

## The Prediction of DEA based Efficiency Rating for Venture Business Using Multi-class SVM\*

Jiyoung Park\*\*, Taeho Hong\*\*\*

For the last few decades, many studies have tried to explore and unveil venture companies' success factors and unique features in order to identify the sources of such companies' competitive advantages over their rivals. Such venture companies have shown tendency to give high returns for investors generally making the best use of information technology. For this reason, many venture companies are keen on attracting avid investors' attention. Investors generally make their investment decisions by carefully examining the evaluation criteria of the alternatives. To them, credit rating information provided by international rating agencies, such as Standard and Poor's, Moody's and Fitch is crucial source as to such pivotal concerns as companies stability, growth, and risk status. But these types of information are generated only for the companies issuing corporate bonds, not venture companies. Therefore, this study proposes a method for evaluating venture businesses by presenting our recent empirical results using financial data of Korean venture companies listed on KOSDAQ in Korea exchange. In addition, this paper used multi-class SVM for the prediction of DEA-based efficiency rating for venture businesses, which was derived from our proposed method. Our approach sheds light on ways to locate efficient companies generating high level of profits. Above all, in determining effective ways to evaluate a venture firm's efficiency, it is important to understand the major contributing factors of such efficiency. Therefore, this paper is constructed on the basis of following two ideas to classify which companies are more efficient venture companies: i) making DEA based multi-class rating for sample companies and ii) developing multi-class SVM-based efficiency prediction model for classifying all companies.

First, the Data Envelopment Analysis(DEA) is a non-parametric multiple input-output efficiency technique that measures the relative efficiency of decision making units(DMUs) using a linear programming based model. It is non-parametric because it requires no assumption on the shape or parameters of the underlying production function. DEA has been already widely applied for evaluating the relative efficiency of DMUs.

Recently, a number of DEA based studies have evaluated the efficiency of various types of companies,

---

\* This work was supported for two years by Pusan National University Research Grant

\*\* Ph.D. Candidate, College of Business Administration, Pusan National University

\*\*\* Corresponding author, Associate Professor, College of Business Administration, Pusan National University

such as internet companies and venture companies. It has been also applied to corporate credit ratings. In this study we utilized DEA for sorting venture companies by efficiency based ratings.

The Support Vector Machine(SVM), on the other hand, is a popular technique for solving data classification problems. In this paper, we employed SVM to classify the efficiency ratings in IT venture companies according to the results of DEA. The SVM method was first developed by Vapnik (1995). As one of many machine learning techniques, SVM is based on a statistical theory. Thus far, the method has shown good performances especially in generalizing capacity in classification tasks, resulting in numerous applications in many areas of business. SVM is basically the algorithm that finds the maximum margin hyperplane, which is the maximum separation between classes. According to this method, support vectors are the closest to the maximum margin hyperplane. If it is impossible to classify, we can use the kernel function. In the case of nonlinear class boundaries, we can transform the inputs into a high-dimensional feature space. This is the original input space and is mapped into a high-dimensional dot-product space. Many studies applied SVM to the prediction of bankruptcy, the forecast a financial time series, and the problem of estimating credit rating. In this study we employed SVM for developing data mining-based efficiency prediction model. We used the Gaussian radial function as a kernel function of SVM. In multi-class SVM, we adopted one-against-one approach between binary classification method and two all-together methods, proposed by Weston and Watkins(1999) and Crammer and Singer(2000), respectively.

In this research, we used corporate information of 154 companies listed on KOSDAQ market in Korea exchange. We obtained companies' financial information of 2005 from the KIS(Korea Information Service, Inc.). Using this data, we made multi-class rating with DEA efficiency and built multi-class prediction model based data mining. Among three manners of multi-classification, the hit ratio of the Weston and Watkins method is the best in the test data set. In multi classification problems as efficiency ratings of venture business, it is very useful for investors to know the class with errors, one class difference, when it is difficult to find out the accurate class in the actual market. So we presented accuracy results within 1-class errors, and the Weston and Watkins method showed 85.7% accuracy in our test samples.

We conclude that the DEA based multi-class approach in venture business generates more information than the binary classification problem, notwithstanding its efficiency level. We believe this model can help investors in decision making as it provides a reliably tool to evaluate venture companies in the financial domain. For the future research, we perceive the need to enhance such areas as the variable selection process, the parameter selection of kernel function, the generalization, and the sample size of multi-class.

**Keywords :** Multi-class Support Vector Machines(SVM), Data Envelopment Analysis(DEA), Business Efficiency, Venture Business

# 다분류 SVM을 이용한 DEA기반 벤처기업 효율성등급 예측모형

박지영, 홍태호

## I. 서론

최근 들어 핵심지식을 바탕으로 하는 벤처기업에 대한 관심이 더욱 높아지면서 벤처기업의 성공요인에 대한 분석, 개념적 특성 및 본질에 관한 연구 등이 지속적으로 행해지고 있다. 벤처기업은 일반기업보다 자본 대비 높은 수익성을 가지기 때문에 투자자들에게는 매우 매력적인 투자처이다. 반면에 벤처기업은 기술력과 창의적인 아이디어를 기반으로 하는 모험적인 기업 특성으로 인해 태생적으로 위험이 높은 투자처이기도 하다. 이처럼, 벤처기업이 처한 환경적 특성 중 경영환경의 불확실성 때문에 많은 벤처기업들은 자사의 기술 및 자산 등에 대한 효율적 운영 문제와 같은 경영성과와 관련된 사항에 대한 다각적인 논의를 필요로 하며, 이에 근거한 체계적인 전략수립을 도모하고 있다. 경영자의 입장이 아닌 벤처기업의 높은 수익성을 기대하는 투자자들의 입장에서는 어떤 벤처기업이 경쟁적 이점이 있는지를 알고 싶어하며, 최종적으로 수익을 얻을 수 있는 기업에 투자하려고 한다. 벤처기업의 경우 수익성과 위험성을 동시에 내포하고 있는 특성 때문에, 특히 투자 대상으로 평가하기 위한 도구의 필요성이 더 커질 수 있다. 그러나 벤처기업들은 대외적으로 가장 일반적인 평가 잣대가 될 수 있는 신용등급 정보를 가지고 있지 않은 경우가 많기 때문에 신용정보를 이용한 기업 평가는 용이하지 않은 것이 사실이다. 그러나, 투입자원 대비 산출량에 대한 관점에서 볼 때, 동일한 자원을 투입하여 더 많은 산출이 발생되거나 또는 동일한 산출이 발생하는 기업들 중에서 더 적은 자원 투입이 이루어지는 조직이 상대적으로 효율성이 높다고 평가될 수 있다. 기업에서

운영 효율성을 높이기 위해 투자를 실행하려는 의사 결정 문제에 있어서, 투자대비 성과에 대한 효율성 분석이 필요한 것처럼, 어느 기업에서 얼마나 더 높은 효율성을 가지는가 하는 문제는 결국 투자에 대한 효과를 분석할 수 있는 상대적인 도구가 될 수 있음을 의미한다. 또한 같은 업종으로 분류되는 기업들에 대한 평가 문제에 대해서는 절대적인 관점이 아닌 상대적인 관점에서의 평가가 이루어지도록 하는 것이 보다 합리적일 수 있다. Fortune 500 기업을 평가한 연구[Zhu, 2000]를 보더라도, 기업의 효율성 분석은 여러 가지 다양한 변수들을 통해 기업들의 수익성이나 시장성에 대한 상대적인 평가를 가능하게 하며, 결과적으로 이러한 분석 정보가 기업 투자에 직·간접적으로 영향을 줄 수 있다.

