

## Support Vector Machine을 이용한 문맥 민감형 융합

허경용\*

### Context Dependent Fusion with Support Vector Machines

Gyeongyong Heo\*

#### 요약

문맥 종속형 융합(CDF, Context Dependent Fusion)은 여러 분류기의 결과를 종합하여 성능을 향상시키는 융합 방법으로 주어진 문제의 문맥을 균일한 여러 문맥으로 나누고 각 문맥에서 문맥 종속적인 융합을 시도함으로써 기존 융합 방법에 비해 향상된 성능을 보여주었다. 하지만 CDF는 학습해야할 파라미터의 개수가 많아 학습 데이터가 적은 경우 잡음에 민감한 문제점이 있으며, 선형 알고리즘이라는 한계로 인해 문맥 추출 및 지역적 융합 과정에서 성능 저하의 원인이 된다.

본 논문에서는 CDF의 문제점을 완화할 수 있는 방법으로 SVM(Support Vector Machine)과 커널 주성분 분석을 이용한 CDF-SVM을 제안하였다. 커널 주성분 분석은 입력 벡터에 비선형 변환을 가함으로써 타원형이 아닌 비정형의 클러스터 생성이 가능하도록 해주며, SVM은 융합 과정에서 비선형 경계의 생성을 가능하게 해주어 CDF의 선형성 제약을 극복하도록 해준다. 또한 목적함수에 정규화 항을 추가함으로써 잡음 민감성을 줄이도록 하였다. 제안한 CDF-SVM은 기존 CDF 및 그 변형들에 비해 나은 성능을 보여주었으며 이는 실험 결과를 통해 확인할 수 있다.

▶ Keywords : 융합, 퍼지, 문맥, SVM

#### Abstract

Context dependent fusion (CDF) is a fusion algorithm that combines multiple outputs from different classifiers to achieve better performance. CDF tries to divide the problem context into several homogeneous sub-contexts and to fuse data locally with respect to each sub-context. CDF showed better performance than existing methods, however, it is sensitive to noise due to the large number of parameters optimized and the innate linearity limits the application of CDF.

In this paper, a variant of CDF using support vector machines (SVMs) for fusion and kernel

•제1저자 : 허경용 •교신저자 : 허경용

•투고일 : 2013. 2. 21, 심사일 : 2013. 5. 7, 게재확정일 : 2013. 6. 23.

•이 논문은 2012년 동의대학교 교내연구비 지원으로 연구되었음 (과제번호:2012AA168)

\* 동의대학교 전자공학과(Dept. of Electronic Engineering, Dong-Eui University)

principal component analysis (K-PCA) for context extraction is proposed to solve the problems in CDF, named CDF-SVM. Kernel PCA can shape irregular clusters including elliptical ones through the non-linear kernel transformation and SVM can draw a non-linear decision boundary. Regularization terms is also included in the objective function of CDF-SVM to mitigate the noise sensitivity in CDF. CDF-SVM showed better performance than CDF and its variants, which is demonstrated through the experiments with a landmine data set.

▶ Keywords : Fusion, Fuzzy, Context, Support vector machine

## I. 서 론

분류(classification)는 데이터 분석의 주요한 부분을 차지하는 문제로 패턴 인식, 기계 학습, 통계적 추론을 포함하여 최근에는 데이터 마이닝에 이르기까지 다양한 분야에서 활용되고 있다. 분류는 교사 학습(supervised learning) 방법의 일종으로 입력값에 따라 종속 변수의 값을 예측하는 과정으로 생각할 수 있다[1]. 분류 기법들은 다양한 문제들에 적용되어 성공적인 결과를 보여주었지만, 모든 문제를 풀 수 있는 최선의 한 가지 방법은 존재하지 않으며 어떤 분류기가 최상인지를 이야기한다는 것은 무의미하다는 사실이 오래전부터 알려져 있다[2]. 모든 분류기들은 나름의 장점과 단점을 가지고 있어 문제에 따라 적합한 분류기를 선택하여야 하지만 하나의 분류기로는 더 이상 성능을 높이기 힘든 수준에 이르러 있다. 이러한 상황에서의 개선 방법 중 하나가 다중 분류기 시스템(multiple classifier system) 또는 융합(fusion)으로 불리는 방식이다. 융합은 여러 개의 분류기를 결합함으로써 단일 분류기로는 보일 수 없는 성능을 쉽게 달성할 수 있음이 증명됨으로써 각광을 받게 되었다. 최근 융합이라는 용어는 다양한 의미로 사용되고 있으며 여러 개의 분류기를 사용하는 융합에서도 여러 단계에서 융합이 적용될 수 있다. 이 논문에서 다루고자 하는 융합은 신뢰도 융합(confidence fusion)으로, 각 분류기는 분류하고자 하는 대상에 대한 신뢰도를 계산하고 이들 신뢰도를 바탕으로 새로운 최종 신뢰도를 계산하는 방식을 융합이라 지칭한다. 이는 일반적인 패턴 인식의 문제로도 여겨질 수 있으며 이 때 입력은 각 분류기가 계산한 신뢰도값이 되며 출력은 최종 신뢰도가 된다[3].

