

퍼지-유전자 알고리즘을 이용한 무감독 특징 선택 방법

Unsupervised Feature Selection Method Using a Fuzzy-Genetic Algorithm

이 영 제
삼성전자 통신연구소
nezumi@samsung.co.kr

이 정 훈
한양대학교 전자공학과
Computational Vision & Fuzzy Systems Lab
frhee@fuzzy.hanyang.ac.kr

요약

본 논문에서는 퍼지-유전자 접근방법을 이용한 무감독 특징 선택방법에 대하여 나타내었다. 이 방법은 각각의 특징들의 중요도에 따라 순서를 정하기 위해 사용되는 weighted distance 를 포함하는 특징 평가 지표 (feature evaluation index)를 최소화 시키는데 있다. 또한 특징 평가 지표에서 사용되는 각 패턴들의 쌍에 대하여 근접함의 정도를 퍼지 멤버쉽 함수를 이용하여 결정하고 유전자 알고리즘은 평가 지표를 최소화 시킴으로써 각 특징의 중요도를 나타내는 최적의 weighting 계수의 집합을 찾기 위하여 적용하였다.

1. 서론

특징 선택(feature selection)은 패턴 인식 학문에서 중요한 부분으로 인식되어 왔다. 특징 추출의 결과로 인식 과정에서 중요하게 영향을 미치지 않는 몇몇의 중복적이고 상관관계가 없는 특징들[2]이 존재할 수 있다. 더욱이 많은 특징들이 존재한다면 그의 따른 특징 공간은 증가하게 된다. 그리고 패턴 분석을 위한 계산량도 증가하게 되고 비효율적이며 경제적인 측면에서도 손해를 얻게 된다. 따라서 패턴 분석을 위한 최적의 뚜렷한 특성을 가지고 있는 특징들을 중요도에 따라 순서를 정하여 특징들을 선택하는 방법이 필요하게 된다. 이렇게 되면 중요하지 않는 특징들을 제거함으로써 본래의 특징 공간이 줄어들게 됨으로써 결과적으로

효과적인 알고리즘으로 발전할 수 있다.

많은 경우에서 특별한 목적을 위해서 추출된 특징들은 센서의 부정확한 측정에 의해 정확하지 않는 특징들이 포함될 수 있다. 퍼지 집합 이론 접근 방법은 부정확하고 애매 모호한 특징들을 효과적으로 다룰 수 있기 때문에 기존에 나왔던 알고리즘보다 좀 더 향상된 인식결과를 나타내고 있다[3].

신경망 접근 방법이 많은 장점이 있음에도 불구하고 gradient descent 의 느린 학습 속도와 local minima 문제 때문에 원하는 결과를 얻을 수 없는 단점을 갖게 된다.

요즘에 유전자 알고리즘은 빠른 학습 속도와 최적화 능력에 의해서 인기있는 탐색 방법으로써 폭 넓게 사용되고 있다. 특별히 유전자 알

고리즘에서 탐색 전략은 random 을 기초로 하기 때문에 gradient descent 학습에서의 local minima 문제를 어느 정도 해결할 수 있다.

본 논문에서는 무감독 특징 선택을 위하여 퍼지-유전자 접근 방법을 제안하였다. 이 방법의 목적은 특징들의 중요도에 따라서 순서를 정하기 위하여 퍼지 멤버쉽 함수를 포함하고 있는 평가 지표[1]를 유전자 알고리즘을 이용하여 최소화 시키는 것이다. 제안하는 방법은 두 가지 장점이 있다.

- 1) 학습 속도 향상
- 2) Local minima 문제 해결

2. Feature Evaluation Index

평가 지표는 아래와 같이 정의한다.

$$E(w) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N [\mu_{ij}(1-\mu_{ij}) + \mu'_{ij}(1-\mu'_{ij})] \quad (2.1)$$

여기서 μ_{ij} 와 μ'_{ij} 는 각각 N차 특징 공간에서 euclidean distance 와 weighted distance 를 이용하여 i 번째와 j 번째 패턴 사이의 유사도의 정도를 의미한다. 또한 N 은 특징 공간에 대응하는 패턴의 수를 나타낸다. 위에서 언급한 평가 지표는 각 패턴 쌍에 대하여 다음과 같은 특성들을 가지고 있다.

- 1) $\mu_{ij} = \mu'_{ij} = 0$ 또는 1 일 때, E 는 최소값인 0 이 된다.
- 2) $\mu_{ij} = \mu'_{ij} = 0.5$ 일 때, E 는 최대값인 0.25 가 된다.
- 3) $\mu_{ij} > 0.5$ ($\mu_{ij} < 0.5$) 이고 μ'_{ij} 가 1(0)에 접근할 때, index E 는 감소하게 된다.

