

Support Vector Fuzzy Inference System을 이용한 Pattern Recognition 에 관한 연구

김 용균*°, 정은화*
천안대학교 정보기술대학원

A Study on the Pattern Recognition Using Support Vector Fuzzy Inference System

Yongkyun Kim*°, Eunhwa Jeong*
Graduate School of Information Technology, Chonan Univ.

요 약

본 논문에서는 pattern recognition을 위하여 support vector fuzzy inference system을 제안하였다. Fuzzy inference system의 structure와 parameter를 identification 하기 위하여 Support vector machine을 이용하였으며 여러 최소화 기법으로는 gradient descent 방법을 사용하였다. 제안된 SVFIS 방법의 성능을 파악하고자 COIL 이미지를 이용한 3차원 물체 인식 실험을 수행 하였다.

1. 서론

Input 과 Output 의 관계를 Learning 을 통해 Pattern classification 또는 Recognition 하는 기법은 지금까지 수많은 연구자들이 연구를 해 왔으며 지금도 연구를 수행중에 있다. Learning 을 통한 pattern classification 또는 Recognition 기법 중 대표적인 것으로 Fuzzy inference system, Neural network, 그리고 최근에 많은 연구가 진행 중인 Support vector machine 이 있다. 위에서 말한 각각의 기법들은 저마다 해당되는 장점을 가지고 있는 반면에 단점들을 가지고 있어 최근에는 많은 연구자들이 위의 기법들을 연계하여 더 나은 classification 방법을 연구하고 있다. 특히 Fuzzy inference system 의 경우는 경험에 의한 지식을 이용하여 직접 얻거나 수치 데이터를 이용한 learning machine 기법을 이용하여 생성되어질 수 밖에 없는데 복잡하고 불확실한 pattern 의 경우 인간 경험의 부족으로 인한 불확실성이 존재 할 수 밖에 없다. 이러한 이유 때문에 input-output data 로부터 지식을 얻어 classification 을 수행하는

neuro-fuzzy system 이 제안되었다[7]. Neural Network 과 Fuzzy inference system 을 합친 neuro-fuzzy system 은 classification 을 수행하고자 하는 대상 system 을 fuzzy system 의 fuzzy logic 과 reasoning 같은 경험 지식으로 표현할 수 있을 뿐만 아니라, neural network 의 learning 과 adaptive 성질을 이용하여 fuzzy system 의 parameter 들을 결정할 수 있다. 일반적으로 input-output data 를 이용한 neuro-fuzzy classification 은 structure identification 과 parameter identification 으로 구성되는데, structure identification 은 fuzzy 규칙 또는 변수의 수를 결정하고 parameter identification 은 소속 함수들의 parameter 들과 후반부의 선형 방정식의 계수를 결정한다. 즉 neuro-fuzzy classification 에서 중요한 점은 classification 을 수행하고자 하는 대상 시스템의 주어진 input-output data 로부터 가장 좋은 structure 와 parameter 를 결정하는 것이다. 만약 fuzzy model 이 너무 많은 수의 규칙을 갖는다면 주어진 output data 와 fuzzy output data 사이의 error 는 감소하지만 너무 주어진 output data 에 꼭 맞추어 규칙을 생성한 관계로 주어지지 않은

output data 에 관하여서는 error 가 커진다. 이와 반대로 너무 작은 수의 규칙을 갖는다면 위와 같은 현상은 막을 수 있으나 원래 존재하는 error 가 커지게 되는 것이다.

그럼 structure identification 에서 퍼지 규칙의 수를 결정하는 것은 input 공간을 나누는 방법에 관련이 있는데, input 공간을 나누는 방법은 두가지 종류가 있다. 첫번째 방법은 grid partition 방법으로 이 방법은 course of dimensionality, 규칙 수의 기하급수적 증가, Unprdictable completeness, computation cost 등의 단점들을 가지고 있기 때문에 거의 사용을 하지 않는다. 두번째 방법으로 structure identification 에서 주로 쓰이는 clustering 방법이 있는데, 이 방법에서는 cluster 의 개수를 미리 알아야만 하거나, 아니면 grid point 가 cluster 중앙으로서의 기능을 수행할 수 있어야 한다.

