

다중 패턴 분류를 위한 Import Vector Voting 모델

Import Vector Voting Model for Multi-pattern Classification

최준혁* · 김대수** · 임기욱***

Jun-Hyeog Choi, Dae-Su Kim, Kee-Wook Rim

*김포대학 컴퓨터계열

**한신대학교 컴퓨터학과

***신문대학교 지식정보산업공학과

요 약

일반적으로 Support Vector Machine은 이진 분류 모형에 있어 우수한 성능을 보이지만 모델의 한계로 인하여 다중 패턴의 분류 문제에는 쉽게 적용하기가 어렵다. 본 논문에서는 이진 분류를 포함한 다중 레이블을 갖는 데이터의 정확한 패턴 분류를 위하여 Zhu가 제안한 Import Vector Machine에 커널 Bagging 전략을 적용하여 분류의 정확성을 향상시키기 위한 Import Vector Voting 모형을 제안한다. 이러한 Import Vector Voting 모형은 다수의 커널함수를 적용한 결과 중에서 가장 성능이 우수한 커널함수를 이용하여 최종 분류를 수행하기 위한 voting 전략으로 사용한다. 본 논문에서 제안하는 Import Vector Voting 모형은 이진 분류를 포함한 3개 이상의 다중 패턴 데이터에 대한 분류 문제에 있어 매우 정확한 분류 성능을 보임을 실험을 통해 입증한다.

Abstract

In general, Support Vector Machine has a good performance in binary classification, but it has the limitation on multi-pattern classification. So, we proposed an Import Vector Voting model for two or more labels classification. This model applied kernel bagging strategy to Import Vector Machine by Zhu. The proposed model used a voting strategy which averaged optimal kernel function from many kernel functions. In experiments, not only binary but multi-pattern classification problems, our proposed Import Vector Voting model showed good performance for given machine learning data.

Key words : Kernel Bagging, Multi-pattern Classification, Import Vector Voting Model

1. 서 론

패턴 분류(pattern classification)를 수행하기 위한 기계학습(machine learning) 알고리즘들은 많이 연구되어 왔으며, 지금도 수많은 알고리즘들이 연구 개발 중에 있다. 이러한 알고리즘들 중에서 Vapnik이 제안한 Support Vector Machine(SVM)은 통계적 학습 이론(statistical learning theory)을 기반으로 한 매우 우수한 패턴 분류 알고리즘이다 [3]. 특히, SVM은 두 개의 패턴이 있는 이진 레이블의 목표 변수(target variable)를 갖는 데이터의 분류 작업에 있어서는 매우 좋은 성능을 보인다는 것이 많은 연구를 통해 확인되었다[5, 12].

그러나 SVM을 목표 변수의 레이블의 개수가 3개 이상인 다중 패턴 분류의 문제에 적용하기는 어렵다. Support Vector Machine의 이러한 문제점을 해결하기 위하여 Zhu는 3개 이상의 부류(class)를 갖는 다중 패턴(multi-pattern) 분

류를 위하여 IVM(Import Vector Machine)을 제안하였다 [15]. 이 모형은 SVM에서 다루기 어려운 다중 패턴 분류를 가능케 하는 것으로 IVM에서는 커널 로지스틱 기반의 함수를 이용하여 이를 해결한다[14, 15]. 하지만, IVM은 특정 커널함수 1개에만 의존하여 모형을 만들기 때문에 분류의 성능을 향상시키는 데는 한계가 있다. 따라서, 본 논문에서는 다중 패턴의 분류 문제에 있어 예측의 정확성을 향상시키는 전략으로 Import Vector Voting(IVV) 모델을 제안한다. 이 모델에서는 한 개의 커널함수 대신에 다수의 커널함수를 적용함으로써 가장 성능이 우수한 커널함수를 찾아내어 최종 분류를 수행하는 bagging 전략을 사용한다.

