
부스팅 인공신경망 학습의 기업부실예측 성과비교

김명종* · 강대기**

An Empirical Analysis of Boosting of Neural Networks for Bankruptcy Prediction

Myoung-Jong Kim* · Dae-Ki Kang**

요 약

최근 기계학습 분야에서 분류자의 정확도 개선을 위하여 제안된 다양한 방법들 중 가장 큰 주목을 받고 있는 학습방법 중 하나는 앙상블 학습이다. 그러나 앙상블 학습은 의사결정트리와 같이 불안정한 학습 알고리즘의 성과 개선 효과는 탁월한 반면, 인공신경망과 같이 안정적인 학습알고리즘의 성과 개선 효과는 응용 분야와 구현 방법에 따라 서로 상반된 결론들을 보여주고 있다. 본 연구에서는 국내 기업의 부실화 예측문제를 활용하여 인공신경망 분류자 및 대표적 앙상블 학습기법인 부스팅 분류자를 적용한 결과 앙상블 학습은 기업부실 예측문제에 있어 전통적 인공신경망의 성과를 개선할 수 있음을 검증하였다.

ABSTRACT

Ensemble is one of widely used methods for improving the performance of classification and prediction models. Two popular ensemble methods, Bagging and Boosting, have been applied with great success to various machine learning problems using mostly decision trees as base classifiers. This paper performs an empirical comparison of Boosted neural networks and traditional neural networks on bankruptcy prediction tasks. Experimental results on Korean firms indicated that the boosted neural networks showed the improved performance over traditional neural networks.

키워드

부스팅, 인공신경망, 기업부실예측

Key word

Boosting, Neural Networks, Bankruptcy Prediction

* 동서대학교 경영학부 전임강사
** 동서대학교 컴퓨터정보공학부 조교수

접수일자 : 2009. 07. 15
심사완료일자 : 2009. 08. 27

I. 서 론

최근 기계학습 분야에서 분류자(classifier)의 정확도 개선을 위하여 제안된 다양한 방법들 중 가장 큰 주목을 받고 있는 학습방법 중 하나는 앙상블 학습이다. 앙상블 학습은 학습 데이터를 정확하게 표현할 수 있는 하나의 가설을 선택하는 것이 아니라 가설들의 집합을 구성하고 새로운 데이터를 예측할 때에는 그 가설들의 예측을 결합하여 최종 결정을 내린다. 앙상블 학습은 복수의 기저 분류자(base classifiers) 집합인 앙상블을 구성하고, 앙상블에서 추론된 복수의 학습결과를 단일 강분류자(a single strong classifier)를 통하여 결합한 최종 결과를 산출하게 된다. 현재까지 제안된 앙상블 학습 알고리즘 중 보편적으로 사용되고 있는 알고리즘은 Breiman (1994)에 의하여 제안된 배깅 알고리즘(Bagging algorithm)과 Freund(1995) 및 Freund & Schapire(1997)에 의하여 제안된 부스팅 알고리즘(Boosting algorithm)이다.

특히, 앙상블 학습은 의사결정트리(Decision Tree: DT)와 같은 기계학습과 결합하여 학습 알고리즘의 예측력을 크게 개선시킬 수 있음을 보여주었다(Freund & Schapire, 1996; Drucker & Cortes, 1993; Quinlan, 1996; Bauer & Kohavi, 1999; Macline & Optiz, 1997). 이러한 연구를 기초로 앙상블 학습과 관련된 대부분의 연구는 DT를 기저분류자로 활용한 학습 알고리즘의 성과개선에 초점을 맞추고 있다.

최근에는 DT 부스팅 기법을 기업부실예측에 적용한 연구들이 진행되었다 (Alfrazo, Gámez, & García, 2007; Alfrazo, García, Gámez, & Elizondo, 2008). 스페인의 기업을 대상으로 DT 앙상블과 인공신경망(Neural Networks: NN)의 성과를 비교하여 DT 부스팅 기법이 NN에 비하여 오분류율을 약 30% 감소시키는 등 탁월한 예측력을 보유하고 있음을 보여주었다.