융합을 위해 지금까지 제안된 많은 방법들은 크게 분류기 융합(classifier fusion)과 분류기 선택(classifier selection)의 두 가지로 나눌 수 있다. 분류기 융합은 융합에

사용된 모든 분류기들이 문제의 전체 영역에서 유용한 결론을 내릴 수 있다고 가정하고 전체 영역에서 모든 분류기의 신뢰도를 결합할 수 있는 방법에 집중한다. 하지만 문제 영역은 일반적으로 특성이 다른 몇 개의 영역으로 나눌 수 있으며 각 영역에서 특히 뛰어난 성능을 보이는 분류기가 있을 수 있음을 분류기 융합에서는 간과하고 있다.

분류기 선택은 분류기 융합과는 달리 문제 영역을 균일한 여러 부분으로 나누고 선택된 분류기를 각 영역에 할당하는 방법으로 분할 정복 기법(divide and conquer)의 일종이라 할 수 있다. 분류기 선택에서는 특징 공간을 분할하고 각 영역에서 분류기의 성능을 추정하는 방법이 중요하다[5-6]. 하지만 최적의 분할 및 국부적으로 분류기의 성능을 판별하는 것은 쉽지 않으며 다양한 특징 공간 분할 및 국부적 성능 평가 방법이 존재한다는 사실이 이를 보여준다.

문맥 민감형 융합(context dependent fusion, CDF)은 분류기 융합과 분류기 선택을 결합한 방법이다[6]. CDF는 특징 공간을  $K$ 개의 균일한 영역으로 나누며 이처럼 특징 공간을 분할하는 점에서는 분류기 선택에서와 동일하다. 특징 공간에서의 균일한 영역은 문맥(context)이라 정의되며 특정 문맥에서는 문맥에 적합한 융합이 이루어진다. 반복 최적화 기법을 이용하여 CDF는 문맥을 자동으로 결정하고 동시에 특정 문맥에서 융합에 필요한 파라미터들을 학습할 수 있다. CDF는 여러 문제에 적용하여 성공적인 결과를 보였지만 몇 가지 문제점이 있다. 그 중 가장 큰 문제점은 잡음에 민감하다는 점이다. CDF에는 최적화시켜야 할 파라미터들이 다수 존재하므로 반복 최적화 기법에 의한 최적화는 국부 최적해에 수렴할 가능성이 높다. CDF의 여러 변형들이 제안되었으며 그 중 하나가 CDF에 정규화 기법을 도입한 CDF-R(CDF with Regularization)이다[7]. 정규화는 파라미터가 가질 수 있는 값에 조건을 부여함으로써 잡음 민감성을 줄이는 대표적인 방법 중 하나로[8] 정규화를 통해 잡음 민감성을 줄일 수 있다는 사실이 입증되었다. 또 다른 CDF의 문제점은

CDF가 선형 알고리즘이라는 점이다. CDF는 문맥을 결정하기 위해 퍼지 클러스터링을, 융합을 위해서는 선형 조합을 사용하며 이들은 모두 대표적인 선형 알고리즘들이다. 비선형의 문맥은 선형 문맥의 조합으로 표현될 수 있지만 문맥의 개수가 증가할수록 학습하여야 할 파라미터의 개수가 증가하게 되어 학습 민감성이 증가하는 문제점이 있다.

이 논문에서는 위에서 언급한 CDF의 두 가지 문제점을 완화하기 위해 CDF의 비선형 변형인 CDF with Support Vector Machine(CDF-SVM)을 제안하였다. CDF-SVM은 비선형 문맥 분할을 위해 커널 주성분분석을 전처리로 사용하였으며, 융합을 위해서는 SVM을 이용함으로써 CDF의 선형성을 제거하였다. 또한 정규화를 포함함으로써 CDF-R의 장점을 살리도록 하였다.

이 논문의 구성은 다음과 같다. 2장에서는 CDF와 위의 변형인 CDF-R을 설명한다. 3장에서는 CDF와 CDF-R의 단점을 보완할 수 있는 CDF-SVM을 제안한다. 4장에서는 실험 결과를 통해 제안한 방법의 실효성을 보이며 향후 연구 방향은 5장에서 제시한다.

