

정보이득 분할을 이용한 분류기법의 지배적 초월평면 생성기법

이형일*

A dominant hyperrectangle generation technique of classification using IG partitioning

Hyeong-il Lee*

요약

중첩형 일반화 사례 (NGE, Nested Generalized Exemplar) 기법은 거리 기반 분류를 최적 일치 규칙으로 사용하며, 노이즈에 대한 내구력을 증가시켜 주는 동시에 모델 크기를 감소시키는 장점이 있다. NGE 학습 중 생성된 교차(cross)나 중첩(overlap) 현상은 분류성능을 저해하는 요인으로 작용한다. 따라서 본 논문은 NGE 학습 중 생성된 교차나 중첩 현상이 발생한 초월 평면에 대해 상호정보가 가장 큰 구간을 분리하여, 새로운 초월평면을 구성하게 하여, 분류성능 향상시키고 초월평면의 개수를 감소시키는 기법인 DHGen(Dominant Hyperrectangle Generation) 알고리즘을 제안하였다. 제안한 DHGen은 분류성능면에서 kNN과 유사하고 NGE이론으로 구현한 EACH보다 우수함을 UCI Machine Learning Repository에서 벤치마크 데이터를 발췌한 실험 자료로 입증하였다.

▶ Keywords : 사례기반학습, 중첩형 일반화 사례기법, 메모리 기반 학습, 정보이득

Abstract

NGE(Nested Generalized Exemplar Method) can increase the performance of the noisy data at the same time, can reduce the size of the model. It is the optimal distance-based classification method using a matching rule. NGE cross or overlap hyperrectangles generated in the learning has been noted to inhibit the factors. In this paper, We propose the DHGen(Dominant Hyperrectangle Generation) algorithm which avoids the overlapping and the crossing between hyperrectangles, uses interval weights for mixed hyperrectangles to be splitted based on the mutual information. The DHGen improves the classification performance and reduces the number of hyperrectangles by processing the training set in an incremental manner.

•제1저자 : 이형일

•투고일 : 2013. 12. 29, 심사일 : 2014. 01. 10, 게재확정일 : 2014. 01. 17.

* 김포대학교 인터넷정보과(Dept. of Internet Information, Kimpo College)

•이 논문은 2013학년도 김포대학교의 연구비 지원에 의하여 연구되었음

The proposed DHGen has been successfully shown to exhibit comparable classification performance to k-NN and better result than EACH system which implements the NGE theory using benchmark data sets from UCI Machine Learning Repository.

▶ Keywords : Case-Based Learning, Nested Generalized Exemplar Method, Memory-Based Learning, Information Gain

I. 서 론

사례 기반 학습은 추상화 모델을 따로 구성하지 않고 학습 사례 중 발췌된 사례를 개념 모델로 직접 이용한다. 이들의 경우에는 어느 정도 유효한 것으로 입증되긴 하지만 그와 관련된 명시적 일반화 모델이 존재하지 않으므로써, 그 판독 능력을 제한 받게 된다. 그에 반해, 규칙 기반 학습은 필요 모델을 설정하고, 그것에 기반해서 결론을 도출함으로써, 납득할 만한 규칙 집합을 제공해 줄 수 있다[1,2].

혼합형 사례 기반 학습은 위 두 가지 기법을 혼합한 개념으로 일반화 사례 개념에 기반하고 있으며, 거리 기반 분류 방안을 최적 일치 규칙 개념으로 사용한다. 이것은 일종의 규칙으로 해석될 수 있을 뿐 아니라, 사례 기반 분류기 노이즈에 대한 내구력을 증가시켜 주는 동시에, 모델 크기를 감소시킬 수 있다[2]. 이 혼합형 사례 기반 학습에 기여한 주요 이론 중 한 가지는 중첩형 일반화 사례(NGE, Nested Generalized Exemplar) 기법이다. 이 기법에서 사례는 점 또는 초월평면의 형태를 취하게 되며, 테스트 사례는 가장 가까운 예제의 클래스로 분류한다. 예제가 점(point)일 경우에는 점과의 거리를 계산하며, 초월평면일 경우에는 가까운 면과의 거리를 계산한다[1,3,4].