DT 앙상블 중심의 연구가 이루어지는 주요원인은 앙상블 학습이 DT와 같이 학습 데이터에 따라 생성되는 분류자가 변하는 불안정한 학습 알고리즘에 대해서는 탁월한 성과 개선의 효과가 있기 때문이다(Bauer & Kohavi, 1999; Drucker & Cortes, 1996; Freund & Schapire, 1996; Maclin & Optiz 1997; Quinlan, 1996). 반면에 NN과 같이 안정적인 학습 알고리즘에 대한 앙상블 학습의 적용은 단일 분류자의 정확도를 개선시킬 수 있다는 결과

를 보고한 연구(Kim et al., 2003; Schwenk & Benjio, 2000)도 있는 반면, 이들의 성능 향상은 DT 앙상블 기법과 비교하여 일부 응용 분야에 제한되어 나타나고 있으며, 심지어 앙상블 학습의 성과가 단일 분류자보다 떨어진다는 연구 결과도 보고되고 있다(Buciu et al., 2001; Dong & Han, 2004; Evgeniou et al., 2000; Optiz & Maclin, 1999). 즉, NN 앙상블은 응용 분야와 구현 방법에 따라 서로 상반된 결론들을 보여주고 있기 때문에 앙상블 기법을 효과적으로 적용하기 위해서는 응용분야별로 사전적 실험을 통한 효과검증이 선행되어야 한다.

최근 DT 앙상블 중심의 연구에 대한 많은 비판이 제기되고 있으며 NN 앙상블이나 SVM 앙상블의 성과를 개선하기 위한 연구들이 진행되고 있다(Optiz & Maclin, 1999; Schwenk & Benjio, 2000). 이러한 연구들은 첫째, NN 앙상블이나 SVM 앙상블의 성과 저하의 원인을 분석하고 둘째, 전통적인 앙상블 학습과는 차별화된 학습 전략을 이용하여 앙상블 학습의 성과를 개선하는데 초점을 맞추고 있다. 마지막으로, 이러한 연구 결과를 기초로 각 분류자의 일반적 학습 특성 및 향후 앙상블 학습에 고려해야 할 요인에 대한 통찰력을 제공하는 것을 목적으로 하고 있다. 본 연구에서는 전통적 인공신경망 분류자와 앙상블 학습의 대표적 기법인 부스팅 알고리즘을 부실예측 문제에 적용하여 성과 측면에서 각 분류자들의 효과를 분석하고자 한다.

본 논문은 구성체계는 다음과 같다. 제 2장에서는 앙상블 알고리즘에 대하여 소개하고자 한다. 제 3장에서는 표본수집 및 채무비율 선정과 관련된 연구설계 과정을 설명하고자 하며, 제 4장에서는 본 연구에서 벤치마킹 대상이 되는 인공신경망 분류자, 배깅 분류자 및 부스팅 분류자의 설계 프레임워크를 설명하고자 한다. 제 5장에서는 각각 주요 연구 결과와 향후 연구방향을 제시하고자 한다.

II. 앙상블 학습

앙상블 학습은 여러 개의 분류자로 구성된 분류자 앙상블을 생성하여 단일 분류자의 정확성을 개선하기 위한 목적으로 제안된 학습방법이다. 앙상블 학습을 하는 이유는 단일 분류자만으로는 학습이 어려운 복잡한 패

턴들을 여러 개의 하위 패턴으로 나누어 효과적인 학습을 진행할 수 있으며 여러 하위 패턴에 대한 학습 결과를 결합하여 단일 분류자 보다 정확한 예측 결과를 제공하기 때문이다.

