

데이터 마이닝에서 패턴 분류를 위한 다중 SVM 분류기

Multiple SVM Classifier for Pattern Classification in Data Mining

김만선*, 이상용**

Man-Sun Kim* and Sang-Yong Lee**

* 공주대학교 컴퓨터공학과

** 공주대학교 정보통신공학부 교수(교신저자)

요 약

패턴 분류는 실세계의 객체를 표현한 다양한 형태의 패턴 정보를 추출하여, 이것이 어떤 부류(클래스)인가를 결정하는 것이다. 패턴 분류 기술은 데이터 마이닝, 산업 자동화나 업무자동화를 위한 컴퓨터 응용 소프트웨어 기술로서 현재 다양한 분야에서 활용되고 있다. 패턴 분류 기술의 최대 목표는 분류 성능 향상이며 이것을 위해 지난 40년간 많은 연구자들이 다양한 접근 방법들을 시도해 왔다. 주로 이용되는 단일 분류 방법들로는 패턴들의 확률적 추론에 기반한 베이스 분류기, 결정 트리, 거리함수를 이용하는 방법, 신경망, 군집화 등이 있으나 대용량 다차원 데이터를 분석하기에는 효율적이지 못하다. 따라서 상호 보완적인 여러 분류기들을 사용해 결합을 통하여 성능 향상에 도움을 주고 있는 다중 분류기 시스템에 대한 연구가 활발하게 진행되고 있다.

본 논문에서는 다중 SVM(Support Vector Machine) 분류기에 관한 기존 연구의 문제점을 지적하고 새로운 모델을 제안한다. SVM을 다중 클래스 분류기로 확장하기 위해 일대다 정책을 기반으로 하여 각각의 SVM 출력값을 비선형 패턴을 갖는 신호로 간주하고 이를 신경망에 학습하여 최종 분류 성능 결과를 결합하는 모델인 BORSE(BOotstrap Resampling SVM by Ensemble)를 제안한다.

Abstract

Pattern classification extracts various types of pattern information expressing objects in the real world and decides their class. The top priority of pattern classification technologies is to improve the performance of classification and, for this, many researches have tried various approaches for the last 40 years. Classification methods used in pattern classification include base classifier based on the probabilistic inference of patterns, decision tree, method based on distance function, neural network and clustering but they are not efficient in analyzing a large amount of multi-dimensional data. Thus, there are active researches on multiple classifier systems, which improve the performance of classification by combining problems using a number of mutually compensatory classifiers.

The present study identifies problems in previous researches on multiple SVM classifiers, and proposes BORSE, a model that, based on 1:M policy in order to expand SVM to a multiple class classifier, regards each SVM output as a signal with non-linear pattern, trains the neural network for the pattern and combine the final results of classification performance.

Key words : 다중 SVM 분류기, 데이터마이닝, 배깅, 패턴 분류, MLP

1. 서 론

패턴 분류 기술의 최대 목표는 분류 성능 향상이며 이것을 위해 지난 40년간 많은 연구자들이 다양한 접근 방법들을 시도해 왔다. 최근 패턴 분류를 적용하는 데이터들의 특징으로는 10,000건 이상의 대용량 데이터나 수십 개 이상의 속성으로 구성된 다차원 데이터로 구성되어 있다. 패턴 분류를 위한 단일 분류기 시스템은 전체 학습 데이터로부터 얻은 하나의 결정규칙만을 이용하기 때문에 복잡한 실세계를 표현한 다양한 형태의 측정값을 하나의 분류기만으로는 효과적으

로 분류하기 어렵다. 주로 이용되는 단일 분류 방법들로는 패턴들의 확률적 추론에 기반한 베이스 분류기, 결정 트리, 거리함수를 이용하는 방법, 신경망, 군집화 등이 있으나 대용량 다차원 데이터를 분석하기에는 효율적이지 못하다. 따라서 상호 보완적인 여러 분류기들을 사용하여 결합을 통하여 성능 향상에 도움을 주고 있는 다중 분류기 시스템에 대한 연구가 활발하게 진행되고 있다.

