

유전자 알고리즘과 정보이론을 이용한 속성선택

Feature Selection by Genetic Algorithm and Information Theory

조재훈* · 이대종** · 송창규** · 김용삼* · 전명근**

Jae Hoon Cho*, Dae Jong Lee**, Chang Kyu Song** and Myung Geun Chun**

* 충북 청주시, 충북대학교 전기전자컴퓨터공학부

** 충북 청주시, 충북대학교 BK21 충북정보기술사업단

요 약

속성선택(Feature Selection)은 패턴분류 문제에서 분류기들의 성능을 향상시킬 수 있는 중요한 기법이다. 특히, 많은 속성들을 가지는 데이터의 분류문제에서 관련이 적은 데이터, 중복되거나 또는 노이즈 있는 데이터를 제거한 주요 속성부분집합을 선택하여 이용함으로써 분류기의 정확도를 향상시킬 수 있다. 본 논문에서는 유전자 알고리즘과 정보이론의 상호정보량을 이용하여 속성선택을 하는 기법을 제안하였다. 실험을 통하여 제안된 알고리즘이 패턴인식문제에서 다른 방법들보다 성능이 우수함을 보였다.

키워드 : 속성선택, 패턴분류, 유전자 알고리즘, 상호정보량

Abstract

In the pattern classification problem, feature selection is an important technique to improve performance of the classifiers. Particularly, in the case of classifying with a large number of features or variables, the accuracy of the classifier can be improved by using the relevant feature subset to remove the irrelevant, redundant, or noisy data. In this paper, we propose a feature selection method using genetic algorithm and information theory. Experimental results show that this method can achieve better performance for pattern recognition problems than conventional ones.

Key Words : Feature selection, Pattern classification. Genetic algorithm, Mutual information.

1. 서 론

속성선택(Feature Selection)은 주어진 문제에서 효과적이고 개선된 해를 얻기 위해 유용한 속성들을 선택하는 처리과정이다. 일반적으로 모든 이용 가능한 속성들이 분류기의 좋은 성능에 기여하지는 않는다. 속성들 중에는 서로 다른 속성들 사이에서 혼란을 야기 시키는 중복성이나 분류기 성능을 저하 시키는 잡음들을 가지고 있는 속성들이 존재하기 때문에 주어진 문제를 해결하기 위해 사용될 수 있는 데이터 집합으로부터 최적의 속성부분집합을 선택하는 것은 패턴 분류 문제에서 중요하게 인식되어지고 있다.

속성선택기법은 검색 특징을 기반으로 exhaustive 기법, heuristic 기법 그리고 random 기법으로 나뉘질 수 있다[1]. 또한, 속성들의 선택에 이용되는 성능 평가 함수들을 기반으로 distance [2], information [3], dependence [4], consistency [5], 그리고 분류기의 오차비율을 이용하는 5개의 기법으로 구분할 수 있다.

속성선택의 조건으로서 분류기 오차 비율을 사용한 속성선택은 랩퍼(wrapper) 기법으로 알려져 있다. 랩퍼 기법에

기반을 둔 속성선택기법은 반복적인 속성들의 선택을 통한 분류기의 성능 평가로서 속성들을 선택하기 때문에 일반적으로 필터(filter) 기법으로 불리는 속성선택기법보다 우수한 성능을 보이는 장점이 있다. 반면에, 랩퍼 기법은 필터 기법에 비하여 연산속도가 느린 단점을 가지고 있다. 필터 기법에 기반을 둔 속성 선택은 학습알고리즘에 독립적으로 행해지고, 전처리 과정에서 적절한 속성들을 선택하기 때문에 연산 시간이 빠르다.

최근에는 정보이론을 이용한 속성선택기법들이 많이 연구되어지고 있다. Battiti[6]는 상호정보량(Mutual Information)을 이용한 MIFS(Mutual Information Feature Selection) 제안하였다. Kwak and Choi은 Battiti의 MIFS의 파라미터를 조절하여 성능이 개선된 방법을 제안하였다. 그러나 상호정보량만을 사용하는 것은 일반적으로 성능이 좋지 못한 것으로 알려져 있다[3].