Hansen and Salamon의 연구(1990)에서는 앙상블이 단일 분류자와 비교하여 높은 성과를 나타내는 이유를 분석하였다. 앙상블의 분류 결과가 상호 독립적이고 정확성 수준을 p 라 하면 K 개의 분류자가 구성된 앙상블의 분류 정확성은 다음과 같이 계산된다.

$$\sum_{k>K/2}^K \binom{K}{k} p^k (1-p)^{K-k}$$

이진분류 문제에 있어 만일 p 가 0.5보다 크다면 K 가 증가함에 따라 앙상블의 분류 정확도가 높아지게 된다. 결과적으로 앙상블 학습을 통하여 단일 분류자의 정확도가 개선되며 일반화 성능(generalization performance) 또한 크게 향상된다.

현재까지 제안된 앙상블 학습 알고리즘 중 보편적으로 사용되고 있는 알고리즘은 배깅 알고리즘(Breiman, 1994) 및 AdaBoost 알고리즘(Freund, 1995; Freund & Schapire, 1997)이다.

2.1 배깅 알고리즘

배깅 알고리즘은 n 개의 원시 데이터에서 복원추출 방법으로 K 개의 붓스트랩(bootstrap) 분석용 데이터를 생성하고 각 붓스트랩 분석용 데이터에 특정한 분류 알고리즘을 적용하여 K 개의 분류자를 구성한다. i 번째 관측치에 대한 최종 결과는 강분류자인 배깅 분류자를 통하여 앙상블의 결과와 결합함수 ak 를 결합하여 다음과 같이 산출된다. 여기에서 xi 는 i 번째 관측치의 예측변수 벡터이며 $C_k(xi)$ 는 예측변수 벡터 xi 에 대한 k 번째 분류자의 분류결과이다.

$$C(x_i) = \text{sign} \left(\sum_{k=1}^K \alpha_k C_k(x_i) \right)$$

2.2 AdaBoost 알고리즘

AdaBoost 알고리즘은 가장 보편적으로 사용되는 부스팅 알고리즘이다. n 개의 학습 표본과 K 개의 기저 분류

자로 구성된 앙상블 $C = \{C_1, C_2, \dots, C_K\}$ 을 가정하면 k 번째 분류자의 오류율(e_k)은 다음과 같이 단순평균으로 계산된다.

$$e_k = \frac{1}{n} \sum_{i=1}^n L(C_k(x_i), y_i)$$

$$L(C_k(x_i), y_i) = \begin{cases} 1 & C_k(x_i) \neq y_i \\ 0 & C_k(x_i) = y_i \end{cases}$$

여기에서 xi 는 i 번째 관측치의 예측변수 벡터이고 y_i 는 i 번째 관측치의 범주를 나타내며 $C_k(xi)$ 는 예측변수 벡터 xi 에 대한 k 번째 분류자의 분류결과이다. $k+1$ 번째 분류자에서 i 번째 관측치에 부여되는 가중치는 다음과 같이 조정되어 오분류된 관측치에 더 높은 가중치가 부여된다.

$$w_{k+1}(i) = w_k(i) \exp(\alpha_k L(C_k(x_i), y_i))$$

여기에서 α_k 는 분류자의 결합가중치로 활용되며 $\ln((1-e_k)/e_k)$ 로 계산된다. $k+1$ 번째 분류자의 학습표본을 구성할 때 가중치가 높은 오분류 관측치가 많이 포함되기 때문에 부스팅 알고리즘은 오분류 관측치에 초점을 맞춘 학습을 진행할 수 있게 된다. i 번째 관측치의 최종 결과는 강분류자인 부스팅 분류자를 통하여 다음과 같이 계산된다.

$$C(x_i) = \text{sign} \left(\sum_{k=1}^K \alpha_k C_k(x_i) \right)$$

2.3 앙상블 학습의 성과 비교

많은 실증연구에서 활용된 DT 앙상블의 경우 Drucker and Cortes의 연구(1996)에서는 부스팅 알고리즘을 DT에 적용하였으며, Quinlan(1996)은 부스팅 알고리즘과 배깅 알고리즘을 C4.5와 결합하였으며, Bauer and Kohavi의 연구(1999)에서는 대규모의 UCI데이터를 대상으로 C4.5와 naïve-Bayes에 배깅과 부스팅을 적용하였다. 이들의 연구 결과에서는 앙상블 학습이 CART (Classification and Regression Tree)와 C4.5 등의 예측력을 크게 개선시킬 수 있음을 보여주었다.

