# 퍼지 유사 척도에 관한 연구

## A Study on the Fuzzy Similarity Measure

김 용 수*
Yongsoo Kim*

요 약

본 논문에서는 퍼지 유사 척도가 제시된다. 제시된 퍼지 유사 척도는 유사도를 결정하기 위해서 유크리디안 거리와 함께 데이터와 클러스터 대표값들 사이의 상대적 거리를 고려한다. 클러스터의 경계선은 경쟁이 심한 곳에서는 축소되며 경쟁이 심하지 않은 곳에서는 확장된다. 본 논문의 결과는 상대적 거리를 유사 척도로 사용하는 가능성을 보인다.

ABSTRACT

In this paper a fuzzy similarity measure is proposed. The proposed fuzzy similarity measure considers the relative distance between data and cluster centers in addition to the Euclidean distance to decide the degree of similarity. The boundary of a cluster center is constracted on the competitive region and expanded on the less competitive region. This result shows the possibility of using relative distance as a similarity measure.

## Ⅰ. 서 론

패턴 인식(pattern recognition)은 데이터에 내재되어 있는 구조를 찾는 탐색으로 정의 될 수 있다[1]. 하나의 데이터 집합은 데이터의 구조에 따라서 몇 개의 그룹으로 나뉘어 질 수 있는데, 나뉘어진 결과는 사용한 유사척도에 따라서 영향을 받는다. 즉 같은 그룹에 속해 있는 데이터들은 유사도가 다른 그룹에 속해 있는 데이터에 대한 유사도보다 크도록 데이터들이 나뉘어 진다. 따라서 같은 두개의 데이터에 대한 유사도는 사용한 유사척도에 따라서 다른 값을 가질 수 있다. 그러므로 패턴인식에 사용되는 클러스터링 (clustering) 알고리즘의 성능은 사용한 유사척도에 의해서 좌우될 수 있다[2].

패턴 인식에 사용되는 모델들 중에서 신경회로망은 계산능력 때문에 널리 사용되어 왔고[3], 최근에는 신경회로망 구조에 퍼지 논리를 조합한 퍼지 신경회로망이 연구되고 있다[4]. 퍼지 논리는 실제 상황에서 발생하는 모호함을 잘 표사하는 장점이 있어 신경회로망의 학습 규칙에 퍼지 소속 함수(fuzzy membership function)을 사용하면, 연결 강도(weight)를 더 정교하게 조절할 수 있다. 신경회로망 및 퍼지 신경회로망중 ART(Adaptive Resonance Theory)[5], Fuzzy ART[6], Fuzzy min-max[7], IAFC(Integrated Adaptive Fuzzy Clustering)[8] 등은 승자를 결정한 후에, 승자의 대표값과 데이터사이의 유사도가 사용자가 설

*대전 대학교 컴퓨터 공학과

정한 치수를 만족하는 지를 점검하는 과정을 가지고 있다. 이 점검 과정을 통하여 더 나은 다른 대안 또는 출력 뉴론이 없어 데이터가 특정한 그룹에 속하는 것으로 결정하기 보다는 경쟁에서 이긴 출력 뉴런이 나타내는 그룹의 대표값과 데이터의 유사도가 설정치 보다 클 때 비로서 데이터가 그 특정 그룹에 속하는 것으로 결정하는 것이다. 따라서, 이 점검 과정이 없을 때 발생하는 문제, 즉 outlier에 의해서 그룹을 대표하는 대표값이 악화되는 문제를 해결할 수 있다. 위의 (퍼지) 신경회로망들이 점검 과정에서 사용하는 유사도를 결정하는 유사 척도는 각기 다르다.