## II. 관련 연구

### 1. 문맥 민감형 융합 (CDF)

$N$ 개의 학습 데이터와 기대 출력(expected output)이  $O = \{o_i | i = 1, \dots, N\}$ 으로 주어졌다가 가정해보자. 데이터는  $T$ 개의 알고리즘이 개별적으로 처리하여 신뢰도 값  $y_t = \{y_{ti} | i = 1, \dots, N\}$ 을 출력하며 각 알고리즘은 학습 데이터로부터 알고리즘에서 사용하는  $L_t$  차원의 특징 벡터  $X_t = \{x_{i,t} | x_{i,t} \in R^{L_t}, i = 1, \dots, N\}$ 를 추출한다. 추출한 각 특징 벡터는 융합에서 사용할 전역 특징 벡터  $X$ 를 식 (1)에서와 같이 구성한다.

$$X = \bigcup_{t=1}^T X_t = \{x_i | x_i \in R^L, i = 1, \dots, N\} \quad (1)$$

이 때  $L$ 은 전역 특징 벡터의 차원으로 각 알고리즘의 특징 벡터 차원을 더한 값으로 주어진다.

$$L = \sum_{t=1}^T L_t \quad (2)$$

문맥은 특정 공간에서의 균일한 영역으로 특징 공간은  $K$ 개의 문맥으로 구성되며 각 문맥은 특정 공간에서 하나의 클러스터(cluster)로 표현된다. CDF의 목적 함수는 식 (3)과 같다.

$$J = \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m \|x_i - v_k\|^2 + \alpha \sum_{i=1}^n \sum_{k=1}^K u_{ki}^m \left( \sum_{t=1}^T w_{kt} y_{ti} - o_i \right)^2 \quad (3)$$

이 때  $v_k$ 는  $k$ 번째 클러스터의 중심을,  $u_{ki}$ 는  $i$ 번째 데이터가  $k$ 번째 클러스터에 소속되는 소속도를,  $m$ 은 퍼지화 상수를,  $w_{kt}$ 는  $t$ 번째 알고리즘이  $k$ 번째 클러스터 또는 문맥에서 융합에 참여하는 가중치를 나타낸다. 소속도와 분류기의 가중치에 대한 제약 조건은 식 (4) 및 (5)와 같다.

$$U_{K \times N} = \left\{ u_{ki} | u_{ki} \in [0, 1] \forall i \text{ and } k; \sum_{k=1}^K u_{ki} = 1 \forall i \right\} \quad (4)$$

$$W_{K \times T} = \left\{ w_{kt} | w_{kt} \in R \forall k \text{ and } t; \sum_{t=1}^T w_{kt} = 1 \forall k \right\} \quad (5)$$

식 (3)의 첫 번째 항은  $N$ 개의 데이터 포인트를  $K$ 개의 클러스터로 FCM(fuzzy c-means)을 이용하여 나누기 위해 사용되며 FCM의 목적 함수와 동일하다(9). 두 번째 항은 각 클러스터에서 분류기의 신뢰도를 조합하여 기대 출력과 차이를 최소화 하기 위해 사용된다. CDF에서는 각 분류기의 신뢰도에 클러스터별 가중치를 부여한 가중치 합을 사용하며  $\alpha$ 는 두 항의 반영 비율을 결정하는 상수이다. 식 (3)은 라그랑지(Lagrange) 방정식을 이용하여 갱신식을 얻어낼 수 있으며 이를 이용한 반복 최적화 기법을 통해 문맥을 나누기 위한 소속도  $U$ 와 융합을 위한 분류기 가중치  $W$ 를 얻을 수 있다.

### 2. 정규화를 이용한 문맥 민감형 융합

CDF-R은 CDF에 정규화를 도입하여 학습 민감성을 줄인 CDF의 변형이다. 이외에도 CDF-R이 CDF와 다른 점은 클러스터 중심에서 데이터 포인트까지의 거리 계산 방식이다. CDF의 경우 거리는 유클리드 거리로 주어졌다. 하지만 특징 값들이 정규화되지 않은 경우를 고려하여 CDF-R에서는 마할라노비스 거리(Mahalanobis distance)로 변경되었다. CDF-R의 목적 함수는 식 (6)과 같다.

$$\begin{aligned}
 J = & \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m \sum_{l=1}^L r_{klt}^q d_{kil}^2 & (6) \\
 & + \alpha \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m \left( \sum_{t=1}^T w_{kt} y_{ti} - o_i \right)^2 \\
 & + \beta \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m + \gamma \sum_{k=1}^K \sum_{l=1}^L r_{klt}^q + \delta \sum_{k=1}^K \sum_{l=1}^L w_{kt}^2
 \end{aligned}$$

식 (6)에서의 거리는 식 (7)과 같이 특징 별로 계산된다.

$$d_{kil} = |x_{il} - v_{kl}| \quad (7)$$

특징에 대한 가중치  $R$ 의 제약 조건은 식 (8)과 같다.

$$R_{K \times T} = \left\{ r_{kt} | w_{kt} \in [0, 1] \quad \forall k \text{ and } l; \sum_{t=1}^T r_{kt} = 1 \quad \forall k \right\} \quad (8)$$