본 논문에서는 다중 SVM(Support Vector Machine) 분류기에 관한 기존 연구의 문제점을 지적하고 새로운 모델을 제안한다.

접수일자 : 2005년 2월 7일
완료일자 : 2005년 5월 18일

2. 관련 연구

2.1 패턴 분류에 사용된 기존의 방법

2.1.1 신경망 구축 모델

Jordan[1]에 의해 소개되어진 Mixture-of-experts는 다수의 전문가 신경망을 구축하고, 이를 경험적 위험을 최소화하거나 교차검증과 유사한 리샘플링 방법을 사용하여 결합하였다. 두 단계를 거치는데, 첫 번째 방법에서는 학습 데이터를 여러 후보 모델들을 생성하는데 사용하고 이들의 출력으로부터 가중치 평균을 계산하여 결합 결과를 얻는다. 두 번째 방법에서는 모델들의 결합을 위해 리샘플링을 수행하게 된다. 각각의 근사 함수를 구하는데 사용된 데이터 샘플들이 결합 계수를 평가하는데 사용되지 않도록 리샘플링이 이루어진다.

2.1.2 RBF와 의사결정 트리의 결합 모델

Guttag[2]는 FERET 얼굴 이미지 데이터로 성별, 인종 및 얼굴 포즈 인식을 위하여 RBF와 의사결정 트리의 결합 모델을 제안하였다. 그는 원본 이미지, 가우시안 잡음을 섞은 이미지, 5°회전을 가한 이미지를 RBF의 입력으로 사용하여 60개의 테스트 데이터에 93.3%의 성별 분류율을 얻었다. 이는 얼굴 인식을 위하여 여러 가지 특징과 분류기(RBF 및 의사결정 트리)들을 결합했다는 점에서 기존 연구와 구분된다.

2.1.3 특징-분류기 쌍을 갖는 앙상블 모델

Park[3]은 유전발현 데이터를 통하여 질병을 분류하기 위하여 유전 알고리즘을 이용한 유전자발현 데이터상의 특징-분류기쌍 최적 앙상블 모델을 탐색하는 연구를 하였다.

유전발현 데이터는 생명체의 특정 조직에서 채취한 샘플을 microarray 상에서 측정된 것으로, 유전자들의 발현 정도를 수치로 나타낸 데이터이다. 일반적으로 정상조직과 이상조직에서 관련 유전자들의 발현정도는 차이를 보이기 때문에, 유전발현 데이터를 통하여 질병을 분류할 수 있다. 이러한 분류에 모든 유전자들이 관여하지는 않으므로 관련 유전자를 선별하는 작업인 특징선택이 필요하며, 선택된 유전자들을 적절히 분류하는 방법이 필요하다. 상관계수, 유사도, 정보이론 등에 기반을 둔 7가지 특징 선택 방법과 대표적인 6가지 분류기에 대하여 특징-분류기 쌍의 최적 앙상블을 탐색하기 위하여 유전자 알고리즘 기반의 방법을 제안하였다.

2.1.4 동적 분류기 모델

Song[4]은 입력 패턴의 지역적 특성을 살려서 동적으로 분류기를 선택할 수 있는 새로운 다중 분류기 시스템을 제안하였다. 기존에 연구된 지역적 학습방법의 문제점은 분류시간에 학습이 이루어지기 때문에 시간이 많이 걸린다는 단점이 있었다. 따라서 시간을 단축하기 위하여 학습 단계에서 미리 지역적인 분류기들을 설계하였다

학습단계에서는 마이크로 인식기를 설계하는 부분으로 모든 학습패턴에 대하여 여러 개의 마이크로 인식기를 생성하게 된다. 마이크로 인식기는 각 학습 패턴을 둘러싼 주변의 학습패턴들을 이용하여 선형 SVM과 RBF 커널함수를 이용한 비선형 SVM 학습방법을 사용하여 결정 함수를 유도한다.

