

퍼지 분류기를 위한 통계적 정보 기반의 퍼지 함수 설정 기법

신상호* · 한수환** · 우영운**

*동의대학교 디지털미디어공학과

**동의대학교 멀티미디어공학과

Creation Methods of Fuzzy Membership Functions Based on Statistical Information for Fuzzy Classifier

Sang-Ho Shin* · Soowhan Han** · Young Woon Woo**

*Dept. of Digital Media Engineering, Dong-Eui University

**Dept. of Multimedia Engineering, Dong-Eui University

E-mail : shsh@deu.ac.kr, swan@deu.ac.kr, ywoo@deu.ac.kr

요 약

패턴 인식에서 분류기 모형으로 많이 사용되는 퍼지 분류기는 퍼지 소속 함수를 적절히 설정함으로써 보다 향상된 분류 성능을 얻을 수 있다는 장점이 있다. 그러나 일반적으로 함수 설정은 인식 문제 분야의 특성이나 해당 전문가의 지식과 주관적 경험을 기반으로 설정되므로 설정된 소속도 함수의 일관성과 객관성을 보장하기가 어려운 문제점을 갖고 있다. 따라서 이 논문에서는 퍼지 분류기의 소속도 함수를 설정하기 위한 객관적 기준을 제시하기 위하여 특징값들 간의 통계적 정보를 이용한 소속도 함수 설정 기법들을 제안하였다. 제안한 기법들을 이용하여 UCI machine learning repository 사이트에서 제공되는 표준 데이터 중에 Iris 데이터 세트를 이용하여 실험하고 그 결과를 비교, 분석하였다.

키워드

퍼지 분류기, 통계적 정보, 소속 함수 설정, Iris 데이터 세트

1. 서 론

퍼지집합은 애매모호한 값을 표현하는데 적합하여 그동안 많은 다양한 분야에서 다루기 위한 여러 가지 연구가 진행되어 왔다. 퍼지 집합은 애매모호한 값을 가능성 분포를 이용하여 표현한 것으로, 관심 있는 모든 원소들을 포함한 전체집합에서 어떤 특정한 성질을 갖는 원소들과 그 원소들의 집합에 대한 소속 가능성 만족도를 이용해서 나타낸 집합이다. 이에 따라 퍼지집합은 각 원소들의 집합에 대한 만족도를 임의로 설정할 수 있어서, 전통적인 고전집합(crisp set)이 표현하지 못하는 불분명성과 애매성을 표현할 수 있다 [1].

이 논문은 퍼지 집합의 만족도를 나타내기 위하여 사용 되는 이론적인 소속 함수에 대한 통계적인 정보를 이용하는 함수를 제안한다. 본 연구의 동기는 일반적인 퍼지집합의 소속 함수에서, 소속 함수 값에 따른 결과에 영향을 미치는 중요

도에 차이를 어떻게 둘 것인지에 대한 문제로부 터 제시되었다.

이러한 관점에서 접근하고자 하는 기존의 연구는 대부분 퍼지집합의 형태에 대한 것과 문제분야의 특성이나 전문가의 지식과 주관적 경험에 의해 소속 함수의 정도를 결정하고 있으며, 그 결과로는 기본적인 사다리꼴 퍼지 집합, 삼각 퍼지 집합, 가우시안 퍼지 집합 등이 있다[2]. 그러나 어떤 시스템을 구성하는 데 있어, 이러한 퍼지집합들 중 경험적인 성분으로만 정하여 사용하는 것은 시스템 사용 중 소속 함수의 특정 부분에 대한 선호도가 달라지는 경우에 대처하기 힘들게 된다.

이 논문에서는 퍼지집합의 소속 함수의 정도 산출을 위해 통계적인 정보를 이용하여 기본 형태를 변형하지 않더라도 보다 나은 결과를 이끌 수 있는 배경이 되기 위한다는 점에서 의미가 있다고 할 수 있다.