2. 관련 연구

일반적인 패턴 분류의 목적은 입력 벡터(input vector)와 목표 변수로 이루어진 데이터를 학습하여 최적의 분류기(classifier)를 만드는 것이다. 이를 통해 새로운 입력이 주어졌을 때 가장 적절한 클래스를 발견하는 작업은 기계 학습의 여러 응용 분야에서 필요로 하는 작업이다.

많은 분류 모형 중에서 SVM은 목표 변수가 두 개의 클래스를 갖는 이진 분류 문제에 있어서 매우 우수한 성능을 보

접수일자 : 2003년 1월 16일

완료일자 : 2003년 11월 26일

본 연구는 2003학년도 김포대학의 연구비 지원에 의해 연구되었음

인다[7]. 그림 1은 SVM의 기본적인 구조를 나타낸다.

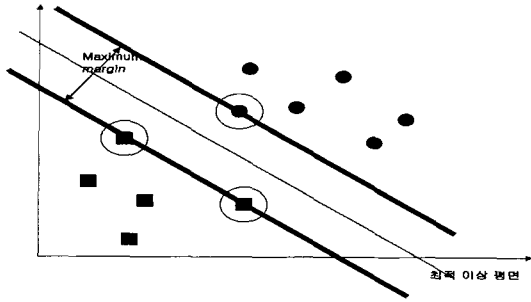


그림 1. Support Vector Machine의 기본 구조
Fig. 1. Basic Architecture of Support Vector Machine

그림 1은 두 개의 패턴을 갖는 데이터를 가장 잘 분류하는 최적 이상 평면(optimal hyper-plane)을 구하는 전략을 나타낸다. 이는 두 개의 각 클래스를 가장 잘 대표하는 Support Vector들을 대상으로 두 그룹을 동시에 가장 잘 분리할 수 있는 분류 평면을 구한다. 그림 1에서는 두 집단의 Support Vector를 지나고 두 개의 평면으로부터 동시에 가장 멀리 위치하는 평면을 결정하게 되는데, 이 평면을 최적 이상 평면이라고 정의한다. 새로운 데이터에 대한 패턴 분류는 이 최적 이상 평면이라는 분류기를 이용하게 된다. 그러나 SVM을 통한 다중 패턴의 분류 작업은 분류를 위한 많은 어려움을 내포하고 있으며, 이러한 문제점을 해결하려는 연구 또한 계속 진행 중에 있다.

SVM은 식 (1)과 같은 분류기 구조를 이용하여 이진 패턴을 분류한다. 여기서 목표 변수의 레이블은 $\{+1, -1\}$ 로 설정한다.

$$\text{Sign}[p(x) - \frac{1}{2}] \quad (1)$$

식 (1)에서 $\text{Sign}(g)$ 함수는 함수값이 임계값 이상이 되면 1로 임계값 이하이면 -1로 분류하는 역할을 수행한다. 여기서, $p(x)$ 는 주어진 입력 x 가 클래스 1로 분류될 확률이다. Zhu의 IVM을 이용한 다중 분류는 $p(x)$ 가 직접 다중 패턴의 분류에 이용된다[15]. SVM의 support vector와 유사하게 IVM은 학습 데이터의 fraction을 이용하는데, 이를 IVM에서는 import points라고 부른다.

기존의 연구에서 전체 학습 데이터 중에서 SVM의 모형 구축에 사용되는 support points 수보다 IVM의 import points의 수가 항상 적기 때문에, 학습 시간에 대한 비용 측면에서도 IVM이 SVM보다 더 효율적임을 알 수 있다.

임의의 커널함수 K 를 갖는 SVM은 식 (2)와 같이 표현할 수 있다.

$$\sum_{i=1}^n \hat{\alpha}_i K(x, x_i) \quad (2)$$

식 (2)에서, y_i 는 다중 클래스 레이블 중의 하나를 나타내는 것으로, $y_i = 1, 2, \dots, K$ 이다. 또한 $f = b + h$ ($h \in H_K, b \in R$)로서 b 는 편이(bias)를 나타내고 h 는 분류를 위한 초평면식의 모수(parameter)들을 나타낸다. 또한, H_K 는 커널함수 K 에 의해 생성된 함수들의 Reproducing Kernel Hilbert