본 연구의 주요 목적은 크게 두 가지로 나타낼 수 있다. 첫 번째 목적은, 대상 기업들의 평가를 위해 상대적 효율점수를 기반으로 신용등급과 같은 형태의 새로운 다분류 등급 정보를 제시하는 것이다. 비교대상의 상대적인 평가를 위해 대표적인 비모수적 접근 방법인 DEA(Data Envelopment Analysis)를 사용하며, 연구 대상에 대한 단순 효율성 평가를 실시하는 문제에서 벗어나, 단계적으로 DEA를 실행하여 신용등급과 같은 형태의 효율점수 기반 등급화 정보를 제시하고, 이를 다분류 문제에 적용하고자 한다. 두 번째 목적은, DEA 기반으로 생성된 다분류 등급화 정보에 대한 예측을 용이하게 하도록 다분류 SVM(Support Vector Machines)을 적용하여 효율성 기반 다분류 등급 예측 모형을 개발하는 것이다. SVM은 명백한 이론적 근거에 기반하므로 결과 해석이 용이하고, 실제 응용에서도 높은 성과를 내는 것으로 알려져 있다. DEA를 통한 전체 기업의 효율점수 결

과를 이용하여 상대적으로 효율적인 기업을 식별할 수 있음에도 불구하고, 이와 같이 등급 예측 모형을 만들기 위해 다분류 SVM을 적용한 이유는 크게 두 가지로 설명될 수 있다. 첫 번째 이유는 DEA를 적용할 때의 문제점으로 만약 새로운 평가 대상이 생기면 재차 효율성 평가를 실행해야 한다는 번거로움이 있다는 것이다[Hong *et al.*, 1999]. 즉, 기존의 평가 대상 기업에 대해 효율성 평가를 실시한 이후, 정보를 얻고자 하는 새로운 기업에 대한 효율성 정보를 구하기 위해서는 이전 평가 결과를 버리고, 기존 평가 대상 기업에 새로운 평가 대상 기업을 포함시켜 다시 DEA를 실행하는 등 매번 새로운 DEA 모형을 개발해야 하는 번거로움이 존재하는 것이다. 뿐만 아니라 DEA의 특성이 상대적 효율성의 측정이기 때문에 추가되는 기업을 포함하여 DEA 분석을 수행하게 되면, 기존에 효율적이라고 판단되던 기업이 비효율적인 기업으로 평가될 수도 있고, 또한 등급정보마저 변경되는 결과가 초래될 수 있다. 즉, 최종적으로 특정 기업에 대한 보편 타당한 단독 평가가 가능하도록 SVM 기반 효율성 등급 모형을 개발하게 된 것이다. 두 번째 이유는 현업에서 사용할 때의 유용성 때문이다. DEA는 대표적인 비모수 기법에 기인한 것이지만, 모수를 가진 모형을 사용할 경우 현업에서 사용할 때 실시간 적용이 용이할 수 있다. 즉, 현업에서 특정기업에 대한 효율성 등급을 산정하고자 할 때, SVM과 같은 모수를 가진 등급 예측 모형을 적용함으로써 기업의 평가가 용이하게 된다.

결과적으로 이 연구를 통해 현실적으로 더욱 유용하다고 할 수 있는 다분류 문제에 대해, 기업의 효율성 측면에서 기존의 이진분류 문제뿐만 아니라 다분류 문제를 해결할 수 있는 방안을 제시할 수 있다. 또한 기업의 효율성 정보를 투자자, 대출심사자, 벤처기업 지원을 위한 정부 등에 제공함으로써 기업투자자와 같은 직접적인 의사결정 문제에 반영이 가능하게 함으로써 벤처기업 활성화에 기여할 수 있다.

본 연구의 구성은 다음과 같다. 제 II장에서는 기업의 상대적 효율성 분석을 위해 사용되는 DEA 모형 및 관련 연구와 데이터마이닝 기법 가운데 본 연구에서 적용한 다분류 SVM에 대해 알아본다. 제 III장에서는 연구 프레임워크를 제안하고, 제 IV장에서는 국내 벤처기업에 대한 기업 효율성을 다분류 문제에 적용하기 위한 등급화 과정을 제시하며, DEA 기반 SVM 방법을 적용한 다분류 예측 모형에 대한 실증분석 및 결과를 제시한다. 마지막으로 제 V장에서는 연구결과 및 향후 연구과제에 대해 논의하기로 한다.

## II. 선행연구

### 2.1 DEA(Data Envelopment Analysis)와 기업의 효율성 분석에 대한 연구

효율성은 다양한 방법으로 측정이 가능하며, 투입요소에 대한 산출요소의 비로 정의될 수 있다. 본 연구에서는 대상 기업들의 효율성을 측정하기 위해 대표적인 비모수 방법인 DEA를 적용하였으며, 이 방법은 어떤 의사결정단위가 효율적인지, 어느 정도 효율적인지 등 의사결정단위에 대한 상대적인 효율성 측정이 가능하다. DEA는 투입요소 및 산출요소에 대한 가정을 필요로 하지 않으며, 단일 요소뿐만 아니라, 다수의 투입요소와 산출요소에 대한 평가도 가능하다. 최종적으로 효율적인 의사결정단위와 비효율적인 의사결정단위를 구분하며, 비효율적인 의사결정단위의 경우 비효율의 원인을 규모 및 내부 운영 측면으로 분석이 가능하도록 한다. DEA는 단일 투입요소와 단일 산출요소의 효율성을 측정하기 위한 개념[Farrell, 1957]으로 시작하여, Charnes *et al.*[1978]에 의해 CCR(Charnes-Copper-Rhodes)이라는 기본모형이 개발되었고, 이후 규모의 불변성을 가정하고 있는 CCR 모형의 대안으로 많은 확장 모형이 개발되었는데, Banker *et al.*[1984]의 BCC(Banker-Charnes-Cooper) 모형이 대표적

인 개선 모형이라 할 수 있다.

DEA의 기본모형이라고 할 수 있는 CCR 모형은 각 의사결정단위의 규모수익성(return to scale)을 고려하지 않고 효율성을 평가하는 방법으로 규모효율성(scale efficiency)과 기술효율성(technical efficiency)을 구분하지 못한다. 반면에 BCC 모형은 규모효율성과 순수기술효율성을 분리하여 측정할 수 있다. 여기서, 규모효율성은 규모의 불변을 가정한 CCR 모형의 효율성 값을 규모의 가변을 가정한 BCC 모형의 효율성 값으로 나누어줌으로써 얻을 수 있다. DEA에 의해 상대적으로 비효율적으로 판명된 의사결정단위는 참조집합(reference set)을 통한 벤치마킹으로 효율성을 개선할 수 있다.