II. 퍼지 함수 설정 기법

퍼지 함수를 이용하여 퍼지 분류기를 구현하기 위해서는 퍼지 소속 함수를 개념화 하여야 한다. 또한 이 개념을 기존의 패턴 분류에 사용되는 함수와 퍼지 소속 함수를 정의하기 위해 그림 1에서 나타나 있는 것 처럼 일반적으로 많이 사용되는 '싱글톤(Singleton)', '선형(Linear)', '삼각형(Triangular)', '사다리꼴(Trapezoidal)', '가우시안(Gaussian)', '시그모이드(Sigmoid)'와 같은 함수의 형태를 개념화하고 표현한다. 퍼지 소속 함수의 속성 중 'min', 'center', 'max' 중에서 반드시 설정되어야 하며, 설정된 퍼지 소속 함수의 속성값들로 클래스를 구분할 수 있다.

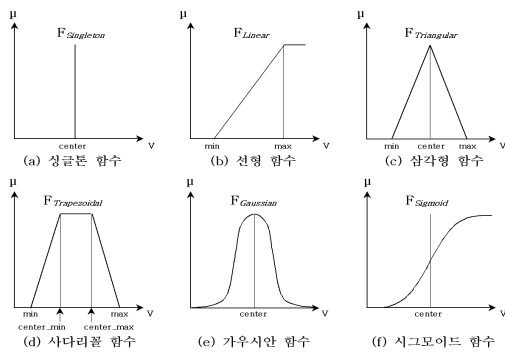


그림 1. 퍼지 소속 함수의 종류

그 중에서도 삼각형과 사다리꼴 퍼지 함수는 중심과 기울기(center, min, max)를 설정하는데 학습 데이터의 클래스별 특징에 대한 최소값, 최대값, 평균, 분산을 이용하여 퍼지 함수를 구성하게 된다.

아래 그림들에 나타나 있는 것처럼 특징값 벡터들의 평균값만이 퍼지 함수의 중심값으로 사용되었다.

그림 2의 경우에는 Euclidean distance로 된 수식 (1)로 퍼지 함수 소속도를 산출하였다.

