

계층적 구조를 가진 퍼지 패턴 분류기 설계

A Design of Fuzzy Classifier with Hierarchical Structure

안태천* · 노석범* · 김용수**

Tae-Chon Ahn*, Seok-Beom Roh*, and Yong Soo Kim**

*원광대학교 전자융합공학과

**대전대학교 컴퓨터공학과

Dept. of Electronics Convergence Engineering, Wonkwang University, Chonbuk, Korea

† Dept. of Computer Engineering, Daejeon University, Daejeon, Korea

요 약

본 논문은 단순한 후반부 구조를 가진 퍼지 모델을 계층적 구조로 결합한 퍼지 패턴 분류기를 제안한다. 계층적 구조를 가진 퍼지 패턴 분류기의 기본 구조는 단순한 후반부 구조를 가진 퍼지 모델을 사용하여 전체 패턴 분류기의 구조적 복잡성을 높이지 않도록 설계 하였다. 입력공간을 계층적으로 분할하기 위하여 대표적인 퍼지 클러스터링 알고리즘인 Fuzzy C-Means clustering 기법을 이용하였다. 분할된 퍼지 입력 공간의 하위 구조를 분석하기 위하여 conditional Fuzzy C-Means 클러스터링 기법을 이용하였다. 계층적으로 분할된 퍼지 입력공간에 간단한 구조를 가진 퍼지 패턴 분류기를 적용하여 계층적 구조를 가진 패턴 분류기를 설계한다. 계층적으로 퍼지 모델들을 결합함으로써 입력 공간의 정보 분석을 거시적인 관점에서 시작하여 세부적으로 분석이 가능하게 되었다. 제안된 퍼지 패턴 분류기의 성능을 평가하기 위하여 다양한 기계 학습 데이터를 사용하였다.

키워드 : 퍼지 패턴 분류기, TSK 퍼지 모델, 계층적 구조, 퍼지 클러스터링, 퍼지 KNN

Abstract

In this paper, we proposed the new fuzzy pattern classifier which combines several fuzzy models with simple consequent parts hierarchically. The basic component of the proposed fuzzy pattern classifier with hierarchical structure is a fuzzy model with simple consequent part so that the complexity of the proposed fuzzy pattern classifier is not high. In order to analyze and divide the input space, we use Fuzzy C-Means clustering algorithm. In addition, we exploit Conditional Fuzzy C-Means clustering algorithm to analyze the sub space which is divided by Fuzzy C-Means clustering algorithm. At each clustered region, we apply a fuzzy model with simple consequent part and build the fuzzy pattern classifier with hierarchical structure. Because of the hierarchical structure of the proposed pattern classifier, the data distribution of the input space can be analyzed in the macroscopic point of view and the microscopic point of view. Finally, in order to evaluate the classification ability of the proposed pattern classifier, the machine learning data sets are used.

Key Words : Fuzzy Pattern Classifier, TSK Fuzzy model, Hierarchical Structure, Fuzzy Clustering, Fuzzy KNN

1. 서 론

현재는 스마트폰, 태블릿 컴퓨터와 같은 모바일 기기의 보급이 일반화 되었다. 이러한 모바일 기기의 일반화로 인

하여 일반 사용자들에 의하여 SNS와 같은 어플리케이션 프로그램을 통해 생산되는 다양하고 복잡한 정보가 폭증하고 있다. 이와 같은 다양한 데이터를 분류하여 의미를 부여하는 데이터 처리과정을 통해 양질의 정보를 획득하여 개인의 삶의 질을 개선할 수 있다. 위에 언급한 바와 같이 다양한 형태의 데이터들을 특성별로 분류하여 데이터 군집들이 가지는 의미를 파악하는 것은 데이터 처리의 주요한 업무들 중 하나이다. 이러한 데이터의 특성을 분석하여 특성에 따라 분류하기 위한 다양한 패턴 분류기들이 연구되어 왔다 [1].