벤처기업 운영에 있어서, 산출요소에 대한 여러 가지 투입요소들을 개선함으로써 효율성을 추구하는 것이 용이할 것이므로 본 연구에서는 BCC 모형을 벤처기업의 상대적인 효율성을 평가에 적용하였다. BCC 모형은 Banker *et al.*[1984]이 개발한 것으로 CCR 모형의 경우 일반적으로 규모의 수익(Return-to-scale, RTS)이 일정하다고 가정하고 DMUs에 대한 효율성을 평가한 반면, BCC 모형에서는 효율적인 프론티어가 주어진 DMUs의 볼록집합으로 구성되므로 규모의 수익성이 변동한다고 보고 있으며, DMUs를 평가함에 있

어 규모를 고려함으로써 효율성을 조금 더 의미 있게 평가할 수 있다. 투입요소 중심의 BCC 모형은 식 (1)과 같다.

$$\begin{aligned}
 & \min_{\theta_B, \lambda} \quad \theta_B \\
 & \text{subject to} \quad \theta_B x_0 - X\lambda \geq 0 \\
 & \quad \quad \quad Y\lambda \geq y_0 \\
 & \quad \quad \quad e\lambda = 1 \\
 & \quad \quad \quad \lambda \geq 0
 \end{aligned} \tag{1}$$

DEA 결과는 상대적으로 효율적인 집단과 비효율적인 집단으로 제시되며, DEA를 통한 상대적인 효율성 평가에서 중요한 사항은, 평가대상 집단에 대해 CCR 모형 및 BCC 모형에 따른 기술 효율성(TE), 순수기술효율성(PTE), 규모효율성(SE) 등을 통합 분석하는 것이다. 또한, DEA에서는 분석에 사용되는 투입요소와 산출요소를 선정하는 것이 매우 큰 쟁점이 된다. 만약 평가 대상기업 가운데 일부 기업에게만 유리한 투입요소와 산출요소를 선정하게 될 경우 유리하게 선정된 요소를 근거로 하여 해당 기업이 다른 비교 대상 기업 보다 상대적으로 더 효율적이라는 결과를 초래할 수 있게 되므로, DEA에서 효율성 평가결과에 대한 설득력 있는 설명을 위해 평가

<표 1> DEA를 이용한 기업의 효율성 분석에 관한 선행연구

연구자	투입요소	산출요소	응용분야
Lee <i>et al.</i> [2007]	자본금, 유형자산, 급여, 광고선전비	순방문자수, 매출액	인터넷 기업
Hong <i>et al.</i> [2007]	종업원수, 총자산, 개발비, 판매비와 관리비	매출액	IT 벤처 기업
Koo <i>et al.</i> [2006]	직원수, 총자산, 매출원가, 판매관리비	매출액	코스닥 소프트웨어 기업
Kim[2006]	종업원수, 경영자산, 투입비용	매출액	국내 SI 기업
Kim[2005]	직원수, 총자산, 매출원가	매출액	코스닥/외감 소프트웨어 기업
Kim[2004]	투입인건비, 유형고정자산, 연구개발비	매출액	코스닥제조 기업
Carlos <i>et al.</i> [2005]	종업원수, 총운영비용, 총자산	방문자수, 매출액	인터넷 기업
Barua <i>et al.</i> [2004]	IT자본, 비IT자본, 노동력, 사업연수	매출액, 매출총이익	인터넷 기업

관련 문헌이나 대상 기업이 속한 산업의 특성을 토대로 투입요소와 산출요소를 신중하게 선별할 필요성이 있다.

DEA는 상대적인 효율성 및 생산성의 비교연구에 있어서 다양한 분야에 적용되고 있으며, <표 1>에 제시된 연구들은 DEA를 이용한 기업의 효율성 분석에 대한 기존 연구들 중 인터넷, 소프트웨어, IS 기업 등 본 연구의 대상 기업들과 유사한 업체들을 대상으로 한 연구들이다. DEA는 은행 지점의 상대적 평가에도 적용되고 있으며, 해운산업, 철강산업 등 동종업계의 상대적 평가에 많이 응용되고 있다. 최근에는, 기업의 신용위험 평가[Lee, 2006] 및 도산예측[Paradi *et al.*, 2004; Min and Jeong, 2006] 등에도 활용되고 있으며, 특히 이들 연구는 층화기법을 분류방식으로 적용하였다. 기존의 DEA를 통한 기업의 효율성 분석에 있어서 새로운 의사결정단위가 추가되었을 경우에 대한 분석 및 예측의 용이성을 위해, 기존의 DEA 방법에 데이터마이닝 기법을 추가한 연구도 있다[Hong *et al.*, 1999; Hong *et al.*, 2007].

## 2.2 다분류 Support Vector Machine(SVM)

SVM은 Vapnik[1995]에 의해 개발된 것으로 입력공간과 관련된 비선형문제를 고차원의 특징공간의 선형문제로 대응시켜 나타내기 때문에 수학적 분석이 용이하다[Hearst *et al.*, 1998]. 특히 SVM이 분류기법으로서 각광을 받는 이유는 대부분의 학습 알고리즘은 경험적 위험최소화에 기반하는 모형인데 반해, SVM은 구조적 위험 최소화에 충실하므로, 분류문제에 있어 일반화 능력이 높아 많은 분야에서 응용되고 있다[Tay and Cao, 2001].

기존의 분류문제는 일반적으로 이진분류 기법을 적용한 연구들이 대부분이며, SVM도 기업의 부도예측 등의 이진분류 문제에 주로 응용되었다. 다분류 기법은 통계적 방법 및 인공지능 방법으로 채권등급예측 및 평가에 관한 연구에서 많이

적용되고 있으며[Garavaglia, 1991; Kim, 1993; Kwon *et al.*, 1997; Shin and Han, 2001; Huang *et al.*, 2004; Ahn *et al.*, 2006], 특히 채권등급모형은 순차적이고 다분류 문제가 적용되어야 하는 분야이며, 이러한 특징을 이용하여, Kwon *et al.*[1997]은 ordinal pairwise partitioning(OPP) 접근법을 제안하였는데 이는 산출 등급에 따라 순차적으로 짝짓는 방법으로 데이터 셋을 나누어 채권등급평가에 적용한 것이다. SVM은 이진분류 문제에 있어서 매우 강력한 도구가 되고 세 개 이상의 다분류 문제에 대한 적용도 가능하다. 다분류 SVM을 적용하기 위해서는 두 가지 유형의 접근법을 고려할 수 있는데, 하나는 여러 개의 이진분류 SVM을 결합하는 방법이며, 다른 하나는 한번에 모든 데이터를 고려하여 최적화 문제로 해결하는 것이다[Hsu and Lin, 2002].

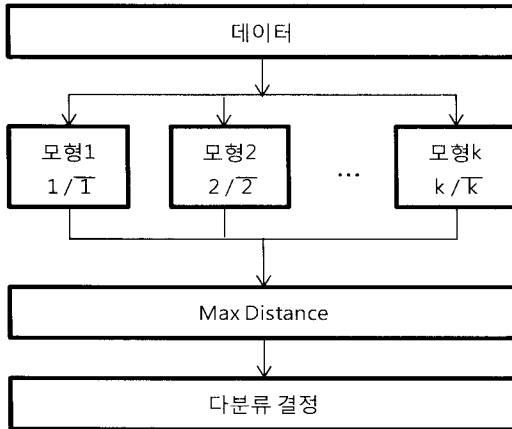
본 연구대상과 같이 벤처기업의 효율성 예측 문제에 있어서는 기업 채권등급 예측 문제와 마찬가지로 효율성 등급이 다양하게 존재할 수 있으므로 다분류 기법을 적용하는 것이 적절하다. 이러한 다분류 문제를 해결하기 위해서는 이분류 예측모형을 여러 개 결합하는 방법과, 모든 등급을 한 번에 고려하여 하나의 최적화 문제로 해결하는 방법이 있다. 이진분류 기반 방법(Method based on binary classification)으로는 one-against-all, one-against-one, directed acyclic graph SVM(DAGSVM), OPP(Ordinal pairwise partitioning) 등 4가지 방법과 all-together 방법(Weston and Watkins[1999], Crammer and Singer[2000])은 다음과 같다.