지금까지 패턴 인식에서 많이 사용되어 왔던 유사 척도에는 유크리디안 거리와 Mahalanobis 거리가 있다. 유크리디안 거리는 특징 공간(feature space)에서 특징점(feature point)들 간의 거리를 나타내며 특징점들이 클러스터(cluster)를 이루고 있을 때 방향이나 위치 또는 통계적 성질에 따른 가중치를 주지 않는다. 이에 비해 Mahalanobis거리는 특징점들이 클러스터를 이루고 있을때 특징점들의 통계적 성질에 따라 가중치를 준다. 즉 공분산(convariance)이 크게 되면 공분산 행렬의 역(inverse)을 곱하므로 Mahalanobis 거리는 작고, 공분산이 작게되면 Mahalanobis 거리는 크다. 즉 유크리디안 거리와는 달리 특징점들의 분포에 따라서 특징점들간의 거리가 변형된다. 위와 같은 성질 때문에 유크리디안 거리나 Mahalanobis 거리를 승자가 결정된 후에 점검하는 과정에서 사용할 때, 잘 적용되는 클러스터의 모양이 결정되게 된다. 유크리디안 거리는 원의 모양을 갖는 클러스터들을 분류하는데 잘 적용되고, Mahalanobis 거리는 타원형의 모양을 갖는 클러스터들을 분류하는데 잘 적용된다. 그러나 우리가 처리하고자 하는 데이터들은 위와 같은 특정한 모양의 클러스터들로 이루어진 것이 아니다.

위와 같은 문제를 개선하고자 본 논문에서 제안하는 퍼지 유사 척도는 인식하고자 하는 데이터와 현존하는 그룹들의 대표점들간의 상대적인 위치를 고려했다. 데이터가 그룹들 간의 경쟁이 심한 곳에 존재할 때는 데이터와 대표점들 간의 거리의 변화가 조금만 있어도 이것이 중요한 영향을 미치므로 가중치를 크게 주고, 데이터가 그룹들간의 경쟁이 약한 곳에 존재할 때는 데이터와 대표점들 간의 거리의 변화가 있어도 이것이 중요한 영향을 미치지 못하므로 가중치를 작게 준다. 위와 같은 방법으로 하여 데이터와 대표점들 사이의 거리뿐만 아니라 상대적인 위치까지도 고려하는 것이며, 이렇게함으로써 유크리디안 거리나 Mahalanobis거리를 유사척도로 사용할 때 발생되는 문제를 해결할 수 있다.

## Ⅱ. 퍼지 유사 척도

본 논문에서 제안하는 퍼지 유사 척도는 $\|\mathbf{X}-\mathbf{V}_i\| e^{-\gamma\mu_i}$이다. 여기서 $\mathbf{X}$는 데이터이고, $\mathbf{V}_i$는 i번째 그룹의 대표점이며, $\mu_i$는 데이터의 i번째 그룹에 소속되는 정도를 나타내는 퍼지 소속 함수 값이며, $\gamma$는 상수로서 상대적인 위치가 유사도에 미치는 영향의 정도를 결정한다. 퍼지 소속 함수는 Bezdek[1]이 제안 한 것과 유사하며 그 식은

$$\mu_i = \frac{\left(\dfrac{1}{\|\mathbf{X}-\mathbf{V}_i\|^2}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^{n}\left(\dfrac{1}{\|\mathbf{X}-\mathbf{V}_j\|^2}\right)^{\frac{1}{m-1}}}$$

이고, Bezdek이 제한 한 것과 다른 점은 그룹의 수 n은 고정된 값이 아니고 변화한다는 것이다. $e^{-\gamma\mu_i}$는 데이터의 특정한 클러스터에 대한 퍼지 소속도가 크면 그 값이 작아 지므로, 이것이 유크리디안 거리와 곱해져서 전체값이 작아진다. 이는 통계적 성질에 따라서 가중치를 변화시키는 방법과 비교할 수 있다[2]. 통계적 방법은

$$D(a, b) = \sqrt{\sum_{k=1}^{p} w^2_{kk}(a_k-b_k)^2}$$

이며, 여기서 $W_{kk}$는 가중치 상수이고

$$w_{kk} = \frac{\dfrac{1}{\sigma^2_k}}{\sum_{j=1}^{p}\dfrac{1}{\sigma^2_j}}$$ 인데, 분포도가 클 경우에는 가중치를 작게 하고, 분포도가 작을 경우는 가중치를 크게 한다. 통계적 방법은 데이터에 대한 통계적 성질을 알아야 하나, 본 논문에서 제시하는 퍼지 유사 척도를 사용할 때는 존재하는 클러스터의 대표값들만 알아도 된다. 현재의 데이터와 존재하는 클러스터의 대표