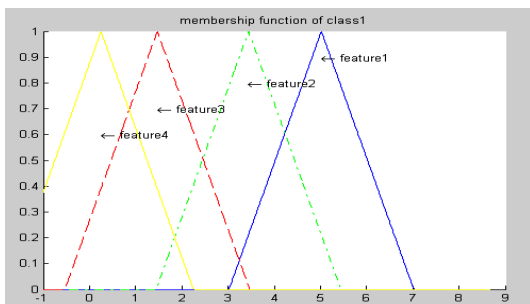


그림 2. 기울기가 동일한 퍼지 함수

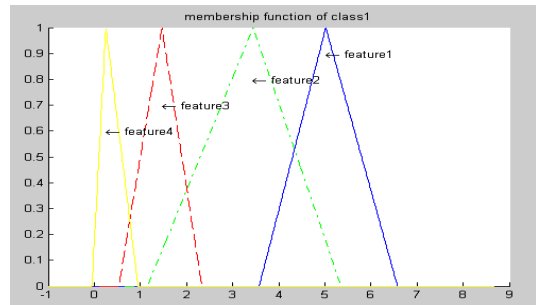


그림 3. 입력의 min-max를 이용한 퍼지 함수

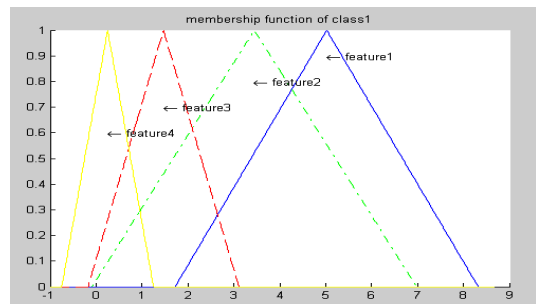


그림 4. 입력의 분산을 기울기로 이용한 퍼지 함수

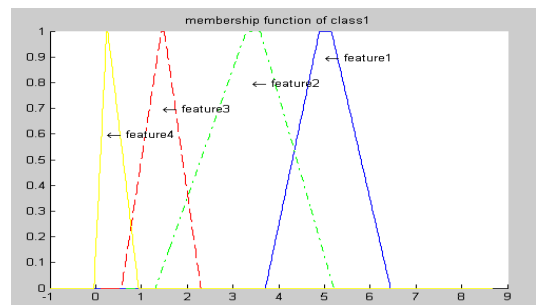


그림 5. 표준편차의 범위를 최대 소속정도로 min-max를 이용한 퍼지 함수

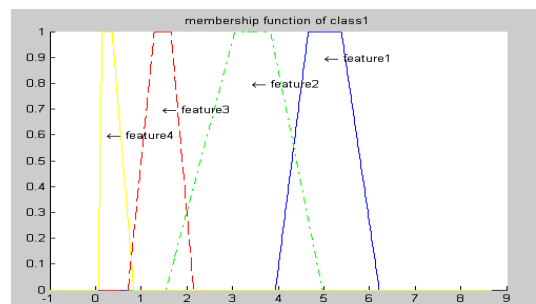


그림 6. 분산의 범위를 최대 소속정도로 min-max를 이용한 퍼지 함수

$$\begin{cases} u_{ij}(x_j) = 1 + 0.5(x_j - m_{ij}) & \text{if } x_j < m_{ij} \\ u_{ij}(x_j) = 1 - 0.5(x_j - m_{ij}) & \text{if } x_j \geq m_{ij} \\ u_{ij}(x_j) = 0 & \text{if } u_{ij}(x_j) < 0 \end{cases} \quad (1)$$

여기서 m_{ij} 는 i 번째 클래스의 j 번째 특징값 평균을 나타낸다.

그림 3의 경우에는 수식 (2)로 퍼지 함수 소속도를 산출하였다. 그 방법은 입력 데이터들의 특징별 최소값 및 최대값을 이용하는 것이다.

$$\begin{cases} u_{ij}(x_j) = 0.5 + 0.5(x_j - \min(\forall x_j)) / (m_{ij} - \min(\forall x_j)) & \text{if } x_j < m_{ij} \\ u_{ij}(x_j) = 1 - 0.5(x_j - m_{ij}) / (\max(\forall x_j) - m_{ij}) & \text{if } x_j \geq m_{ij} \\ u_{ij}(x_j) = 0 & \text{if } u_{ij}(x_j) < 0 \end{cases} \quad (2)$$

그림 4의 경우 수식 (3)에서 보는 것처럼 각각 특징별 분산을 구하여 각각의 분산간의 정규화하여 기울기로 사용하였다. 분산 대신 표준편차를 이용할 수도 있다.

$$var_j = \frac{1}{c} \sum_{i=1}^c (x_j - m_{ij})^2, \quad std_j = \sqrt{var_j} \quad (3)$$

여기서 c 는 각 클래스 수, x_{ij} 는 i 번째 클래스의 j 번째 특징값, m_j 는 j 번째 특징값들의 평균을 의미한다.

$$nvar_j = \min(var) / var_j \quad (4)$$

$$\begin{cases} u_{ij}(x_j) = 1 + nvar_j(x_j - m_{ij}) & \text{if } x_j < m_{ij} \\ u_{ij}(x_j) = 1 - nvar_j(x_j - m_{ij}) & \text{if } x_j \geq m_{ij} \\ u_{ij}(x_j) = 0 & \text{if } u_{ij}(x_j) < 0 \end{cases} \quad (5)$$

그림 5의 경우에는 클래스별 각 특징별로 평균에서 분산값 범위만큼을 소속 정도 1로 주고 나머지부분은 입력데이터들의 특징별 최소 및 최대값에 소속 정도 0.5로 퍼지 함수 소속도를 산출하였다. 그림 6의 경우에는 수식 (6), (7)에서 분산 대신 표준편차를 사용하였다.

$$tvlow_j = f_{ij} - var_j, \quad tvhigh_j = f_{ij} + var_j \quad (6)$$

$$\begin{cases} u_{ij}(x_j) = 1 & \text{if } tvlow_j < x_{ij} < tvhigh_j \\ u_{ij}(x_j) = 0.5 + 0.5(x_j - \min(\forall x_j)) / (tvlow_j - \min(\forall x_j)) & \text{if } x_j \leq tvlow_j \\ u_{ij}(x_j) = 1 - 0.5(x_j - tvhigh_j) / (\max(\forall x_j) - tvhigh_j) & \text{if } x_j \geq tvhigh_j \\ u_{ij}(x_j) = 0 & \text{if } u_{ij}(x_j) < 0 \end{cases} \quad (7)$$

퍼지 소속 함수가 결정된 후 클래스 판별은 각 특징별 소속도의 산술평균으로 산출하여 가장 큰 값을 가지는 클래스로 판별하였다.