### 2.2.1 이진분류기반 방법

#### (1) one-against-all 방법

One-against-all 방법은 이진분류모형을 이용하여 다분류 문제를 해결하기 위해 사용된 방법이다[Hsu and Lin, 2002]. k개의 집단을 분류하는 문제에 대해서 이 방법을 적용하면 다음과 같이

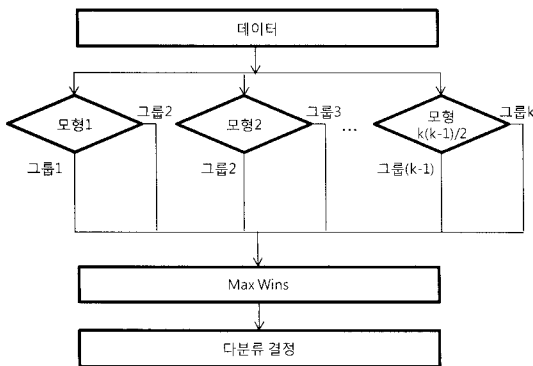
해결할 수 있다.  $k$ 개의 이진분류모형을 이용하여  $i$ 번째 모형에서 분류결과에 따라  $i$ 에 소속되는 데이터와  $i$  외의 다른 상태에 소속되는 데이터로 분류하는데, 이와 같은 단계를  $k$ 번 학습한 이후, 테스트 단계에서 최종적으로 새로운 데이터가 어떤 집단에 소속되는지를 예측할 수 있다(<그림 1>).



<그림 1> SVM의 one-against-all 방법(Nguyen and Rajapakse, 2003)

**(2) one-against-one 방법**

$k$ 개의 집단을 분류하는 문제에 대해 one-against-one 방법을 적용하면  $kC_2 = k(k-1)/2$ 개의 이진분류모형이 만들어진다(<그림 2>). 이는, 첫 번째 집단과 두 번째 집단, 그리고 첫 번째 집단과 세

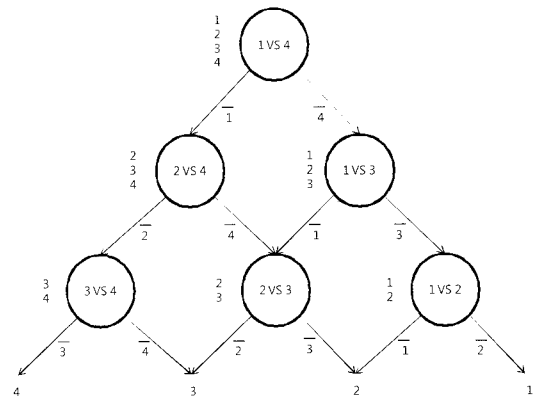


<그림 2> one-against-one 방법(Hwang and Yang [2004]에서 수정)

번째 집단 등  $k$ 개의 집단 가운데 두 개의 집단에 대해서만 고려하게 된다. 이 방법에서는 “max wins” 알고리즘을 사용하게 되며, 좀 더 우세한 집단에 대한 표결(voting)을 통해 최종 결과를 얻는다. 즉, 새로운 데이터는 최종적으로 가장 선택을 받은 집단으로 예측할 수 있다.

**(3) DAG 방법**

DAG(Directed Acyclic Graph)는 각 사이클이 없는 방향을 갖는 그래프이다. 각 노드를 출발하는 아크가 0 또는 2개이며, 트리와 다른 점은 각 부모 노드(parent node)가 2개 이상일 수 있다는 점이다. DAG의 방법은 Platt *et al.*[2000]이 제시하였으며, 학습은 one-against-one 방법과 동일하고 테스트 데이터 셋에서는 <그림 3>과 같이 최상위 노드에서 부터 이진분류를 시작한다. 각 노드에서 이진분류에 따라  $i$  집단이 아니라면 다음 노드에서  $i$  집단을 제거하고 다시 이진분류를 수행한다. 하나의 집단이 남을 때까지 각 노드를 거치는 과정을 통해서 최종적으로 하나의 집단이 남으면 이진분류를 멈추고 최종 집단을 결정한다.

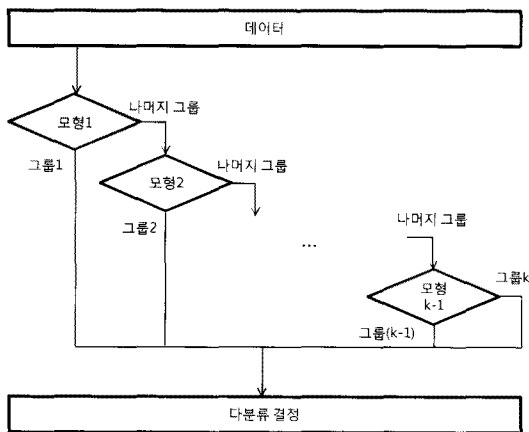


<그림 3> DAG 방법: 4개 집단 분류의 예(Platt *et al.*, 2000)

**(4) OPP 방법**

다분류 문제를 풀기 위한 이진분류 방법인 one-against-all, one-against-one, DAGs 방법은 모두

분류할 집단이 명목척도인 경우를 전제로 한다. 하지만, 경영의사결정에서 보면 분류는 대부분 등급과 같은 서열척도인 경우가 많다. Kwon *et al.*[1997]은 분류집단이 서열척도인 경우에 적용할 수 있는 OPP(ordinal pairwise partitioning) 방법을 제안하였다. Kwon *et al.*[1997]은 분류집단 N에 대해서 분류방법 및 분류순서에 따라 모두 4가지의 OPP 방법을 제안하였는데 즉, OPP1\_forward와 OPP1\_backward 방법 그리고 OPP2\_forward와 OPP2\_backward 방법 등이 그것이다. 여기서 OPP1은 N and(N-1)의 형태로 그리고 OPP2는(N) and (나머지 집단) 형태로 정의되며, 각각의 등급 결합 순서에 따라 다시 forward 및 backward 등으로 정의된다. 예를 들어 4개의 분류 등급이 있다면, OPP1\_forward 방법은 (1 and 2), (2 and 3), 그리고 (3 and 4)의 분리된 데이터를 이용하여 각각 학습하고, OPP2\_forward 방법은 (1 and 2, 3, 4), (2 and 3, 4), (3 and 4)로 학습한다(<그림 4>). 이와 같은 단계를 k-1번 학습한 이후, 테스트 단계에서 최종적으로 새로운 데이터가 어떤 집단에 소속되는지를 예측할 수 있다.



<그림 4> OPP2\_forward 방법

### 2.2.2 all-together 방법

이진분류 기반 방법과 달리 ‘all-together’ 방법

은 다분류 문제를 하나의 의사결정함수를 이용하여 해결하는 것이다. 이에 대해서는 Vapnik[1998], Weston and Watkins[1999], 그리고 Crammer and Singer[2000]가 제시한 방법들을 고려할 수 있다. Vapnik[1998]은 모든 집단들에 대해 단일 최적화 문제를 해결할 수 있는 접근법을 제시하였는데, 이는 one-against-all 방법과 유사하다. 그러나, one-against-all 방법에서는 k-1개의 서로 다른 최적화 문제가 있지만, 이 방법은 모두가 하나의 최적화 문제로 해결될 수 있다는 것이다. Weston and Watkins[1999]와 Crammer and Singer[2000] 또한, 단일 최적화 문제를 해결함으로써 다분류 문제를 해결하는 접근법을 제안하였으며, 이 두 방법은 해를 찾는 과정에 있어 분해기법에 차이가 있다.

