

특정 공간을 사선 분할하는 퍼지 결정 트리

이우항, 이건명
충북대학교 컴퓨터 과학과

A Fuzzy Decision Tree to Partition Feature Space with Oblique Planes

Woo Hang Lee, Keon-Myung Lee
Dept. of Computer Science, Chungbuk National University

요 약

결정 트리는 실제 세계에서 얻어지는 많은 사례들로부터 분류 정보를 얻기 위해 사용되는 유용한 방법중의 하나이다. 분류를 목적으로 사용되는 사례, 즉 데이터들은 실제 현장에서 얻어지기 때문에 관측오류, 불확실성, 주관적인 판단 등의 원인으로 참 값이 아닌 근사 값으로써 기술되는 경우가 많으며, 이러한 잠재적 오류로 인해 잘못된 결정 트리가 생성될 수 있다. 한편, 트리를 생성하는 각각의 과정에서 하나의 특징 값만을 고려하지 않고 두 가지 이상의 특징 값을 동시에 고려하여 결정 트리를 생성할 경우 보다 정확한 분류 정보를 기대할 수 있다. 본 논문에서는 수치 특징 값으로 기술된 데이터로부터 보다 정확한 분류 정보를 얻을 수 있고, 작은 오류에 강건한 사선형 분할 퍼지 결정 트리를 제안한다. 또한 제안된 사선형 분할 퍼지 결정 트리의 생성 절차 및 생성된 결정 트리를 이용하여 새로운 데이터에 분류 정보를 부여하는 추론 과정을 소개한다.

1. 서 론

결정 트리는 사례들로부터 분류 정보를 추출하기 위해 사용되는 방법들 중의 하나로써 이해하기 쉽다는 큰 장점을 갖고 널리 사용되어왔다[1][2]. 한편 분류를 목적으로 사용되어지는 사례, 즉 데이터들은 실제 현장에서 얻어지기 때문에 관측오류, 불확실성, 주관적인 판단 등의 원인으로 참 값이 아닌 근사 값으로써 기술되는 경우가 많다. 이런 잠재적 오류는 잘못된 결정 트리 생성의 원인이 되기도 한다. 이런 문제를 해결하고자 오류에 강건하고 이해하기 쉬운 퍼지 결정 트리 생성에 대한 많은 연구가 진행 되어왔다.[7]

ID3와 같은 전통적인 결정 트리는 트리를 확장해 가는 각 과정에서 하나의 특징 값을 고려한다. 즉, 현재 상태에서 가장 좋은 분류 정보를 얻을 수 있는 특징을 선택하고 이 특징을 기준으로 트리를 확장해 간다. 따라서 이해하기 쉽고 하나의 규칙 집합(rule set)으로 전환하기 쉽다. 그러나 실제세계의 데이터를 사용할 경우 트리의 깊이가 매우 깊어질 수 있으며, 또는 트리의 정확성이 떨어질 수 있다. 이런 문제점을 해결하기 위해 데이터를 기술하는 각각의 특징 값을 가공하여 사용하는 방법이 연구되었다.[4][5][6] 즉, 여러 특징 값을 동시

에 고려하여 트리를 생성하는 연구가 진행되었으며, OC1은 이런 연구 중의 하나이다.[3]

본 논문에서는 각 트리 생성 과정에서 사례를 기술하는 특징 값을 동시에 두 개 이상 사용하여 결정 트리의 구조를 단순화시킨 OC1의 알고리즘을 변형 사용하고 또한 퍼지 이론을 적용하여 잠재된 오류에 강건한 사선형 분할 퍼지 결정 트리를 제안한다. 또한 제안된 사선형 분할 퍼지 결정 트리의 생성 절차 및 생성된 결정 트리를 이용하여 새로운 데이터에 분류 정보를 부여하는 추론 과정을 소개한다.

본 논문은 다음과 같이 구성된다. 2절에서는 퍼지 언어항에 대한 소속 함수 결정을 위한 특정 공간 변환과 소속 함수 정의에 대해서 기술하고, 3절에서는 제안된 퍼지 결정 트리에 사용된 혼잡 측도를 소개하며, 4절에서는 제안된 결정 트리의 생성과정을 보이고, 5절에서는 제안된 퍼지 결정 트리를 이용하여 새로운 데이터에 분류정보를 부여하기 위한 추론과정을 보인 후, 끝으로 6절에서 결론을 맺는다.