Salzberg에 의해 제시된 EACH (Exemplar-Aided Constructor of Hyperrectangles) 알고리즘은 해당 사례를 점진적 추가하는 방식으로 일반화시켜 주는 동시에, 초월평면이 중첩 및 내포 방식으로 생성될 수 있도록 허용해 준다. EACH에서 신규 입력 사례는 속성 및 초월평면별 동적 가중치를 포함한 일반화된 거리를 기준으로 분류된다. 최근접(Nearest Neighbor) 분류기와 EACH 사이의 비교 분석 자료에서 EACH가 NN보다 성능이 떨어지는 것은 주로 중첩/내포형 초월평면이 존재함으로써 발생하게 된 것임을 지적하

였고, 그 결과로 인해 NGE 접근 방법에 대한 몇 가지 변형 모델이 제안되었다[2,5].

그 중 가장 효과적으로 검증된 것은 초월평면 사이에 중첩 현상이 발생하지 않도록 작용하는 BNGE(Batch Nested Generalized Exemplars) 기법이 있다. 이 기법에서는 상호 정보 기반의 속성별 가중치의 사용과 학습 집합 내 존재하는 사례를 일괄 배치 방식으로 처리한다. 또한 거리 및 규칙 기반 귀납법을 통합한 RISE이 있다[2-8].

일괄 배치 방식은 사례 제시 순서에 따라 결정되며, 표준 분류시 결과는 좋지 않으나, 온라인 학습 시나리오에서 사용할 수 있다. 배치 모드 방법은 휴리스틱스(heuristics) 방식을 채용하여 매 단계마다 병합 또는 일반화할 사례의 선택을 결정한다. 제공되는 결과는 매우 흥미롭고, 일반적으로 기준선 방법(baseline method)으로 사용되는 1NN(1-nearest neighbour) 분류자를 통해 얻는 결과를 부분적으로 능가한다.

본 논문은 NGE 학습 중 생성된 교차나 중첩 현상이 발생한 초월 평면에 대해 상호정보가 가장 큰 구간을 분리 초월평면을 구성케 함으로써, 각 초월평면에는 하나의 클래스로 표시하도록 하여 분류 시 정확성을 증가시키는 기법을 제안하였다. 이 접근법을 1NN와 EACH, 그리고 BNGE, RISE등과 비교하여 분류성능을 개선시키고 최종 집합에 저장된 초월 평면의 수는 상당히 줄어들었다.

본 논문의 구성은 다음과 같다. II장은 관련연구로 초월평면 학습에 관한 설명이다. III장은 이 문제를 해결하기 위한 상호정보기반의 초월평면 분리 알고리즘을 설명한다. IV장에는 실험적 프레임워크와 그 결과 및 분석을 제시했다. 마지막으로 V장에는 결론을 제시했다.

II. 관련 연구

이번 장에서는 NGE 기법의 비교분석으로 사용하는 최근접(Nearest Neighbor) 분류기와 NGE 기법, 그리고 초월평면 학습에 관한 변형 모델인 BNGE(5)와 RISE(7)를 학습 측면에서 간략하게 설명하였다.

1. 최근접(Nearest Neighbor) 분류

최근접 분류기는 먼저 전체 사례를 단순히 메모리에 저장한다. 그리고 시험할 사례와 메모리에 저장된 사례들과의 거리를 수식 (1)을 이용하여 계산한 다음 계산한 거리를 기준으로 테스트 사례와 근접한 k개의 저장된 사례를 선정한다. 이 선정된 k개 중에서 가장 많은 개수의 학습사례를 포함하는 클래스로 시험 사례를 분류하는 알고리즘이다.

$$D_{EQ} = \sqrt{\sum_{i=1}^n (E_i - Q_i)^2} \quad (1)$$

이때, E 는 메모리에 저장된 학습사례를 나타내며, Q 는 주어진 입력사례이다. 또한 n 은 사례를 구성하는 특징의 개수이며, E_i, Q_i 는 각각 학습사례와 입력사례의 i 번째 특징 값을 나타낸다. 이 때 k값은 분류기의 성능을 최적화하기 위하여 일반적으로 Cross Validation기법을 사용하여 결정하며, $k=1$ 인 경우를 NN 분류기라 한다(10,11). 이미 다양한 분야에 응용되고 있다.

2. NGE 기법

NGE기법은 먼저 각각의 사례를 하나의 클래스와 n 속성 값으로 구성되는 일련의 L 사례 $\{E^1, E^2, \dots, E^L\}$ 로 메모리에 저장한다. 속성 값은 숫자 형태이다. NGE기법은 그림 1과 같은 절차에 따라 학습된다. 이 학습 과정의 목표는 일단의 초월평면 집합 $\{H^1, H^2, \dots, H^K\}$ 을 구성하는 것이다.