Space(RKHS)이고, $\|f\|_{H_K}$ 는 모형의 roughness에 대한 패널티이다. 최종적인 모형은 라그랑지(Lagrange) 확장을 통하여 식 (3)과 같이 $K(x, x_i)$ 의 유한한 확장으로 표현된다[6, 9].

$$\sum_{i=1}^n \hat{\alpha}_i K(x, x_i) \quad (3)$$

식 (3)에서 $\hat{\alpha}_i$ 는 라그랑지 확장 계수(Lagrange multiplier)를 의미하며, SVM의 커널 전략으로 아래 식 (4)와 같은 KLR(Kernel Logistic Regression)을 사용하면 다중 패턴의 분류를 가능케 한다[8, 11].

$$\sum_{i=1}^n \hat{\alpha}_i K(x, x_i) \quad (4)$$

이러한 KLR의 사용이 바로 IVM의 기본 개념이다.

3. 커널 Bagging 기반의 IVM

3.1 커널 Bagging 예측

Bagging은 예측 모형을 위한 여러 개의 변형 형태를 가능하게 하고 이를 통하여 집합적인 예측 모형을 구축한다[2]. 구축된 예측 모형에서 집합적 예측 모형들의 평균은 새로운 데이터에 대한 분류를 위한 예측에 사용된다.

다수의 변형 모형을 위한 학습 데이터는 전체 데이터를 사용하지 않고 전체 데이터로부터 붓스트랩 반복(bootstrap replicates)의 대표본(resampling) 기법을 사용한다. 붓스트랩이란 통계적 추론을 위한 방법으로 분포에 대한 정보없이 주어진 데이터만을 사용하여 대표본 추출 방법을 통해 컴퓨터 모의실험을 수행하는 방법이다. 즉, 전체 학습 데이터로부터 일정량의 표본을 추출하고 이로부터 모형을 만들고 난 후에는 다시 이 표본을 전체 데이터에 포함시켜 또 다시 같은 크기의 표본을 추출하여 모형을 만드는 방법을 말한다. 이러한 작업을 n 번 반복한 후 최종적으로 n 개의 모형에 대한 평균 모수를 취하여 최적의 분류 식을 찾게 된다.

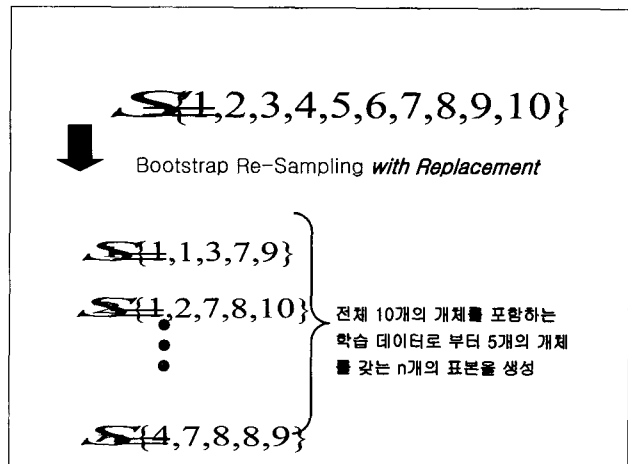


그림 2. 붓스트랩-대표본 과정
Fig. 2. Bootstrap-Resampling with Replacement

붓스트랩 추정량은 추정량 자체가 하나의 통계량 (statistic)으로 표현되는 것이 아니라 알고리즘의 형태로 표현된다. 붓스트랩은 자료만을 가지고 자료의 분포에 대한 특성을 찾아내는 방법으로 본 논문에서는 이러한 붓스트랩 전략 중에서 재표본 기법을 사용한다.

전체 집합 $S=\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ 에 대하여 재표본 기법을 적용하여 개체의 수가 5개인 n 개의 표본을 새롭게 만들면 그림 2와 같이 표현할 수 있다.

본 논문에서 제안하는 커널 Bagging 예측 모형은 일반적인 Bagging 전략에서 사용되는 집합적 예측 모형 대신에 커널함수를 사용하여 Zhu의 IVM에 적용한다.