Optiz and Maclin의 연구(1999)에서는 23개 UCI 데이

터 집합(data set)을 대상으로 DT와 NN에 배깅 알고리즘과 부스팅 알고리즘을 적용하였다. 그 결과 배깅 알고리즘과 부스팅 알고리즘은 DT의 성과 개선에 효과적임을 보고하고 있다. 반면에 NN에 대하여 배깅 알고리즘은 성과 개선에 효과적이거나, 부스팅 알고리즘의 경우 일부 데이터 집합에서 단일 분류자보다 성과가 저하될 수 있음을 보여주었다. Schwenk and Bengio의 연구(2000)에서는 학습방법을 차별화한 세 가지 NN과 부스팅 알고리즘을 결합하였으며, 그 결과 안정적인 성과 개선을 보여주었다.

앙상블 학습은 DT와 같은 불안정한 학습 알고리즘의 예측력을 크게 개선시킬 수 있는 반면에 NN과 같은 안정적인 학습 알고리즘에 대해서는 불안정한 학습 알고리즘과 비교하여 일부 분야에 한정되어 성능 향상을 보여주었다. 일부 연구에서는 앙상블 학습의 성과가 단일 분류자보다 성능이 떨어진다는 연구 결과를 제시하기도 하였다(Buciu et al., 2001; Dong & Han, 2004; Evgeniou et al., 2000; Optiz & Maclin, 1999; Valentini & Dietterich, 2004).

앙상블의 성과 저하는 분류자의 다양성이 부족하고 분류자 사이에 높은 상관관계가 존재하는 경우에 나타나는 것으로 분석되고 있다.

Hansen and Salamon(1990)은 성과개선을 위한 앙상블의 필요충분조건으로 분류자들이 임의 추측보다 정확해야 하고 다양하게 구성되어야 함을 주장하였다. 이는 분류자의 정확도가 50% 이상이며 오분류 패턴이 서로 다른 다양한 분류자로 앙상블을 구성하였을 때 성과가 크게 개선될 수 있음을 의미한다. 예를 들어, 다양한 분류자로 구성된 앙상블은 특정 분류자가 부정확한 예측을 하더라도 다른 분류자들이 정확히 예측할 수 있다면 최종 결과에서 정확하게 분류될 수도 있다. 그러나 분류자의 다양성이 부족한 경우, 특정 분류자가 부정확하게 예측한 관측치에 대하여 나머지 분류자들도 부정확하게 예측할 가능성이 높기 때문에 최종 결과 역시 부정확하게 된다.

Brieman(1996)은 배깅과 같은 앙상블 학습은 불안정한 학습 알고리즘의 성능을 크게 향상시킬 수 있는 반면에 안정적인 학습 알고리즘에 대해서는 뚜렷한 성능 향상이 관찰되지 않는다고 밝혔다. 앙상블 학습은 분류자 풀(classifier pool)의 다양성을 확보하기 위한 학습 전략으로 배깅, 부스팅 및 random forest 등의 다양한 샘플링

방법을 이용하고 있다. DT와 같은 불안정한 학습 알고리즘은 학습 데이터의 변화에 민감하게 반응하여, 생성되는 분류자에도 큰 변화가 나타나므로 분류자의 다양성을 어느 정도 확보할 수 있다. 반면, NN과 같은 안정적인 학습 알고리즘은 학습 데이터가 변하더라도 유사한 분류자를 반복적으로 생성하므로 분류자 사이의 상관관계가 높아진다. 이러한 높은 상관관계로 인하여 다중공선성 문제가 발생되며, 결과적으로 앙상블의 성과가 저하될 수 있다.

III. 연구 설계

3.1 표본선정방법

부실 기업은 2003~2005년 중 은행연합회 신용정보등록, 당좌부도발생, 회사정리절차개시, 기업구조조정절차개시 사유에 해당하는 729개 외부감사 제조기업으로 구성하였으며 부실 기업에 대응되는 정상 기업은 외부감사 제조기업 중 2005년 말 기준으로 부실 사유에 해당하지 않는 729개의 기업으로 구성하였다.

