

퍼지 소속도 함수와 가중치 평균을 이용한 지도 학습 기반 분류기 설계

우영운*

Design of a Classifier Based on Supervised Learning Using Fuzzy Membership Function and Weighted Average

Young Woon Woo*

*Full Professor, Division of Creative Software Eng., Dong-eui University, Busan, 47340 Korea

요약

본 논문에서는 지도 학습 기반의 분류기 제안을 위해, 분류 데이터의 각 특징별 소속도를 결정하는 3가지 종류의 퍼지 소속도 함수를 제안하였다. 또한 각 특징별 소속도들의 평균값을 이용하여 분류 결과를 도출하는 과정에 사용되는 평균값 산출 기법을 단순 산술평균이 아닌 다양한 가중치를 활용한 가중치 평균을 이용함으로써 분류기 성능을 향상시킬 수 있는 가능성을 제시하였다. 제안한 기법들의 실험을 위해 Iris, Ecoli, Yeast의 3가지 표준 데이터 세트를 사용하였다. 실험 결과, 서로 다른 특성의 데이터 세트들에 대해서도 고르게 우수한 분류 성능이 얻어질 수 있음을 확인하였고, 기존에 발표된 다른 기법들에 의한 해당 데이터 세트들의 분류 성능과 비교했을 때, 퍼지 소속도 함수의 개선과 가중치 평균 기법의 개선을 통해 더욱 우수한 분류 성능이 가능함을 확인할 수 있었다.

ABSTRACT

In this paper, to propose a classifier based on supervised learning, three types of fuzzy membership functions that determine the membership of each feature of classification data are proposed. In addition, the possibility of improving the classifier performance was suggested by using the average value calculation method used in the process of deriving the classification result using the average value of the membership degrees for each feature, not by using a simple arithmetic average, but by using a weighted average using various weights. To experiment with the proposed methods, three standard data sets were used: Iris, Ecoli, and Yeast. As a result of the experiment, it was confirmed that evenly excellent classification performance can be obtained for data sets of different characteristics. It was confirmed that better classification performance is possible through improvement of fuzzy membership functions and the weighted average methods.

키워드 : 지도 학습, 분류기, 퍼지 소속도 함수, 가중치 평균, UCI

Keywords : Supervised learning, Classifier, Fuzzy membership function, Weighted average, UCI

Received 19 March 2021, Revised 19 March 2021, Accepted 1 April 2021

* Corresponding Author Young Woon Woo(E-mail: ywwoo@deu.ac.kr, Tel:+82-51-890-1712)

Full Professor, Division of Creative Software Engineering, Dong-eui University, Busan, 47340 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.4.508>

print ISSN: 2234-4772 online ISSN: 2288-4165

©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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

인공지능 기술이 발전함에 따라 다양한 형태의 빅데이터들을 적절히 표현하는 것은 물론, 데이터의 특징에 따라 자료를 분류하고 처리하기 위한 기술이 급격히 발전하고 있다. 그 중에서도 머신러닝을 이용한 인식과 관련된 기술들은 정보 검색, 데이터 분류, 데이터 분석 및 가공 등을 처리하기 위한 널리 사용되고 있는 기술이다 [1]. 머신러닝 기법을 이용한 데이터 인식은 인간의 학습 능력과 추론 능력을 인공적으로 프로그래밍하여 외부 대상을 파악하는 능력, 나아가 자연어와 같은 문장까지 이해하는 능력을 컴퓨터 프로그램으로 구현하는 인공지능의 한 분야이다. 특히 머신러닝은 생체인식, 음성인식, 문자인식, 의료영상 분석, 진단시스템, 행동 패턴 분석, 도면 인식, 예측 시스템 등에 활용되고 있으며, 최근에는 군사, 보안 분야뿐만 아니라 비정보 기술도 포함한 다양한 분야에서 활용되고 있다[2].