그림 3은 본 논문에서 사용하는 커널 Bagging 예측 모형에 대한 구조를 나타낸다.

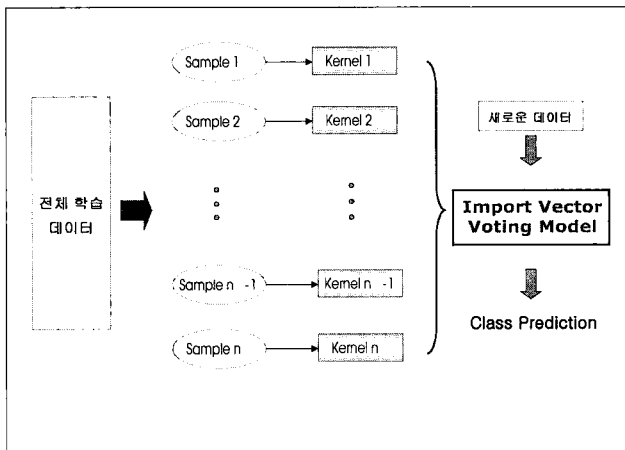


그림 3. 커널 Bagging 예측 모형
Fig. 3. Kernel Bagging Prediction Model

그림 3의 단계에서 우선적으로 전체 데이터로부터 반복이 있는 재표본 방법에 의해 n 개의 표본을 만든다. 이렇게 생성된 n 개의 표본에는 n 개의 커널함수를 적용하게 되고, 이때 사용되는 커널함수들은 모두 다중 패턴 분류를 위한 Import Vector의 결정(voting)을 위한 함수들을 사용한다. 최종적으로 n 개의 커널함수들의 평균을 이용하여 최적의 Import Vector를 결정하고 새로운 다중 패턴 데이터에 대한 분류를 수행하는데, 이를 위해 본 논문에서는 임의의 6개의 커널함수를 결정하여 사용한다. 커널함수의 개수와 종류에 따른 모형의 성능은 그 경우의 수가 매우 많기 때문에 실험의 고려 대상에서는 생략하였다.

3.2 Import Vector Voting 모형

IVM은 식 (2)에 KLR을 적용한 식 (5)로 정의된다.

$$\tilde{f}(x) = \frac{1}{n} \sum_{i=1}^n H(x; \mathbf{r}_i) \quad (5)$$

식 (5)는 SVM과 마찬가지로 라그랑지 확장을 통하여 모형의 최적 계수(parameter)를 구한다[12].


최적의 다중 패턴 분류를 위해 본 논문에서 제안하는 IVV 모델은 알고리즘 1에서와 같은 4단계의 과정을 수행하게 된다.

알고리즘 1. 제안하는 IVV 모델 알고리즘
Algorithm 1. Proposed IVV model algorithm

(step 1) Initialization

$S = \emptyset, R = \{x_1, x_2, \dots, x_N\}, k=1$, 여기서, S : import point들의 집합

(step 2) Lagrange Expansion

와 같은 분류 회귀 모형에 대해 식 (5)를 최소화 하는 모수를 구함

(step 3) Kernel Bagging

다음의 조건을 만족하는 x_r 를 구함

a1. $\arg \min_{x \in R} H(x_i)$, 여기서, $H(x_i)$ 은 식 (5)임

a2. $average[bagging_{k_i \in \{k_1, \dots, k_M\}} B_{k_i}(H(x_i))]$, 다수의 커널함수에 대한 bagging 전략을 수행한 후의 average를 이용하여 최종 모수를 결정 함

(step 4) Repeat

모수의 변화량이 0.001 이하의 조건이 만족될 때 까지 (step 2)와 (step 3)를 반복함

본 논문에서 제안하는 방법은 로지스틱 커널함수에 의존하는 Zhu의 IVM 모형에 다수의 커널함수를 적용하여, 이에 대한 각각의 학습을 통하여 최적의 모형을 결정하는 방법을 사용한다. 이는 표 1에 나타나 있는 6개의 커널함수와 같이 각각의 함수에 대하여 IVM 모형에 대한 학습을 수행한 후, 이들 6개의 모형을 결합하는 전략을 취한다. 이를 위해 본 논문에서는 모형 결합의 방법으로 6개의 모형들의 계수에 대한 평균을 구하는 방법을 선택한다[1].