이러한 단점들을 극복하고자 본 논문에서는 퍼지 규칙의 수를 결정할 뿐만 아니라, 파라미터 값도 결정하는 Support Vector Fuzzy Inference Classification System 을 제안한다. 여기서 제안된 Support Vector Fuzzy Inference Classification System 은 input 공간을 높은 차수의 특성 공간 (High dimensional feature space) 으로 사상하는 비선형 사상 (mapping)을 수행함으로써 주어진 복잡하고 non-separable data 를 선형적으로 분석할 수 있다. 먼저 SVFIS 의 구조는 model error 를 이용한 constrained quadratic programming 을 최소화함으로써 얻을 수 있다. Chan 등 [1]은 radial basis function network 에 기반한 Support Vector Neural Network (SVNN) 을 제안하였고, Jeng 등 [2]은 Fuzzy inference system 을 향상시키기 위한 SVM 을 제안하였다. 하지만 위의 두개 모두 fixed Gaussian Kernel parameter 를 가지기 때문에, completeness 가 보장이 되지 않는다. 이것은 퍼지 룰의 개수가 완전히 최소화 되지 않는 것을 의미한다. 본 논문에서는 Learning process 과정 중에 gradient descent algorithm 을 이용하여 소속함수의 parameter 값을 조정함으로써 퍼지 규칙의 개수를 줄일 수 있다. 일단 Structure 가 결정되면 TS fuzzy model 의 추론 부분의 parameter 값들은 최소 제곱법을 통해 결정된다.

본 논문에서 TS fuzzy model 에 기반한 제안된 SVFIS 를 이용하며 pattern recogniton 을 수행해 본다. Learning 및 test 데이터로는 100 개의 3 차원 물체를 5 도씩 돌려가면서 image file 로 생성한 COIL-100(Columbia Object Image Library 100)을 사용하였으며 총 128*128 의 RGB 형태의 7200 장 image 들로 구성되어져 있는 데이터를 사용하였다.

본 논문은 2장에서 관련 연구로 SVM의 근간이 된 statistical learning theory과 input space의 구분 방법을 살펴보고 3장에서 SVFIS시스템에 대하여 설명하고, 4장에서는 COIL image를 이용한 실험

결과에 대하여 설명한 뒤, 5장에서 결론을 맺는다.

2. 관련연구

2.1 Statistical Learning Theory[3]

이 부분에서는 Statistical Learning Theory 에 대해 설명한다. 좀 더 자세한 설명은 다음 서적들을 참고하기 바란다[3].

1. Generalization Error

추정 오차와 근사화 오차를 합친 generalization 오차는 Steve Gunn[24] 에 의해 소개 되었다. Generalization 오차는 다음 두 가지 경우로 구성된 다.

- 근사화 오차 : 근사화 오차는 hypothesis 공간 에서 target 공간으로 근사화 되는 과정에서 발생하는 오차이다.

- 추정 오차 : 추정 오차는 hypothesis 공간 으로부터 최적화 되지 못한 모델을 선택할 때 발생하는 오차로서 learning 과정 중에 발생하는 오차이다.

2. Empirical Risk Minimization principle

Statistical learning theory 에서, learning 문제를 푸는 가장 표준적인 방법은 추정기에 의한 오차의 평균적 양을 측정하는 risk function 을 정의하여 사용하는 방법이다[1]. 예를 들어 집합 $X \times Y$ 에서 샘플링 시간 l 시간에서 얻어진 training data $D_l = \{(x_i, y_i) \in X \times Y\}_{i=1}^l$ 이 있을 때, $V(y_i, f(x_i))$ 가 우리가 $f(x)$ 로부터 y 를 추정할 때 만드는 오차를 측정하는 loss 함수라고 할 때 평균 오차를 expected risk 라 부르고 정의는 다음과 같다

$$R(f) = \int V(y, f(x))P(x, y)d(x)dy. \quad (2.1)$$

그리고 F 에서 expected risk 를 최소화 하는 함수를 f_0 라 정하면 정의는 다음과 같다.

$$f_0 = \arg \min_f R[f] \quad (2.2)$$

여기서 이 함수 f_0 를 우리는 ideal 추정기라 하고 보통 target 함수라 부르기도한다. 하지만 expected risk 를 정의하는 확률 분포 함수 $P(x, y)$ 를 모르기 때문에 ideal 추정기를 구할 수 없고, 그래서 이러한 문제점을 극복하기 위하여 Vapnic[1]은 empirical risk minimization principle 을 제안한다.