부실 예측에 사용되는 재무비율은 일차적으로 기존의 기업부실 예측연구에 사용된 비율 및 실무에서 부실 예측의 지표로 유용하게 활용되는 비율을 중심으로 30개의 재무비율을 수집하였다. 수집된 재무비율을 수익성, 부채상환능력, 레버지리, 자본구조, 유동성, 활동성 및 규모의 7개 재무비율 군으로 재분류하였다. 최종 입력변수는 t-test 등의 전통적 변수선정 기법은 데이터 노이즈(noise)의 영향을 직접적으로 받을 수 있기 때문에 이를 대체하여 데이터 노이즈의 영향을 완화할 수 있는 장점을 가진 ROC 분석을 이용하여 선정하였다. ROC 분석은 수평축에는 1-특이도, 수직 축에는 민감도를 표시하여 연결한 ROC 곡선에 기초하여 모형 또는 설명변수의 정확성을 분석하는 방법이다. 여기에서 민감도와 특이도는 <표 1>의 오분류표 (Confusion matrix)에서 각각 TP/(TP+FN)과 TN/(FP+TN)으로 측정된다.

ROC 곡선은 설명변수(또는 모형)의 정확도를 ROC 곡선의 면적 (Area Under ROC: AUROC)로 계산한다. AUROC가 1에 근접할수록 정확도가 높은 설명변수로 평가된다(Fawcett, 2006).

표 1. 오분류표
Table 1. Confusion matrix

		예측 범주	
		부실 (Positive)	정상 (Negative)
실 제 범 주	부실 (Positive)	True Positive (TP)	False Negative (FN)
	정상 (Negative)	False Positive (FP)	True Negative (TN)

ROC 곡선에 의하여 산출된 AUROC를 이용하여 각 분류군별로 AUROC가 높은 7개 재무비율을 선정하였다. 일차적으로 선정된 30개 비율 및 최종 선정된 7개 재무비율의 AUROC는 <표 2>에 제시되어있다.

비록 변별력과 직접적인 관련성은 없으나, 다중공선성 문제는 모형 개발 시 필수적으로 고려해야 할 문제이다. 본 연구에서는 7개 재무비율 사이의 다중공선성의 존재여부를 확인하기 위하여 분산팽창요인(Variance Inflation Factors: VIF) 분석을 실시하였다. VIF는 회귀계수 추정치의 분산이 이상적인 경우(서로 직교하는 경우)에 비해 얼마만큼 계수 추정치들의 분산이 팽창하는가를 알려주는 척도이며, k번째 변수의 VIF는 다음과 같이 계산된다.

$$VIF(k) = \frac{1}{1 - R_k^2}$$

여기서 R_k^2 는 k번째 변수를 반응변수로 하여 나머지 변수에 대해 회귀분석을 하였을 때 얻어지는 결정계수의 값이다. k번째 변수가 다른 변수들과 밀접한 관계에 있으면 R_k^2 는 1에 가깝게 되고 따라서 VIF(k)의 값이 커지게 된다. 만일 $5 < VIF(k) < 10$ 이면 k번째 변수의 다중공선성 가능성이 있으며, $VIF(k) < 10$ 이면 k번째 변수의 다중공선성이 심각하다고 평가할 수 있다. 분산 팽창요인의 분석결과 역시 <표 2>에 제시되어 있으며 최종 선정된 7개 재무비율 간에는 다중공선성이 실질적으로 존재하지 않음을 나타내고 있다.