속성선택기술은 다양한 분류기에 적용되어져왔다. Mao는 가지치기(pruning) 분석과 SVM (support vector machine)을 이용한 속성선택을 제안하였고[7], Hsu 등은 신경회로망에서 가중치를 기반으로 하는 ANNIGMA(artificial neural net input gain measurement approximation)-랩퍼 기법을 제안하였다[8]. Pal 과 Chntalpu디는 신경회로망을 학습하는 동안 중요 속성을 온라인으로 선택하는 개선된 기법을 제안하였다[9]. 또한, 무작위 기법으로서 진화 알고리즘의 속성선택 적용에 대한 연구들도 시도되어져왔다. 일반적으로, 유전자 알고리즘 기반 속성선택 기법에서 집단의 각각의 개체(영

접수일자 : 2007년 10월 25일

완료일자 : 2008년 1월 10일

감사의 글 : 본 연구는 산업자원부의 지원에 의하여 기초전력연구원(R-2007-2-046) 주관으로 수행된 과제임.

+ : 교신저자

색체)들은 속성의 부분집합으로 표현한다. n-차원 속성 공간에 대해서 각각의 염색체는 n-bit 이진스트링으로 표현되고, i번째 속성이 염색체에 의해 1로 표현되면, 그 속성은 속성부분집합으로서 선택되어지는 기법들이 일반적으로 사용되어 왔다. [10]에서, 유전자 알고리즘 기반 속성선택기술이 고차원데이터에 대해 다양한 고전적인 속성선택 기술보다 더 우수한 성능을 가지는 것을 보였으며, Siedlecki 와 Sklansky는 유전자 알고리즘을 이용한 속성선택에 대해 분기한정법(branch and bound) 기술을 이용하였다[11]. Pal 등은 속성선택에 대해 self-crossover로 불리는 새로운 유전 연산자를 제안하였다[12]. 현재까지도 다양한 진화 알고리즘을 이용한 속성선택 기법들이 연구되어지고 있다.

본 논문에서는 랩퍼와 필터 기법의 두 특징을 가지는 융합된 구조의 속성선택기법을 제안하였다. 먼저 상호정보량과 필터 기법을 이용하여 속성들을 전처리하고 우수한 속성들을 선택한 후 유전자 알고리즘과 신경회로망을 이용하여 랩퍼 기법으로 가장 적절한 속성들을 선택한다. 제안된 기법의 유용성과 성능을 평가하기 위하여 UCI Machine-Learning Repository [13]데이터에 적용하고 기존 기법들의 결과들과 비교하여 그 타당성을 보이고자 한다.

2. 상호정보량을 이용한 속성선택

속성선택문제에서 주요한 피쳐들은 출력에 대하여 중요한 정보들을 많이 포함하고 있고 반대로 그렇지 못한 속성들은 출력에 관해 적은 정보들을 포함한다. 분류 문제를 해결하기 위해서는 입력 속성에서 가능한 한 많은 정보들을 포함하도록 속성들을 선택 하여야한다. 이런 목적을 달성하기 위해 랜덤한 변수들의 정보를 측정하는 새논의 정보이론에서는 엔트로피(Entropy)와 상호정보량을 소개하였다. 기본적으로 엔트로피는 랜덤 변수들의 무질서도(uncertainty)를 측정하는 것이다. 만약 랜덤변수 X 가 $p(x) = \Pr\{X=x\}, x \in \lambda$ 의 소스 알파벳 λ 을 가진다면 X 의 엔트로피는 아래 식(1)과 같이 계산된다.

$$H(X) = - \sum_{x \in \lambda} p(x) \log p(x) \quad (1)$$

두 개의 랜덤 변수 X 와 Y 의 결합 엔트로피(joint entropy)는 아래 식 (2)로 계산될 수 있다.

$$H(X, Y) = - \sum_{x \in \lambda} \sum_{y \in \delta} p(x, y) \log p(x, y) \quad (2)$$

위 식(2)에서 $p(x, y)$ 는 X 와 Y 의 결합확률밀도함수(joint probability density function)이다. 또한, Y 를 알고 있을 때의 X 의 엔트로피는 수식(3)으로 계산될 수 있다.