식 (6)에는 3개의 다항식 정규화 항이 추가되었으며 이들은 각각 소속도  $U$ , 분류기 가중치  $W$ , 특징 가중치  $R$ 에 대응한다.  $\beta, \gamma, \delta$ 는 정규화 상수들이며  $q$ 는 특징 가중치에 대한 퍼지화 상수이다. 라그랑지 방정식을 이용하여 갱신식을 구하면 다음과 같다[7].

$$v_{kl} = \frac{\sum_{i=1}^N u_{ki}^m x_{il}}{\sum_{i=1}^N u_{ki}^m} \quad (9)$$

$$r_{kt} = \frac{\left( \frac{1}{D_{kt}} \right)^{\frac{1}{q-1}}}{\sum_{a=1}^L \left( \frac{1}{D_{ka}} \right)^{\frac{1}{q-1}}} \quad (10)$$

$$w_{kt} = \frac{\sum_{i=1}^N u_{ki}^m \left( o_i - \sum_{a=1, a \neq t}^T w_{ka} y_{ai} \right) y_{ti} - \zeta_k}{\sum_{i=1}^N u_{ki}^m y_{ai}^2 + \frac{\delta}{\alpha}} \quad (11)$$

$$u_{ki} = \frac{\left( \frac{1}{D_{ki}} \right)^{\frac{1}{m-1}}}{\sum_{a=1}^K \left( \frac{1}{D_{ai}} \right)^{\frac{1}{m-1}}} \quad (12)$$

이 때  $\bar{D}, \zeta, \tilde{D}$ 는 다음과 같다.

$$\bar{D}_{kl} = \sum_{i=1}^N u_{ki}^m d_{kil}^2 + \gamma \quad (13)$$

$$\begin{aligned}
 \zeta_k = & \frac{\sum_{a=1}^T \frac{\sum_{i=1}^N u_{ki}^m \left( o_i - \sum_{t=1}^T w_{kt} y_{ti} \right) y_{ai} - w_{ka} \frac{\delta}{\alpha}}{\sum_{i=1}^N u_{ki}^m y_{ai}^2 + \frac{\delta}{\alpha}}}{\sum_{a=1}^T \frac{1}{\sum_{i=1}^N u_{ki}^m y_{ai}^2 + \frac{\delta}{\alpha}}} & (14)
 \end{aligned}$$

$$\tilde{D}_{ki} = \sum_{l=1}^L r_{klt}^q d_{kil}^2 + \alpha \left( \sum_{t=1}^T w_{kt} y_{ti} - o_i \right)^2 + \beta \quad (15)$$

구해진 갱신식을 이용하여 CDF-R의 학습 알고리즘은 그림 1과 같이 나타낼 수 있다.

입력 :

- $X$  : 데이터 ( $N \times L$ )
- $Y$  : 분류기 출력 ( $N \times T$ )
- $O$  : 기대 출력 ( $N \times 1$ )

- 1:  $U$ 와  $W$ 를 무작위 값으로 초기화한다.
- 2:  $R$ 을  $1/L$  값으로 초기화한다.
- 3: repeat
- 4: 클러스터 중심  $V$ 를 식 (9)로 갱신한다.
- 5: 특징 가중치  $R$ 을 식 (10)으로 갱신한다.
- 6: 분류기 가중치  $W$ 를 식 (11)로 갱신한다.
- 7: 소속도  $U$ 를 식 (12)로 갱신한다.
- 8: until  $V, R, W, U$  수렴
- 9: return  $V, R, W, U$

그림 1. CDF-R 학습 알고리즘  
Fig. 1. CDF-R training algorithm

### III. SVM을 이용한 문맥 민감형 융합

CDF는 일반화된 융합 기법으로 이전의 연구들을 포괄할 수 있는 방법으로 그 효용성이 입증되었지만 잡음 민감성과 선형성의 한계가 있다. 잡음 민감성의 문제는 정규화를 도입함으로써 일부 해결이 가능하였지만 선형성의 문제는 여전히 남아있다. CDF는 크게 문맥 감지와 융합의 두 부분으로 나눌 수 있다. 문맥 감지는 FCM과 마할라노비스 거리를 이용하여 이루어지므로 모든 클러스터는 볼록한(convex) 것으로 가정하고 있다. 하지만 특성 벡터의 차원이 높아질수록 균일한 클러스터가 볼록한 형태를 가지는 경우는 많지 않다. 오목

한(concave) 형태의 클러스터는 볼록한 클러스터의 조합으로 표현할 수 있지만 클러스터의 개수가 늘어나면 학습해야 할 파라미터의 개수가 증가하므로 바람직하지 않다. 융합의 경우 CDF는 선형 조합을 사용하고 있다. 이는 클러스터 내에서 선형 분류기에 의해 분류가 가능함을 가정한 것이다. 하지만 이 역시도 실제 데이터의 경우 선형으로 분리가 불가능한 경우가 많다. 비선형 분리의 경우도 볼록한 클러스터 가정의 경우와 동일한 해결책이 가능하지만 역시 동일한 이유로 바람직하지 않다.