2.공간 변환과 소속 함수 정의

사례, 즉 데이터를 묘사하는 특정 공간을 분할하는 사선형 분할 퍼지 결정 트리를 생성하기 위해 다음과 같은 공간 변환을 하며, 변환된 거리 공간에서 퍼지 언어항에 대한 소속 정도를 나타내는 소속 함수를 정의한다.

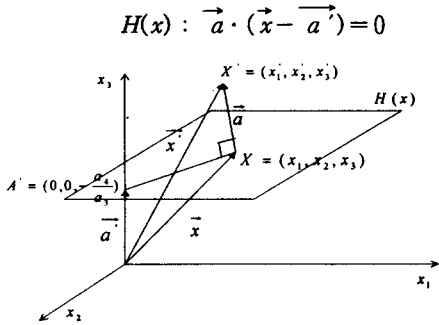
*본 연구는 첨단정보기술 연구센터(AITrc)를 통하여 과학재단의 지원을 받았음.

먼저, d 개의 특징 값으로 기술되는 데이터를 분할하는 초평면 $H(x)$ 와 법선 벡터 \vec{a} 를 다음과 같이 정의한다.

$$H(x) = a_1x_1 + a_2x_2 + \dots + a_dx_d + a_{d+1}$$

$$\vec{a} = (a_1, a_2, \dots, a_d)$$

위의 초평면에 대한 벡터 식은 아래 [그림 1]을 이용하면 다음과 같다.



[그림 1] 초평면 $H(x)$

또한, 하나의 데이터인 점 X' 을 지나고 평면에 수직인 직선의 벡터 식은 다음과 같다.

$$\vec{x} = \vec{x}' - t \frac{\vec{a}}{|\vec{a}'|}$$

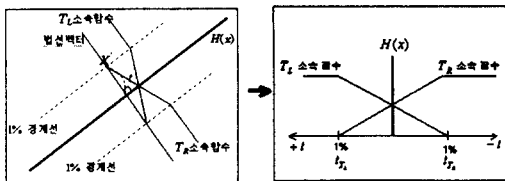
여기서, t 는 방향성을 갖고 있는 평면과 점 X' 사이의 거리를 나타낸다. 즉, t 의 값이 양수이면, 초평면 위에 데이터 X' 이 존재하게 되고, t 의 값이 음수이면 초평면 아래쪽에 데이터 X' 이 존재하게 된다. 초평면 벡터 식과 직선 벡터 식에서 직선과 평면과의 교점을 이용하여 거리 t 에 관한 식으로 정리하면 다음과 같은 수식을 얻는다.

$$t = \frac{\vec{a} \cdot (\vec{x}' - \vec{a}')}{|\vec{a}'|}$$

예로써, 위의 [그림 1]과 같이 3개의 특징 값을 갖는 데이터에 대한 t 식은 아래와 같다.

$$t = \frac{\vec{a} \cdot (\vec{x}' - \vec{a}')}{|\vec{a}'|} = \frac{a_1x_1 + a_2x_2 + a_3x_3 + a_4}{\sqrt{a_1^2 + a_2^2 + a_3^2}}$$

따라서 각각의 데이터에 대하여 t 값을 계산함으로써 다차원 특징 공간을 초평면과 데이터 사이의 일차원 거리 공간으로 변환할 수 있다.



[그림 2] 공간 변환

앞에서 소개한 d 차원 특징 공간을 일차원 거리 공간으로 변환한 후, 각 데이터를 t 값을 기준으로 내림차순으로 정렬한다. 정렬된 데이터를 이용하여 평면으로부터 1% 지점에 가장 근접한 데이터를 찾고 이를 기준으로 [그림 2]와 같이 소속 함수를 정의한다. 단, 데이터의 전체 개수가 부족하여 1% 지점 데이터를 찾을 수 없는 경우는 초평면에 가장 가까운 데이터를 선택한다.