인식단계에서는 테스트패턴과 가까운 학습패턴들을 이용하여 지역적 영역을 정의하고, 이 영역에 속하는 각 학습패턴들이 가지는 마이크로 인식기들의 성능을 평가하여 가장 좋은 인식기를 선택한다. 동적으로 선택된 마이크로 인식기

를 이용하여 테스트패턴의 최종 클래스 레이블을 결정하게 된다.

2.1.5 다중 SVM 분류기 모델

Scholkopf[6]은 일대다 정책의 다중 SVM 분류기를 제안하여, 모음 음성 분류 문제에 적용하여 54.9%~79.6%의 분류율을 보였다[5]. 그리고 Clarkson은 더 높은 분류율을 얻기 위해 일대다 정책을 개선하여 일대일 정책을 제안하여 다중 SVM 분류기를 사용하였다. 일대다 정책 방법보다 일대일 정책 방법이 4~6% 정도 분류율이 향상되었다. 알고리즘의 효율성과 최적화 문제를 해결하지 못했다[8].

2.2 기존 다중 SVM 분류 방법의 문제점

다중 SVM 분류기는 단일 SVM 분류기를 사용했을 때 보다 더 안정적이고 좋은 성능을 기대할 수 있으나, 다음과 같이 문제점이 있다.

① 대용량의 분류 문제에서 전체 입력 데이터를 그대로 사용한다면 다중 클래스로 확장함에 있어서 학습 시간이 많이 걸린다. 따라서 전체 입력 데이터 세트를 추출하여 최소한의 입력 데이터만을 사용하여야 한다.

② 다중 클래스를 확장할 경우 다른 클래스 간에 선형 분리가 곤란하거나(one-against-all 방법), 클래스가 중복되는 문제(one-against-one 방법)가 발생한다. 예를 들어 3개의 클래스로 나뉘어질 때, one-against-all 방법의 경우 이들 중 하나의 클래스(클래스-1)는 다른 클래스(클래스-2, 클래스-3)와 선형적으로 분리가 가능하지만, 다른 두 클래스(클래스-2, 클래스-3)는 서로 선형적으로 분리가 불가능하다. 그래서 이러한 문제는 one-against-one 방법으로 해결하였으나, 이 경우 $n(n-1)$ 개 만큼의 SVM을 구성해야 하므로 분류기의 개수가 많아 많아지는 현상이 일어난다.

③ 현재 사용되고 있는 다중 SVM의 최종 결합 방법은 문제점을 가지고 있다. 최종 결과를 통합하는 방법은 추상 레벨(abstract level), 순위 레벨(rank level), 측정치 레벨(measurement level)로 나누어진다[7]. 현재 주로 사용되는 추상 레벨의 다수결 결합 방법은 각 SVM의 성능이 모두 우수할 경우에만 좋은 결과를 도출할 수 있다. 그리고 2단계 계층적 SVM으로 구성하는 최종 결합 방법도 제안되었으나, 다수 SVM들의 출력을 통합하기 위해 상위 레벨에 또 다른 하나의 SVM을 추가한 복잡한 구성으로 인하여 최종 결과의 신뢰도를 보장할 수 없게 된다.

3. BORSE

본 논문에서는 기존 다중 SVM 분류 방법의 문제점을 해결하기 위하여 효율적인 입력세트 재구성, 간단한 구성의 일대다 정책 기반의 다중 SVM, 신뢰성 있는 최종 결합 방법의 특징을 가진 BORSE를 제안한다.

3.1 BORSE의 특징

본 논문에서는 SVM을 다중 클래스 분류기로 확장하기 위해 일대다 정책을 기반으로 하여 신경망을 결합하는 방법을 제안한다. 각각의 SVM 출력값을 비선형 패턴을 갖는 신호로 간주하고 이를 신경망에 학습하여 최종 분류성능을 향상시키고자 한다.