$$H(X|Y) = - \sum_{x \in \lambda} \sum_{y \in \delta} p(x, y) \log p(y|x) \quad (3)$$

X 와 Y 의 공통의 정보량은 두 변수 사이의 상호정보량으로 정의 될 수 있고 아래 식 (4)로서 계산될 수 있다.

$$I(X; Y) = \sum_{x \in \lambda} \sum_{y \in \delta} p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)} \quad (4)$$

수식 (4)에서, 두 변수 사이의 상호정보량 $I(X; Y)$ 크면 두 변수는 가까운 연관성을 가지고, 그렇지 않으면 두 변수는 연관성이 적거나 독립일 수 있다. 또한 불확실도와 상호정보량

의 관계는 아래의 수식 (5)로 표현될 수 있다. 그림 1은 엔트로피 와 상호정보량의 관계를 벤다이어그램으로 보였다.

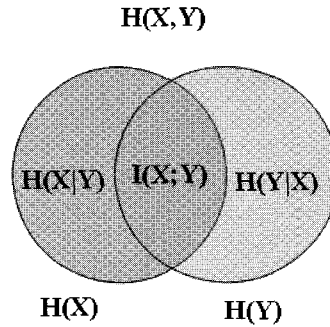


그림 1. 엔트로피와 상호정보량의 관계

Fig. 1. Relation between entropy and mutual information

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y), \end{aligned} \quad (5)$$

$$\begin{aligned} I(X; Y) &= I(Y; X), \\ I(X; X) &= H(X) \end{aligned}$$

패턴인식 문제의 속성 선택에서 연속적인 속성을 F , 클래스를 C 하면 속성과 클래스간의 상호정보량은 다음과 같이 정의 된다.

$$I(F; C) = \sum_{f \in \lambda} \sum_{c \in \gamma} p(f, c) \log \frac{p(f, c)}{p(f) \cdot p(c)} \quad (6)$$

상호정보량 $I(F; C)$ 이 클수록 속성 F 가 클래스 C 에 대해 많은 정보를 포함하는 것으로 정의되고 작으면 작을수록 속성 F 가 클래스 C 에 대해서 작은 정보를 포함하는 것으로 간주할 수 있다. 클래스에 대해 정보를 많이 포함한 속성을 선택함으로써 확률적으로 분류기의 성능을 우수하게 할 수 있고, 적은 정보량을 포함한 속성들을 제거함으로써 노이즈나 잘못된 데이터에 의한 분류기의 오차를 줄일 수 있다.

본 논문에서는 필터 기반의 전처리 선택으로서 이러한 상호정보량의 특징을 이용하여 각각의 속성들과 클래스 간의 상호정보량을 계산하고 순위별로 정리하여 각 속성들의 중요도를 평가하였다. 평가된 속성들을 기반으로 설정된 비율에 따라 하위 순위의 속성들을 제거하였다. 남은 속성들은 유전알고리즘의 초기 집단 생성에 있어 후보 집단으로 사용하였다.

3. 유전알고리즘을 이용한 속성선택

유전자 알고리즘의 기본 구조는 크게 초기화, 적합도 평가와 재생산, 교배, 돌연변이의 4단계로 구분된다. 초기화 단계에서는 최적화 문제의 해가 될 가능성이 있는 개체들의 집단이 형성된다. 해가 될 수 있는 해공간상의 초기점들은 무작위로 분포 되도록 선택되거나 아니면 경험적인 기법으로 선택된다. 경험적인 초기집단 생성은 문제마다 다르기 때문에 해공간에 대한 정보가 없을 경우에는 거의 사용하지 않는다. 그 다음 단계는 적합도 평가이다. 구해진 해들은 목적함수를 제공하고 이로부터 적합도를 평가한다. 적합도가 우수한 개체들은 다른 개체들보다 더 많이 선택되어 교배를 통해 재결합되는데 서로간의 유전정보를 교환함으로써 집단에 새로운 개체를 도입하게 되고 현 집단 내에 존재하는 정보만을 이용하여 변화를 시도하게 된다.