이 절에서 제안하는 CDF-SVM(CDF with SVM)은 위의 두 가지 문제를 비선형 알고리즘을 도입함으로써 해결하였다. 클러스터링은 특징 공간이 아닌 커널 특징 공간(kernel feature space)에서 수행하는 K-FCM(Kernel FCM)[10]을 이용하면 비선형 클러스터링이 가능하다. 하지만 특징 공간을 커널 특징 공간으로 사상하는 경우 커널 특징 공간의 차원은 일반적으로 무한대가 되므로 각 특징값에 가중치를 부여하기가 어렵다. 따라서 이 논문에서는 특징 집합에 전처리 과정으로 비선형 변환을 가한 후 FCM을 수행함으로써 비선형 클러스터링의 효과를 얻도록 하였다. 비선형 변환은 PCA(주 성분 분석, Principal Component Analysis)를 비선형으로 변환한 RKF-PCA(Robust Kernel Fuzzy PCA)[11]를 이용하였으며 유한 차원  $L$ 로 변환하였다. (수식의 일관성을 위하여  $L$ 을 사용하였으며,  $L$ 로 표시되던 특징 공간에서의 차원은  $L'$ 으로 표시하도록 한다.)

융합의 위해서는 SVM(Support Vector Machine)[12]을 사용하였다. SVM은 전통적인 원형(prototype) 기반의 분류 방식이 아니라 분류 경계에 놓이는 샘플들을 기반으로 분류하는 방식으로 대표적인 비선형 분류기 중 하나이다. SVM은 그 자체로도 많은 분야에서 응용이 되고 있지만, CDF-SVM에서는 잡음 민감성을 줄이기 위해 데이터의 소속도를 추가한 Fuzzy SVM(F-SVM)[13]을 사용하였다. RKF-PCA와 F-SVM을 이용하여 식 (6)을 수정한 CDF-SVM의 목적 함수는 식 (16)과 같다.

$$J = \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m \sum_{l=1}^L r_{kil}^q d_{kil}^2 + \alpha \sum_{k=1}^K \left( \frac{1}{2} \|w_k\|^2 + C \sum_{i=1}^N u_{ki}^m \xi_{ki} \right) + \beta \sum_{i=1}^N \sum_{k=1}^K u_{ki}^m + \gamma \sum_{k=1}^K \sum_{l=1}^L r_{kil}^q \quad (16)$$

이 때  $w_k$ 는  $k$ 번째 F-SVM의 경계를,  $C$ 는  $K$ 개 F-SVM이 공통으로 사용하는 정규화 상수를,  $\xi_{ki}$ 는  $k$ 번째

데이터 포인트가  $i$ 번째 F-SVM에서의 오류를 나타낸다. 식 (16)의 첫 번째 항은 CDF와 마찬가지로  $K$ 개의 균일한 영역으로 나누는 역할을 하며, 두 번째 항은 F-SVM의 목적 함수로 클러스터 별 융합에 해당하며 나머지 두 항은 정규화 항이다.  $d_{kil}$ 은 식 (7)과 동일하지만 CDF에서는 특징 공간에서 거리를 계산하는 반면 CDF-SVM에서는 커널 특징 공간에서 거리가 계산된다는 점에서 차이가 있다. 식 (16)에 라그랑지 방정식을 적용하여 갱신식을 구하면 식 (9), (10), (12)와 같은 식을 구할 수 있다.  $\bar{D}$  역시 식 (13)과 동일하며  $\tilde{D}$ 는 식 (17)과 같이 구해진다.

$$\tilde{D}_{ki} = \sum_{l=1}^L r_{kil}^q d_{kil}^2 + \alpha C \xi_{ki} + \beta \quad (17)$$

CDF-SVM의 경우 CDF와 달리 분류기 융합 가중치  $W$ 가 존재하지 않는다. CDF-SVM에서는 분류기의 출력값이 SVM의 입력으로 주어지므로  $W$ 는 SVM의 분류 경계를 나타내는  $w$ 로 대체되었다. CDF-SVM에서는 기대 출력  $O$  역시 사용되지 않으며 SVM에서의 오류를 나타내는  $\xi$ 로 대체되었다. CDF-SVM에서  $w$ 와  $\xi$  값들은 F-SVM의 최적화를 통해 얻어진다. CDF-SVM의 학습 알고리즘은 그림 2와 같이 나타낼 수 있다. 그림 2에서 알 수 있듯이 CDF-SVM 알고리즘은 그림 1의 CDF-R 알고리즘과 입력 및 전체적인 구조는 동일하며, 비선형 특징 벡터 추출을 위한 전처리 과정이 도입된 점과 F-SVM 파라미터 최적화 과정이 분류기 가중치 최적화 과정과 대체된 점에서 차이가 난다.