본 논문에서는 획득된 데이터들의 입력 공간을 퍼지 공간으로 분할하고, 분할된 공간에 간단한 형태의 패턴 분류기를 설계하고, 각각의 지역적 패턴 분류기를 퍼지 결합하는 형태의 퍼지 패턴 분류기를 제안하였다. 제안된 패턴 분류기의 설계 과정에서, 데이터의 입력공간 분할은 Bezdek

접수일자: 2014년 3월 9일

심사(수정)일자: 2014년 4월 1일

게재확정일자 : 2014년 8월 4일

† Corresponding author

본 논문은 2014학년도 원광대학교의 교비지원에 의해서 수행 됨

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

에 의해 제안된 Fuzzy C-Means Clustering(FCM)[2] 기법과 Pedrycz가 제안한 Conditional Fuzzy C-Means Clustering 기법을 사용하였다[5]. FCM에 의해 분할된 각각의 입력공간에서, 주어진 데이터들의 패턴 분류를 위하여 간단한 형태의 Prototype 기반 패턴 분류기를 사용한다. 일반적인 Prototype 기반 패턴 분류기들은 새로운 데이터가 주어질 경우, 미리 정의된 Prototype들의 위치 정보를 이용하여 새롭게 주어진 데이터와 가장 가까운 곳에 위치한 Prototype을 결정한다. 주어진 데이터와 가장 근접한 Prototype의 특성이 주어진 데이터의 특성과 유사할 가능성이 가장 높다는 가정 하에서 주어진 데이터의 소속 클래스를 근접한 Prototype의 소속 클래스로 결정하게 된다.

여러 가지 종류의 Prototype 기반 패턴 분류기들 중에서 Nearest Neighbor Quantization(NNQ) 기법을 사용하였고, NNQ에서 각 클래스를 대표하는 prototype을 정의하는 방법을 제안하였다.

Prototype 패턴 분류기는 prototype의 위치가 패턴 분류기의 패턴 분류 성능을 결정하는 중요한 요소이다. 일반적으로 prototype 패턴 분류기의 prototype의 위치는 Learning Vector Quantization (LVQ) [6]과 같은 알고리즘을 이용하여 결정한다.

본 논문에서는 prototype의 위치를 결정하기 위하여 퍼지 분할된 부 공간에서 각 클래스에 속한 데이터들의 특성을 분석하여 prototype의 위치를 결정하게 된다.

본 연구에서 제안된 패턴 분류기의 성능을 평가하기 위하여 여러 개의 머신러닝 데이터를 이용하여 기존의 패턴 분류기의 패턴 분류 성능을 비교, 평가한다.

2. 퍼지 Prototype 패턴 분류기

제안된 퍼지 prototype 기반 패턴 분류기는 입력 공간의 퍼지 분할 단계와 분할된 퍼지 부공간에서 prototype 기반 패턴 분류기 설계 단계로 구성되어 있다. 입력 공간의 퍼지 공간 분할은 퍼지 클러스터링 기법인 Fuzzy C-Means clustering 기법을 이용한다.

2.1 Fuzzy C-Means Clustering

FCM 클러스터링은 n 개의 벡터 $x_i (i = 1, \dots, n)$ 집합을 c 개의 클러스터로 분할하고, 목적함수가 최소가 일 때 생성된 각 클러스터에서 중심 값을 찾는다. FCM 클러스터링의 목적함수는 식(1)와 같다.

$$J(u^{fcm}, v^{fcm}) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik}^{fcm})^m (\|x_k - v_i^{fcm}\|)^2 \text{ s.t.} \quad (1)$$

$$\sum_{i=1}^c u_{ik}^{fcm} = 1$$

목적함수를 최소화하는 u^{fcm} 와 v^{fcm} 는 식(2), (3)과 같이 얻을 수 있다.

$$u_{ik}^{fcm} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{fcm}\|}{\|x_k - v_j^{fcm}\|} \right)^{2/m-1}} \quad (2)$$

$$v_{ij}^{fcm} = \frac{\sum_{k=1}^N (u_{ik}^{fcm})^m x_{ki}}{\sum_{j=1}^N (u_{ik}^{fcm})^m} \quad (3)$$

여기서, u^{fcm} 는 데이터의 퍼지 집합 소속 멤버십 함수를 의미하며, v^{fcm} 는 해당 하는 퍼지 집합의 중심점을 의미한다.

Fuzzy C-Means Clustering 알고리즘의 흐름은 아래와 같다.

step 1. 클러스터 개수 ($2 \leq c \leq N$) 결정, 퍼지화 계수 ($1 < m < \infty$)를 선택, 소속함수 $U^{(0)}$ 를 초기화 한다.

step 2. 각 클러스터의 중심 벡터를 식 (3)를 이용하여 결정한다.

step 3. 각 클러스터의 중심벡터와 주어진 데이터와의 거리를 계산하여 새로운 소속 행렬 $U^{(r+1)}$ 을 식 (2)을 이용하여 결정한다.

step 4. 오차가 허용범위 안에 도달하면 종료하고, 그렇지 않으면 [Step 2]로 돌아간다.