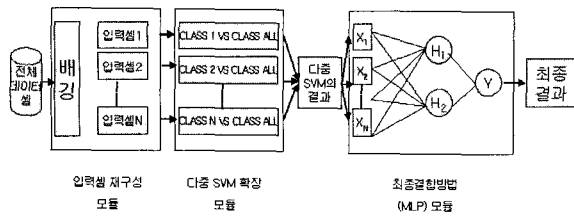


그림 1. BORSE의 개념도

위 그림 1. 에서 볼 수 있듯이 제안하는 모델은 크게 세 가지 특징을 갖는다.

첫 번째 특징으로는 대용량 데이터에 적용 가능하도록 배깅 알고리즘을 사용해 최소한의 입력으로 데이터 세트를 재구성한다. 배깅은 학습자료로부터 표본추출을 위한 분포의 가정 없이 동일한 방법으로 추출된 여러 개의 붓스트랩 리샘플링을 통하여 분류기를 생성하고 그 결과를 Voting Method를 통해 최종적인 하나의 분류결과를 얻어내는 방법으로 분산을 줄여 분류 예측률을 높여준다. 본 연구에서는 대용량 다차원 데이터를 분류하기 위하여 표본추출을 하는 과정에서만 배깅기법을 적용한다.

두 번째 특징인 다중 SVM 확장에서는 SVM의 생성개수는 클래스 분류 개수와 동일하다. 즉, 10개의 분류 클래스를 다루는 문제는 10개의 SVM이 필요하다. 일대다 방식을 기반으로 하기 때문에 일대일 방식보다 간단한 다중 분류기의 구성을 갖는다. 본 논문에서 사용한 데이터는 3, 6, 18 개의 클래스 분류 개수로 분리된 데이터로 3, 6, 18개의 SVM을 구성한다.

세 번째 특징인 최종 결합 방법에서는 일대다 정책을 기반으로 클래스를 +1, -1값을 얻는 것이 아니라 실수값을 얻어서 각 분류기의 결과를 신경망의 입력으로 제공한다. 분류기의 의존도에서 벗어나 최종 결과의 신뢰도를 보장할 수 있다. 기존의 방법에서는 올바른 클래스의 SVM이 조금이라도 낮은 값을 출력하면 잘못 분류하는 경향이 있었다. 즉 SVM의 성능에 매우 의존적인 면을 탈피할 수 있다.

3.2 입력세트의 재구성

배깅(bagging: bootstrap aggregating)은 Breiman(1996)에 의해 처음 소개된 알고리즘으로 여러 개의 붓스트랩 리샘플로부터 얻어진 분류기의 결과를 결합하여 하나의 분류기를 얻어내는 기법이다.

배깅에서는 원래 자료에서 m개의 관찰치를 갖는 학습자료 T와 검증용자료 Test를 설정한 뒤, T로부터 붓스트랩 리샘플로 m개의 관찰치의 복원추출을 k개만큼 실행한다. 이때 추출된 붓스트랩 리샘플은 $T^{(1)}, T^{(2)}, \dots, T^{(k)}$ 이며 이에 의해 각각의 분류기 C_k 가 학습되어지고 이에 의해 C_1, C_2, \dots, C_k 가 생성된다. 결국 각각의 C_k 의 결과를 수집하여 앙상블하는 방법에 의해 최종 분류기인 C^* 가 생성된다.

3.3 다중 SVM 확장

주어진 N클래스를 두 부분으로 분리할 하는데 한쪽에는 클래스 i만 존재하고 나머지 부분에는 i를 제외한 나머지 클래스들을 할당하는 형태로 문제를 재 정의한다.

이와 같은 방법으로 만들 수 있는 짝의 수는 N가지의 경우가 존재한다.

각각의 짝에 대해 SVM을 이용하여 학습을 시키면 각 문제로부터 N개의 함수를 얻게 된다.

$$f1(x) = \sum_{i=1}^L a_i y_{i1} k(x_i, x) + b_1$$

$$f2(x) = \sum_{i=1}^L a_i y_{i2} k(x_i, x) + b_2$$

$$fN(x) = \sum_{i=1}^L a_i y_{iN} k(x_i, x) + b_N$$

이것이 SVM의 결정 함수(decision function)가 된다.