### Ⅲ. 연구 프레임워크

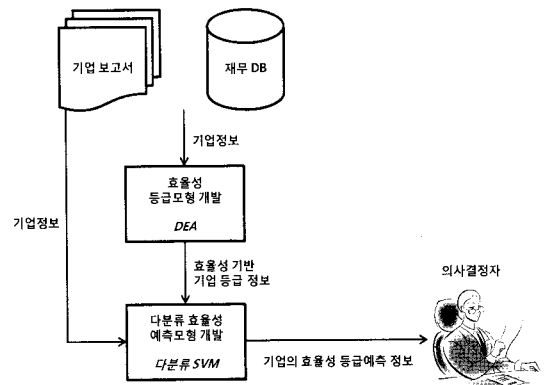
본 연구의 주요 목적은 크게 두 가지로 구분된다. 첫째, 벤처기업들은 대외적으로 가장 일반적인 평가 잣대가 될 수 있는 신용등급과 같은 정보를 가지고 있지 않은 경우가 많기 때문에 신용등급을 이용한 기업의 객관적인 평가는 용이하지 않다. 이러한 점에 기인하여, 대상 기업들의 평가를 위해 상대적 효율점수를 기반으로 신용등급과 같은 형태의 새로운 다분류 등급 정보를 제시하는 것이며, 이를 위해 비모수적 접근방법인 DEA를 사용하여 비교대상의 상대적인 평가를 실시한다. 이는 DEA를 이용하여 연구 대상에 대한 단순 효율성 평가를 실시하는 문제에서 벗어나, 층화기법을 적용한 DEA 실행을 통해 신용등급과 유사한 형태의 효율점수 기반 등급 정보를 제공하여 의사결정문제에 반영할 수 있도록 하는 것이 주요 목적이다. 둘째, DEA 기반으로 생성된 다분류 등급 정보에 대한 예측을 용이하게 할 수 있도록 다분류 SVM(Support Vector Machines) 기법을 적용하여 효율성 기반 등급 예측 모형을 구축하는 것이다. DEA를 통한 전체 기업의 효율점수 결과를 이용하여 상대적으로 효율적인 기



업을 식별할 수 있음에도 불구하고, 이와 같이 SVM을 적용하여 등급 예측 모형을 구축하는 이유는 크게 두 가지로 설명될 수 있다. 첫 번째 이유는, 평가 대상 기업들의 분석을 위해 DEA를 적용할 때의 문제점에서 기인한 것으로 만약 새로운 평가 대상이 생길 경우, 기존 대상과 함께 효율성 평가를 재차 실행해야 한다는 번거로움 [Hong et al., 1999]을 해결하기 위해서이다. 또한, 평가해야 할 기업들이 재차 추가됨으로써, 기존에 효율적이라고 판단되던 기업이 비효율적인 기업으로 평가될 수도 있고, 또한 등급정보 마저 변경되는 결과가 초래될 수 있다. 즉, 최종적으로 특정 기업에 대한 보편 타당한 단독 평가가 가능하도록 SVM을 사용한 효율성 등급 모형을 개발하게 된 것이다. 두 번째 이유는 연구에서 제안하는 예측모형이 현업에서 보다 유용하게 사용될 수 있도록 하기 위함인데, 즉 현업에서 특정기업에 대한 효율성 등급을 산정하고자 할 때, 비모수 접근법인 DEA를 전적으로 이용하여 평가하는 것 보다 SVM과 같은 모수를 가진 등급 예측 모형을 적용함으로써 기업의 단독평가가 더욱 더 용이하게 된다. DEA를 이용하여 일괄적인 상대평가가 가능하기는 하지만, 위에서 제시한 이유들로 인해 한계점을 가지게 되며, 결과적으로 이

연구에서 제안하는 바가 이러한 문제점들을 개선하는데 좋은 방안이 될 것이다.

<그림 6>은 다분류 SVM을 이용하여 최종적으로 DEA 기반 벤처기업 효율성 등급 예측모형을 개발하기 위한 전체 연구프레임워크를 나타낸 것이다. 앞서 제시했듯, 새로운 기업의 효율성 평가에 따른 DEA 재실행 문제점을 해결하기 위해 DEA와 데이터마이닝을 통합한 효율성 예측모형 [Hong et al., 2007]을 확장하여, DEA 기반 다분류 예측모형을 구축하여 향후 특정 의사결정단위에 대한 효율성 등급 예측이 용이하도록 한다.

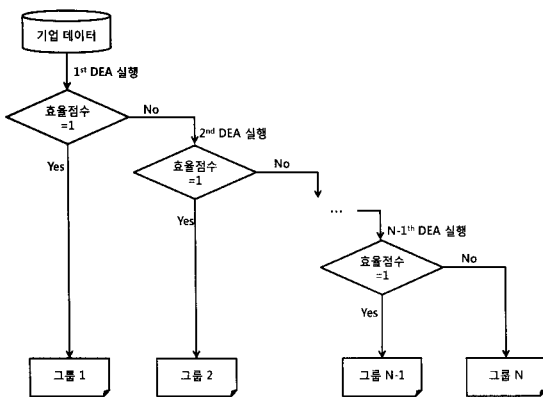


<그림 6> DEA 기반 다분류 효율성 예측모형 개발을 위한 연구 프레임워크

## IV. 실증분석 및 결과

### 4.1 데이터

본 연구에서는 벤처기업의 효율성 분석을 실시한 후, 이를 기반으로 한 다분류 효율성 분석모형 개발을 위해 코스닥에 등록된 벤처기업들 가운데, 사업서비스 영역에 해당되는 164개의 벤처기업 자료를 수집하였다. 이 가운데 자료가 불충분하거나 연구목적에 부합되지 않는 10개의 기업을 제외한 후 최종적으로 154개의 기업을 연구 대상으로 분석을 실시하였다. 분석에 사용된 기업 정보는 한국신용평가에서 제공하는 2005년도 공시



<그림 5> DEA 기반 다분류 효율성 등급모형 개발 프로세스

자료를 활용하였으며(<http://www.kisvalue.com>), DEA 분석을 위해 Zhu[2003]의 DEA Solver를 사용하였다.

#### 4.2 DEA 효율성 평가

DEA 분석에 있어서는 우선적으로 투입요소와 산출요소를 선정하는 것이 매우 큰 이슈가 된다. 만약, 평가 대상기업 가운데 일부 기업에게만 유리한 투입요소와 산출요소를 선정했을 경우, 해당 기업이 다른 비교 대상들 보다 상대적으로 더 효율적이라는 결과를 초래할 수 있다. 그러므로, DEA에서 설득력 있는 효율성 평가결과를 설명하기 위해서는 평가 관련 문헌이나 대상 기업이 속한 산업의 특성을 토대로 투입요소와 산출요소를 선별하여야 한다. 본 연구에서는 DEA를 이용한 연구 가운데, 코스닥 기업, 소프트웨어 기업, SI 기업, 인터넷 기업 등을 대상으로 한 DEA 효율성 분석 연구를 토대로 하여, 벤처기업의 특성을 잘 반영할 수 있는 투입요소와 산출요소를 선택하였다. 연구대상이 벤처기업임을 고려하여, 본 연구에서는 DEA를 이용한 기업효율성 평가 연구 가운데 인터넷기업, 소프트웨어 기업, 그리고 SI기업 등에 적용된 선행 연구[Kim, 2004; Barua et al., 2004; Kim, 2005; Carlos et al., 2005; Koo et al., 2006; Kim, 2006; Lee et al., 2007; Hong et al., 2007; Hong and Park, 2007]를 기반으로 종업원수, 자산총계, 매출원가, 판매비와 관리비 등 총 4

가지의 투입요소와 매출액을 산출요소로 하여 연구를 진행하였다. 종업원수, 자산총계, 그리고 매출액은 DEA를 적용한 기업평가에 대표적으로 사용되는 투입 및 산출요소이며, 판매비와 관리비는 기업 관리 및 유지 등으로 발생하는 비용으로 연구비, 정상개발비 등 매출원가에 속하지 않는 모든 영업비용을 포함하는 것으로 코스닥 벤처기업의 특성을 반영할 수 있는 것이라 판단되어, DEA 효율성 평가에 투입요소로 적용하였다. 아래의 <표 2>에서 본 연구에 사용된 DEA 투입요소와 산출요소의 기초통계량을 보면, 투입요소와 산출요소에 대해 기업들의 편차가 크음을 알 수 있다. 이는 상대적으로 운영 규모가 큰 기업도 있으나, 벤처기업의 특성상 적은 자원으로 운영되고 있는 기업들 또한 많이 분포해있기 때문인 것으로 추정된다.