표 1. 커널 bagging 함수
Table 1. Kernel bagging function

커널 ID	함수 이름
k1	Running-mean smoothers
k2	Running medians and enhancements
k3	Equivalent kernels
k4	Regression splines
k5	Cubic smoothing splines
k6	Locally-weighted running smoothers

표 1의 6개의 커널함수를 bagging 전략으로 사용하게 되면 Zhu의 IVM에 비해 6배의 학습 시간을 요구하는 문제가 발생한다. 이러한 문제를 해결하기 위해 본 논문에서는 붓스트랩의 재표본 전략을 통하여 전체 학습 데이터를 모두 사용하지 않고 추출된 표본만을 이용한다. 표 1의 6개의 커널함수들은 [1]에 의한 각 수식을 적용하였다.

4. 실험 및 결과

본 논문에서 제안하는 IVV 모형과 기존의 분류 알고리즘들과의 성능 평가를 위한 다중 패턴의 실험 데이터로 German의 Glass Identification 데이터와 Fisher의 Iris 데이터를 이용하였다. 그리고 "Data Archive of Journal of Statistical Education"의 Fish Species 데이터를 이용하여 Kernel Bagging에 이용되는 커널함수의 개수에 따른 오분류율의 결과를 비교하였다.

실험을 수행하기 위한 Class Identification과 Iris 데이터는 UCI, Machine Learning Repository[16]로부터 구하였고, Fish Species 데이터는 SAS(Strategic Application Software)의 데이터 예제에 있는 것을 이용하였다[17].

4.1 Glass Identification 데이터를 이용한 성능 평가

Glass Identification 데이터는 총 214개의 인스턴스로 구성되어 있다. 입력 변수는 굴절률 변수와 성분을 나타내는 8개의 변수들(나트륨, 마그네슘, 알루미늄, 실리콘, 칼륨, 칼슘, 바륨, 철)로서 총 9개의 변수로 구성되어 있으며, 각 클래스를 나타내는 1개 목표 변수는 glass의 형태를 나타내는 변수로서 6개의 레이블을 갖는다. 표 2는 Glass identification 데이터에 대한 요약 정보를 나타낸다.

표 2. Glass Identification 데이터 요약
Table 2. Summary of Glass Identification data

변수명	Min	Max	Mean	S.D.
굴절률	1.51	1.53	1.52	0.003
나트륨	10.73	17.38	13.41	0.817
마그네슘	0	4.49	2.68	1.442
알루미늄	0.29	3.5	1.44	0.499
실리콘	69.81	75.41	72.65	0.775
칼륨	0	6.21	0.50	0.652
칼슘	5.43	16.19	8.96	1.423
바륨	0	3.15	0.18	0.492
철	0	0.51	0.06	0.097

본 논문에서 제안하는 알고리즘의 성능을 비교 평가하기 위하여 기계 학습 알고리즘의 대표적인 분류 모형으로 사용되는 의사결정 나무(decision tree) 모형과, 대표적인 통계 분류 모형인 로지스틱 판별(logistic discriminant) 모형을 사용하였다.

표 3은 제안 알고리즘을 포함한 4개의 분류 모형에 대한 Glass identification 데이터에 대한 분류 실험을 수행한 결과이다. 여기서, 성능 평가의 척도로는 분류모형에서 가장 많이 사용되고 있는 오분류율(misclassification ratio)을 사용하였다[10].

표 3의 결과로부터 본 논문에서 제안하는 방법에 의해 얻어진 분류 결과에 대한 오분류율이 가장 낮음을 알 수 있다. 이를 통해 제안하는 알고리즘이 통계적 분류 모형에 비해 약 2배의 성능 향상을 보이고 있음을 알 수 있다.