표 2. ROC 및 분산팽창요인 분석 결과
Table 2. The result of ROC and VIF analysis

	재무비율	AUROC	VIF
수익성	총자산경상이익율	52.1	1.66
부채상환	EBITDA/이자비용	53.2	2.31
레버리지	총부채/총자산	52.4	1.97
자본구조	이익잉여금/총자산	53.6	2.74
유동성	현금비율	46.5	1.64
활동성	재고자산회전율	31.5	1.69
규모	총자산	25.2	1.61

3.2 실험설계

본 연구에서 벤치마킹으로 사용될 인공신경망은 입력층에 7개의 입력노드와 출력층에 2개의 출력노드 및 1개의 은닉층을 가진 다층 퍼셉트론 (Multi-Layer Perceptron: MLP)이다.

배경 분류자는 Breiman (1994)에 의하여 제안된 배깅 알고리즘을 사용하였으며, 부스팅 분류자는 Freund & Schapire(1997)에 의하여 제안된 AdaBoost.M1 알고리즘을 활용하였다. 부스팅 분류자의 새로운 기저 분류자는 이전 분류자의 성과에 기초하여 생성되며, 분류자의 가중평균 오류가 0.5보다 작을 때까지 순차적으로 분류자를 생성하게 된다.

Opitz와 Maclin(1999)에 의하면 앙상블 학습은 분류자 생성회수가 25번이 될 때까지는 오분류율이 크게 개선이 되었지만 그 이후에는 오분류의 감소에 별 영향을 받지 않는다고 제안하였다. 이러한 연구결과를 기초로 본 연구에서는 배깅 분류자 및 부스팅 분류자 생성 회수를 25회로 제한하였다

IV. 연구 결과

모형 적합과 모형의 성능평가에 모두 같은 자료를 써서 얻어지는 오분류율을 견보기 오분류율 (apparent misclassification rate)라고 하며 견보기 오분류율을 이용하는 경우 참오분류율(true misclassification rate)을 과소 추정한다고 알려져 있다.

이를 해결하기 위하여 가장 널리 쓰이는 분석방법은 교차타당성에 의한 추정방법으로 보통 주어진 자료를

10등분하는 10-fold 검증방법 (10-fold cross validation)을 일반적으로 사용하고 있다 (Optiz & Macline, 1999). 본 연구에서는 인공신경망과 부스팅 인공신경망의 성과차이가 우연한 결과가 아님을 확인하기 위하여 10-fold 검증을 5회 반복 수행하였다. 이를 위하여 전체 1458개의 기업을 표본수가 동일한 10개 fold로 구성하고 9개의 집합은 분석용 데이터로 활용하고 나머지 1개 fold를 검증용 데이터로 활용하게 된다. 이러한 방법으로 50회의 교차타당성 검증을 수행하였다.

<표 3>은 50회 교차타당성 검증결과로서 학습용 데이터와 검증용 데이터에 대한 인공신경망과 부스팅 인공신경망의 평균 정확도를 보여주고 있다. 부스팅 인공신경망은 전통적 인공신경망에 비하여 10개의 모든 신경망구조에서 약 3% 이상의 정확성 개선을 일관성있게 보여주고 있다. 50개의 실험 결과를 대상으로 t-test 분석 결과 부스팅 인공신경망은 1% 유의수준에서 인공신경망보다 높은 정확성을 보여주었다.

표 3. 예측정확성의 비교 (단위: %)
Table 3. Comparison of prediction accuracy

구조	인공신경망		부스팅	
	학습 표본	검증 표본	학습 표본	검증 표본
7-5-2	74.71	71.06	75.58	75.04
7-6-2	74.81	70.78	75.29	75.03
7-7-2	74.73	71.12	75.51	75.24
7-8-2	74.84	71.12	75.95	74.90
7-9-2	74.81	70.92	75.92	75.24
7-10-2	74.82	70.92	75.54	74.62
7-11-2	74.84	70.85	75.76	75.58
7-12-2	74.80	71.19	75.66	74.96
7-13-2	74.84	71.12	75.87	75.18
7-14-2	74.84	71.12	75.96	75.18
평균	74.80	71.02	75.70	75.10

V. 결론 및 향후 연구 방향

본 연구에서는 부실예측 인공신경망의 성과개선을 위하여 부스팅 기법과 인공신경망이 결합된 부스팅 인공신경망을 제안하였다. 실험결과들은 부스팅 인공신경망은 10개의 다양한 인공신경망의 성과를 안정적으로 개선하였고 과도적합에 대한 견고성을 강화하였음