코스닥 벤처기업 중 사업서비스 관련 154개 기업에 대한 상대적인 효율성을 평가하기 위해 BCC 모형에 대한 DEA를 실행하였으며, BCC 모형에서 효율성이 1인 기업은 모두 33개로 효율성 등급 1을 부여하였다. 그리고, 전체 154개 기업 중 효율기업으로 분류된 33개 기업을 제외한 121개 기업자료에 대해 다시 DEA를 실행한 후, 결과로 제시된 BCC 효율점수가 1인 기업에 효율성 등급 2를 부여하였다. 이와 같은 방법으로 모두 3회의 DEA를 실행하여 각 단계마다 효율점수가 1인 기업을 제외하고 나머지 기업들에 대해 DEA로 재분석하면, 154개 기업은 1등급에서 4등급까지 분

<표 2> DEA에 사용된 투입요소와 산출요소의 통계량(단위: 명, 천원)

구 분		평 균	표준편차	최대값	최소값
투입요소	종업원수	188	355	2,819	5
	자산총계	46,115,384	56,987,490	373,726,916	2,973,996
	매출원가	28,723,577	46,931,188	319,699,733	126,968
	판매비와 관리비	7,706,322	10,621,591	101,455,489	743,379
산출요소	매출액	38,076,654	57,770,163	357,494,015	931,550

<표 3> DEA 효율점수로 분류 된 기업의 수

효율성 등급	설 명	기업 수
1	첫 번째 DEA 실행 결과 효율점수가 1인 기업	33
2	첫 번째 DEA 실행 결과 중 그룹 1을 제외하고 두 번째 DEA 실행 결과 효율점수가 1인 기업	44
3	두 번째 DEA 실행 결과 중 그룹 2를 제외하고 세 번째 DEA 실행 결과 효율점수가 1인 기업	34
4	세 번째 DEA 실행 결과 중 그룹3을 제외한 나머지 기업	43

류되었다. 최종적으로 분류된 그룹별 기업 수는 <표 3>과 같다.

### 4.3 DEA 기반 다분류 등급 예측 모형

본 절에서는 DEA 효율성 분석 모형 개발에 사용되었던 154개의 사업서비스 영역 기업자료를 사용하여 DEA 기반 다분류 모형을 개발한다. 모형개발에 사용할 변수는 모두 44개의 재무변수를 대상으로 다중판별분석과 분산분석을 통해 최종적으로 영업이익, 경영자본 영업 이익률, 매출액 총 이익률, 총 C/F대 부채비율, 총자본회전율, 경영자본회전율 등 6개를 선정하였다. SVM을 이용하여 DEA 기반 다분류 예측모형을 구축하기 위해, SVM에서 비선형함수로 많이 사용되고 있는 커널함수인 가우시안 RBF를 사용하였으며, 이는 주요 모수에 따라 성과가 달라진다. 최적화 모형을 구축하기 위해 모수인 C와  $\gamma$ 의 조절이 필요하며 본 연구에서는 모수 C에 대해서는 1, 20, 40, 60, 80, 100, 200, 1000 등의 8가지 경우를,  $\gamma$ 는 1, 0.7, 0.5, 0.2, 0.1, 0.05, 0.01, 0.005 등의 8가지 경우로 다분류 SVM의 실험을 수행하였다. 데이터마이닝 기법에서 흔히 발생하는 과다학습(over-fitting)을 제어하기 위해 표본은 학습용, 평가용, 검증용 셋으로 나누어 평가용 셋에서 가장 성과가 좋은 모수를 이용하여 SVM 모형을 최적화 하였다. 또한, 신뢰할 만한 실험결과를 얻기 위해서 5-fold cross validation 기법을 통한 5개의 실험

표본을 구성하였다. K-fold cross validation 기법은 실험결과가 표본추출에 따른 변동에 강건하고 성과가 매우 신뢰할 만한 것으로 알려져 있다 [Hu *et al.*, 1999]. 실험에는 공개 소프트웨어인 B SVM(<http://www.csie.ntu.edu.tw/~cjlin/bsvm>)을 사용하였다.

SVM 모형을 이용하여 다분류 문제를 해결하기 위해, 본 연구에서는 기존의 이진 분류 기반 one-against-one 방법과, all-together 방법으로 분해기법에 차이가 있는 Weston and Watkins[1999]의 방법과 Crammer and Singer[2000]의 방법 등 세 가지 기법을 적용하였으며, 결과는 다음의 <표 4>와 같다. SVM의 세 가지 분류 방법에 따른 벤처기업의 다분류 예측 결과를 보면, one-against-one 방법은 검증용 데이터 셋에서 28.6%에서 56.3% 정도의 예측 정확성을 나타내며 평균적으로 약 39.9%의 정확성을 보이고 있다. Weston and Watkins 방법에 대한 예측 정확성은 검증용 데이터에서 33.3%에서 56.3% 정도이며, 평균적으로 약 45.2%의 예측 정확성을 나타내었다. 그리고, Crammer and Singer 방법의 예측 정확성은 검증용 데이터에서 33.3%에서 50.0% 정도이며 평균적으로 40.4%의 예측 성과를 보였다.

특히, 다분류 문제는 이진분류 문제와 달리 각 범주에 대해서 적은 표본 수의 영향으로 예측모형의 성과가 급격하게 떨어져 통계분석의 적합성이 결여될 수 있으며, 실제로 예측모형의 성과가 급격하게 떨어지는 경향이 있다. 다분류 등급

<표 4> 다분류 SVM의 적중률

	OAO <sup>1)</sup>				W&W <sup>2)</sup>				C&S <sup>3)</sup>			
	학습	평가	검증	$\gamma, C^4)$	학습	평가	검증	$\gamma, C$	학습	평가	검증	$\gamma, C$
Set 1	49.2	56.3	37.5	0.7, 1	63.9	56.3	43.8	0.2, 80	66.4	56.3	31.3	0.2, 200
Set 2	50.0	43.8	56.3	0.05, 40	51.6	43.8	56.3	1, 1	55.7	62.5	43.8	0.1, 1
Set 3	41.0	62.5	43.8	0.005, 60	54.2	62.5	50.0	0.7, 1	50.0	56.3	43.8	0.01, 100
Set 4	66.9	53.3	33.3	0.7, 20	66.1	53.3	33.3	0.5, 40	69.4	53.3	33.3	0.5, 60
Set 5	69.8	64.3	28.6	0.5, 80	74.6	64.3	42.9	1, 80	67.5	57.1	50.0	0.5, 20
평균	55.4	56.0	39.9		62.1	56.0	45.2		61.8	57.1	40.4	

주) 1) One-against-one, 2) Weston and Watkins(1999), 3) Crammer and Singer(2000), 4)  $\gamma, C$ : 가우시안 RBF의 모수.