초월평면에는 통상 일련의 사례가 포함되어 있으며, 그 각각의 영역에는 일정 범위의 값으로 지정되어 있게 된다. 그리고 초월평면에 단지 하나의 사례만 포함되어 있는 특별한 경우에는 비일반화된 표본으로 간주된다. 각 초월평면에는 또한 하나의 클래스 라벨이 표시되어 있게 되는데, 만일 동일한 초월평면에 다른 클래스 라벨이 표시된 사례가 포함되어 있는

경우 혼합 초월평면으로 간주된다(2,11).

그림 1은 NGE기법의 알고리즘을 순서도로 나타낸 것으로 식 (2)에서 초월평면 H 와 example $E = (E_1, E_2, \dots, E_n)$ 사이의 거리 $D(E, H)$ 계산 결과에 근거하여 유사도 계산이 진행된다.

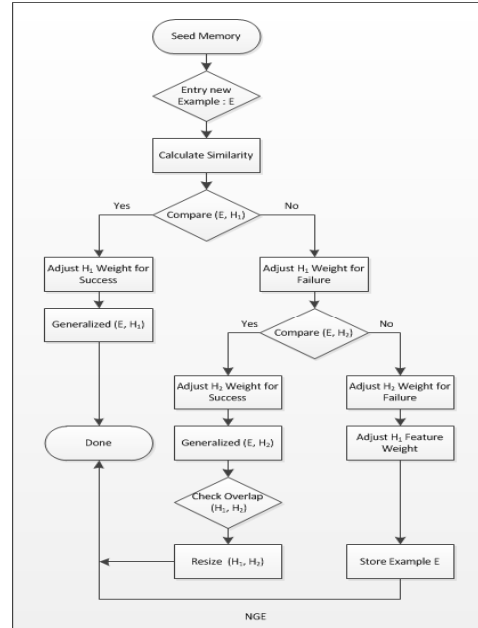


그림 1. NGE 기법의 알고리즘
Figure 1. Algorithm of NGE

$$D(E, H) = \sqrt{\sum_{i=1}^n \left(w_i \frac{d(E_i - H_i)}{E_i^{\max} - E_i^{\min}} \right)^2} \quad (2)$$

식 (2)의 H_i 는 만일 E_i 가 숫자형 속성으로 구성되어 있으므로 interval $[H_i^{\min}, H_i^{\max}]$ 을 나타내며 식 (3)와 같이 나타낸다.

$$d(E_i, H_i) = \begin{cases} 0 & (\text{그외}) \\ H_i^{\min} - E_i & (H_i^{\min} > E_i) \\ E_i - H_i^{\min} & (E_i > H_i^{\min}) \end{cases} \quad (3)$$

3. BNGE와 RISE

BNGE(Batch Nested Generalized Exemplar)는 NGE의 배치 버전이고, 초기의 NGE에서 보이는 몇 가지 단

점을 개선한다. 즉 BNGE에서 NGE의 증분법을 배치 모드(batch mode)로 바꾸고, 초월 평면의 구성 시 제외했던 범주형 속성을 포함시키고, 그리고 결측값(missing value)을 처리하도록 하는 등 일치 규칙이 일부 수정되었다. 초월 평면의 일반화는 그 한계를 확장하여 원하는 사례를 포함시키기만 하면 된다(2). 이때 식 (2)의 H_i 는 만일 E_i 가 E_i 가 범주형 속성으로 구성된 경우가 추가되어 일정 목록의 값을 식 (4)와 같이 나타낸다.

$$d_{범주}(E_i, H_i) = \begin{cases} 0 & (E_i \text{가 } H_i \text{에 속하면}) \\ 1 & (\text{그외}) \end{cases} \quad (4)$$

$$d_{수치}(E_i, H_i) = \begin{cases} 0 & (\text{그외}) \\ H_i^{\min} - E_i & (H_i^{\min} > E_i) \\ E_i - H_i^{\min} & (E_i > H_i^{\min}) \end{cases}$$

RISE는 거리 기반 학습과 규칙 기반 귀납법을 통합화하여 이들의 한계를 극복하는 접근법이다. 위의 설명과 유사한 절차를 따르지만, 명목적 속성에 SVDM 거리[12]가 사용되므로, 거리 계산이 더욱 개선된다. RISE는 사례를 포함하는 최소 규칙을 선택하지 않고 가장 높은 정확도를 가진 규칙을 선택한다(7).

