

# Interpretability Comparison of Popular Decision Tree Algorithms

Jung-Sik Hong\*<sup>†</sup> · Geun-Seong Hwang\*\*

\*Dept. of Industrial Engineering, Seoul National University of Science and Technology

\*\*Dept. of Data Science, Graduate School, Seoul National University of Science and Technology

## 대표적인 의사결정나무 알고리즘의 해석력 비교

홍정식\* · 황근성\*\*

\*서울과학기술대학교 산업공학과

\*\*서울과학기술대학교 일반대학원 데이터사이언스학과

Most of the open-source decision tree algorithms are based on three splitting criteria (Entropy, Gini Index, and Gain Ratio). Therefore, the advantages and disadvantages of these three popular algorithms need to be studied more thoroughly. Comparisons of the three algorithms were mainly performed with respect to the predictive performance. In this work, we conducted a comparative experiment on the splitting criteria of three decision trees, focusing on their interpretability. Depth, homogeneity, coverage, lift, and stability were used as indicators for measuring interpretability. To measure the stability of decision trees, we present a measure of the stability of the root node and the stability of the dominating rules based on a measure of the similarity of trees. Based on 10 data collected from UCI and Kaggle, we compare the interpretability of DT (Decision Tree) algorithms based on three splitting criteria. The results show that the GR (Gain Ratio) branch-based DT algorithm performs well in terms of lift and homogeneity, while the GINI (Gini Index) and ENT (Entropy) branch-based DT algorithms performs well in terms of coverage. With respect to stability, considering both the similarity of the dominating rule or the similarity of the root node, the DT algorithm according to the ENT splitting criterion shows the best results.

Keywords : Interpretability Comparison, Stability Measure, Decision Tree

### 1. 서론

의사결정나무 알고리즘(Decision Tree algorithm, 이하 DT 알고리즘)은 분류 알고리즘중 가장 대표적인 알고리즘이며 지난 수십년간 성능 측면의 개선을 다룬 많은 논문들이 제시되었고[7, 9, 14, 16, 23, 28, 29], 이중 다수의 논문들은 새로운 분기기준을 제시한 논문이다[9, 14, 16, 28, 29]. 그러나 R이나 Python 기반의 오픈 소스 프로그램들은 여전히 세 개의 분기 기준(엔트로피, 지니지수, 게인 비율)

을 토대로 한 알고리즘이 대부분이다. 따라서 가장 널리 활용되는 이 3가지 분기기준을 토대로 한 DT 알고리즘에 대해서는 장점과 단점이 충분히 연구될 필요가 있다.

그런데 이 3가지 알고리즘의 비교는 주로 정확도 측면에서 이루어졌다[4, 26, 30]. 해석력 측면에서의 이들 알고리즘에 대한 비교는 제대로 수행되지 못하였다. 해석력은 실제로 매우 다양한 현실 문제에서 머신러닝 모델이 사용될 때 매우 중요한 요소이다[2, 24, 27]. 가령, 대출 과정이 편향되지 않음을 고객에게 보이기 위해서, 설명 가능한 대출 규칙이 필요하고[2], 의료 분야에서도 환자에게 설명을 해야 할 필요성이 있으며[24], 이외에도 다양한 비즈니스에서 고객 유지 등을 위해 고객 이탈 현상을 설명해주는 머신러닝 모델을 필요로 한다[27]. 따라서,

머신러닝 모델에서 해석력은, 매우 중요한 이슈로 부상하고 있으며, DT 알고리즘을 포함하여 머신 러닝 알고리즘에 대한 해석력을 분석하는 연구는 최근에야 집중적으로 이루어지고 있다[11, 13, 17, 18]. 머신 러닝 모델의 해석력은 설명력과 이해 가능성들의 용어와 혼용되기 시작하며 다양한 연구가 수행되었다. 이러한 용어들의 혼용성을 지적하며, 머신 러닝 모델에서의 해석력 문제를 비교적 포괄적으로 다룬 논문이 최근 1~2년 사이에 제시되고 있다[3, 22].