유전 알고리즘에서의 최적 속성 부분집합 선택은 N개의 속성을 유전자의 염색체 길이로 표현할 수 있고, N개의 속성

들을 1에서 N까지의 염색체 인덱스로서 표현할 수 있다. 따라서 유전 알고리즘에서의 속성 선택 문제는 전체의 임의의 해집합으로부터 분류기의 성능을 우수하게 하는 최소의 염색체(1로 표현되는 염색체 수)를 가지는 최적의 개체를 선택하는 문제이다. 속성선택을 위한 유전 알고리즘의 적합도 함수는 일반적으로 선택된 속성 수와 분류기의 성능으로 선정할 수 있고, 각 문제의 특성들에 따라 다양한 조건들을 적용할 수 있다.

일반적인 유전 알고리즘의 처리과정은 다음과 같다.

```

GA()
{
  initialize population P;
  repeat{
    select two parents p1 and p2 form P;
    offspring=crossover(p1,p2);
    mutation(offspring);
    replace(P, offspring);
  } until (stopping condition);
}
    
```

4. 상호정보량과 유전알고리즘을 이용한 속성선택 알고리즘

제안된 방법은 크게 전처리과정인 필터기법 단계와 유전 알고리즘을 이용한 램퍼기법 단계로 나누어진다. 필터기법 단계에서는 램퍼기법 단계를 위한 전처리 단계로서 상호정보량의 순위를 기반으로 속성들을 제거한다. 램퍼기법단계에서는 전처리된 속성들을 유전알고리즘과 신경망을 이용하여 최적의 속성부분집합을 선택한다.

그림 2에서는 본 논문에서 제안된 기법의 순서도를 나타냈으며 단계별 처리과정은 아래와 같다.

[단계 1] 문제의 입력데이터를 각각의 속성에 대하여 속성과 클래스 간의 상호정보량을 앞서 설명한 수식 (6)를 이용하여 계산한다.

[단계 2] 계산된 상호정보량을 기반으로 순위를 결정한다. 순위가 높은 순서대로 다시 정리하여 상호정보량이 낮은 속성들을 제거한 후, 유전 알고리즘의 초기 집단을 생성하기 위한 후보 속성들만을 저장한다. 이 과정까지가 필터 형태의 속성선택과정이다.

[단계 3] 단계 2에서 선택된 후보속성들을 가지고 초기 집단을 생성한다.

유전 알고리즘에서의 속성들은 염색체로서 이진 스트링으로 표현된다. 각각의 비트는 하나의 속성들을 표시하고 1이면 그 속성을 선택하고 0이면 선택하지 않는다. 초기 집단을 생성할 때 선택되는 속성의 비율(1의 개수)과 선택되지 않는 비율(0의 개수)는 설계자에 의해 생성과정에서 조절될 수 있다. 선택되어지는 속성의 비율이 커지면 유전 알고리즘의 연산시간이 증가할 수 있고, 너무 작으면 조기수렴이 발생할 수 있으므로 적절한 조절이 필요하다. 또한 염색체의 길이는 입력데이터의 총 속성 수와 동일한 크기로 코딩된다.

그림 3은 본 논문에서 사용된 n-차원 이진 염색체를 나타내었다. 초기 집단 생성 과정에서 랜덤변수와 β 는 0과 1사의 값을 가지며, 0과 1의 비율은 β 를 조절하여 원하는 비율로 생성할 수 있다. 초기집단 생성과정은 아래와 같이 수행하였다.