입력 :

$X$  : 데이터 ( $N \times L'$ )  
 $Y$  : 분류기 출력 ( $N \times T$ )  
 $O$  : 기대 출력 ( $N \times 1$ )

- 1:  $L$  차원의 비선형 특징 벡터를 RKF-PCA로 구한다.
- 2:  $U$ 를 무작위 값으로 초기화한다.
- 3:  $R$ 을  $1/L$  값으로 초기화한다.
- 4: repeat
- 5: 클러스터 중심  $V$ 를 식 (9)로 갱신한다.
- 6: 특징 가중치  $R$ 을 식 (10)으로 갱신한다.
- 7: 소속도  $U$ 를 식 (12)로 갱신한다.
- 8: F-SVM 파라미터를 F-SVM 최적화 방법으로 갱신한다.
- 9: until  $V, R, U, F$ -SVM 파라미터 수렴
- 10: return  $V, R, U, F$ -SVM 파라미터

그림 2. CDF-SVM 학습 알고리즘  
 Fig. 2. CDF-SVM training algorithm

그림 2의 학습 알고리즘으로 CDF-SVM의 파라미터들이 결정되고 테스트 벡터  $x$ 가 주어지면 먼저 RKF-PCA를 이용하여  $L$ 차원의 비선형 특징 벡터  $x'$ 으로 사상하고  $T$ 개의 알고리즘은 자기  $x$ 의 부분집합을 이용하여 알고리즘 출력값  $y = [y_1, \dots, y_T]$ 를 계산한다. 테스트 데이터의  $K$ 개 클래스터 즉 문맥에 대한 소속도 값은 식 (18)과 같이 계산된다.

$$u_{k^*} = \frac{\left(\frac{1}{\bar{D}_{k^*}}\right)^{\frac{1}{m-1}}}{\sum_{a=1}^K \left(\frac{1}{\bar{D}_{a^*}}\right)^{\frac{1}{m-1}}} \quad (18)$$

이 때  $\bar{D}_{k^*}$ 는 식 (17)을 이용할 수 없다. 식 (17)은 클러스터링의 비교사 학습에 의한 거리와 F-SVM의 교사 학습에 의한 거리로 이루어져 있으며 이 중  $\xi_{k^*}$ 는 F-SVM에서 사용되는 오류값, 즉 교사 거리로 테스트 데이터의 경우 기대 출력  $o_k$ 가 주어지지 않으므로 계산할 수 없다. 따라서 테스트 데이터의 경우 오류는 없는 것으로 가정하고 식 (19)와 같이 비교사 학습에 의한 거리만을 계산한다.

$$\bar{D}_{k^*}' = \sum_{l=1}^L r_{kl}^q d_{kl}^2 + \beta \quad (19)$$

$K$ 개의 F-SVM이 자기 출력값  $y' = [y_1', \dots, y_K']$ 을 계산하면 식 (18)에 의해 계산된 소속도와 함께 최종 출력  $z$ 는 식 (20)과 같이 계산된다.

$$z(x) = \sum_{k=1}^K u_{k^*}^m y_k' \quad (20)$$

그림 3은 테스트 벡터  $x$ 와 분류기 출력  $y$ 가 주어진 경우 최종 출력을 내는 테스트 과정을 나타낸 것이다.

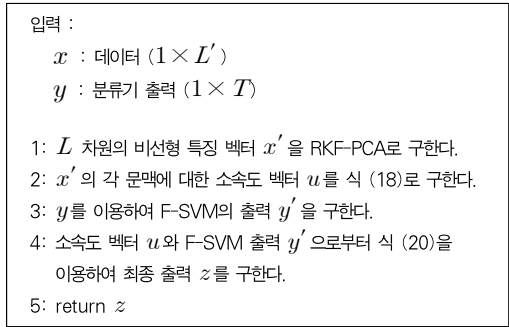


그림 3. CDF-SVM의 테스트 단계  
Fig. 3. Test phase of CDF-SVM

### IV. 실험 결과

제안한 알고리즘의 유효성을 보이기 위해 CDF-SVM은 지뢰 탐지 데이터에 적용되었다. 사용한 데이터는 비행기에서 하이퍼 스펙트럴 센서를 통해 획득한 하이퍼 스펙트럴 영상(hyperspectral image)으로 원격 탐사에서 많이 사용된다. 데이터는 482개의 샘플로 구성되며 167개의 지뢰와 315개의 지뢰 이외의 대상 물체에 대한 데이터로 구성되어 있다. 하이퍼 스펙트럴 영상을 이용하여 3개의 분류기를 작성하였으며 자기 스펙트럴 영상 분석에 많이 적용되는 기법인 WDW(whiten and de-whiten)[14], SAM(spectral angle mapper)[15], SCM(spectral correlation mapper)[16]을 사용하여 신뢰도 값을 얻었다. 얻어진 신뢰도는 하이퍼 스펙트럴 영상이 더불어 CDF-SVM의 입력으로 사용되어 최종 신뢰도 값을 얻기 위해 사용되었다. CDF-SVM의 결과를 얻기 위해 10-fold cross-validation 기법을 적용하였으며 CDF-SVM의 학습 및 테스트 알고리즘은 Matlab으로 구현하여 실험하였다.