값들로부터 퍼지 소속도를 계산한 후, 이 퍼지 소속도를 이용하여 퍼지 소속도가 클 경우에는 가중치를 작게하고, 퍼지 소속도가 작을 경우에는 가중치를 크게한다. 퍼지 소속도가 클 경우에는 데이터가 특정 클러스터의 대표값에 가까이 존재한다는 것을 의미하며, 이 경우 데이터의 위치가 조금 변화해도 클러스터링 결과에는 영향을 미치지 못한다. 그러므로 퍼지 소속도가 클 경우에는 상대적인 위치를 고려한 새 유사척도를 사용한 유사도는 크게 된다. 이에 비해 퍼지 소속도가 작을 경우에는 데이터가 클러스터의 대표값들 사이에 있다는 것을 의미하며, 이 경우에는 데이터의 위치가 조금 변화해도 클러스터링 결과에 영향을 미친다. 그러므로 퍼지 소속도가 작을 경우에는 상대적인 위치를 고려한 새 유사척도를 사용한 유사도는 작게 된다. 이에 반해 유크리디안 거리를 유사척도로 사용하면 데이터가 어디에 위치하느냐를 고려하지 않고 데이터와 하나의 특정한 클러스터의 대표값과의 거리만을 고려한다.

## Ⅲ. 모의 실험

모의 실험은 클러스터의 대표값들이 임의의 위치에 있을 때 형성 가능한 경계선을 유클리디안 거리를 유사 척도로 사용했을 때와 퍼지 유사 척도를 유사척도로 사용했을 때를 비교하였다. 또한 이 모의 실험에서는 두 개의 크러스터가 형성되어 있고 숫자로 표시된 특징점이 2차원일 때의 각 클러스터의 경계선이 어떠한 모양으로 형성되는 지를 $\gamma$값을 변화시키면서 살펴보았다. x축과 y축은 추출해 낸 특징점의 2차원 요소를 나타내며 크러스터링하는 과정에서 클러스터의 대표값들이 변화하게 되는데, 대표값들의 변화에 따른 클러스터의 경계선이 어떠한 영향을 받는지를 살펴보았다.

그림 1에서와 같이 유크리디안 거리를 유사 척도로 사용할 때에는 경계선이 원으로 나타나는 것에 비해서 퍼지 유사척도를 사용하면 데이터가 한쪽에 치우친 부분에서는 클러스터의 경계선이 확장되는 것을 알수 있다. 이는 데이터가 클러스터의 대표값들 중에서 어느 하나의 대표값에 치우쳐 있으면 그 클러스터에 대한 퍼지 소속도가 크기 때문에, 유사도가 커지는 관계로 경계선이 확장되는 것이다. 경쟁이 심한 부분