머신러닝 기법을 이용한 데이터 분류 기술에서 분류의 정확성을 향상시키기 위해서는 데이터의 특성을 파악하고, 그 특성에 따른 변별력이 높은 특징값을 추출하는 것이 가장 중요한 요소이며, 그 다음으로 추출된 특징값들을 하나의 값으로 통합, 비교하여 분류 결과를 도출하기 위한 분류 기법을 결정하고 세부 요인들을 적절히 구성하는 것이 요구된다[3]. 최근에 널리 연구되어 활용되고 있는 분류 기법들로는 신경망(neural network)[4], SVM(support vector machine)[5], 퍼지 집합 이론(fuzzy set technique)[6], 베이스 이론(Bayes theory)[7] 등의 기법들이 대표적이다.

이 논문에서는 지도 학습 기반의 머신러닝 기법을 이용한 분류기를 설계함에 있어서, 분류 데이터의 각 특징별 소속도(membership)를 결정하는 퍼지 소속도 함수로 3가지 종류의 함수들을 제안하였다. 또한 각 특징별 소속도들의 평균값을 이용하여 분류 결과를 도출하는 과정에 사용되는 평균값 산출 기법을 단순 산술평균이 아닌 다양한 가중치를 활용한 가중치 평균을 이용함으로써 분류기 성능을 향상시킬 수 있는 가능성을 제시하였다. 이를 위하여 특징값들 간의 관계를 파악할 수 있는 다양한 통계적 정보들을 활용하였으며, 제안된 기법들을 UCI(University of California, Irvine) 머신러닝 저장소(machine learning repository) 사이트[8]에서 제공되는 표준 데이터들 중 Iris 데이터 세트, Ecoli 데이터 세

트, Yeast 데이터 세트의 3가지 데이터 세트를 이용하여 실험하고 그 결과를 비교, 분석하였다.

논문의 전체 구성은 다음과 같다. 2장에서 제안한 퍼지 소속도 함수들에 대하여 설명하고 3장에서는 가중치 평균에 사용되는 가중치들을 산출하기 위한 다양한 통계적 기법에 대하여 설명한다. 4장에서는 제안 기법들을 이용한 실험 및 분석 결과를 제시한 후 5장에서 결론을 맺는다.

II. 제안한 퍼지 소속도 함수

이 논문에서 제안하는 지도학습 기반의 분류기는 분류를 위한 데이터들의 개별 차원 특징값들을 각각 1차원 데이터로 보고 각 특징값에 따른 클래스 소속 정도를 퍼지 소속도 함수[9]에 의해 결정한 후, 모든 소속 정도를 다양한 가중치를 사용한 가중치 평균에 의해 최종 클래스를 결정하는 방식의 분류기이다. 따라서 어떠한 퍼지 소속도 함수를 사용하는가, 어떠한 가중치를 사용하는가에 따라 성능이 달라진다.

먼저 2장에서는 이 논문에서 제안한 퍼지 소속도 함수의 세가지 종류를 제안하며 그 3가지는 다음의 그림 1부터 그림 3까지와 같다.

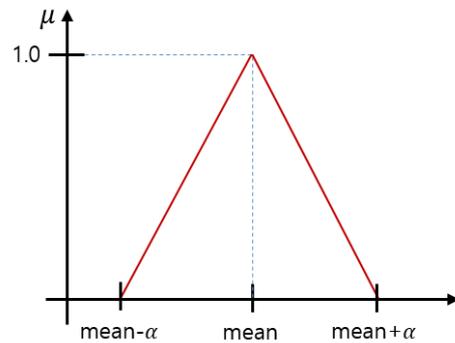


Fig. 1 Proposed fuzzy membership function #1

첫 번째 제안한 퍼지 소속도 함수는 그림 1과 같이 삼각형 모양으로 학습 데이터에 대한 각 클래스의 차원별 특징값들의 평균값을 꼭지점으로 설정하고, 임의의 기울기(좌우 동일)를 선정하여 생성한 퍼지 소속도 함수이다. 이 경우에는 분산이나 다른 데이터의 특성을 반영하지 않고 오로지 한 차원의 학습 데이터들의 1차원 평

균값을 해당 데이터의 발생 중심으로 가정하고 양쪽으로 펼쳐지는 데이터 값들의 밀도나 분산 등을 전혀 고려하지 않는다. 이러한 방식으로 퍼지 소속도 함수를 제안한 배경에는 가중에는 실제 발생된 학습 데이터의 평균에 가장 가까울수록 그 클래스에 소속될 가능성이 가장 높을 것이라는 가정을 포함하고 있다.