즉, 각각의 퍼지 언어항 T_L 과 T_R 에 대한 소속 함수를 다음과 같이 정의한다.

$$\mu_{T_L}(x) = \begin{cases} 1 & \text{if } t_x \geq t_{T_L} \\ \frac{1}{t_{T_L} - t_{T_R}}(t_x - t_{T_R}) & \text{if } t_{T_L} > t_x \geq t_{T_R} \\ 0 & \text{if } T_{T_R} > t_x \end{cases}$$

$$\mu_{T_R}(x) = \begin{cases} 1 & \text{if } t_{T_R} > t_x \\ \frac{1}{t_{T_R} - t_{T_L}}(t_x - t_{T_L}) & \text{if } t_{T_L} > t_x \geq t_{T_R} \\ 0 & \text{if } t_x \geq T_{T_L} \end{cases}$$

위 식에서

$\mu_{T_L}(x)$: 데이터 x 의 퍼지 언어항 T_L 에 대한 소속정도

$\mu_{T_R}(x)$: 데이터 x 의 퍼지 언어항 T_R 에 대한 소속정도

t_x : 데이터 x 의 t 값

t_{T_L} : 왼쪽에 있는 데이터들 즉, t 의 값이 양수인 데이터들의 1% 경계 위치

t_{T_R} : 오른쪽에 있는 데이터들 즉, t 의 값이 음수인 데이터들의 1% 경계 위치

이다.

3. 퍼지 혼잡 측도

제안된 사선형 분할 퍼지 결정 트리를 만들어 가는 각 과정에서 혼잡도를 계산하고 이를 이용하여 정확하고 간단한 결정 트리 생성과정을 수행한다. 이 때 사용되는 측도로써 두 원률(twoing rule)[3]을 퍼지 이론에 적용할 수 있도록 변형한 퍼지 투원률(twoing rule)을 다음과 같이 정의한다. 즉, 노드 N 의 혼잡도 I_N 은

$$I_N = \frac{\sum_{x \in N} \mu_{T_L}(x)}{\sum_{x \in N} \mu_N(x)} \times \frac{\sum_{x \in N} \mu_{T_R}(x)}{\sum_{x \in N} \mu_N(x)} \times \left(\sum_{j=1}^k \left| \frac{\sum_{x \in N, x.class=j} \mu_{T_L}(x)}{\sum_{x \in N} \mu_{T_L}(x)} - \frac{\sum_{x \in N, x.class=j} \mu_{T_R}(x)}{\sum_{x \in N} \mu_{T_R}(x)} \right| \right)^2$$

이다.

4. 제안된 퍼지 결정 트리 생성

제안된 사선형 분할 퍼지 결정 트리는 특징 공간을 반복적으로 분할함으로써 트리를 생성한다. 즉 주어진 모든 데이터를 근 노드에 위

치시키고, 초 평면 H 에 의해 분할이 이루어지면 각각 분할된 영역에서 다시 초평면을 이용한 분할을 시도한다.

다음은 제안된 사선형 분할 퍼지 결정 트리의 각 노드에서 초평면 H 를 결정하는 알고리즘이다.

procedure 초평면 H 찾기
begin

모든 특징 값들에 대하여, 단 하나의 특징 값만을 순차적으로 고려하여 혼잡도가 가장 낮은 초평면을 찾는다.

현재 초평면 H_0 와 혼잡도 I_0 를 저장한다.

다음 과정을 r 번(기본 값 또는 입력 값)만큼 반복한다.

입의 초평면 H 을 선택한다.

(첫 수행시, H 는 현재 노드에서 가장 좋은 축 평행 초평면으로 초기화)

과정 1: 혼잡도가 더 이상 낮아지지 않을 때까지 다음을 반복 수행

순차적으로 H 의 각 계수 a_i 를 Perturb 한다.

과정 2: 적어도 j (기본 값 또는 입력 값)번 다음을 반복한다.

입의 방향을 선택하고 H 의 각 계수 a_i 를 Perturb 한다.

H 의 혼잡도가 낮아지면, 과정 1로 간다.

현재 초평면의 혼잡도 I_1 이 I_0 보다 작으면, I_0 를 I_1 으로 대체한다.

I_0 에 대응되는 H 를 반환한다.

end

위 알고리즘에서 사용된 Perturb 과정은 입의 숫자 발생을 이용한 방법으로써 [3]을 참고하였다.