본 논문에서는 SVM을 다중 클래스 분류기로 확장하기 위해 간단한 구성을 갖는 일대다 정책을 기반으로 하여 학습을 하고, 신뢰성을 보장하기 위한 방법으로 최종 결함을 결정하는 방법을 제안한다.

3.4 최종 결합 방법

일반적으로 여러 분류기의 결과를 결합하는 방법은 결합에 사용되는 정보의 정도에 따라 다음 표 1. 과 같이 3가지 레벨로 나눌 수 있다[8].

표 1. 여러 분류기의 결과를 결합하는 방법

레벨	이용하는 정보(결합방법)
추상 레벨	분류기가 출력하는 하나의 분류 결과 (다수결, BKS, 베이시안)
순위 레벨	분류기의 모든 클래스별 분류 순위 (Borda, Condorect 함수)
측정치 레벨	분류기의 모든 클래스에 대한 분류신뢰도 (신경망, 신뢰값합산)

기존의 일대다 방식은 각 SVM 출력값의 서열로서 클래스를 구분하는 단순 분류기(simple classifier)이다. $F(x)$ 의 출력값 중 최고 높은 값은 단순히 선택해 주는 것이기 때문에, 올바른 클래스의 SVM이 조금이라도 낮은 값을 출력하면 잘못된 분류를 하게 된다. 즉, 분류기의 성능을 해당 클래스를 구분하는 개별 SVM 분류기의 성능에 매우 의존적이게 된다. 따라서 측정치 레벨의 신경망 방법을 적용하여 잘못된 분류를 최소화한다.

신경망의 MLP(Multilayer Perception)는 감독학습(Supervised Learning)으로 입력값과 출력값이 제시되므로 입력층에 주어진 입력값이 중간층을 거쳐 출력층으로 나온 출력값과 이미 제시된 기대값을 쉽게 비교할 수 있다. 이 때 출력값과 기대값이 일치하는 경우에는 학습이 일어나지 않으며, 일치하지 않는 경우에는 얻어진 출력값과 기대값의 차이를 줄이는 방향으로 연결강도를 조정하여 학습된다.

4. 실험 결과

제안된 방법의 현실문제에의 적용을 위해 기계학습에 적용되는 벤치마크 데이터를 이용한다. UCI Machine Learning Repository DB에는 실세계의 데이터 103종류(ABALONE~ZOO)가 있는데 샘플개수(instance), 부류개수(features)가 각기 다른 다양한 형태로 제공된다. 그 중에서 대용량(150~60,000개 이상의 instance)이며, 다차원(3개 이

상의 features)을 갖는 데이터를 선택하여 실험데이터로 사용하였다[9].

4.1 실험 데이터

아래의 표 2. 는 다섯 가지의 UCI 데이터 집합의 샘플 개수, 속성 개수, 부류 개수를 정리한 것이다.

표 2. 실험에 사용된 5개의 UCI 데이터 집합

실험 데이터	샘플 개수	속성 개수	부류 개수
Iris Plants DB	150	4	3
Glass Glass Identification DB	214	10	7
Letter Recognition DB	20,000	16	26
Chess DB	28,056	6	18
Connect-4 Opening DB	67,557	42	3

4.2 평가 기준

정확도(accuracy) 측정은 검증자료에서 나온 결과를 훈련 자료와 비교하여 일치하는 데이터의 수 c를 전체 비교 데이터의 수 N으로 나누어서 계산하였다.

$$accuracy = \frac{c}{N} \times 100\% = \frac{\text{분류한데이터의개수}}{\text{총데이터의개수}} * 100$$

배깅에 의하여 입력세트를 4단계로 단계별로 증가시켜 나눴다. Iris Plants DB의 20%, 40%, 60%, 80%에 해당하는 데이터의 수는 30, 60, 90, 120개, Glass Glass Identification DB의 20%, 40%, 60%, 80%에 해당하는 데이터의 수는 43, 86, 128, 172개, Letter Recognition DB의 20%, 40%, 60%, 80%에 해당하는 데이터의 수는 4,000, 8,000, 12,000, 16,000개, Chess DB의 20%, 40%, 60%, 80%에 해당하는 데이터의 수는 5,611, 11,222, 16,833, 22,444개, Connect-4 Opening DB의 20%, 40%, 60%, 80%에 해당하는 데이터의 수는 13,511, 27,022, 40,534, 54,045개이다.