위에서 언급한 특성들은 평가 지표값이 감소할수록 그에 따른 intercluster distance 가 증가하게 되고 intracuster distance 가 감소하게 됨을 의미한다. 즉 이 의미는 각기 다른 cluster 에 속해 있는 패턴들의 거리(intercluster distance)가 멀어지고 같은 cluster 에 속해 있는 패턴들의 거리(intracuster distance)가 가까울수록

패턴들의 유사도를 정확하게 구분할 수 있는 특성을 갖게 된다는 것이다. 따라서 각 패턴들의 특징들 중에서 이러한 특성을 만족하면 가장 중요한 특징들이 되는 것이다.

식(2.1)에서 멤버쉽 함수, μ_{ij} 와 μ'_{ij} 는 다음과 같이 정의한다.

$$\mu_{ij} = \frac{d_{\max} - d_{ij}}{d_{\max} - d_{\min}} \quad (2.2)$$

$$\mu'_{ij} = \frac{d_{\max} - d'_{ij}}{d_{\max} - d_{\min}} \quad (2.3)$$

여기서 d_{ij} 와 d'_{ij} 는 i 번째 와 j 번째 패턴의 유사도를 측정하기 위한 거리 측정값이고 d_{\min} 는 각 패턴 쌍의 최소 거리를 나타내며 d_{\max} 는 각 패턴 쌍의 최대 거리를 나타낸다. 각 쌍의 패턴들이 서로 가까울수록 그들의 유사도는 높아지게 되고 반대로 멀어질수록 유사도는 낮아지는 것을 알 수 있다.

식(2.2)에서 이용되는 거리측정은 패턴 쌍 사이의 euclidean distance 로 표현되어진다.

$$d_{ij} = \left[\sum_{k=1}^n (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \quad (2.4)$$

위의 식에서 x_{ik}, x_{jk} 는 각각 i 번째와 j 번째 패턴들에 대한 k 번째 특징 값이다.

패턴의 모든 쌍들의 최대 거리는 아래와 같이 정의한다.

$$d_{\max} = \left[\sum_{k=1}^n (x_{\max k} - x_{\min k})^2 \right]^{\frac{1}{2}} \quad (2.5)$$

변수 $x_{\max k}$ 와 $x_{\min k}$ 는 k 번째 특징에 대한 최대값과 최소값을 나타낸다.

거리 d_{ij} 는 euclidean distance 이기 때문에 이

런 방법은 패턴의 cluster 가 hyperspherical 일 때를 가정하고 있다. 따라서 좀 더 실제적인 환경을 modeling 하기 위해서 weighted distance 측정 방법을 사용하였다.

$$d'_{ij} = \left[\sum_{k=1}^n w_k^2 (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (2.6)$$

Weight w_k 는 0 과 1 사이의 실수값을 갖는 k 번째 특징에 대한 weight 계수로 나타내어 진다. 따라서 w_k 는 k 번째 특징에 대한 중요도를 나타내는 weight 값으로 간주 할 수 있다. $w_k=1$ 이면 특징 k 가 가장 중요하다는 것을 나타내고 $w_k=0$ 이면 그 반대의 의미를 지닌다.

2. 유전자 알고리즘을 이용한 Feature Ranking

유전자 알고리즘(genetic algorithm)은 자연적 유전 과정을 기초로 한 적응적 탐색 방법이라고 할 수 있다[6]. 무작위로 population 집합을 선택한 후 유전자 연산들을 이용하여 현재의 집합으로부터 새로운 집합이 생성된다. 이런 과정은 사전에 문제의 해를 찾기 위해서 정한 fitness function 이 만족될 때까지 반복된다.

위에서 설명한 바와 같이 논문의 목적은 식 (2.1)의 평가 지표를 최소화 시킴으로써 최적의 weight 계수의 집합을 찾는 것이기 때문에 유전자 알고리즘에서 사용되는 population 은 식 (2.6)에서 weight 계수인 w_k 로 정의하였다. n 개의 특징들로 이루어진 패턴의 대하여 염색체의 길이는 $m \times n$ 이 된다.

여기서 유전자 알고리즘을 이용하여 평가 지표를 최소화 시키는 전체적인 알고리즘은 다음과 같다.

Fuzzy Genetic Feature Ranking Algorithm

Initialize genetic operations, population number, weight coefficients

REPEAT

FOR pattern i DO

FOR pattern j ($i \neq j$) DO

 Calculate sum of distance of individual features for pattern pair using (2.4) and (2.6);

 Calculate membership values using (2.2) and (2.3);

END FOR

END FOR

Calculate Evaluation Index;

UNTIL (Evaluation Index is minimized)

4. 실험 및 결과

각 특징들의 중요도에 따라 순서를 정하기 위하여 제안한 방법에 대한 실험결과를 살펴보기로 하자. 먼저 실험을 위한 유전자 알고리즘에서 초기치 조건은 다음과 같다.