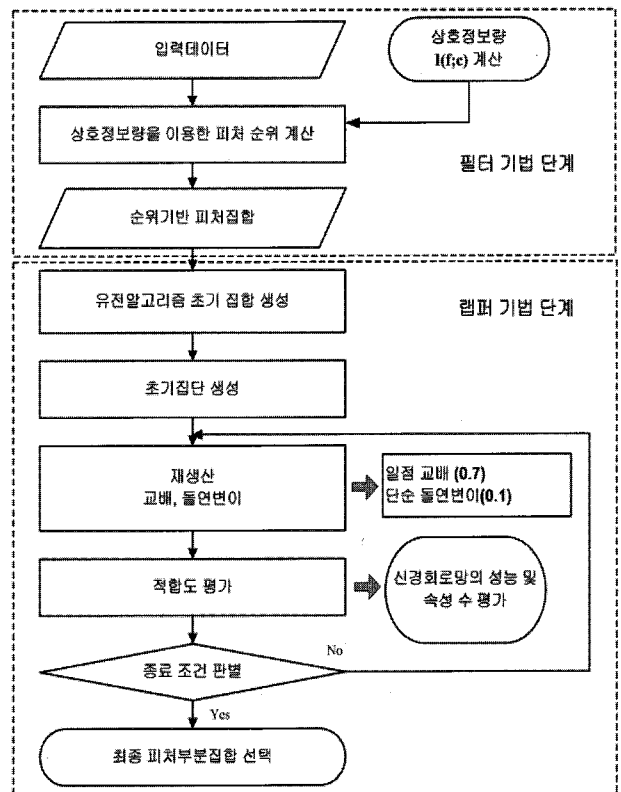


그림 2. 제안된 기법의 순서도
Fig. 2. Flowchart of proposed method.

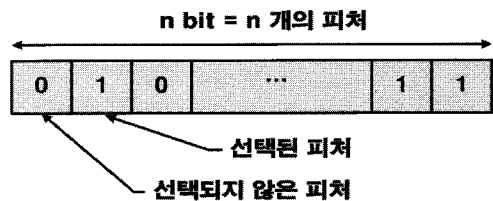


그림 3. n차원 이진 염색체.
Fig. 3. n-dimension binary chromosome

```

초기집단 생성
for (i = 1 to 집단크기)
  for (j = 1 to 염색체 길이)
    if (Generation random number(<math>\beta</math>))
      i번째 개체의 j번째 염색체 = 1
    else
      i번째 개체의 j번째 염색체 = 0
    end
  end
end
    
```

[단계 4] 유전 알고리즘의 선택 과정, 교배, 돌연변이를 통하여 속성들을 재생산한다.

[단계 5] 적합도 함수를 이용하여 각각의 염색체(선택된 속성들)에 대한 평가를 수행한다.

유전자 알고리즘의 적합도 계산은 분류기의 성능과 선택된 속성들의 수들을 이용하는 기법이 일반적이다. 두 조건들 사이에는 설계자에 의해 비율이 조정되는 가중치들을 사용하기도 한다. [14]에서는 위에서 설명한 적합도 평가의 조건을 아래 수식(7)로 정의하였다.

$$F = w * c(x) + (1 - w) * (1/s(x)) \quad (7)$$

위 식에서 F는 적합도, w는 분류기의 성능과 선택된 속성 사이의 가중치, c(x)는 분류기의 성능(분류기의 정확도), s(x)는 선택된 속성들의 수를 나타낸다. [15]에서는 아래의 수식을 이용하였다.

$$fitness(z) = \lambda \cdot acc(z) - (1 - \lambda) \cdot \frac{feats(z)}{totalfeat} \quad (8)$$

식(8)에서 fitness(z)는 선택된 속성부분집합 z에서의 적합도, λ는 두 척도의 가중치, acc(z)는 분류기의 정확도, feats(z)는 선택된 속성 수, totalfeat는 입력데이터의 총 속성 수를 나타낸다. 위 두 수식에서 설계자가 두 조건을 만족하는 속성들을 선택하고자 한다면, 즉, 속성들의 수가 가장 적으면서 분류기의 성능이 가장 우수한 속성부분집합을 선택하고자 한다면 첫째항과 두 번째 항의 비율을 동일하게 적용하여야 한다.

[단계 6] 반복횟수, 속성들의 수, 분류기의 정확도 등의 종료 조건들의 판별하여 종료를 하거나 [단계 4]에서부터 반복적으로 속성선택과정을 수행한다. 본 논문에서는 반복횟수에 대한 종료조건만을 사용한다.

5. 시뮬레이션 및 결과 고찰

제한된 기법의 성능을 평가하기 위하여 UCI의 기계학습 연구에서 제공하는 연구용 데이터들 사용하여 다른 기법들과 비교하였다. 모든 분석과 검증은 타당성의 확보를 위해서 데이터들을 10-겹 교차검정(10-fold cross validation) 방식에 의해 수행하였다. 실험 환경은 Pentium4 3.2GHz, 2G 메모리에서 수행하였다.