BNGE와 RISE는 비슷한 절차를 통해 초월 평면을 생성한다. 이들은 완벽한 학습 집합에서 시작하며, 전역 정확성이 감소하지 않는 경우 최단 사례/초월 평면을 병합한다. RISE는 교차 검증 방법론을 사용하여 훈련 정확도를 계산하고, 초월 평면 간의 중첩 회피는 사용되지 않는다(2,10).

III. 정보이득 분할을 이용한 지배적 초월평면 생성기법

기존 NGE 기법은 학습 후 생성된 초월평면이 교차나 중첩 현상이 발생하여 성능이 떨어지는 것이 문제점으로 지적되어 왔다. 본 논문에서는 NGE 기법의 문제점들을 정보이득(Information Gain) 분할기법을 통해 초월평면을 구성하여 성능이 떨어지는 것을 방지하였다.

그림 2는 본 논문에서 제안한 정보이득 분할을 이용한 분류기법의 지배적 초월평면 생성 (이하 DHGen, Dominant Hyperrectangle Generation) 기법의 모델이다. 그 구성은 4 단계로 첫 번째 단계는 학습 자료에 대한 정규화 단계와 두 번째 단계는 II장 2의 NGE 단계, 그리고 세 번째 단계인 지배적 초월평면의 정보이득 분할단계, 마지막 단계인 분류단

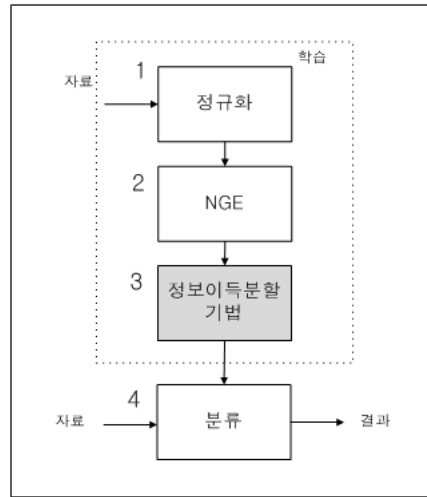


그림 2. DHGen 기법의 모델
Figure 2. Model for DHGen Method

계이다. 두 번째 단계인 II장 2의 NGE 단계(5)는 학습사례에 의해 생성된 일단의 초월평면 집합이 구성되는 단계이며, 세 번째 단계인 지배적 초월평면의 정보이득 분할단계는 본 논문에서 제안한 기법으로 NGE 학습 중 생성된 교차나 중첩 현상이 발생한 초월 평면에 대해 상호정보가 가장 큰 구간을 분리하여 초월평면을 구성케 함으로써 성능이 떨어지는 것을 방지하였다. 마지막 네 번째 단계인 분류단계는 주어진 자료에 대해 생성된 초월평면들과 식 (7)로 거리 계산을 하여 가장 가까운 초월평면의 클래스로 출력을 결정하였다.

1. 특징 정규화(Feature Normalization)

DHGen에서 출력 클래스의 결정은 입력 사례와 메모리에 저장된 초월평면 사이의 거리를 이용하게 된다. 이 기법에서는 사례를 구성하는 특징들이 갖는 값의 범위가 편이하게 다를 경우 문제가 발생하게 된다. 예를 들어 (0.9, 400, 0.0004), (0.8, 410, 0.02)와 같은 특징으로 구성된 사례에서, 두 번째 특징은 다른 두 개의 특징에 비하여 상대적으로 큰 값으로 구성되어있다. 따라서 두 번째 특징이 조금만 차이가 나더라도 나머지 특징간의 차이에 관련 없이 출력 클래스가 결정된다. 이러한 문제점의 해결을 위하여 DHGen에서는 다음의 식 (5)를 이용하여 특징 값을 정규화 한다. 이 기법은 식 (5)에 의하여 사례를 구성하는 모든 특징 값을 0과 1사이의 값으로 정규화 함으로써, 모든 특징의 변화가 사례의 소속 클래스 결정에 미치는 영향력을 동일하게 한다.

$$f_{i_n} = \frac{f_i - f_{i_{\min}}}{f_{i_{\max}} - f_{i_{\min}}} \quad (5)$$

이때 f_i 는 i 번째 특징 값, $f_{i_{\max}}, f_{i_{\min}}$ 는 f_i 가 가질 수 있는 최대값 및 최소값을 나타낸다.