을 보여주고 있다. 그러나 본 연구의 결과를 다양한 분야에 적용하기 위하여 향후 데이터 특성에 따른 부스팅 기법의 효과와 같은 심층분석이 추가되어야 한다. 예를 들어 데이터에 내재된 노이즈 (noise)가 학습결과에 미치는 영향에 대한 분석을 통하여 다양한 분류기법과 부스팅 기법의 결합에 대한 연구가 더욱 활발해질 수 있을 것으로 기대된다.

참고문헌

- [1] Alfaro, E., Gámez, M., & García, N. (2007). Multiclass corporate failure prediction by AdaBoost.M1. *Advanced Economic Research*, 13, 301-312.
- [2] Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, 45, 110-122.
- [3] Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36, 105-139.
- [4] Breiman, L. (1994). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- [5] Breiman, L. (1996). *Bias, variance, and arcing classifiers* (Tech.Rep.No.460). Berkeley: Statistics Department, University of California at Berkeley.
- [6] Buciu, I., Kotrooulos, C., & Pitas, I. (2001). Combining support vector machines for accuracy face detection. *Proc. ICIP*, 1054-1057.
- [7] Dong, Y. S., & Han, K. S. (2004). A comparison of several ensemble methods for text categorization, *IEEE International Conference on Service Computing*.
- [8] Drucker, H., & Cortes, C. (1996). Boosting decision trees, *Advanced Neural Information Processing Systems*, 8.
- [9] Evgeniou, T., Perez-Breva, L., Pontil, M., & Poggio, T. (2000). Bound on the generalization performance of kernel machine ensembles, *Proc. ICMI*, 271-278.
- [10] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.

[11] Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2), 256-285.

[12] Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Machine Learning: Proceedings of Thirteenth International Conference*, 148-156.

[13] Freund, Y., & Schapire, R. E. (1997). A decision theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Science*, 55(1), 119-139.

[14] Hansen, L., & Salamon, P. (1990). Neural network ensembles, *IEEE Trans. PAMI*, Vol.12, 993-1001.

[15] Kim, H. C., Pang, S., Je, H. M., K, D. J., & Bang, S. Y. (2003). Constructing support vector machine ensemble, *Pattern Recognition*, 36, 2757-2767.

[16] Maclin, R., & Opitz, D. (1997). An empirical evaluation of bagging and boosting. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, 546-551.

[17] Opitz, D., & Maclin, R. (1999). Popular ensemble methods: an empirical study. *Journal of Artificial Intelligence*, 11, 169-198.

[18] Quinlan, J. R. (1996). Bagging, boosting and C4.5. *Machine Learning: Proceedings of the Fourteenth International Conference*, 725-730.

강대기 (Kang, Dae-Ki)



1992년 한양대학교 전자계산학과 졸업 (공학사)

1994년 서강대학교 전자계산학과 졸업 (이학석사)

1994년~1999년 : 한국전자통신연구원 (연구원)

2006년 : Iowa State University Dept. of Computer Science 졸업 (PhD in Computer Science)

2007년 2월~2007년 8월 : 국가보안기술연구소 (선임연구원)

2007년 9월~현재 : 동서대학교 컴퓨터정보공학부 조교수

※ 관심분야 : 기계학습, 관계학습, 통계 적그래피컬 모델, 스마트폰, 온톨로지학습, 침입탐지, 웹방화벽, 웹마이닝, 컴퓨터비전

저자소개



김명종 (Kim, Myoung-Jong)

1991년 성균관대학교 회계학과 졸업 (경영학학사)

1993년 성균관대학교 일반대학원 졸업 (경영학석사)

1994년~2004 : 한국과학기술원 경영공학과 졸업 (경영학박사)

2008년3월~현재 : 동서대학교 경영학부 전임강사

※ 관심분야 : 지능형시스템, 의사결정지원시스템, 지식관리시스템