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^l V(y_i, f(x_i)) \quad (2.3)$$

$$\lim_{l \rightarrow \infty} R_{emp}[\hat{f}_l] = \lim_{l \rightarrow \infty} R[\hat{f}_l] = R[f_0] \quad (2.4)$$

\hat{f}_l 이 수식 (2.3)을 최소화하는 함수라 할 때 수식 (2.4)는 empirical risk minimization principle 이 만족되는 조건을 보여 준다.

3. Structure Risk Minimization Principle

Vapnik Chervonenkis(VC) dimension h 를 다음과 같이 정의 한다.

정의 2.1 [3] : The capacity of a set of function with logarithmic bounded growth function can be characterized by the coefficient h . The coefficient h is called the VC dimension of a set of indicator functions. It characterizes the capacity of a set of functions. When the growth function is linear the VC dimension is defined to be infinite.

Vapnik Chervonenkis 의 중요한 특징은 다음 theorem 에서 보여지는 것 처럼, Hypothesis 공간상에서 empirical risk 와 expected risk 사이의 uniform deviation 이 VC-dimension 으로 제한된다는 것이다.

Theorem 2.1 [3] : Let $A \leq V(y, f(x)) \leq B, f \in F$, with A and $B < \infty$, F be a set of bounded functions and h the VC-dimension of V in F . Then, with probability at least $1 - \eta$, the following inequality holds simultaneously for all the elements f of F :

$$R_{emp}[f] - (B-A) \sqrt{\frac{h \ln \frac{2l}{h} - \ln \frac{\eta}{4}}{l}} \leq R[f] \leq R_{emp}[f] + (B-A) \sqrt{\frac{h \ln \frac{2l}{h} - \ln \frac{\eta}{4}}{l}} \quad (2.5)$$

보통 $R[f] - R_{emp}[f]$ 를 추정 오차라 부른다. Loss function V 가 정의된 공간 F 는 매우 크기 때문에 보통 좀더 작은 hypothesis 공간 H 를 고려한다. 이 제한된 공간에 관련한 cost 를 근사화 오차라 부른다.

hypothesis 공간들을 $H_1 \subset H_2 \subset \dots \subset H_{n(l)}$ 로 정의하고 각각의 공간 H_i 의 VC dimension 을 h_i 라 하면 $h_1 \leq h_2 \leq \dots \leq h_{n(l)}$ 이 성립하여 수식(2.5)가 수식 (2.6)같이 바뀌어 써질 수 있다.

$$R[f] \leq R_{emp}[f] + (B-A) \sqrt{\frac{h \ln \frac{2l}{h} - \ln \frac{\eta}{4}}{l}} \quad (2.6)$$

그래서 Structural risk minimization induction principle 의 기본적인 idea 는 다음과 같이 이야기 할 수 있다.

용납 가능한 structure 를 가지는 함수들의 집합을 제공하고, guaranteed risk 를 최소화하는 함수를 찾기 위하여 structure 의 주어진 element 들 상에서 수식 (2.6)을 만족 시키도록 최소화하는 것이다.

2.2 Partitioning of Input Space[4]

data로부터 퍼지 규칙을 만들어내기 위한 방법은 크게 partitioning 방법과 clustering 방법의 이 두가지 종류로 나누어 진다.

Partitioning 방법은 다시 다음 4가지 경우로 방법으로 나누어 진다.

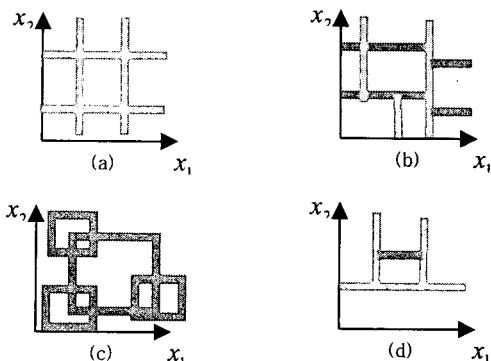


그림 2.1 여러가지 input 공간 구분 방법. (a) Grid partition, (b) Tree partition, (c) Scatter partition, (d) GA algorithm based partition

두 번째로 Clustering 방법으로 다음 방법들로 세분화 할 수 있다.

- Fuzzy C-mean Clustering [8] : Fuzzy C-mean clustering 방법은 n 개의 벡터 $x_j, j=1, \dots, n$ 를 c 개의 groups $G_i, i=1, \dots, c$ 로 구분하고, 각 group 에서 cost 함수를 최소화 하는 cluster 중앙을 찾는 방법이다.
- Mountain Clustering [9] : Mountain clustering 방법은 밀도 측정 기반 상에서 cluster 중앙을 예측하는 비교적 간단하고 효과적인 방법이다.