III. 실험 및 결과 고찰

실험을 위하여 UCI(University of California, Irvine) machine learning repository 사이트[3]에서 제공되는 표준 데이터들 중의 하나인 Iris 데이터 세트를 사용하였다. Iris 데이터 세트는 3개의 클래스와 4개의 특징값으로 구성된 벡터가 클래스 별로 50개씩 존재하는 구조이다.

위에서 언급한 5가지 경우의 퍼지 함수 설정 방법으로 Iris 데이터를 분류한 실험 결과는 표 1과 같다. 표 1의 결과값들은 각각의 경우에 대하여 10-fold cross validation 방법[4]으로 실험한 평균 결과를 나타낸다. 이 때 퍼지 함수에 따른 결과들만을 비교하기 위하여 90%의 학습 데이터들로부터 얻어지는 최소값, 최대값, 분산, 표준편차만 이용하였으며, 학습 데이터들의 평균값만을 클래스 기준 데이터로 사용하여 분류 실험을 수행하였다.

표 1. Iris 데이터 세트에 대한 각 기법의 결과

처리 기법	인식률
Euclidean distance	92%
min-max 삼각형	94.7%
분산 이용한 삼각형	95.3%
표준편차 이용한 삼각형	96.7%
min-max와 분산 이용한 사다리꼴	94.7%
min-max와 표준편차 이용한 사다리꼴	94.7%

이 논문에서는 퍼지 함수 설정에만 중점을 두었지만, 가중치 평균이나 클래스간 분산의 상관관계 등을 활용하여 학습데이터와 테스트 데이터들 간의 클래스 소속도를 특징값 벡터 요소에 대해 보다 가변적으로 할당할 수 있는 퍼지 기법이 함께 적용된다면 더욱 우수한 성능이 나올 것으로 기대한다.

III. 결론

이 논문에서는 퍼지 함수 분류기의 성능을 문제 분야의 특성에 관계없이 개선시킬 수 있는 한 가지 방법으로 퍼지 함수 기법을 제안하였다. 수학적으로 합리적이라고 판단될 수 있는 분산 및 표준편차를 산출하기 위하여 특징값들 간의 관계를 나타내는 통계적 정보를 활용하였다. 이를 위하여 min, max, 평균, 분산, 표준편차를 활용하여 통계적 척도 몇 가지를 제안하였다. 첫 번째 척도

는로 “분산 혹은 표준편차를 이용한 삼각형 퍼지 함수”, 두 번째 척도로 “min-max와 평균에 따른 분산의 범위를 이용한 사다리꼴 퍼지 함수” 제안하였으며, 이 2가지 기법들과 기존에 사용되었던 Euclidean distance 이용한 퍼지 함수와 min-max 퍼지 함수의 2가지를 실험하고 그 결과를 비교, 분석하였다.

실험에 사용된 표준 데이터 세트로는 UCI(University of California, Irvine) machine learning repository 사이트에서 제공되는 Iris 데이터 세트를 이용하였다. 실험 결과 클래스 소속도를 나타내는 거리 척도로 다른 퍼지 함수의 소속도 값을 사용함으로써, 기존에 사용되는 기법들과 유사하거나 그 이상의 성능이 달성됨을 알 수 있었다.

향후 연구 과제로는 클래스 내 통계 정보를 활용하여 더욱 합리적인 클래스 소속도를 할당하는 기법을 보완하는 것과 더욱 다양한 표준 실험 데이터 세트를 이용하여 제안한 기법의 정당성과 일반성을 검증하는 것이 필요할 것으로 생각한다.

참 고 문 헌

- [1] Timothy J. Ross, Fuzzy Logic With Engineering Applications, Second Edition, 2004.
- [2] 이광형, 오길록, 퍼지 이론 및 응용: I 권 이론, 홍릉과학출판사, 1991.
- [3] A. Asunion and D. Newman, UCI machine learning repository, <http://archive.ics.uci.edu/ml>, School of Information and Computer Science, University of California, Irvine, 2007.
- [4] Ron Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 1137-1143, 1995.