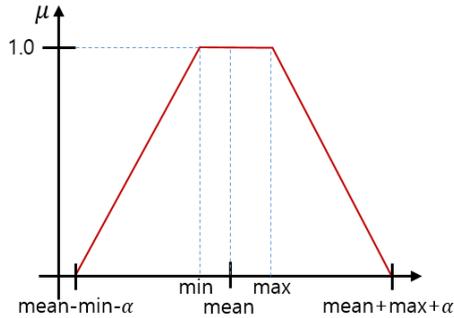


Fig. 2 Proposed fuzzy membership function #2

두 번째 제안한 퍼지 소속도 함수는 그림 2와 같이 사다리꼴로 학습 데이터에 대한 각 클래스의 각 차원별 특징값들의 최소값, 최대값 구간은 모두 소속도 1.0으로 설정하고 임의의 기울기(좌우 동일)를 선정하여 생성한 퍼지 소속도 함수이다. 이 퍼지 소속도 함수는 학습 데이터들의 실제 발생이 이루어지는 전 구간 내에는 테스트 데이터들이 모두 존재할 수 있다는 가정하에 모두 동일한 1.0의 소속도를 가지도록 하며, 혹시라도 그 범위를 벗어나서 나타나게 되는 테스트 데이터의 특이한 값들이라도 어느 정도의 소속도는 가지는 것이 합리적일 것이라는 가정하에 제안한 퍼지 소속도 함수이다.

세 번째 제안한 퍼지 소속도 함수는 그림 3과 같이 삼각형 함수로 꼭지점은 각 클래스의 특징값들의 평균 값을 사용하지만 좌우 기울기를 갖는 함수 부분이 그림 1처럼 임의로 정한 좌우 대칭의 일정한 기울기가 아니고, 학습 데이터의 최소값, 최대값에 해당하는 곳의 소속도를 일정한 0.x로 사전에 정의된 값을 갖도록 기울기를 설정한 퍼지 소속도 함수이다. 세 번째로 제안한 이 방식의 퍼지 소속도 함수는 첫 번째 제안한 방식의 특징과 두 번째 제안한 방식의 특징을 함께 사용하여 제안한 함수 형태로, 학습 데이터의 평균에 가까울수록 그 클래스에 속할 가능성이 높을 것이라는 합리적 가정하에 실제 학습 데이터의 좌우 분포 형태가 대칭적으로 분포하지 않는 경우 데이터의 좌우 분포에 따른 서로 다른 소속도

를 할당할 수 있을 것이라는 점에서 성능 향상을 기대할 수 있는 퍼지 소속도 함수로 판단된다.

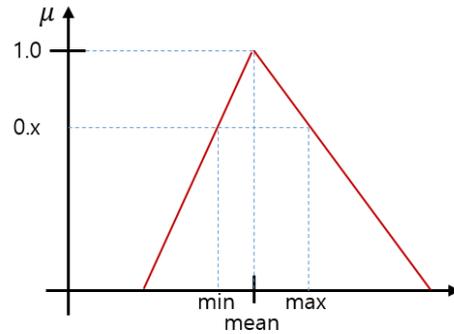


Fig. 3 Proposed fuzzy membership function #3

III. 제안한 가중치 설정 기법

이 논문에서는 지도학습 기반의 분류기를 설계하기 위하여 2장에서는 각 클래스의 차원별 특징값에 대한 소속 정도를 산출하기 위한 퍼지 소속도 함수들에 대하여 살펴보았다. 3장에서는 그러한 퍼지 소속도 함수에 의해 산출된 각 소속도 값들의 가중치 평균값을 도출하여 클래스별 가중치 평균값들을 비교하여 가장 높은 값을 갖는 클래스가 소속 클래스가 되도록 하는 방식의 분류기를 위해 제안한 가중치 설정 기법들에 대하여 설명한다.