그림 4는 3개의 개별 분류기(WDW, SAM, SCM)를 이용한 인식 결과와 CDF-R 및 CDF-SVM을 이용하여 인식한 결과를 비교한 ROC(Receiver Operating Characteristic) 커브를 나타내고 있다. 그림 4에서 PD(Probability of Detection)는 인식 확률로 지뢰 중 지뢰로 분류된 대상체의 수를 전체 지뢰의 개수로 나눈 값이며, PFA(Probability of False Alarm)는 오인식 확률로 지뢰가 아닌 대상체 중 지뢰로 분류된 대상체의 개수를 전체 지뢰 이외의 대상체 개수로 나눈 값이다.

ROC를 통해 알고리즘의 성능을 비교하는 가장 간단한 방법은 커브 아래쪽의 면적을 비교하는 것이다. 그림 3에서 알 수 있듯이 대부분의 영역에서 CDF-SVM의 그래프는 개별

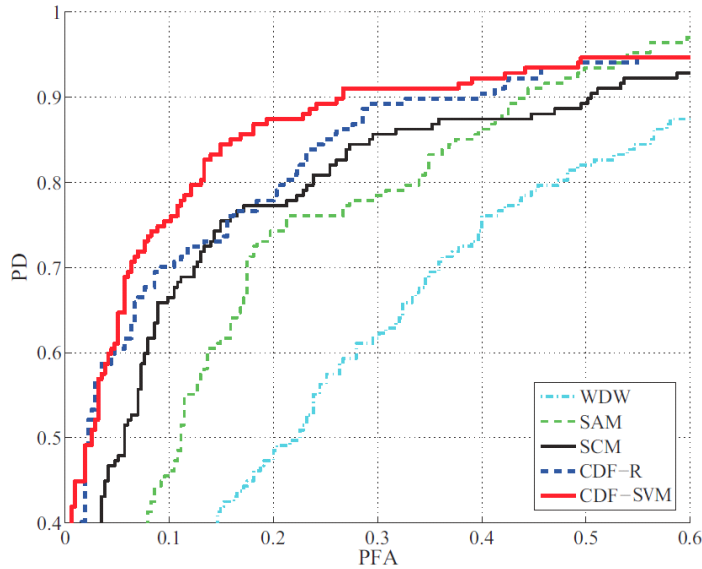


그림 4. 하이퍼 스펙트럴 영상 데이터의 ROC 커브  
Fig. 4. ROC curves for a hyperspectral image data set

분류기나 CDF-R의 그래프 위에 놓여있어 그 면적이 개별 분류기나 CDF-R에 비해 넓은 것을 알 수 있다. 즉, CDF-SVM은 개별 분류기나 CDF-R에 비해 나은 성능을 보이고 있음을 알 수 있다. 이외에도 다양한 ROC 커브 비교 방법이 있지만 이 논문에서는 특정 PD에서 PFA의 신뢰 구간을 비교함으로써 성능을 비교하였다. 지뢰 탐지의 경우 일반적인 인식 문제에서와 같이 최종 결과를 인식률로 나타내는 경우는 드물며 ROC를 통해 지정한 PD에서의 PFA를 비교함으로써 알고리즘의 성능을 비교하는 것이 일반적이다. 비교를 위해서는 지뢰 탐지 문제에서 일반적으로 사용되는 85%, 90%, 95% PD에서 PFA의 신뢰 구간을 신뢰 수준 90%에서 비교하였으며 그 결과는 표 1과 같다. 표 1에 나타난 바와 같이 각 PD 값에서 신뢰 구간은 겹치지 않는다. 즉 90% 신뢰 수준에서 CDF-SVM은 CDF-R에 비해 지정한 PD 값에서 나은 성능을 보여주는 것으로 알 수 있다.

표 1. CDF-R과 CDF-SVM의 신뢰 구간 비교  
Table 1. Comparison of the confidence intervals of CDF-R and CDF-SVM

PD(%)	신뢰 구간	
	CDF-R	CDF-SVM
85	(44.28 78.70)	(49.39 50.59)
90	(125.19 126.81)	(83.28 84.74)
95	(201.22 202.80)	(190.18 191.79)

## V. 결론

CDF(Context Dependent Fusion)는 기존의 융합 알고리즘을 포괄하는 알고리즘으로 서로 다른 문맥에서 동작하는 이종의 분류기들을 결합하여 보다 신뢰성 있는 출력값을 낼 수 있는 일반화된 융합 알고리즘으로 그 유용성이 입증된 방법이다. 하지만 CDF는 선형 알고리즘이라는 한계가 있으며 학습해야 할 파라미터의 개수가 많아 잡음에 민감한 단점을 가지고 있다. 이 논문에서는 이러한 CDF 및 그 변형들의 단점을 극복하기 위해 커널 주성분 분석과 SVM을 이용한 비선형 알고리즘인 CDF-SVM을 제안하였으며 실험 결과 기존의 CDF 및 그 변형들에 비해 우수한 성능을 보임을 확인할 수 있었다.