2.2 Conditional Fuzzy C-Means Clustering

Conditional Fuzzy C-Means 클러스터링은 n 개의 벡터 $x_i (i = 1, \dots, n)$ 집합을 c 개의 클러스터로 분할하고, 목적함수가 최소가 일 때 생성된 각 클러스터에서 중심 값을 찾는다. CFCM 클러스터링의 목적함수는 식(4)와 같다.

$$J(\mu, v) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik}^{cfcm})^m (\|x_k - v_i^{cfcm}\|)^2 \text{ s.t.} \quad (4-1)$$

$$\sum_{i=1}^c u_{ik}^{cfcm} = F \quad (4-2)$$

여기서, F 는 사전에 정하여진 상수 값이다. 본 논문에서는 $F_{ik} = u_{ik}^{fcm}$ 으로 설정하여 한 번 분할된 입력 공간을 다시 한번 퍼지 분할 한다.

목적함수를 최소화하는 u^{cfcm} 와 v^{cfcm} 는 식(5), (6)과 같이 얻을 수 있다.

$$u_{ik}^{cfcm} = \frac{F}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{cfcm}\|}{\|x_k - v_j^{cfcm}\|} \right)^{2/m-1}} \quad (5)$$

$$v_{ij}^{cfcm} = \frac{\sum_{k=1}^N (u_{ik}^{cfcm})^m x_{ki}}{\sum_{j=1}^N (u_{ik}^{cfcm})^m} \quad (6)$$

여기서, u^{cfcm} 는 데이터의 퍼지 집합 소속 멤버십 함수를 의미하며, v^{cfcm} 는 해당 하는 퍼지 집합의 중심점을 의미한다.

Conditional Fuzzy C-Means Clustering 알고리즘의 흐름은 아래와 같다.

step 1. 클러스터 개수 ($2 \leq c \leq N$) 결정, 퍼지화 계수 ($1 < m < \infty$)를 선택하고, (4-2)를 만족하는 소속함수 $U^{(0)}$ 를 초기화 한다.

step 2. 각 클러스터의 중심 벡터를 식 (6)를 이용하여 결정한다.

step 3. 각 클러스터의 중심벡터와 주어진 데이터와의 거리를 계산하여 새로운 소속 행렬 $U^{(r+1)}$ 을 식 (5)을 이용하여 결정한다.

step 4. 오차가 허용범위 안에 도달하면 종료하고, 그렇지 않으면 [Step 2]로 돌아간다.

2.3 Prototype 패턴 분류기

제안된 패턴 분류기는 퍼지 클러스터링 기법에 의해 분할된 부공간의 데이터 패턴들을 패턴 분류하기 위하여 Prototype 기반 패턴 분류기의 일종인 Nearest Neighbor Quantization(NNQ) 기법을 사용한다.

일반적인 Prototype 기반 패턴 분류기는 전체 입력 공간에 미리 지정된 개수의 prototype을 설정하고 새롭게 주어진 데이터와 prototype들과의 거리를 비교하여 주어진 데이터의 소속 클래스가 가장 가까운 prototype의 소속 클래스와 동일하게 설정된다. 그러나 제안된 패턴 분류기는 전체 입력공간의 특성을 분석하여 prototype의 위치를 설정하기 않고, 공간 분할된 부공간의 특성을 분석하여 prototype의 위치를 결정한다.

CFCM에 의해 분할된 i 번째 공간에서 j 번째 클래스를 대표하는 prototype의 위치는 (7) 와 같다.

$$v_{ij} = \frac{\sum_{k=1, y_k \in class_j}^m u_{ik}^{cfcm} \cdot x_k}{\sum_{k=1, y_k \in class_j}^m u_{ik}^{cfcm}} \quad (7)$$

여기서, m 은 데이터 개수를 나타내며, $class_j$ 는 j 번째 클래스에 속한 데이터 집합을 의미한다.

새롭게 주어진 데이터 x_n 와 i 번째 부공간에서 정의된 prototype들 간의 유사도는 (8) 같이 정의한다.

$$s_{ij} = \frac{1}{\sum_{k=1}^l \left(\frac{\|x_n - v_{ik}\|}{\|x_n - v_{jk}\|} \right)^{2/m-1}} \quad (8)$$

제안된 퍼지 패턴 분류기는 (8)과 같이 정의된 유사도들이 퍼지 결합된 형태의 분류기이며, 제안된 퍼지 패턴 분류기의 퍼지 규칙은 (9)과 같다.