2. 지배적 초월평면의 정보이득분할기법

그림 3은 초월평면 정보이득 분할기법의 상세 순서도이다. 그림 3의 순서도에 번호가 붙은 푸른색으로 표시된 4개의 단계가 주요 개념을 이루는 단계이다. 순서도의 1) Store E 단계는 사례 E가 해당 초월평면 H에 소속되는 것으로 평가되어 저장된다. 문제는 저장되는 초월평면의 상한과 하한 값이 사례 E의 수치가 교차되거나 해당 초월평면이 가진 클래스가 서로 다른 경우로 기법의 성능을 저해하는 주요한 원인으로 지적되어 왔다.

따라서 교차나 중첩 현상이 발생한 초월평면에 대해 포함된 사례들에 대해 적절한 분할점을 찾아 분할하는 것이 제안한 기법을 주요 개념 중의 하나이다. 따라서 사례들의 분할점을 선택하기 위하여 그림 4와 같이 사례 값을 오름차순으로 정렬하고 값이 변화하는 위치를 경계값(Boundary Value)으로 선정한다. 예를 들어, 그림4에서 53과 55 사이의 경계값은 두 특징 값의 평균인 54가 된다. 순서도의 2)의 단계

(Find boundary value)에 해당된다.

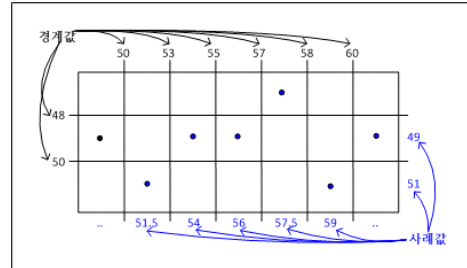


그림 4. 경계 값 선정
Figure 4. select the boundary value

구한 경계값들 중, 결정트리 알고리즘의 결정 노드 (Decision Node)에서 특징의 비교기준을 선정할 때 사용하는 IG(Information Gain) 값을 이용하여 가장 변별력이 좋은 경계값을 분할점으로 선택한다. 여기서 IG값은 수식 (6), (7)을 이용하여 계산한다.

$$I = - \sum_{i=1}^C p_i \log_2 p_i \quad (6)$$

p_i 는 학습사례 집합에서 클래스 i 에 소속되는 사례의 비율이며, C 는 클래스의 개수를 의미한다.

$$IG(f) = I - \sum_{i=1}^N P_i I_c \quad (7)$$

I 는 분할 이전의 정보량이며, P_i 는 분할 이전의 학습사례 중, 분할된 각 초월평면에 포함된 학습사례의 비율이다. I_i 는 특정 경계값 f 를 기준으로 분할했을 때 분할된 각 초월평면의 정보량을 의미하며, 수식 (6)을 이용하여 계산한다. 이때 I 값이 크다는 사실은 올바르게 분류하기 위하여 많은 양의 정보가 필요하다는 것을 의미하며, IG값은 분할 이전의 정보량과 경계값을 기준으로 초월평면을 분할했을 경우 정보량의 차이를 의미한다. 즉, IG값은 초월평면을 분할한 이후의 정보량이 작아질 경우에 큰 값을 가지게 되며, 결국 IG값이 큰 경계값을 분할점으로 선택할 때 초월평면의 효율적인 분할이 가능하다. 순서도의 3) 단계(Compute IG value for every boundary values)에 해당된다.

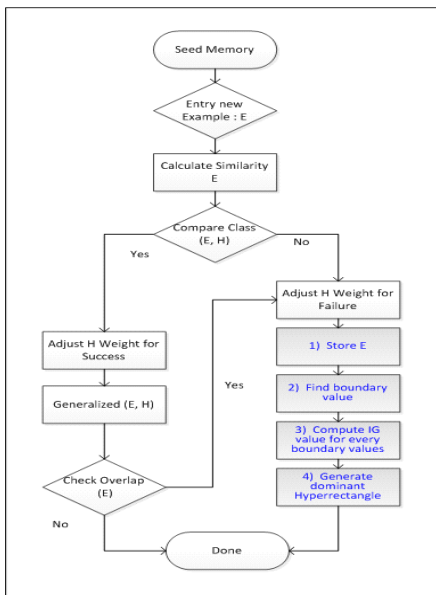


그림 3. 지배적 초월평면 생성기법 순서도
Figure 3. Detail Flowchart for DHG Method

순서도의 4) 단계(Generate Dominant Hyperrectangle) 인

지배적 초월평면의 생성은 분할된 초월평면의 클래스를 선택된 초월평면에 포함된 Majority Class Selection으로 결정한다.