특징값들의 통계적 분포를 이용하여 가중치(weight)를 합리적으로 산출하기 위한 기법을 제안하기 위하여 특징값들의 분포에 따른 다양한 경우들을 고려하였다 [10]. 첫 번째로 고려한 경우는 학습 데이터의 클래스 별로 추출될 수 있는 대표 특징값 간격이 일정하나 그 간격이 달라지는 경우이다. 이 경우에 대표 특징값 간격의 변화에 따른 공통점과 차이점들을 분석하여 가중치를 결정하는데 변별적인 요인이 있을지를 판단하였다.

두 번째로 고려한 경우는 학습 데이터의 클래스들 간의 특징값 간격이 특징값들을 정렬하였다고 가정할 때 간격이 모두 같은 것과 하나 차이가 나는 것, 2개 차이가 나는 것 등 차이의 종류와 개수가 달라지는 모든 경우를 예상하여, 그 변화에 따른 공통점과 차이점을 분석하여 새로운 가중치 도출을 위해 활용하였다.

첫 번째로 고려한 경우에 대해서는, 학습 데이터의 각

클래스 간의 간격이 일정하므로 간격이 크면 클수록 변별력이 우수한 것으로 판단할 수 있으며, 이러한 특징을 고려하였을 때 특징값들의 상호 간격이 큰 특징값의 요소가 높은 가중치를 산출할 수 있도록 하는 통계적 척도를 제안할 필요가 있다. 또한 두 번째로 고려한 경우에 대해서는, 학습 데이터의 각 특징값들의 클래스 간 상호 간격이 변하는 경우이므로 각 특징값들의 상호 간격이 넓으면서도 그 분포가 고르게 나타나는 경우 분류 대상의 학습 데이터나 테스트 데이터의 클래스 소속도 간 변별력이 높아져서 분류 성능을 향상 시키는 것에 우수한 요소로 작용할 수 있으므로, 상호 간격이 넓으면서 고른 특징값 벡터가 더 높은 가중치를 갖도록 하는 통계적 척도를 제안하여 비교, 실험을 할 필요가 있을 것으로 판단한다.

따라서 이 논문에서는 학습 데이터들의 벡터 요소별 대표 특징값들의 분포를 이상의 2가지 측면에서 분석하여, 분류 성능을 높일 수 있을 것으로 판단되는 가중치를 생성하기 위하여 통계적 기본 척도를 4가지 고안하였으며 이 4가지는 다음과 같다.

첫 번째 고안한 통계적 기본 척도로는 식 (1)과 같이 학습 데이터 클래스의 각 특징값들의 분산 (V)을 제안하였다. 이는 대표 특징값들의 상호 간격이 넓음을 의미하는 분산이 크게 나타날수록 우수한 변별력을 가질 수 있는 것으로 기존 연구[11]에서 밝혀진 바 있어 가중치로 사용하기에 적절한 것으로 판단하였다.

$$V_j = \frac{1}{c} \sum_{i=1}^c (x_{ij} - m_j)^2 \quad (1)$$

이 수식에서 c 는 클래스 개수이며 j 는 특징값 벡터의 차수를 나타낸다. 또한 x_{ij} 값들은 각 클래스별 특징값으로서 i 를 기준으로 오름차순으로 정렬되어 수식에 적용된다.

두 번째 고안한 통계적 기본 척도로는 식 (2)과 같이 학습 데이터 클래스의 각 특징값들 간에 발생하는 차이들의 평균값(MD)을 제안하였다. 이는 앞에서 설명한 첫 번째 고려한 데이터 분포를 분석하여 나온 것으로, 우수한 변별력을 가질 수 있는 형태의 특징값 분포의 경우에 더욱 높은 값을 가지는 특성이 있으므로 가중치로 사용하기에 적절한 것으로 판단하였다.

$$MD_j = \frac{1}{c-1} \sum_{i=1}^{c-1} (x_{(i+1)j} - x_{ij}) \quad (2)$$

세 번째 고안한 통계적 기본 척도로는 식 (3)와 같이 학습 데이터 클래스의 각 특징값들 간 차이값들의 분산 (VD)을 제안하였다. 이는 앞에서 설명한 두 번째 고려한 데이터 분포를 분석하여 나온 것으로, 우수한 변별력을 가질 수 있는 형태의 특징값 분포에 따라 비례하는 결과를 얻을 수 있었기 때문에 가중치로 사용하기에 적절하다고 판단하였다.

