), "Comparing university departments", *OMEGA Int. J. of Management Science* 18/2, pp.171-183.
- Bitran. G. R. and L. Chang, "Productivity Measurement at the Firm Level," *Interfaces*, 14(1984), pp.29-40.
- Charnes, A., Cooper, W.W., and Rhodes, E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research* 2, pp.429-444.
- Christopher Deephouse, Tridas Mukhopadhyay, Dennis R. Goldenson, and Marc I. Kellner (1996), "Software Processes and Project Performance", *Journal of Management Information Systems*, Winter 1995-96, 12/3, 187-205.
- E Thanassoulis (1996), "A Data Envelopment Analysis Approach to Clustering Operating Units for Resource Allocation Purposes", *Omega, Int. J. Mgmt Sci*, 24/4, pp.463-476.
- Golany, B. (1988) , "An interactive MOLP procedure for the extension of DEA to effectiveness analysis", *Journal of the Operational Research Society* 39/8, pp 725-734.
- M. Adam Mahmood, Karen J. Pettingell, and Alexander I. Shaskevich (1996), "Measuring Productivity of Software Projects: A Data Envelopment Analysis Approach", *Decision Sciences*, 27/1, pp.56-77.
- Rajiv D. Banker, and Chris F. Kemerer (1992), "Performance Evaluation Metrics for Information Systems Development: A Principal-Agent Model", *Information Systems Research*, December 1992, pp.379-398.
- Roll, Y. and B. Golany, "Alternate Methods of Treating Factor Weights in DEA," *OMEGA International Journal of Management Science*, 21(1989), pp.99-109.
- Rutledge, R., Parsons, S., and Knaebel, R.(1995), "Assessing hospital efficiency over time: An empirical application of data envelopment analysis", *Journal of Information Technology Management* 6/1, pp.13-23.
- Schaffnit, C., Rosen, D., and Paradi, J.C.(1997), "Best practice analysis of bank branches: An application of DEA in a large Canadian bank", *European Journal of Operational Research* 87, pp.269-289.
- Sherman, H.D., and Ladino, G. (1995), "Managing bank productivity using data envelopment analysis (DEA)", *Interfaces* 25/2, pp.60-73.
- Stewart, T.J.(1996), "Relationships between data envelopment analysis and multiple criteria decision making", *Journal of the Operational Research Society* 47/5, pp 654-665.
- Thanassoulis, E. (1995), "Assessing police forces in England and Wales using data envelopment analysis", *European Journal of Operational Research* 87, 641-657.