3. 사례의 분류

제안한 DHGen기법은 테스트 사례를 분류하기 위하여 초월 평면들과 수식 (8)로 거리 계산을 하여 가장 가까운 초월평면의 클래스로 출력을 결정하였다. 이때 거리의 계산에는 분류 성능 향상을 위하여 사례의 최종적으로 생성된 분할공간에 대응하는 수식 (7)의 IG값 구해 입력사례와 메모리에 저장된 초월평면간의 거리계산에 있어 특징의 가중치로 사용한다.

$$D_{EQ} = \sqrt{\sum_{i=0}^n IG(i)(E_{f_i} - Q_{f_i})^2} \quad (8)$$

IV. 실험 및 분석

본 논문에서 제안한 DHGen기법의 성능을 Stratified 10-fold Cross-validation 기법을 사용하여 1NN, EACH, BNGE, RISE 등과 비교 검증하였다[1,11].

1. 실험 데이터

본 논문에서는 기계 학습의 벤치마크 자료로 많이 사용되는 UCI Machine learning Database Repository에서 9 개의 데이터 셋을 발췌하여 사용하였다[11]. 이들 데이터는 모든 특징이 실수 값을 갖는다. 다음의 표 1은 실험 자료의 분포를 보여주고 있다.

표 4. 클래스별 학습패턴의 분포
Table 4. Distribution of learning Patterns in Classes

데이터 셋	패턴 개수	특징 개수
bupa	286	9
cleveland	297	13
glass	214	9
haberman	306	3
iris	150	4
newthyroid	215	5
pima	768	8
wine	178	13
wisconsin	683	9

2. 실험결과

분류 성능 실험에서 각 알고리즘의 파라미터는 표 2에서와 같이 EACH는 시드(Seed) 개수 5, 가중치 변화량 0.2를, 그리고 RISE는 Q = 1, S = 2를 적용하였다. 그리고 1NN, BNGE, 그리고 DHGen은 제한하지 않았다.

표 5. 알고리즘의 파라미터
Table 5. parameters of algorithm

algorithms	Parameters
BNGE	-
EACH	시드(Seed) 개수 5
RISE	Q = 1, S = 2
DHGen	-

표 3은 제안한 기법이 k-NN과 EACH, 그리고 BNGE, RISE를 비교한 결과이다. EACH는 bupa와 cleveland에서 저조한 성능을 보이는 것은 무작위(Random)로 설정한 초기 시드(seed)의 영향으로 분석되다.

표 3에서 DHGen기법은 bupa와 cleveland에서 성능이 다소 저조하지만, 그 외의 데이터 셋에서는 EACH보다는 우수하고 BNGE와 RISE와 유사한 성능을 안정적으로 보이고 있다.

표 6. 분류성능 Table
Table 6. Classification performance using different datasets from UCI

데이터셋 알고리즘	1NN	EACH	BNGE	RISE	DHGen
bupa	0.61	0.44	0.64	0.64	0.57
cleveland	0.53	0.45	0.58	0.49	0.40
glass	0.73	0.50	0.66	0.69	0.62
haberman	0.66	0.68	0.68	0.64	0.72
iris	0.93	0.92	0.96	0.94	0.95
newthyroid	0.97	0.91	0.95	0.95	0.91
pima	0.70	0.65	0.73	0.64	0.61
wine	0.95	0.85	0.96	0.94	0.94
wisconsin	0.95	0.87	0.96	0.94	0.95
평균	0.78	0.70	0.79	0.76	0.74

3. 초월평면의 개수 비교

표 4는 표 3에 대한 초월평면의 개수를 측정한 것이다. 측정 시 초월평면의 중복과 교차 등의 문제를 해결 하려고 시도 한 알고리즘인 BNGE, RISE, DHGen을 비교하였다. 제안

한 DHGen은 평균 개수에서 현저하게 초월평면(rule)의 개수를 줄였으며, 특히 iris, newthyroid, wine, wisconsin의 경우 성능 대비 초월평면 개수가 매우 작아 효율적인 것으로 나타났다.