Rule i : IF x is R_i THEN y_i is RW_i ($1 \leq i \leq c$)
(9)

여기서, $RW_i = [s_{i1} \ s_{i2} \ \dots \ s_{il}]$ 이며 i 번째 클러스터에서 주어진 데이터와 각 클래스를 대표하는 prototype과의 유사도를 나타내는 벡터이다.

그림 1은 퍼지 규칙 (9)를 가진 퍼지 prototype 패턴 분류기 구조도 이다.

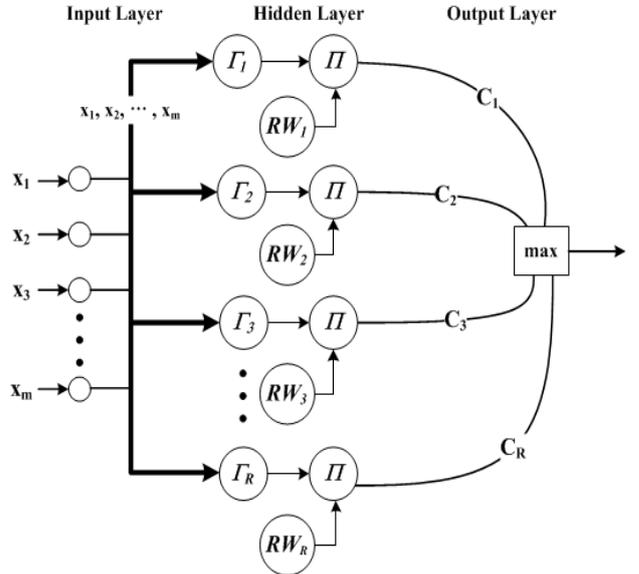


그림 1. 퍼지 prototype 패턴 분류기 구조

Fig. 1. Structure of Fuzzy Prototype Pattern Classifier

위와 같은 퍼지 규칙을 가진 퍼지 패턴 분류기의 최종 출력식은 (10)과 같다.

$$R = [RW_1 \ RW_2 \ \dots \ RW_m] = \frac{\sum_{i=1}^c \mu_i s_i}{\sum_{i=1}^c \mu_i} \quad (10)$$

여기서, R 는 주어진 데이터가 각각의 클래스에 소속할 정도를 의미한다고 할수 있다.

이와 같이 정의된 클래스 유사도 R 를 이용한 데이터의 소속 클래스는 (11)과 같이 결정한다.

$$L(x) = \max_k S_k \quad (11)$$

3. 시뮬레이션 및 결과 고찰

본 논문에서 제안한 패턴 분류기의 성능을 평가하기 위하여 여러 개의 머신러닝 데이터들을 사용한다. 기계 학습 데이터 집합은 대표적인 Benchmark 데이터 집합인 UCI machine learning repository로부터 얻은 데이터 집합들이다. Benchmark 데이터 집합을 이용하여 제안된 패턴 분류기의 성능과 특성을 기존 논문에서 이미 제안된 패턴 분류기의 성능과 비교, 평가 한다. 연구되어진 기존 패턴 분류기와 비교하기 위하여, 전체 데이터를 10 fold cross validation 방법에 따라 학습 데이터와 테스트 데이터로 나누어 실험 한다.

표 1은 제안된 퍼지 결합 다항식 뉴럴 네트워크 기반 패턴 분류기의 설계를 위해 미리 설정되어야 하는 파라미터들을 보인다.

표 1. 설계 파라미터
Table 1. Design Parameters

Parameter	Value
Number of fuzzy rule	2, 3, 5, 7
Number of sub fuzzy Rule	2, 3
Fuzzifier Coefficient (m)	1.2, 1.5, 2.0, 2.5, 3.0

제안된 패턴 분류기의 성능을 평가하기 위한 기계 학습데이터에 대한 개략적인 정보는 표 2에 열거 하였다.

표 2. 실험에 사용된 기계학습 데이터
Table 2. Machine Learning Data used in the experiments

Datasets	Number of features	Number of Data	Number of Classes
Balance	4	625	3
Glass	10	214	7
Thyroid	5	215	3
Wine	12	178	3

표 3은 10 Fold Cross Validation에 의한 데이터 분할 후 실험 결과를 분석, 비교한 결과이다. 패턴 분류 성능을 위하여 패턴 분류기의 오분류율을 평가 지수로 선택 하였다.