표 3. Glass 데이터 성능 평가 결과
Table 3. Performance evaluation of Glass Data

분류 모형	오분류율(%)
Decision tree	3.7
Logistic discriminant	4.3
IVM	2.8
Kernel bagging IVM	1.9

4.2 Iris 데이터를 이용한 성능 평가

Iris 데이터는 총 150개의 데이터로 구성되어 있다. 이는 붓꽃의 외형을 결정하는 4개의 입력 변수들과 3개의 꽃 종류를 결정하는 레이블을 갖는 1개의 목표변수로 이루어져 있다. 이 3개의 레이블에 해당하는 데이터는 모두 50개씩으로 구성되어 있다.

Iris 데이터를 이용한 실험도 Glass Identification을 이용한 실험과 마찬가지로 비교를 위한 3가지의 분류 기법과의 성능 평가 비교를 수행하였다. 표 4는 이러한 결과를 나타낸다.

표 4. Iris 데이터 성능 평가 결과
Table 4. Performance evaluation result of Iris data

분류 모형	오분류율(%)
Decision tree	2.4
Logistic discriminant	3.3
IVM	1.9
Kernel bagging IVM	1.1

표 4의 Iris 데이터를 이용한 실험도 Glass Identification의 경우와 같이 제안하는 IVV 모형이 가장 낮은 오분류율을 보이면서 가장 우수한 성능을 보이고 있음을 알 수 있다. 두 개의 실험 데이터를 통하여 비교된 4개의 다중 패턴의 분류 모형 결과를 살펴보면 본 논문에서 제안한 IVV 모형의 성능이 가장 우수함을 알 수 있다. 다음으로는 IVM 모델, 의사결정 나무, 그리고 로지스틱 판별 모형의 순으로 모형의 정확도가 이루어짐을 알 수 있다.

4.3 Fish Species 데이터를 이용한 성능 평가

Fish Species의 전체 데이터는 총 159개로서 물고기의 외형과 무게를 나타내는 6개의 분류 변수와 7가지 물고기의 종류(Bream, Roach, Whitefish, Parkki, Perch, Pike, Smelt)를 표시하는 1개의 목표 변수로 이루어져 있다.

우선, 표 1의 6개의 커널함수를 순서대로 이용하여 커널 bagging에 사용되는 커널함수의 수에 따른 오분류율의 변화를 계산하였다. 처음에는 커널 ID, k1과 k2의 두 개의 커널 함수만을 사용하여 모형을 만들고 이를 통하여 오분류율을 계산하였으며, 다음으로는 k1, k2 그리고 k3의 3개의 커널 함수를 사용하여 모형의 오분류율을 계산하였다. 이러한 방법으로 최종적으로 6개의 커널함수 모두를 사용한 모형에 대한 오분류율을 계산하게 된다. 표 5는 커널함수의 개수의 변화에 따른 오분류율의 결과를 나타낸다.

표 5. 커널함수의 개수에 따른 Fish Species 데이터의 오분류율

Table 5. Misclassification rate of Fish Species data as the number of Kernel function

커널함수의 개수	Kernel ID	오분류율(%)
2	k1,k2	1.84
3	k1,k2,k3	1.48
4	k1,k2,k3,k4	1.30
5	k1,k2,k3,k4,k5	1.27
6	k1,k2,k3,k4,k5,k6	1.12

표 5의 결과에 의하면 Kernel Bagging에 사용되는 커널함수의 개수가 많아질수록 최종 분류 모형의 오분류율이 낮아짐을 알 수 있다. 다음으로는 표 1의 커널함수의 커널 ID의 순서를 변경하여 오분류율을 계산하였는데, 표 6은 이러한 커널함수의 순서를 변경한 결과를 나타낸다.

표 6. 서로 다른 커널함수의 개수 조합

Table 6. Different combination of the number of Kernel function

커널함수의 개수	Case #1	Case #2
2	k6, k5	k3, k5
3	k6, k5, k4	k3, k5, k1
4	k6, k5, k4, k3	k3, k5, k1, k6
5	k6, k5, k4, k3, k2	k3, k5, k1, k6, k2

그림 4는 표 5의 오분류율 결과와 표 6의 case #1과 case #2의 커널 조합을 이용한 모형의 오분류율 결과를 나타낸다.