표 7. 초월평면 개수
Table 7. number of hyperrectangles

데이터셋 알고리즘	BNGE	RISE	DHGen
bupa	184.3	201	32.4
cleveland	125.9	161.7	99.1
glass	78.5	92.2	21.1
haberman	212.4	132.4	33.0
iris	12	37.1	12.0
newthyroid	17.3	33.9	17.2
pima	326.9	43.9	64.3
wine	10.9	29.7	16.0
wisconsin	66.3	182.9	61.9
평균	114.9	101.6	39.6

V. 결론

본 논문은 혼합형 변형 모델중 대표적인 NGE 기법에서 생성된 교차나 중첩 현상이 발생한 초월 평면에 대해 상호정보가 가장 큰 구간을 분리 초월평면을 구성케 함으로써 각 초월평면에는 하나의 클래스로 표시하도록 하여 분류 시 정확성을 증가시키는 기법을 제안하였다.

제안한 기법은 아주 적은 개수의 초월평면을 이용하여 1NN보다 평균 4% 정도 떨어지지만 EACH시스템 보다 4% 정도 우수한 분류성능을 보이고 있다. 또한 BNGE와 RISE와 비교하면, 분류성능은 3.5% 정도 낮게 보이지만 초월평면 평균 개수에 있어서는 각각 79.3, 62개 정도로 현저하게 줄었다. 특히 iris, newthyroid, wine, wisconsin의 경우 성능 대비 초월평면 개수가 매우 우수한 것으로 나타났다.

참고문헌

[1] Aha, D.W. et al, "Instance-Based Learning Algorithms. Machine Learning," Vol. 6, pp 37-66. 1991.
 [2] D. Zaharie, L. Perian, V. Negru, "A View Inside the Classification with Non-Nested Generalized Exemplars," IADIS European Conference on Data Mining, 24-26 July, Rome, Italy, pg.19-26, 2011

[3] kyoung-jae, kim, "Prediction of KOSPI using Data Editing Techniques and Case-based Reasoning," Journal of the Korea Society of Computer and Information, v.12, no.6, pp.287-295, 2007
 [4] Jeong-hoon, Seu, "The Study for Traffic Signal Control Expert System using Case-based system and Rule-based system," Journal of the Korea Society of Computer and Information, v.11, no.2, pp.121-129, 2006.
 [5] Wettschereck, D. and Dietterich, T.G., "An Experimental Comparison of the Nearest-Neighbor and Nearest-Hyperrectangle Algorithms," Machine Learning, Vol. 19, pp. 1-25, 1995.
 [6] Wettschereck, D., "A hybrid nearest-neighbor and nearest hyperrectangle algorithm," Proceedings of European Conference on Machine Learning, Springer Verlag NY, eds. F. Bergadano, L. De Raedt pp. 323-335, 1994.
 [7] P. Domingos, "Unifying instance-based and rule-based induction," Machine Learning, vol. 24, pp. 141-168, 1996.
 [8] Lee-sang, Jeong, Chang-seung, Ha, "A Study on the Design and Implementation Human Resource Dispatch System of Using Case Based Reasoning," Journal of the Korea Society of Computer and Information, v.12, no.3, pp.95-103, 2007.
 [9] Xiang Y, Jin R, Fuhry D, Dragan F, "Summarizing transactional databases with overlapped hyperrectangles." Data Min Knowl Discov 23(2), pp.215-251, 2011.
 [10] S Garcia et al, "A First Approach to Nearest Hyperrectangle Selection by Evolutionary Algorithms," Proc. of 9th Intern. Conf. on Intelligent System Design and Applications, Pisa, Italy, pp.517-522, 2009.
 [11] A. Asuncion and D. Newman, "UCI machine learning repository," 2007. <http://www.ics.uci.edu/mllearn/MLRepository.html>
 [12] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," Journal of

Artificial Intelligence Research, vol. 6, pp.
1-34, 1997.

저 자 소 개



이 형 일

1985: 명지대학교
전자계산학과 공학사.

1994: 명지대학교
컴퓨터공학과 공학석사.

2000: 명지대학교
컴퓨터공학과 공학박사

현 재: 김포대학교
인터넷정보과 부교수

관심분야: 기계학습, 미디어영상인식,
실내위치추적

Email : hilee@kimpo.ac.kr