예측에 대한 여러 선행연구들이 분류 등급 수에 대한 예측오류 분석을 제시하고 있으며, <표 5>는 다분류 SVM이 실제 등급과 상하 한 개 등급 내의 예측력을 보여준다[Huang et al., 2004; Song et al., 2007]. 실제 등급과 인접한 두 개의 등급을 포함한 예측성적을 살펴보면 W&W 방법이 85.7%의 예측 적중률을 보이며 그 뒤로 OAO가 83.1%, C&S가 79.2%임을 알 수 있다. 이러한 예측성과는 기존의 이진분류모형의 결과와 비교하면, 다분류 모형을 이용하여 실무에서 벤처기업의 효율성 평가를 위한 예측모형을 85.7%(W&W 방법)

의 성과로 구축할 수 있으며 효율성의 여부만을 판단하는 이진분류보다 훨씬 많은 정보를 제공할 수 있다.

### V. 결론 및 향후 연구 방향

본 연구에서는 154개의 국내 벤처기업을 대상으로 하여 상대적인 기업 평가가 용이한 대표적인 비모수 접근법인 DEA를 이용하여 다분류 효율성 등급모형 개발 프로세스를 제시하였다. 이는 신용등급과 같은 형태의 기업평가 관련 주요

<표 5> 상하 한 개 등급을 포함한 예측결과

실제 등급	OAO <sup>1)</sup> 예측등급				W&W <sup>2)</sup> 예측등급				C&S <sup>3)</sup> 예측등급				총합계
	1	2	3	4	1	2	3	4	1	2	3	4	
1	6	9	2	4	7	8	1	5	6	10	2	3	21
2	1	11	3	7	4	7	4	7	1	9	7	5	22
3	0	10	0	6	2	7	1	6	0	6	6	4	16
4	0	0	4	14	1	0	1	16	0	1	3	14	18
총합계	7	30	9	31	14	22	7	34	7	26	18	26	77
예측율	83.1%				85.7%				79.2%				

주) 1) One-against-one, 2) Weston and Watkins[1999], 3) Crammer and Singer[2000].

정보가 미비한 기업들에 대한 상대적 평가 지표로 사용될 수 있다. 또한 DEA 실행문제에 있어서의 새로운 평가 대상이 생길 때 마다 분석을 다시 실행해야 하는 번거로움을 제거하고, 최종적으로 특정 기업에 대한 보편 타당한 단독 평가가 가능하도록 SVM과 같은 데이터마이닝 기법을 적용하여 효율성 등급 예측모형을 개발하였다. DEA는 상대적인 기업 평가에 잘 활용되고 있는 방법론이기는 하지만, 현업에서 실시간으로 사용될 때에는 비모수적 접근법 보다는 모수를 가진 모형을 이용하여 등급예측에 적용함으로써 유용성을 더 극대화할 수 있다는 장점을 가지게 된다.

동일 자원을 투입하여 더 많은 산출이 발생되거나, 또는 동일한 산출이 발생하는 기업들 중에서 더 적은 자원 투입이 이루어지는 조직이 상대적으로 효율성이 높다고 판단될 수 있으므로, 어느 기업에서 얼마나 더 높은 효율성을 확보했는가에 대한 문제는 투자에 대한 효과를 분석할 수 있는 주요 조건이 될 수 있다. DEA를 통한 효율성 분석의 단점일 수 있는 보편 타당한 등급 문제에 대한 어려움, 그리고 현업에서 적용하기 용이한 모수기반 예측 모형의 필요성 등을 해결하기 위해 데이터마이닝 기법의 적용이 필요하며, 이러한 예측 모형은 벤처기업과 같이 주요 신용평가 회사가 제시하지 못하는 등급 정보를 제공함으로써, 의사결정 문제에 유용하게 활용될 수 있다. 결과적으로 이 연구를 통해 현실적으로 더욱 유용하다고 할 수 있는 다분류 문제에 대해, 기업의 효율성 측면에서 기존의 이진분류 문제뿐만 아니라 다분류 문제를 해결할 수 있는 방안을 제시하였고, 또한 신용등급과 같은 정보를 활용하기 어려운 기업에 대해, 효율성을 기반으로 한 기업의 등급 정보를 투자자, 대출심사자, 벤처기업 지원을 위한 정부 등에 제공함으로써 벤처기

업의 활성화에 기여할 수 있을 것으로 기대된다.

향후 연구에서 개선되어야 할 본 연구의 한계점은 다음과 같다. 첫째, 본 연구는 신용등급과 유사한 사례로 효율성 정보를 통한 다분류 등급화를 기반으로 한 예측문제에 기거인하고 있으므로, 향후 연구에서는 모형 구축에 필요한 변수 선정 문제에 대한 다양성을 고려해 보아야 할 것이다. 즉, 보다 영향력 있는 변수를 채택하기 위한 방법에 대한 연구가 필요하다. 둘째, 연구 대상으로 벤처기업 가운데 사업서비스 영역의 기업체들을 선정하여 실험하였으나, 모형의 일반화를 위해 그 적용 영역을 확장시킬 수 있도록 해야 한다. 셋째, 다분류 문제에서는 이진 분류에 비해 등급 내 데이터 개수의 불균형이 문제가 생길 수 있기 때문에 예측결과 및 정확성 검증 여부가 어려울 수 있으므로, 등급 간 정보가 어떠한 중요성을 가지고 있는지 검토하여 개별 등급에 대한 예측력을 고려할 것인가 또는 상하 몇 개의 등급을 포함한 결과를 고려할 것인가에 대한 의사결정 과정에 대한 고민이 필요하다. 넷째, 다분류 SVM의 주요 모수를 결정하는 문제이다. 다분류 SVM을 이용한 예측 모형 개발에 있어서, 커널함수에 대한 주요 모수 선택은 시행착오(trial and error) 방법을 사용한다. 때문에 커널함수의 선택과 함께 최적의 모수를 찾아내기 위한 방법을 탐구하는 노력이 필요하다. 다섯째, 분류 기법의 적용에 대한 것으로, 본 연구에서는 분류기법으로 이진분류 기반 방법인 one-against-one 그리고 all-together 방법인 Weston and Watkins[1999], Cramer and Singer[2000] 등 제한적으로 3가지 방법만을 적용하였다. 새로운 분류방법들에 대한 다양한 시도가 더 필요하며, SVM 기법 이외의 여러 데이터마이닝 기법을 예측모형 개발에 적용해봄으로써 예측모형 성과를 비교하는 시도도 가능할 것이다.