표 1은 본 실험에서 사용된 UCI 데이터들과 각각의 데이터들에 대한 학습데이터와 검증데이터들을 보였다. 표 2는 본 실험에서 사용된 유전자 알고리즘의 파라미터들을 나타냈다. 실험에서 적합도 함수는 식(8)을 이용하였으며 모든 실험에서 사용된 가중치 λ는 분류기의 성능과 선택되는 속성 수를 같은 비율로 반영하기 위하여 0.5로 설정하였다. 초기 집단 비율 β 은 램퍼기법 단계에서 초기에 선택되는 속성들과 선택되지 않는 속성의 비율을 같게 하기 위하여 0.5로 설정하였다. 유전자 알고리즘의 선택 기법은 룰렛휠 기법을 이용하였고, 일점교배와 단순 돌연변이를 사용하였다. 분류기로는 신경회로망을 이용하였다. 그림 4는 Ionosphere 데이터를 이용하여 제안된 기법의 최적속성 부분집합 탐색과정의 적합도 값의 변화를 보였고, 그림 5는 최적속성 부분집합 탐색과정의 선택된 속성 수를 보였다. 그림 4와 그림 5에서 알 수 있듯이 제안된 기법의 최적속성 부분집합 탐색이 효과적으로 수행되는 것을 알 수 있다.

표 1. 실험에서 사용된 UCI 데이터.
Table 1. UCI data used in our experiment.

데이터 명	속성 수	학습데이터	검증데이터
Monk3a	6	122	432
Monk3b	15	122	432
Cancer	9	399	300
Credit	9	490	200
Ionosphere	34	200	151
Pima	8	576	192

표 2. 유전자 알고리즘의 파라미터.
Table 2. Parameters for genetic algorithm.

파라미터	값
집단크기	20
교배 확률	0.7
돌연변이 확률	0.1
세대수(반복횟수)	20
가중치(λ)	0.5
초기집단비율(β)	0.5

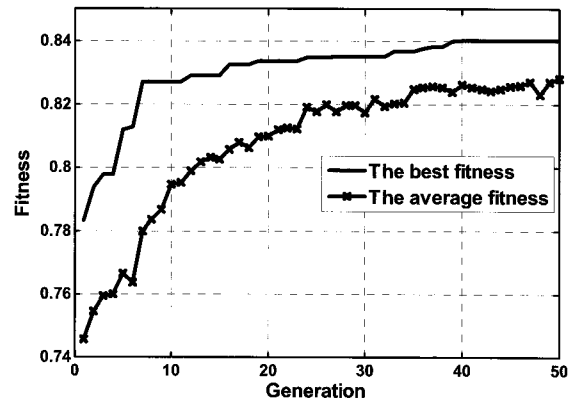


그림 4. Ionosphere 데이터에 대한 최적 속성 부분집합 탐색과정.

Fig. 4. Search process of optimal feature subset for Ionosphere data.

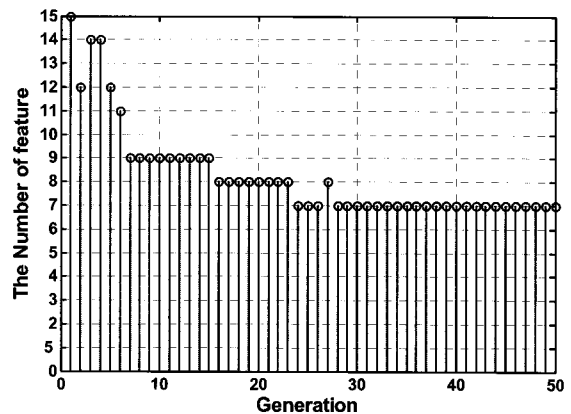


그림 5. 최적속성 부분집합 탐색과정에서 선택된 속성 수.
Fig. 5. The optimal number of selected feature in each generation.