표 3. 제안된 패턴 분류기와 기존 패턴 분류기 성능 비교
Table 3. Result of Comparative analysis

	Balance	Glass	Thyroid	Wine
PFARS[3]	33.1	N/A	N/A	4.0
CCP[4]	N/A	28.51	N/A	N/A
PART(WEKA)	16.83	31.25	N/A	N/A
SMO(WEKA)	12.43	42.64	N/A	N/A
Proposed Classifier	12.15	38.78	12.08	3.40

4. 결론 및 향후 연구 방향

본 논문에서는 prototype 패턴 분류기를 기반으로 하여 입력 공간을 퍼지 분할 한 후 conditional Fuzzy C-Means Clustering 알고리즘을 이용하여 퍼지 부공간을 다시 한 번 공간을 분할 한 후 제안된 prototype 패턴 분류기를 각각의 부공간에 적용하는 퍼지 결합한 형태의 퍼지 패턴 분류기를 제안한다.

제안된 패턴 분류기는 Fuzzy C-Means clustering 기법을 이용하여 입력 공간을 퍼지 분할 하고 분할된 퍼지 부 공간에 다시 한번 Conditional Fuzzy C-means Clustering 기법을 적용하여 분할된 공간의 데이터 분포를 분석한다. 이와 같은 알고리즘의 흐름은 일반적으로 알려진 전역적 pro-

totype 기반 패턴 분류기에 비해 지역적 특성을 강조하는 패턴 분류기를 설계하고자 하는 설계흐름 이다. 제안된 패턴 분류기의 성능을 평가하기 위하여 기계학습 데이터를 이용하여 평가하였고 평가 결과 우수한 결과를 얻었다. 향후 제안된 패턴 분류기의 지역적 성능을 강화하기 위하여 분할된 퍼지 공간의 지역적 특성을 자세히 파악할 수 있는 기법을 개발할 필요가 있다 하겠다.

References

- [1] G. Pajares, M. Guizarro, A. Ribeiro, "A Hopfield Neural Network for combining classifiers applied to textured images," *Neural Networks*, Vol.23, pp. 144-153, 2010.
- [2] J. C. Bezdek, "Pattern Recognition With Fuzzy Objective Function Algorithms", New York: Plenum, 1981.
- [3] A. Bargiela, W. Pedrycz, "Granular Computing: An Introduction," Kluwer Academic Publishers, Dordrecht, 2003
- [4] W. Pedrycz, "Analysis and Design of Intelligent Systems: A Framework of Granular Computing," CRC Press, Boca Raton, FL, 2013
- [5] W. Pedrycz, "Conditional Fuzzy C-Means," *Pattern Recognition Letters*, vol. 17, no. 15, pp 625-631, 1996
- [6] R. Parades, and E. Vidal, "Learning prototypes and distance: A prototype reduction technique based on nearest neighbor error minimization," *Pattern Recognition*, vol. 39, pp. 180-188, 2006.

저 자 소개



안태천 (Tae-Chon Ahn)

1978년: 연세대학교 전기공학과 공학사
 1980년: 연세대학교 전기공학과 공학석사
 1986년: 연세대학교 전기공학과 공학박사
 1981년 ~ 현재: 원광대학교 전자융합공학과 교수
 2013년 ~ 현재: 한국지능시스템학회 이사

관심분야 : Computational Intelligence, Soft Computing
 Fuzzy Control, Pattern Recognition

Phone : +82-63-850-6344
 Fax : +82-63-853-2196
 E-mail : tcahn@wku.ac.kr



노석범 (Seok-Beom Roh)

1994년 원광대학교 제어계측공학과 학사
 1996년 동 대학원 컴퓨터공학과 석사
 2006년 동 대학원 제어계측공학과 박사

관심분야 : Bio-inspired Optimization Algorithm,
Fuzzy Theory, Neural Networks,
Pattern Recognition

Phone : +82-63-850-6344

E-mail : nado@wku.ac.kr



김용수(Yong Soo Kim)

1981년 연세대학교 전기공학과 공학사

1983년 KAIST 전기및전자공학과 공학
석사

1986년 삼성전자종합연구소 주임연구원

1993년 Texas Tech Univ. 공학박사

1995년~현재 대전대학교 컴퓨터공학과
교수

관심분야 : 신경회로망, 퍼지논리, 패턴인식, 영상처리,
침입탐지 등

Phone : +82-42-280-2547

Fax : +82-42-280-2889

E-mail : kystj@dju.kr