$$VD_j = \frac{1}{c-1} \sum_{i=1}^{c-1} [(x_{(i+1)j} - x_{ij}) - MD_j]^2 \quad (3)$$

마지막으로 고안한 통계적 기본 척도는 학습 데이터 클래스의 각 특징값들의 분산을 가장 큰 특징값 차이로 나눈 값($VPDM$)을 제안하였다. 이는 앞에서 설명한 첫 번째, 두 번째 경우를 모두 고려한 것으로, 우수한 변별력을 가질 수 있는 형태의 특징값 분포가 복합적으로 나타날 경우에 더욱 높은 수치를 얻을 수 있다는 판단에 따라 기본 통계적 가중치 산출식으로 사용하였다.

$$VPDM_j = \frac{\frac{1}{c} \sum_{i=1}^c (x_{ij} - m_j)^2}{DiffMax_j} \quad (4)$$

여기서 $DiffMax_j = \max(x_{(i+1)j} - x_{ij}), i = 1..c$

이 논문에서는 이상의 4가지 통계적 기본 척도를 도출한 후 지도학습 기반의 분류기 설계에 적합한 가중치를 각 척도 개념을 활용하여 다음과 같은 4가지 종류의 가중치 설정 기법을 제안하였다.

- ① $VPDM$
- ② $MD + VPDM$
- ③ $MD + (1 - VD)$
- ④ V

또한 제안한 가중치들의 효용성을 판단하기 위해 기존에 널리 사용되는 산술평균식을 함께 실험에 사용하였다.

- ⑤ Arithmetic average

IV. 실험 및 결과 고찰

2장에서 제안한 3가지 형태의 퍼지 소속도 함수와 3장에서 제안한 4가지 가중치 설정 기법과 산술평균식을 포함한 5가지 기법을 조합하여 모두 15가지 경우를 이용하여 Iris, Ecoli, Yeast의 3가지 표준 데이터 세트들에 대한 분류 실험을 진행하였다. 모든 실험에 대하여 분류 정확도는 5-fold cross validation 기법[12]을 이용하여 산출하였다.

첫 번째 실험에 사용한 표준 데이터 세트는 Iris 데이터 세트이다. Iris 데이터 세트를 이용하여 15가지 경우에 대하여 실험한 결과는 표 1과 같다. 이하 모든 결과 표에서 W로 표시된 가로축은 제안한 4가지 가중치 설정 기법과 산술 평균식을 순서대로 나타낸 것이며, F로 표시된 세로축은 제안한 3가지 퍼지 소속도 함수의 종류를 순서대로 나타낸다. 또한 모든 정확도의 단위는 %이다.

Table. 1 Classification Accuracy for Iris dataset

F \ W	①	②	③	④	⑤	Avg.
①	96.67	96.00	94.67	96.00	91.33	94.93
②	94.00	94.00	93.33	94.00	93.33	93.73
③	96.00	96.00	96.00	95.33	96.00	95.87
Avg.	95.56	95.33	94.67	95.11	93.55	94.84

두 번째 실험에 사용한 표준 데이터 세트는 Ecoli 데이터 세트이다. Ecoli 데이터 세트를 이용하여 15가지 경우에 대하여 실험한 결과는 표 2와 같다.

Table. 2 Classification Accuracy for Ecoli dataset

F \ W	①	②	③	④	⑤	Avg.
①	84.92	85.23	85.23	82.46	85.23	84.61
②	75.38	75.38	75.38	75.38	75.38	75.38
③	80.00	80.31	81.54	78.46	80.92	80.25
Avg.	80.10	80.31	80.72	78.77	80.51	80.08

세 번째 실험에 사용한 표준 데이터 세트는 Yeast 데이터 세트이다. Yeast 데이터 세트를 이용하여 15가지 경우에 대하여 실험한 결과는 표 3과 같다.