표 3에서는 제안된 알고리즘과 다른 속성부분집합 선택 기법들과의 성능을 비교하였다. 선택된 속성 수 항목에서는 평균 속성선택수와 표준오차로 나타냈으며, 오차 항목에서는 평균오차와 표준편차로 표시하였다. 신경회로망의 오차는 데이터들의 모든 속성을 이용하였을 때의 오차이다. 램퍼 기법과 ANNIGMA의 램퍼 기법에서는 순차 탐색 기법 중 전방향 순차탐색(Forward Sequential Search)와 역방향 순차 탐색(Backward Sequential Elimination)기법을 적용하여 둘 중 우수한 값을 비교 값으로 사용하였다. 표 3에서 보이는바

표 3. 제안된 기법에 의한 속성선택과 다른 기법들의 성능비교[8]
Table 3. Performance comparisons with other techniques.

데이터	신경회로망		랩퍼 기법		ANNIGMA-랩퍼 기법		제안된 기법	
	원 속성수	오차(%)	선택된 속성수	오차(%)	선택된 속성수	오차(%)	선택된 속성수	오차(%)
Monk3a	6	10.0±5.2	3.4±1.6	5.1±3.4	2.3±0.7	2.9±0.8	2.0±0.3	2.2±0.1
Monk3b	15	2.8±0.0	4.4±1.1	2.8±0.0	2.2±0.4	2.8±0.0	2.2±1.0	2.2±0.1
Cancer	9	4.1±4.7	7.2±1.2	3.6±1.1	5.8±1.3	3.5±1.2	2.1±0.4	1.4±0.3
Credit	9	14.1±1.7	13.4±1.0	14.4±0.8	6.7±2.5	12.0±0.8	3.1±0.5	10.0±0.9
Ionosphere	34	11.4±3.9	32	10.2	9.0±2.5	9.8±1.3	8.2±2.3	4.2±0.6
Pima	8	24.1±5.0	6.9±1.0	23.0±1.3	5.2±1.4	22.2±1.4	3.4±0.5	18.2±1.1

와 같이 제안된 기법이 다른 기법들에 비해 우수함을 알 수 있다. 특히, cancer 와 Ionosphere 데이터의 오차는 기존의 방법들에 비해 약 50[%]의 성능향상을 보였으며, 다른 데이터들의 오차는 평균 18[%] 정도의 성능향상을 보였다.

6. 결 론

속성선택은 분류기의 성능을 높이기 위해 초기 입력데이터 속성들에서 유용한 속성들만을 선택하여 사용하는 기법으로 다양한 기법들이 제안 되어져 왔다. 필터 기법 기반 속성선택은 연산시간은 우수하지만 랩퍼 기법에 비하여 성능이 저하되는 단점을 가지고 있고, 랩퍼 기법기반 속성선택은 반복적인 성능 평가로 인해 성능은 우수하지만 연산시간을 증가시키는 단점을 가지고 있다.

본 논문에서는 랩퍼와 필터 기법의 두 특징을 가지는 융합된 구조의 속성선택 기법을 제안하였다. 필터 선택 단계에 상호정보량을 이용하여 각 속성들의 순위를 결정하였으며, 유전자 알고리즘과 신경회로망을 이용하여 랩퍼 기법으로 가장 적절한 속성들을 선택하였다. 제안된 기법의 성능을 평가하기 위해 UCI 데이터에 적용하고 기존의 기법들과 비교하였다. 실험결과에서 제안된 기법이 기존의 기법보다 속성선택의 수는 작으면서 분류기의 성능이 우수함을 보였다.

참 고 문 헌

[1] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 3, pp. 131-156, 1997.

[2] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature selection," *IEEE Trans. Comput.*, vol. C-26, no. 9, pp. 917-922, Sep. 1977.

[3] N. Kwak and C.-H. Choi, "Input feature selection for classification problems," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 143-159, Jan. 2002.

[4] A. N. Mucciardi and E. E. Gose, "A comparison of seven techniques for choosing subsets of pattern recognition," *IEEE Trans. Comput.*, vol. C-20, pp. 1023-1031, Sep. 1971.

[5] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, pp. 155-176, 2003.