5.분류를 위한 추론

생성된 사선형 분할 퍼지 결정 트리는 새로운 데이터를 분류하기 위해 사용되며, 다음과 같은 과정을 따른다. 분류될 데이터는 초기 소속 정도 1로써, 생성된 트리의 근 노드에 주어지고, 하위 노드로 링크를 따라 내려간다. 입의 노드 N_i 는 데이터 x 를 받을 때마다, 해당 노드에서 하위 노드로 향하는 링크에 부여된 퍼지 언어항에 대한 만족 정도를 계산한다. 데이터 x 의 하위 노드에 대한 소속 정도는 퍼지 언어항이 T_L 일 경우 $f(\mu_{N_i}(x), \mu_{T_L}(x))$ 에 의해 계산되며, T_R 일 경우 동일한 방법으로 계산된다. 여기서 $f(\cdot, \cdot)$ 는 T-norm 연산자의 하나이며, 하위 노드에 대한 소속 정도가 미리 지정된 임계값 이상일 때에만 하위 노드로 데이터 x 를 전달한다. 퍼지 결정 트리는 하나의 데이터가 여러 개의 단말 노드에 도달할 수 있으며, 따라서 모든 단말 노드에서의 각 클래스별로 소속 정도를 계산하고 이를 종합하여 입력된 데이터 x 에 분류 정보를 계산한다. 또한 하나의 단말 노드는 복수 개의 분류 정보를 나타낼 수 있으며, 따라서 다음과 같은 클래스별 분류 정보 c_k 를 계산한다.

$$c_k = \sum_i \chi_{N_i}(k) \times K_{N_i}(x)$$

위 수식에서

$\chi_{N_i}(k)$: 노드 N_i 의 클래스 k 에 대한 소속정도

$K_{N_i}(x)$: 데이터 x 의 노드 N_i 에 대한 소속정도

이다.

6.결론

본 논문에서는 수치 특징 값으로 기술되는 사례 즉, 데이터를 대상으로 사선형 분할 퍼지 결정 트리를 생성하는 방법을 제안하였다. 제안된 퍼지 결정 트리는 특징 공간을 사선형으로 분할함으로써 동시에 복수 개의 특징값을 고려하여 보다 정확한 분류 정보를 얻을 수 있었다. 또한 퍼지 이론을 적용함으로써 잠재적 오류에 대처하는 강건함을 가질 수 있었다. 결정 트리의 큰 장점인 쉬운 이해성이 다소 감소되고 결정 트리 생성 소요 시간이 길어지는 단점이 있으나 앞에서 서술한 장점을 고려할 때 사선형 분할 퍼지 결정 트리가 유용함을 알 수 있다.

참고 문헌

- [1] T. M. Mitchell, *Machine Learning*, The McGraw-Hill Co., 414p, 1997.
- [2] M. -S. Chen, J. Han, P. S. Yu, *Data Mining: An Overview from a Database Perspective*, IEEE Transactions On Knowledge and Data Engineering Vol. 8, No. 6, December 1996.
- [3] S. K. Murthy, S. Kasif, S. Salzberg, *A System for Induction of Oblique Decision Trees*, Journal of Artificial Intelligence Research. Vol. 2, pp.1-33, 1994.
- [4] A. Ittner, J. Zeidler, R. Rossius, W. Dilger, M. Schlosser, *Feature space partitioning by non-linear and fuzzy decision trees*, Proc. of the 7th IFSA world congress 1997.
- [5] A. Ittner, M. Schlosser, *Non-linear decision trees-NDT*, In L. Saitta, editor, Proc. of the 13th International Machine Learning Conference, pp. 252-257, Morgan Kaufman, San Francisco, CA, Bari, Italy, 1996.
- [6] A. Ittner, M. Schlosser, *Discovery of relevant new features by generating non-linear decision trees.*, In E. Simoudis, J. Han, and U.Fayyad, editors, Proc. of 2nd International Conference on Knowledge Discovery and Data Mining, pp. 108-113, AAAI Press, Menlo Park, CA, Portland, Oregon, USA, 1996.
- [7] C. Z. Janikow, *Fuzzy Decision Trees: Issues and Methods*, IEEE transactions on systems, man and cybernetics. Part B: cybernetics (1998), Vol.28, N1, pp. 1-14