Table. 3 Classification Accuracy for Yeast dataset

F \ W	①	②	③	④	⑤	Avg.
①	46.67	47.14	48.82	31.25	48.75	44.53
②	36.09	36.09	35.96	36.16	36.03	36.07
③	40.13	40.40	39.40	36.70	39.39	39.20
Avg.	40.96	41.21	41.39	34.70	41.39	39.93

이상의 실험 결과에서 알 수 있듯이 Iris 데이터 세트에 대한 15가지 경우에서 최고 정확도(accuracy)는 96.67%이며 최저 정확도 91.33%로 나타났다. Ecoli 데이터 세트에 대한 15가지 경우에서는 최고 정확도(accuracy)는 85.23%이며 최저 정확도 75.38%로 나타났다. 마지막으로 Yeast 데이터 세트에 대한 15가지 경우에서는 최고 정확도(accuracy)는 48.82%이며 최저 정확도 31.25%로 나타남을 알 수 있었다.

기존 연구에서 다양한 분류기를 이용하여 Iris 데이터 세트, Ecoli 데이터 세트, Yeast 데이터 세트를 이용하여 실험한 논문이나 실험 결과로 공개되어 있는 각 기법별 분류 성능은 표 4, 표 5, 표 6과 같다.

Table. 4 Classification Accuracy for Iris dataset by other classification methods[13]

Iris dataset	J48	CART	Bayes Net	Naive Bayes
Accuracy	96.0	95.3	92.7	96.0

Table. 5 Classification Accuracy for Ecoli dataset by other classification methods[14]

Ecoli dataset	SVM	Naive Bayes	Decision Tree	1-layer NN	3-layer NN
Accuracy	72.1	76.0	81.4	76.0	91.2

Table. 6 Classification Accuracy for Yeast dataset by other classification methods[15]

Yeast dataset	k-NN	SVM	C4,5	Ripper	BayesNet
Accuracy	51.0	51.1	42.8	46.3	45.0

기존의 연구 결과에서 발표된 Iris 데이터 세트에 대한 분류 정확도의 최고치가 96.0%였음을 감안하면 제안한 기법에 의한 최고 정확도가 더욱 높게 나타나 제안 기법의 의의를 충분히 파악할 수 있었다. 또한 Ecoli 데이터 세트의 경우에는 기존 연구 결과에서 발표된 결과

들 중 가장 높은 정확도인 91.2%를 제외한 나머지 4가지 기법보다 더욱 높은 정확도를 보여 실용적인 측면에서 충분한 가능성을 보였다고 판단한다. Yeast 데이터 세트의 경우에는 기존 연구 결과들 중 2가지 기법이 제안한 기법의 최고 정확도보다 높았고, 나머지 3가지 기법들의 정확도는 제안한 기법의 최고 정확도보다 낮게 나타나 실용적인 분류기로서의 가능성을 충분히 가지고 있는 것으로 판단된다.

전체적으로 결과 분석 내용을 종합해 보면, 기존에 널리 사용되고 있는 다양한 분류 기법들은 데이터 세트의 종류 및 특성에 따라 분류 성능이 크게 달라짐을 볼 수 있었는데, 이 논문에서 제안한 지도 학습 기반의 분류 알고리즘은 서로 다른 특성을 갖는 데이터 세트들에 대해서도 현재까지 가능한 것으로 밝혀진 각각의 최고 분류 정확도에 근접하는 결과를 얻었기 때문에, 추가적인 개선이 이루어질 경우, 더욱 우수한 성능을 보일 수 있을 것이라는 점에서 의의가 큰 연구 결과로 생각한다.

V. 결론

본 논문에서는 분류 데이터의 각 특징별 소속도를 결정하는 퍼지 소속도 함수로 3가지 종류의 함수들을 제안하였고, 각 특징별 소속도들의 평균값을 이용하여 분류 결과를 도출하는 과정에 사용되는 평균값 산출 기법을 단순 산술평균이 아닌 다양한 가중치를 활용한 가중치 평균을 이용한 지도 학습 기반의 분류 알고리즘을 제안하였다. 제안한 기법들의 실험을 위해서는 UCI 머신러닝 데이터 저장소에서 제공되는 표준 데이터들 중 Iris 데이터 세트, Ecoli 데이터 세트, Yeast 데이터 세트의 3가지 데이터 세트를 사용하였다. 실험 결과, 서로 다른 형태의 퍼지 소속도 함수와 가중치 평균 기법을 이용함으로써 데이터 특성에 관계없이 일관성 있는 분류 성능을 얻을 수 있음을 보였고, 기존에 발표된 다른 기법들에 의한 해당 데이터 세트들의 분류 성능과 비교했을 때에도 퍼지 소속도 함수들의 개선과 가중치 평균 기법들의 개선을 통해 더욱 향상된 분류 성능을 달성할 수 있을 것임을 확인할 수 있었다.