4.3 결과 분석

4.3.1 입력세트 추출량에 따른 분류 결과

아래의 표 3, 표 4, 표 5, 표 6, 표 7. 을 보면 BORSE로 실험한 정확도가 타 다중 분류기인 MLP나 KNN의 정확도보다 향상된 정확도가 나왔으며, 배깅을 적용하여 원 데이터 세트의 20%, 40%, 60%, 80% 추출된 입력세트에 따른 분류 정확도를 보여준다.

배깅에 의한 입력세트의 추출을 20%, 40%, 60%, 80 %로 점차 증가시켜서 분류해본 결과 전반적으로 90% 이상의 분류율을 얻었으나, 60%와 80% 추출된 경우에서 거의 비슷한 분류율을 보였다. 학습 시간을 줄이면서 전체 입력세트를 사용하지 않아도 높은 결과를 얻을 수 있었다.

표 3. Iris DB 분류 결과

	BORSE	MLP	KNN
20%	93.33	86.66	80.00
40%	96.66	90.00	88.33
60%	96.66	94.44	91.11
80%	97.50	90.00	88.33

표 4. Glass Identification DB 분류 결과

	BORSE	MLP	KNN
20%	93.02	88.37	90.69
40%	97.67	96.51	95.34
60%	98.43	96.87	97.65
80%	98.83	84.30	85.46

표 5. Letter Image DB분류 결과

	BORSE	MLP	KNN
20%	92.50	90.00	80.00
40%	95.87	92.50	93.12
60%	98.05	84.26	83.39
80%	98.54	89.37	75.00

표 6. Chess DB 분류 결과

	BORSE	MLP	KNN
20%	89.11	90.31	70.12
40%	94.43	94.01	75.31
60%	94.52	87.02	79.12
80%	94.55	94.52	78.03

표 7. Connect-4 Opening DB 분류 결과

	BORSE	MLP	KNN
20%	88.12	67.31	85.12
40%	90.31	71.32	87.00
60%	91.12	73.02	88.12
80%	91.07	57.03	87.13

4.3.2 최종 결합 방법에 따른 분류 결과

아래의 표 8, 표 9, 표 10, 표 11, 표 12 에서는 SVM 학습의 결과를 제안하는 최종결합방법과 기존의 방법인 다수결 원칙, 균등 보팅과의 분류 정확도를 비교하였다. 실험결과 80% 추출된 입력세트를 신경망 MLP 결합방법과 균등 보팅 방법으로 최종 결합한 결과가 가장 우수했다.

표 8. Iris DB 최종 결합 방법에 따른 결과

	BORSE	다수결 원칙	균등 보팅
20%	93.33	93.33	90.00
40%	96.66	96.66	95.00
60%	96.66	95.55	95.55
80%	97.50	96.66	97.50

표 9. Glass Identification DB 최종 결합 방법에 따른 결과

	BORSE	다수결 원칙	균등 보팅
20%	93.02	86.04	81.39
40%	97.67	89.53	89.53
60%	98.43	93.75	85.93
80%	98.83	93.02	94.18

표 10. Letter Image DB 최종 결합 방법에 따른 결과

	BORSE	다수결 원칙	균등 보팅
20%	92.50	90.00	91.75
40%	95.87	95.00	95.37
60%	98.05	91.66	95.83
80%	98.54	89.50	90.62

표 11. Chess DB 최종 결합 방법에 따른 결과

	BORSE	다수결 원칙	균등 보팅
20%	89.11	70.12	60.33
40%	94.43	80.03	70.47
60%	94.52	80.03	71.03
80%	94.55	80.04	72.44