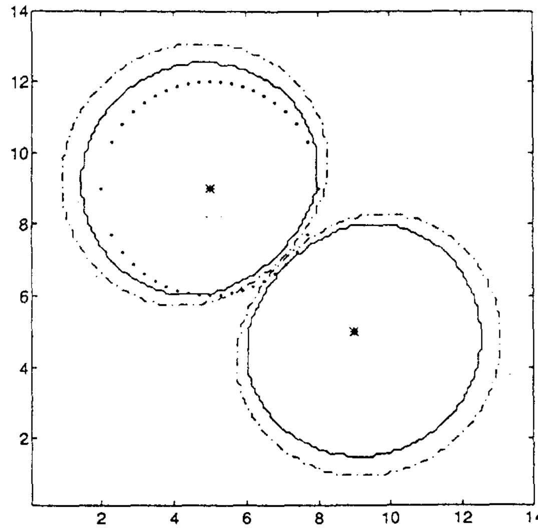


그림 1. 다른 값의 $\gamma$에 따른 경계선

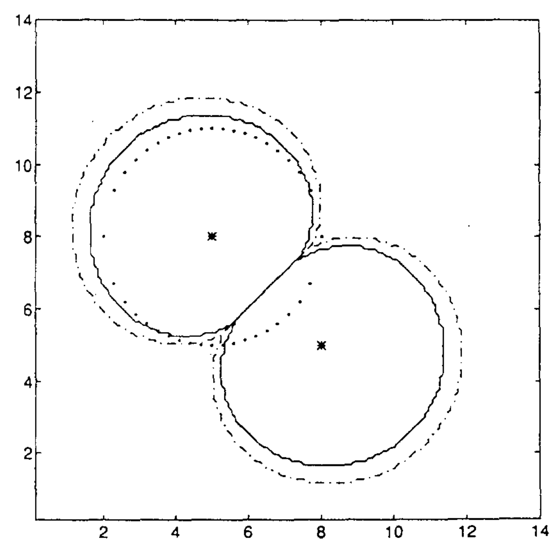


그림 2. 두 클러스터의 대표값이 가까이 있을때의 경계선

에서는 퍼지 소속도 값이 나뉘어 지기 때문에, 어느 한 클러스터에 대한 퍼지 소속도는 작아지고 그 클러스터의 경계선은 축소가 된다. 따라서 퍼지 유사 척도를 사용하면 데이터가 존재하는 위치에 따라서 가중치가 다른 것을 알 수 있다. $\gamma$값이 크면 클수록 상대적 위치에 따른 가중치가 더 급격하게 변화하기 때문에 퍼지 유사 척도가 클러스터의 경계선에 미치는 영향이 크게된다.

그림 2에서와 같이 클러스터들의 대표값들이 아주 가까이 있으면, 클러스터들의 대표값 사이에 있는 영역에서는 경쟁이 심하기 때문에, 클러스터간의 경쟁이 새로운 퍼지 유사 척도의 효과를 압도한다.

위와 같은 효과는 outlier의 문제를 부분적으로 해결할 수 있다. 유크리디안 거리에서와 같이 경계선을 획일적으로 정하는 것은 상대적으로 기존의 클러스터에 가까이 있기 때문에 포함시키는 것이 타당함에도 불구하고 포함시키지 못하고, 새로운 클러스터를 형성하는 문제점을 해결 할수 있다. 그러나 포함되는 정도는 $\gamma$의 값과 점검 과정에서 사용되는 설정치에 달려있다.

## Ⅳ. 결 론

본 논문에서 제시한 퍼지 유사 척도는 상대적 거리를 유사척도로 사용할 수 있는 가능성을 제시하였다. 상대적 거리를 사용함으로써 경쟁이 심한 곳과 경쟁이 심하지 않은 곳을 분리하여, 경쟁이 심한 곳에서는 작은 변화에도 민감하게 반응을 하고 경쟁이 심하지 않은 곳에서는 작은 변화에 대해서는 무시하는 장점이 있다. 이는 사람들이 물체를 비교할 때 비슷한 것들에 대해서는 세밀히 검토를 해야하고 구별이 뚜렷한 것에 대해서는 쉽게 분별할 수 있는 것과 같다. 또한 이 퍼지 유사 척도가 모의 실험을 통하여 경계선을 결정하는 것을 보였다.

## 참 고 문 헌


1. J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," New York, NY, Plenum Press, 1981.
2. J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles," Reading, MA, Addison Wesley, 1974.
3. R. P. Lippmann, "An Introduction to Computing with Neural Net," IEEE ASSP Mag., pp. 4-22, April 1987.
4. J. C. Bezdek and S. K. Pal, "Fuzzy Models for Pattern Recognition-Methods That Search for Structures in Data," New York, NY, IEEE Press, 1992.
5. G. A. Carpenter and S. Grossberg, "A Massively Parallel Architecture for A Self-Organizing Neural Pattern Reconition Machine," Computer Vision, Graphics, and Image Processing, Vol. 37, pp. 54-115, 1987.
6. G. A. Carpenter, S. Grossberg, D. Rosen, "Fuzzy ART: Fast Stable Learning and Categorization of Analog patterns by an Adaptive Resonance Systems," Neural Networks, Vol. 4, No. 6, pp. 759-772, 1992.
7. P. K. Simpson, "Fuzzy min-max neural newworks-part2: Clustering," IEEE Trans. on Fuzzy System, Vol. 1, No. 1, pp. 32-45, 1993.
8. Y. S. Kim and S. Mitra, "An Adaptive Integrated Fuzzy Clustering Model for Pattern Recognition," Fuzzy Sets and Systems, Vol. 65, pp. 297-310, 1994.



김 용 수(Yongsoo Kim) 정회원

1959년 1월 16일생

1981년 2월 : 연세대학교 전기공학과 졸업(공학사)

1983년 2월 : 한국과학기술원 전기 및 전자공학과 졸업(공학석사)

1986년 7월 : 삼성전자

1993년 12월 : Texas Tech Univ. 졸업(공학박사)

현재 : 대전대학교 컴퓨터공학과 전임강사

※주관심분야 : 퍼지 신경회로망, 퍼지 논리, 패턴 인식, 영상 처리