[6] Battiti, R., "Using mutual information for selecting features in supervised neural net learning", *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.

[7] K. Z. Mao, "Feature subset selection for support vector machines through discriminative function pruning analysis," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 1, pp. 60-67, Feb. 2004.

[8] Chun-Nan Hsu, Hung-Ju Huang, and Dietrich Schuschel, "The ANNIGMA-Wrapper Approach to Fast Feature Selection for Neural Nets," *IEEE Trans. on Syst. man and Cybernetics-PART B: CYBERNETICS*, vol. 32, no. 2, 2002.

[9] N. R. Pal and K. Chintalapudi, "A connectionist system for feature selection," *Neural, Parallel, and Sci. Comput.*, vol. 5, pp. 359-381, 1997.

[10] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognit.*, vol. 33, pp. 25-41, 2000.

[11] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for largescale feature selection," *Pattern Recognit. Lett.*, vol. 10, pp. 335-347, 1989.

[12] N. R. Pal, S. Nandi, and M. K. Kundu, "Self-crossover: A new genetic operator and its application to feature selection," *Int. J. Syst. Sci.*, vol. 29, no. 2, pp. 207-212, 1998.

[13] C. J. Merz and P. M. Murphy. UCI repository of machine learning databases. Dept. Computer Science, Univ. California, Irvine. Online available : <http://www.ics.uci.edu/~mlearn/MLRepository.html>.

[14] F. Tan, X. Fu, Y. Zhang and Anu G. Bourgeois, "Improving Feature Subset Selection Using a Genetic Algorithm for Microarray Gene Expression Data", *IEEE Congress on Evolutionary Computation*, pp. 2529-2534, 2006.

[15] J. J. Aguilera, M. Chica, M. J. del Jesus and F. Herrera, "Niche genetic feature selection algorithms applied to the design of fuzzy rule-based classification systems", *IEEE International conference on Fuzzy Systems Fuzz-IEEE2007*, pp. 1-6, 2007.

저 자 소 개



조재훈(Jae Hoon Cho)
2002년 : 한밭대학교 제어계측공학과(학사)
2004년 : 한밭대학교 제어계측공학과(공학석사)
2005년~현재 : 충북대학교 제어계측공학과 박사과정

관심분야 : 지능시스템, 다중생체인식, 퍼지이론
E-mail : mmi8988@lycos.co.kr



이대종(Dae Jong Lee)
1995년 : 충북대학교 전기공학과(학사)
1997년 : 충북대학교 전기공학과(공학석사)
2002년 : 충북대학교 전기공학과(공학박사)
2004년~2005년 : University of Alberta, Postdoc
2006년~현재 : 충북대학교 BK21 충북정보기술단 초빙조교수

관심분야 : 음성신호처리, 얼굴인식, 다중생체인식
E-mail : djmidori@empal.com



송창규(Chang Kyu Song)
1995년 : 충북대학교 전기공학과(학사)
1997년 : 충북대학교 전기공학과(공학석사)
2006년 : 충북대학교 전기공학과(공학박사)
2006년~현재 : 충북대학교 BK21 충북정보기술사업단 Postdoc.

관심분야 : 영상처리, 영상압축, 패턴인식, 지능시스템
E-mail : sckyu@chungbuk.ac.kr



김용삼(Yong Sam Kim)
2006년 : 충북대학교 전기전자공학과(학사)
2006년~현재 : 충북대학교 제어계측공학과 석사과정

관심분야 : 다중생체인식, 신호 처리, 임베디드 프로그래밍
E-mail : yszzang99@paran.com



전명근(Myung Geun Chun)
1987년 : 부산대학교 전자공학과(학사)
1989년 : 한국과학기술원 전기 및 전자공학과(공학석사)
1993년 : 한국과학기술원 전기 및 전자공학과(공학박사)
1993년~1996년 : 삼성전자 자동화연구소 선임연구원

2000년~2001년 : University of Alberta 방문교수
1996년~현재 : 충북대학교 전기전자 및 컴퓨터공학부 교수

관심분야 : Biometrics, 패턴인식, 데이터 마이닝, 얼굴인식
E-mail : mgchun@chungbuk.ac.kr