## ⟨References⟩

- [1] Ahn, H., Kim, K.-j., and Han, I., "Intelligent Credit Rating Model for Korean Companies using Multiclass Support Vector Machines," *Korean Management Review*, Vol. 35, No. 5, 2006, pp. 1479-1496.
- [2] Banker, R.D., Charnes A., and Cooper, W.W., "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, Vol. 30, No. 9, 1984, pp. 1078-1092.
- [3] Barua, P.L., Brouckette, P.L., Cooper, W.W., Deng, H., Parket, B.R. Ruefli, T.W., and Winston, A., "DEA evaluations of long-and short-run efficiencies of digital vs. physical product "dot com" companies," *Socio-Economic Planning Sciences*, Vol. 38, No. 4, 2004, pp. 233-253.
- [4] Carlos, S.C., Yolanda, F.C., and Cecilio, M.M., "Measuring DEA efficiency in Internet companies," *Decision Support Systems*, Vol. 38, No. 4, 2005, pp. 557-573.
- [5] Charnes, W.W., Cooper, and Rhodes, E., "Measuring the Efficiency of Decision Making Units," *European Journal of Operation Research*, Vol. 2, No. 6, 1978, pp. 429-444.
- [6] Crammer, K. and Singer, Y., "On the learnability and design of output codes for multiclass problems," *Comput. Learning Theory*, 2000, pp. 35-46.
- [7] Farrell, M.J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Vol. 120, 1957, pp. 253-281.
- [8] Garavaglia, S., "An application of a Counter-Propagation Neural Networks: Simulating the Standard and Poor's Corporate Bond Rating Systems," *Proceedings of the First International Conference on Artificial Intelligence on Wall Street*, 1991, pp. 278-287.
- [9] Hearst, M.A., Dumais, S.T., Osman, E., Platt, j., and Scholkopf, B., "Support vector machines," *IEEE Intelligent System*, Vol. 13, No. 4, 1998, pp. 18-28.
- [10] Hong, H.K., Shin, C.K., Park, S.C., and Kim, S.H., "Evaluating the efficiency of system integration projects using data envelopment analysis(DEA) and machine learning," *Expert Systems with Applications*, Vol. 16, 1999, pp. 283-296.
- [11] Hong, T. and Park, J., "A DEA-Based Data Mining for the Evaluation of the Efficiency in the IT Venture Business," *Proceedings of the 13th Asia Pacific Management Conference*, 2007.
- [12] Hong, T., Park, J., and Kim, E., "Using DEA-logit Approach to Evaluate Efficiency of IT Venture Business," *The Journal of Internet Electronic Commerce Research*, Vol. 7, No. 3, 2007, pp. 429-449.
- [13] Hsu, C.W. and Lin, C.J., "A Comparison of Methods for Multiclass Support Vector Machines," *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, 2002, pp. 415-425.
- [14] Hu, M.Y., Zhang, G., Christine, X.J., and Patuwo, B.E., "A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting," *Decision Sciences*, Vol. 30, No. 1, 1999, pp. 197-216.
- [15] Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., and Wu, S., "Credit rating analysis with support vector machines and neural network:

- a market comparative study," *Decision Support Systems*, Vol. 37, 2004, pp. 543-558.
- [16] Hwang, W.W. and Yang, B.S., "Fault Diagnosis of Rotating Machinery Using Multi-class Support Vector Machines," *Transactions of the Korean Society for Noise and Vibration Engineering*, Vol. 14, No. 12, 2004, pp. 1233-1240.
- [17] Kim, J.J., "Research Papers: The Effect of DEA Efficiency on ROE of KOSDAQ Firms," *TAX ACCOUNTING RESEARCH*, Vol. 14, 2004, pp. 153-170.
- [18] Kim, J.W., "Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems," *Expert Systems*, Vol. 10, 1993, pp. 167-171.
- [19] Kim, K.J., "Productivity Analysis of Software Industry in Korea," *Productivity Review*, Vol. 19, No. 1, 2005, pp. 95-111.
- [20] Kim, K.S., "Analyzing the Technical Efficiency of Korean System Integration Firms Using DEA and Malmquist Productivity Analysis," *Journal of the Korea society of IT services*, Vol. 5, No. 1, 2006, pp. 1-16.
- [21] Koo, C.O., Kim, G.J., and Park, J.W., "The Productivity Analysis of KOSDAQ Software Companies," *Journal Finance and Accounting Accountiong Information*, Vol. 6, No. 1, 2006, pp. 27-51.
- [22] Kwon, Y.S., Han, I.G., and Lee, K.C., "Ordinal Pairwise Partitioning(OPP) approach to neural networks training in bond rating," *Intelligent Systems in Accounting, Finance and Management*, Vol. 6, 1997, pp. 23-40.
- [23] Lee, K.-J., Kim, J.-J., and Geon, C., "The Evaluation of the Efficiency of Internet Companies Using a DEA Model," *DAEHAN Association of Business Administration*, Vol. 20, No. 1, 2007, pp. 109-136.
- [24] Lee, Y.-C., "Credit Risk Evaluation Using Worst Practice DEA," *DAEHAN Association of Business Administration*, Vol. 19, No. 5, 2006, pp. 183-203.
- [25] Min, J.H. and Jeong, C.W., "Nonparametric Approach to Bankruptcy Prediction: Developing Cross Peeling Technique integrating DEA and Negative DEA," *Korean Management Review*, Vol. 35, No. 4, 2006, pp. 1157-1180.
- [26] Nguyen, M.H. and Rajapakse, J.C., "Multi-Class Support Vector Machines for Protein Secondary Structure Prediction," *Genome Informatics*, Vol. 14, 2003, pp. 218-227.
- [27] Paradi, J.C, Asmild, M., and Simark, P.C., "Using DEA and worst practice DEA in credit risk evaluation," *Journal of Productive Analysis*, Vol. 21, No. 2, 2004, pp. 153-165.
- [28] Platt, J.C., Cristianini, N., and Shawe-Taylor, J., "Large Margin DAGs for Multiclass Classification," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, Vol. 12, 2000, pp. 547-553.
- [29] Shin, K.S. and Han, I., "A case-based approach using inductive indexing for corporate bond rating," *Decision Support Systems*, Vol. 32, No. 1, 2001, pp. 41-52.
- [30] Song, K.M., Bahn, S.S., and Kim, S.T., "The Significance of Financial Model in Venture Credit Rating and Its Comparison with the Case of Large Corporate," *The Korean Venture Management Review*, Vol. 10, No. 2, 2007, pp. 1-23.
- [31] Tay, F.E.H. and Cao, L.J., "Application of support vector machines in financial time series forecasting," *Omega*, Vol. 29, No. 4, 2001, pp. 309-317.
- [32] Vapnik, V., *Statistical Learning Theory*, New

- York, Wiley, 1998.
- [33] Weston, J. and Watkins, C., "Multi-class support vector machines," presented at the Proc. ESAMM99, M. Verleysen, Ed., Brussels, Belgium, 1999.
- [34] Zhu, J., "Multi-factor performance measure model with an application to Fortune 500 companies," *European Journal of Operational Research*, Vol. 123, No. 1, 2000, pp. 105-124.
- [35] Zhu, J., "Multi-factor performance measure model with an application to Fortune 500 companies," *European Journal of Operational Research*, Vol. 123, No. 1, pp. 105-124.
- [36] Zhu, J., *Quantitative Model For Performance Evaluation And Benchmarking*, Kluwer Academic Publishers, 2003.



◆ About the Authors ◆



Jiyoung Park

Jiyoung Park is a Ph.D. Candidate of Management Information Systems/ Operations Management in the College of Business Administration, Pusan National University, Korea. She received a M.A. in Business Administration and a B.S. in Statistics from Pusan National University. Her research interests are in data mining, CRM, and venture business.



Taeho Hong

Taeho Hong is an associate professor in the College of Business Administration, Pusan National University, Korea. He received a Ph.D. from the Korea Advanced Institute of Science and Technology. His research interests include data mining, CRM, and Business Intelligence. He has published in Expert Systems, Expert Systems with Applications, the Journal of MIS Research, Journal of Intelligence and Information Systems, and the Journal of Information Systems.