향후 연구 과제로는 퍼지 소속도 함수 모양을 설정함에 있어서 데이터의 분포 특성 형태를 반영한 소속도 함수를 도출하기 위한 기법 제안이 필요하며, 또한 클래스

간의 데이터 개수 차이를 이용한 사전 데이터 분포 확률 개념을 가중치 설정에 도입하여 더욱 개선된 분류 알고리즘 개발이 필요할 것으로 예상된다.

ACKNOWLEDGEMENT

This work was supported by Dong-eui University Foundation Grant(2017).

REFERENCES

- [1] C. Augenstein, N. Spangenberg, and B. Franczyk, "Applying machine learning to big data streams : An overview of challenges," *2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 25-29, Nov. 2017.
- [2] H. J. Kim, *Pattern Recognition and Machine Learning*, Jpub, Seoul, 2018.
- [3] G.-W. Lee, H. Lee, and H.-W. Cheong, "Object Detection of AGV in Manufacturing Plants using Deep Learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25 no. 1, pp. 36-43, 2021.
- [4] M. Sewak, S. K. Sahay, and H. Rathore, "An Overview of Deep Learning Architecture of Deep Neural Networks and Autoencoders," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 1, pp. 182-188, Jan. 2020.
- [5] S. Behera and M. N. Mohanty, "Classification of EEG Signal Using SVM," *Advances in Electrical Control and Signal Systems*, vol. 665, pp. 859-869, Jul. 2020.
- [6] H. Wang, Z. Xu, and W. Pedrycz, "An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities," *Knowledge-Based Systems*, vol. 118, pp. 15-30, Feb. 2017.
- [7] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kyiv, Ukraine, pp. 900-903, 2017. doi: 10.1109/UKRCON.2017.8100379.
- [8] UCI Machine Learning Repository [internet]. Available: <https://archive.ics.uci.edu/ml/index.php>.
- [9] A. Kandel, *Fuzzy Techniques in Pattern Recognition*, John Wiley & Sons Inc., New Jersey, 1982.
- [10] S. H. Shin, J. H. Cho, and Y. W. Woo, "Weight Adjustment Methods Based on Statistical Information for Fuzzy

- Weighted Mean Classifiers,” *Journal of the Korea Society of Computer and Information*, vol. 14, no. 7, pp. 9-15, Jul. 2009.
- [11] S. Han and Y. W. Woo, “Fuzzy Classifier and Bispectrum for Invariant 2-D Shape Recognition,” *Journal of Korea Multimedia Society*, vol. 3, no. 3, pp. 241-252, Mar. 2000.
- [12] K-fold Corss Validation [Internet]. Available: <https://medium.datadriveninvestor.com/k-fold-cross-validation-6b8518070833>.
- [13] D. Kharche, K. Rajeswari, and D. Abin, “Comparison of Different Data Set Using Various Classification Techniques with WEKA,” *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 4, pp. 389-393, Apr. 2014.
- [14] Which Machine Learning Classifier Is the Best on 18 UCI Datasets? [internet]. Available: <https://abidlabs.github.io/uci-datasets>.
- [15] R. Cerri, R. R. O. da Silva, and A. C. P. L. F. de Carvalho, “Comparing Methods for Multilabel Classification of Proteins Using Machine Learning Techniques,” *Lecture Notes in Computer Science*, vol. 5676, pp. 109-120, Jul. 2009.



우영운(Young Woon Woo)

연세대학교 전자공학과 공학석사(1991)
연세대학교 전자공학과 공학박사(1997)
동의대학교 창의소프트웨어공학부(1997~현재)

※관심분야: 인공지능, 패턴인식, 퍼지이론, 의료정보