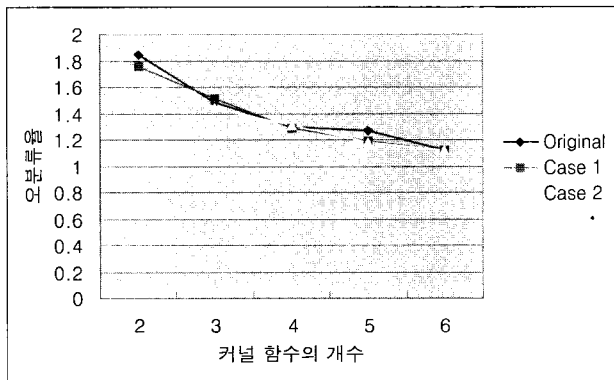


그림 4. 커널함수의 개수 조합에 따른 오분류율의 변화

Fig. 4. Misclassification rate according to combination of kernel function

그림 4의 결과를 보면 커널 bagging에 사용된 커널함수의 개수가 증가할수록 오분류율이 낮아짐을 알 수 있다. 따라서 커널함수의 개수를 증가시키면 분류 모형의 정확도는 더욱 높아진다고 볼 수 있다. 하지만 이때 학습 시간이 더 많이 소요되는 문제는 여전히 남아 있으며 몇 개의 커널함수가 가장 적절한지에 대한 결정이 필요하게 된다.

5. 결론 및 향후 연구과제

본 논문에서는 Vapnik의 SVM 모형이 가지고 있는 다중 패턴 분류의 어려움을 해결하기 위해 Zhu가 제안한 IVM 모델에 커널 bagging 기법을 적용한 IVV 모형을 제안하여 분류 모형의 정확도를 향상시켰다. 또한, bagging을 위한 학습 시간의 증가 문제를 표본 추출을 통하여 해결하였는데, 이는 전체 데이터에 대한 모형을 별도로 구축하지 않고 추출된 표본만을 모형 학습에 적용하는 방법을 사용하였다.

실험에서는 기존의 기계 학습 데이터로 많이 사용되는 Glass Identification 데이터와 Iris 데이터를 사용하여 본 논문에서 제안하는 분류 모형이 기존의 대표적인 분류 모형에 비해 우수한 성능을 보임을 확인하였다. 그리고 Fish Species 데이터를 이용한 커널 bagging에 사용되는 커널함수의 개수에 따른 모형의 정확도에 대한 실험을 통해 커널함수의 개수를 증가시키면 오분류율이 낮아짐을 알 수 있었다.

향후 연구과제로는 SVM의 기본 개념을 기반으로 한 다양한 스플라인 함수의 적용과 베이지안 SVM의 다중 패턴 분류 모형으로의 변환 연구를 통하여 분류 능력이 더욱 우수한 다중 패턴의 분류 모형에 대한 연구와 본 논문에서 제안한 IVV 모형의 커널 bagging에 사용되는 최적의 커널함수의 개수에 대한 결정 방안에 대한 연구도 함께 진행할 예정이다.

참고 문헌

- [1] T. Hastie, R. Tibshirani, Generalized Additive Models, Chapman and Hall, 1990.
- [2] L. Breiman, "Bagging predictors," Mach. Learn. 24(2), 123-140, 1996.
- [3] C. Burges, "A tutorial on support vector machines for pattern recognition," In Data Mining and Knowledge Discovery, 1998.
- [4] A. C. Davison, D. V. Hinkley, "Bootstrap methods and their application," Cambridge University Press, 1998.
- [5] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, MIT Press, 1999.
- [6] P. Green, B. Yandell, "Semi-parametric generalized linear models," Proceedings of 2nd International GLIM Conference, 1985.
- [7] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer, 2001.
- [8] G. Kimeldorf, G. Wahba, Some results on Tchebycheffian spline functions, Math. Anal. Applic. 1971.
- [9] X. Lin, G. Wahba, D. Xiang, F. Gao, R. Klein, B. Klein, "Smoothing spline ANOVA models for large data sets with Bernoulli observations and the randomized GACV", Technical Report 998, Department of Statistics, University of Wisconsin, Madison, 1998.
- [10] B. D. Marx, P. H. C. Eilers, "Direct generalized additive modeling with penalized likelihood", Computational Statistics and Data Analysis 28(2),

pp. 193-209, 1998.