- i) Population Number = 40
- ii) Crossover Probability = 0.6
- iii) Mutation Probability = 0.001

4.1 Two cluster data

그림 4.1 처럼 두 개의 특징들과 cluster 로 이루어진 간단한 예제를 통해 제안한 고리즘의 성능을 살펴보기로 하자.

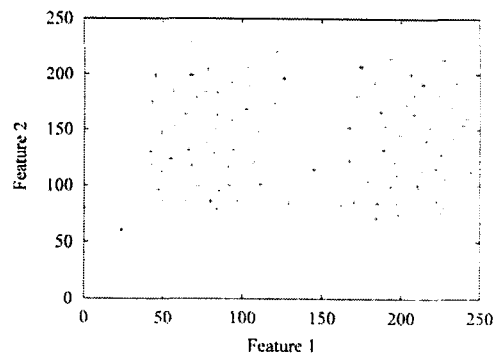


Figure 4.1. Two cluster data 에 대한 분포도.

그림에서 알 수 있듯이 feature 1 이 feature 2 보다 명확하게 구별할 수 있는 특징(characteristic)을 가지고 있음을 알 수 있다. 즉, feature 1 이 feature 2 보다 중요한 특징이라고 할 수 있다. 실험결과는 Table 4.1 에서 보듯이 이를 증명하고 있다.

Table 4.1. Two cluster data 에 대한 결과 weight 계수.

Feature	Weight	Rank
1	0.9909	1
2	0.0706	2

4.2 Iris data set

패턴인식 문제에서 가장 빈번하게 사용되는 iris data [7]는 3 개의 class 와 4 개의 특징들로 이루어져 있고 각 class 마다 패턴의 수가 50 이다.

이 실험결과는 neuro-fuzzy[1] 방법의 결과와 비교하였다. Table 4.2 이 weight 값의 대한 결과를 나타낸다.

Table 4.2. Iris data 에 대한 결과 weight 계수.

Feature	Algorithm			
	Neuro-Fuzzy		GA-Fuzzy	
	Weight	Rank	Weight	Rank
SL	0.0584	4	0.2234	3
SW	0.1944	3	0.0776	4
PL	0.9656	1	0.9741	1
PW	0.6035	2	0.9278	2

실험결과를 살펴보면 neuro-fuzzy 알고리즘보다 좋은 결과를 나타내고 있음을 알 수 있다.

5. 결 론

본 논문에서는 퍼지-유전자 접근방법을 이용한 무감독 특징 선택 방법을 제안하였다. 제안한 방법은 각 특징의 중요도를 결정하기 위하여 weighted distance 를 이용하여 feature evaluation index 를 최소화 시킴으로써 weight 의 계수를 구하는 것이다. 다른 특징 선택 알고리즘들은 데이터의 사전 정보를 기초로 하여 구

현되었지만 제안한 방법은 사전 정보없이 구현되었기 때문에 공학적인 측면에서 보다 효율이 있으며 neuro-fuzzy [1] 방법에 비하여 빠른 계산 속도와 local minima 에 대한 문제점들을 해결하였다.

참고 문헌

- [1] J. Basak, R. De, and S. Pal, "Unsupervised neuro-fuzzy feature selection," in *Proc. 1998 Int. Conf. Neural Network*, Anchorage, AK, vol. I, pp. 18-23, May 4-9, 1998.
- [2] J. Tou and R. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.
- [3] F. C.-H. Rhee and R. Krishnapuram, "Fuzzy rule generation methods for high-level computer vision," *Int. J. Fuzzy Sets Syst.*, Elsevier Science Publishers B. V. (North Holland), vol. 60, pp. 245-258, Dec. 1993.
- [4] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY: Plenum Press, 1981.
- [5] J. Bezdek and P. Castelaz, "Prototype classification and feature selection with fuzzy sets," *IEEE Trans. Syst. Man Cybern.*, vol. 7, pp. 87-92, 1977.
- [6] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [7] C. Merz. and P. Murphy, UCI Repository of machine learning databases, <http://www.ics.uci.edu/~mllearn/MLRepository.html>, Irvine, CA: University of California, Department of Information and Computer Science.
- [8] R. R. Yager, On the measure of fuzziness and negation. Part I: Membership in the unit interval, *Int. J. gen. Syst*, vol. 5, pp. 221-229, 1979.