표 12. Connect-4 Opening DB 최종 결합 방법에 따른 결과

	BORSE	다수결 원칙	균등 보팅
20%	88.12	80.70	70.12
40%	90.31	83.12	70.31
60%	91.12	83.33	70.12
80%	91.07	83.03	71.31

5. 결론 및 향후 과제

본 논문에서는 기존의 다중 SVM 분류기의 문제점을 지적하고 대용량 다차원 데이터를 효율적으로 적용하기 위하여 새로운 모델인 BORSE를 제안하였다.

BORSE의 3가지 특징으로는 효율적인 입력세트 재구성, 간단한 구성의 일대다 정책 기반의 다중 SVM 분류기, 신뢰성 있는 최종 결합 방법이다. 실험 결과에 따라 BORSE가 데이터 마이닝을 비롯한 다양한 패턴 분류의 응용분야에서 효과적으로 활용될 수 있을 것으로 기대된다.

그러나 커널을 무엇으로 선택하는가에 따라 그 성능 차이가 발생한다. 추후 커널 선택 알고리즘은 기계학습의 연구자가 꼭 풀어야 할 난제라 할 수 있다.

참고 문헌

[1] C. Park, S. Kim, "Searching for Optimal Ensemble of Feature-classifier Pairs in Gene Expression Profile using Genetic Algorithm", *Journal of KISS*, Vol.31, No.4, pp. 525-536, 2004.

[2] S. Gutta, H. Wechsler, "Face Recognition Using Hybrid Classifier Systems", *Proc. of International Conference on Neural Networks*. pp. 1017-1022, 1996.

[3] H. Song, "A Multiple Classifier System based on Dynamic Classifier Selection having Local Property", *Journal of KISS*, Vol.30, No.3, pp. 339-346, 2003.

[4] P. Clarkson, P. Moreno, "On the use of support

vector machines for phonetic classification", *Proc. of International Conference on Acoustics, Speech, Signal Processing*, pp. 858-888, 1999.

[5] S. Zahorian, P. Silsbee, and X. Wang. "Phone Classification with Segmental Features and a Binary-Pair Partitioned Neural Network Classifier", *IEEE Conference on Acoustics, Speech, and Signal Processing*, 1997.

[6] M. N. Nguyen and J. C. Rajapakse, "Multi-Class Support Vector Machines for Protein", *Genome Informatics*, Vol. 14, pp. 218-227, 2003.

[7] K. Lee, et al, "Unconstrained Handwritten Numeral Recognition using Multistage Combination of Multiple Recognizers", *Journal of KISS*, Vol. 26, No. 1, pp. 93-105, 1999.

[8] E. A. Patric and F. P. Fisher, "A Generalized k-Nearest Neighbor Rule", *Information and control*. Vol. 16, No. 2, pp. 128-152, 1970.

[9] <http://www.ics.uci.edu/~mlearn/MLRepository.html>.

저자 소개



김만선(Man-Sun Kim)

2000년 홍익대학교 전자전기컴퓨터공학부 (공학사)
 2002년 공주대학교 대학원 전자계산학과 (이학석사)
 2005년 공주대학교 대학원 컴퓨터공학과 (공학박사)

관심분야 : 데이터마이닝, 신경망, 모바일게임, 기계학습

e-mail : mansun@kongju.ac.kr



이상용(Sang-Yong Lee)

1984년 중앙대학교 전자계산학과(공학사)
 1988년 일본동경공업대학 총합이공학 연구과(공학석사)
 1988년~1989년 일본 NEC 중앙연구소 연구원
 1993년 중앙대학교 일반대학원 전자계산학과(공학박사)

1993년~현재 공주대학교 정보통신공학부 교수

1996년~1997년 University of Central Florida 방문교수

관심분야 : 인공지능, 기계학습, 에이전트, 바이오인포매틱스

e-mail : sylee@kongju.ac.kr