- [11] A. Smola, B. Scholkopf, "Sparse Greedy Matrix Approximation for Machine Learning," In Proceedings of the Seventeenth International Conference on Machine Learning, 2000.
- [12] G. Wahba, "Support Vector Machine, Reproducing Kernel Hilbert Spaces and the Randomized," GACV. Technical Report 984, Department of Statistics, University of Wisconsin, Madison, 1998.
- [13] G. Wahba, C. Gu, Y. Wang, R. Chappell, Soft Classification, a.k.a. Risk Estimation, via Penalized Log Likelihood and Smoothing Spline Analysis of Variance, The Mathematics of Generalization, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publisher, 1995.
- [14] C. Williams, M. Seeger, "Using the Nystrom Method to Speed Up Kernel Machines," Advances in Neural Information Processing Systems 13, MIT Press, 2001.
- [15] J. Zhu, T. Hastie, "Kernel Logistic Regression and the Import Vector Machine," NIPS2001 Conference, 2001.
- [16] <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [17] <http://www.sas.com>

저 자 소 개



최준혁(Choi, JunHyeog)
 1990년 : 경기대학교 전자계산학과 졸업(이학사)
 1995년 : 인하대학교 대학원 전자계산공학과 졸업(공학석사)
 2000년 : 인하대학교 대학원 전자계산공학과 졸업(공학박사)
 1997년 3월 ~ 현재 김포대학 컴퓨터계열 조교수

관심분야 : 정보검색, 데이터마이닝, 신경망, 유전자 알고리즘, 유비쿼터스 컴퓨팅 등

e-mail: jhchoi@kimpo.ac.kr



김대수(Kim, DaeSu)

1977년 : 서울대학교 사대수학과 학사
 1986년 : 미국 Univ. of Mississippi, Computer Science, M. S.
 1990년 : 미국 Univ. of South Carolina, Computer Science, Ph. D.
 1991년 - 1993년 : 한국전자통신연구원 컴퓨터연구단 선임연구원
 1993년 - 현재 한신대학교 컴퓨터학과 교수

관심분야 : 신경망, 퍼지, 인공지능, 지능시스템, 에이전트, 융합 모델링

Phone : +82-031-370-6784
 Fax : +82-031-372-3343
 E-mail : daekim@hanshin.ac.kr



임기욱(Rim, KeeWook)

1977년 2월 : 인하대학교 공과대학 전자공학과 졸업
 1987년 2월 : 한양대학교 전자계산학 석사
 1994년 8월 : 인하대학교 전자계산학 박사
 1977년 2월 ~ 1983년 2월 : 한국전자기술연구소 선임연구원
 1983년 3월 ~ 1988년 7월 : 한국전자통신연구소 시스템소프트웨어 연구실장

1988년 8월 ~ 1989년 8월 : 미 캘리포니아 주립대학(Irvine) 방문연구원
 1989년 10월 ~ 1996년 12월 : 한국전자통신연구원 시스템연구부장, 주전산기(타이킵) III, IV개발 사업책임자
 1997년 1월 ~ 1999년 12월 : 정보통신연구진흥원 정보기술전문위원
 2001년 7월 ~ 2003년 2월 : 한국전자통신연구원 컴퓨터소프트웨어 연구소장
 2000년 3월 ~ 현재 : 선문대학교 지식정보산업공학과 교수

관심분야 : 실시간 데이터베이스시스템, 운영체제, 시스템구조

e-mail: rim@omega.sunmoon.ac.kr