CDF-SVM이 기존 방법들에 비해 우수한 성능을 보이기에는 하였지만 개선의 여지는 남아있다. 잡음 민감성의 경우 정규화(regularization)을 도입하여 극복 최적해에 빠질 위험이 줄고 성능이 향상되었지만 여전히 CDF-SVM에는 최적화 시켜야 할 파라미터의 개수가 다수 존재한다. 파라미터의 개수가 많아지면 잡음 민감성이 증가하므로 CDF-SVM에서 학습해야 할 파라미터의 개수를 줄이는 작업이 필요하며 특히 클러스터 별로 계산되는 특징 벡터의 가중치를 제거하는 방법에 대해 현재 연구 중에 있다. 다른 한 가지 개선 방향은 클러스터링의 비선형화이다. 현재 클러스터링은 커널 주성분 분석과

FCM을 통해 비선형 클러스터링의 효과를 내고 있다. 이러한 두 단계의 처리는 특징 벡터의 가중치 학습과도 무관하지 않으므로 특징 벡터의 가중치가 제거된다면 비선형 변환이 CDF-SVM에 포함될 수 있을 것으로 생각되며 특징 벡터의 가중치와 더불어 연구 중에 있다.

## 참고문헌

- [1] C.M. Bishop, "Pattern Recognition and Machine Learning," Springer, Singapore, 2007.
- [2] M. Minsky, "Logical versus analogical or symbolic versus connectionist or neat versus scruffy," *AI Magazine*, Vol. 12, No. 2, pp. 34-51, Jun. 1991.
- [3] L.I. Kuncheva, "Combining Pattern Classifiers," Wiley-Interscience, New York, 2004.
- [4] L.I. Kuncheva, "Clustering-and-selection model for classifier combination," *Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, pp. 185-188, 2000.
- [5] R. Liu and B. Yuan, "Multiple classifiers combination by clustering and selection," *Information Fusion*, Vol. 2, No. 3, pp. 163-168, Sep. 2001.
- [6] H. Frigui, P. Gader, and A.C.B. Abdallah, "A generic framework for context-dependent fusion with application to landmine detection," *Proceedings of SPIE Defense and Security Symposium*, pp. 490-495, 2008.
- [7] Gyeongyong Heo, Paul Gader, and Hichem Frigui, "A Noise Robust Variant of Context Extraction for Local Fusion," *Proceedings of the 2010 IEEE International Conference on Fuzzy Systems*, pp. 1-6, 2010.
- [8] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, "Numerical Recipes: The Art of Scientific Computing," Cambridge University Press, New York, 3rd Ed. 2007.
- [9] N.R. Pal, K. Pal, J.M. Keller, and J.C. Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm," *IEEE Transactions on Fuzzy Systems*, Vol. 13, No. 4, pp. 517-530, Aug. 2005.
- [10] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern Recognition*, Vol. 41, No. 1, pp. 176-190, Jan. 2008.
- [11] Gyeongyong Heo, Paul Gader, and Hichem Frigui, "RKF-PCA: Robust Kernel Fuzzy PCA," *Neural Networks*, Vol. 22, No. 5-6, pp. 642-650, Jul.-Aug. 2009.
- [12] V. Vapnik, "Statistical Learning Theory," John Wiley & Sons, New York, 1998.
- [13] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464-471, Mar. 2002.
- [14] R. Mayer, F. Bucholtz, and D. Scribner, "Object detection by using whitening/dewhitening to transform target signatures in multitemporal hyperspectral and multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 5, pp. 1136-1142, May 2003.
- [15] R.H. Yuhas, A.F. Goetz, and J.W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *Summaries of the 3rd Annual JPL Airborne Geoscience Workshop*, pp. 147-149, 1992.
- [16] O.A. de Carvalho and P.R. Meneses, "Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM)," *Summaries of the 9th Annual JPL Airborne Earth Science Workshop*, pp. 65-74, 2000.



## 저 자 소 개



허 경 용

1994: 연세대학교  
전자공학과 공학사.

1996: 연세대학교  
전자공학과 공학석사.

2009: University of Florida  
컴퓨터공학과 공학박사

현 재: 동의대학교  
전자공학과 교수

관심분야: 인공지능, 패턴인식,  
로봇공학

Email : hgycap@deu.ac.kr