표 2.1 에서 퍼지 규칙을 뽑아내기 위한 다양한 input 공간 구분 방법들이 단점과 함께 정리되어 있다. 현재까지의 structure identification에서 input 공간 구분 방법들은 파라미터의 값들을 결정하는 parameter identification 과는 구분되어져 있다. 더욱이 이러한 처리 방법들은 parameter와 structure 최적화를 동시에 필요한 시스템에 대하여는 불가능 하다. 하지만 앞절에서 설명한 generalization theory 를 기반으로 한 SVFIS 방법은 규칙 기반 최적화를 통해 퍼지 규칙의 개수를 최적화 할 수 있을 뿐만 아니라, Parameter 값들도 최적화 할 수 있다.

3. Support Vector Fuzzy Inference System

이 부분에서는 TS 퍼지 모델[5]을 기반으로 한 SVFIS의 구조를 설명하고, SVFIS의 learning 방법을 설명한다.

TS 퍼지 모델에서 퍼지 규칙의 집합은 수식 (3.1) 과 같이 표현 된다.[5]

R1: If x_{11} is M_{11} and ... and x_{1D} is M_{1D} , Then $f_1 = a_{10} + a_{11}x_{11} + \Lambda + a_{1D}x_{1D}$

R2: If x_{21} is M_{21} and ... and x_{2D} is M_{2D} , Then $f_2 = a_{20} + a_{21}x_{21} + \Lambda + a_{2D}x_{2D}$

Rn: If x_{n1} is M_{n1} and ... and x_{nD} is M_{nD} , Then $f_n = a_{n0} + a_{n1}x_{n1} + \Lambda + a_{nD}x_{nD}$ (3.1)

여기서 n은 퍼지룰의 개수고, D는 입력변수의 차원이며 x 는 입력변수, f 는 출력변수, M은 퍼지 집합이고, a 는 결정 파라미터들이다.

SVFIS의 구조에 대하여 설명하면 총 4개의 계층으로 구성되어 있고, 그림 3.1에 구조가 나타나 있다.

Group	방법	단점
Partition	Grid Partition	Course of dimensionality
	Tree Partition	Number of rule exponential increasing
	Scatter Partition	Completeness not guaranteed
	GA Algorithm based Partition	Long learning time
Clustering	Fuzzy C-mean	Predetermined the number of clustering
	Mountain clustering	Let perception grid points as the candidate for clustering center
	Hybrid clustering	Depending on implementation

표 2.1 퍼지 규칙을 뽑기위한 여러가지 input 공간 구분 방법

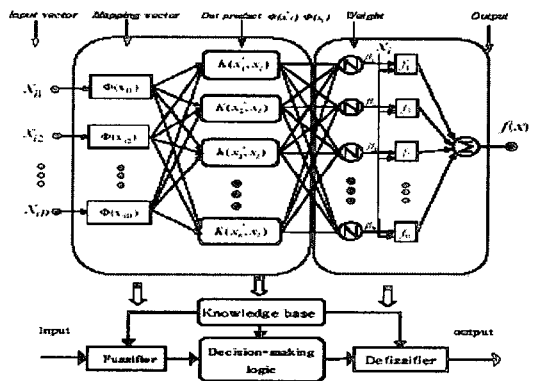


그림 3.1 : SVFIS의 구조

- 1 계층 : 입력 공간이 어떠한 특정 매핑에의하여 특정 공간으로 비선형 사상된다.
- 2 계층 : 사상된 입력 x 와 입력 벡터 x 의 부분집합인 support vector와 dot product를 계산한다. 커널 함수를 생성하는 단계이다. 본 논문에서는 가우시안 커널 함수를 사용하였다.
- 3 계층 : SVM 알고리즘으로부터 비선형 함수를 결정한다. 이 계층에서 최적화 문제를 해결함으로써 퍼지룰의 개수를 정하게 된다.
- 4 계층 : 전체적으로 구성된 퍼지 모델의 결과를 이용하여 디퍼지피케이션이 수행된다.

다음으로 가우시안 커널 파라미터와 결론부 파라메

터를 조정하는 support vector learning 알고리즘을 살펴본다.

1단계 : 정밀도를 초기화하고, 상수와 커널 파라미터들을 정한다.

2단계 : SVM 알고리즘을 이용하여 support vector를 찾는 후에 이 support vector들을 가우시안 소속 함수의 중심으로 정한다.

3단계 : LSE 또는 RLSE방법을 이용하여 수식(3.1)에서의 선형 함수 f의 파라미터 a를 예측한다.

4단계 : gradient descent 알고리즘을 이용하여 에러가 최소화 되도록 커널 파라미터를 고친다.

5단계 : 다시 2단계로 돌아가거나 적절한 에러수준에 도달하면 멈춘다.

4. 실험 및 결과

본 논문에서는 위에서 제안한 SVFIS의 pattern recognition 적용 가능성 및 성능을 파악하기 위하여 100개의 3차원 물체를 5도씩 돌려가면서 image file로 생성한 COIL-100(Columbia Object Image Library 100)을 가지고 실험을 수행해 Recognition의 정도를 파악해 보았다.

실험을 수행하기 위하여 각 물체마다 5도씩 돌려가며 찍은 총 72장의 사진 중에 절반인 36장의 사진을 이용하여 learning data로 사용하였으며 나머지 36장의 데이터를 test data로 사용하였다.

Recognition의 성능 정도를 파악하기 순수하게 SVM 만을 이용한 방법[6]과, SVFIS를 이용한 방법 그리고 neuro-fuzzy classifier를 이용한 방법[7]으로 실험을 해 보았다. SVM을 이용한 방법은 1:1 matching 방법을 이용하였으며 neuro fuzzy classifier의 룰의 개수를 정하기 위해 Scatter partitioning 방법을 이용하였다.

다음 표들은 각각의 경우에 대하여 평균 에러를 백분율로 나타낸 것이며 처음 표는 노이즈가 없는 것에 대한 사진을 이용한 결과이고 두번째 표는 노이즈가 포함된 사진을 이용한 결과 이다. 노이즈를 포함 시킬 때 [-150 +150]의 uniformly distributed 된 random noise를 사용하였다.

방법	평균 에러
SVM	2.5%
SVFIS	3.4%
Neuro Fuzzy	8.7%

표 4.1 : 노이즈가 없는 영상에 대한 결과

방법	평균 에러
SVM	6.2%
SVFIS	5.6%
Neuro Fuzzy	11%

표 4.2 : 노이즈가 존재하는 영상에 대한 결과 위의 결과에서 보이는 것처럼 Neuro Fuzzy classifier의 경우 Scatter partitioning방법을 사용함으로 인해 completeness가 만족되지 않아 오차가 다른 방법에 비해 큼을 알 수 있다. 그리고 노이즈가 없는 경우 순수한 SVM에 의한 방법이 더 좋게 나타남을 볼 수 있고, 노이즈가 포함된 경우 SVFIS방법이 더 좋게 나타남을 볼 수 있다.

5. 결론

본 논문에서는 Support Vector Fuzzy Inference System이라는 hybrid 시스템을 제안하여 이를 pattern recognition에 적용하여 보았다. Fuzzy Inference system의 Structure identification과 parameter identification을 위하여 SVM기법을 이용하였으며 이를 사용함으로써 Completeness를 만족시킬 수 있었다.

[참고문헌]

- [1] W.C.Chan, K.C.Cheung and C.J.Hariss, "On the Modeling of Nonlinear Dynamic Systems Using Support Vector Neural Networks," Engineering Application of Artificial Intelligence 14, pp. 105-113 2001
- [2] Jin-Tsong Jeng and Tsu-Tain Lee, "Support Vector Machines for the fuzzy Neural Networks" IEEE SMC '99 Conference Proceedings, Vol 6, pp. 115-120, 1999

- [3] V.N.Vapnic, “ Statistical Learning Theory,” Springer-Verlag 1999.
- [4] J.S.R.Jang, C.T.Sun, and E. Mizutani, “Neuro-Fuzzy and Soft Computing”, Prentice-Hall, 1997
- [5] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications for modeling and control,” IEEE Trans. On Systems, Man, and Cybernetics, Vol. 15, pp. 116-132, 1985
- [6] Massimiliano Pontil and Alessandro Verri, “Support Vector Machines for 3-D Object Recognition”
- [7] C. T. Sun, and J. S. Jang, “A neuro-fuzzy classifier and its applications,” Proc. IEEE Int. Conf. Fuzzy Syst. San Francisco, Vol I, pp. 94-98,1993