머신 러닝 모델의 해석력에 대한 체계적인 연구를 바탕으로 본 논문은 DT 알고리즘의 해석력을 다루고자 한다. DT 알고리즘의 해석력에 대한 논문은 두 가지 부분으로 나누어 볼 수 있다. 하나는 DT 알고리즘의 해석력을 다른 분류 알고리즘의 해석력과 비교하는 논문이다[8]. 다른 하나는 해석력을 높인 새로운 DT 알고리즘을 제시하고 기존의 DT 알고리즘과 해석력을 비교하는 것이다[17]. 본 논문은 DT 알고리즘의 해석력에 한정하고 있으므로 후자에 좀 더 가까운 연구이다. 그런데 후자에 속하는 연구의 한계점은 해석력의 중요한 척도중 하나인 안정성이 제외되어 있다는 점이며, 또한 새로 제시된 알고리즘에 초점이 맞추어져 있어서 기존의 대표적인 DT 알고리즘의 해석력 비교가 제대로 수행되지 못하였다는 것이다. 특히 안정성은 정확도 측면에서도 다루어지고 해석력 측면에서도 다루어지는 척도로, DT 알고리즘의 경우 독자적으로 다루어지는 경우가 대부분이다[6, 19]. 본 논문은 안정성 척도를 해석력과 연관되는 다른 척도들과 함께 다루고자 한다. 또한, 해석력 측면에서의 DT 알고리즘의 안정성 척도를 기존의 척도를 변형하여 새로운 방식으로 제안하고자 한다. 변형된 안정성 척도와 함께 DT 알고리즘의 해석력을 다루는 다른 척도들(DT의 깊이, 리프 노드의 동질성과 커버리지)을 종합하여 기존의 가장 대표적인 3가지 DT 알고리즘의 해석력을 비교하고 분석하는 작업을 수행하여, 해석력 차원에서 3가지 DT 알고리즘의 장점과 단점을 제시하는 것이 본 논문의 독자적인 기여라고 할 수 있다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 본 연구와 관련된 연구 배경이 다루어진다. 구체적으로는 연구의 자족성을 위해 3가지 DT 알고리즘의 분기 기준이 간단히 소개되고, 머신 러닝 알고리즘의 해석력이 기술된다. 그리고 DT 알고리즘의 해석력에 대한 척도들과 연구들이 소개된다. 3절에서는 DT 알고리즘의 안정성 척도를 해석력 차원에서 고찰하고 기존 척도를 변형한 척도가 기술된다. 4절에서는 UCI 데이터와 kaggle 데이터를 바탕으로 작성된 10가지 데이터를 대상으로 3가지 DT 알고리즘의 해석력을 비교하고 분석하는 작업이 수행된다. 마지막으로 5절에서는 해석력 관점에서 3가지 대표적인 DT 알고리즘의 장점과 단점이 기술된다.

## 2. 연구배경

### 2.1 대표적인 DT 알고리즘의 분기기준

R이나 Python 기반의 오픈 소스 프로그램들은 세 개의 분기 기준(엔트로피, 지니지수, 개인 비율)을 토대로 한 알고리즘이 대부분이다. 이들 분기 기준은 다음과 같다.

#### 2.1.1 지니지수

$$G(S) = 1 - \sum_{k=1}^m p_k \quad (1)$$

$$\Delta G = G(S) - \left( \frac{|S_l|}{|S|} G(S_l) + \frac{|S_r|}{|S|} G(S_r) \right) \quad (2)$$

여기서,  $S$ 는 데이터의 집합을 의미한다. 이 집합의 원소는 설명변수  $(x_1, \dots, x_n)$ 와 목표변수  $y$ 로 이루어져 있다. 목표변수  $y$ 의 범주 개수는  $n$ 이며,  $p_k$ 는 데이터 중 목표변수 범주가  $k$ 인 데이터의 비율이다.  $G(S)$ 는 데이터  $S$ 의 지니지수이다.  $|S|$ 는 집합  $S$ 의 크기이고,  $S_l$ 과  $S_r$ 은  $S$ 를 분기 기준에 따라 둘로 나눌 때, 각각 왼쪽 노드의 데이터 집합과 오른쪽 노드의 데이터 집합을 의미한다.

#### 2.1.2 엔트로피

$$E(S) = - \sum_{k=1}^m p_k \log_2(p_k) \quad (3)$$

$$\Delta E = E(S) - \left( \frac{|S_l|}{|S|} E(S_l) + \frac{|S_r|}{|S|} E(S_r) \right) \quad (4)$$

여기서,  $E(S)$ 는 데이터 집합  $S$ 의 엔트로피를 의미한다.

#### 2.1.3 개인 비

$$N(S) = - \left( \frac{|S_l|}{|S|} \log_2 \frac{|S_l|}{|S|} + \frac{|S_r|}{|S|} \log_2 \frac{|S_r|}{|S|} \right) \quad (5)$$

$$\Delta GR = \frac{\Delta E}{N(S)} \quad (6)$$

여기서,  $N(S)$ 는 데이터 집합  $S$ 의 정규화 인자를 의미한다.

하나의 노드가 자식 노드로 분기될 때, 분기기준은 위의 3가지중 하나를 사용하여 부모 노드와 자식 노드간의 동질성 이득이 최대가 되는 변수와 임계치에 의해 정해진다. 가령 지니 지수를 예로 들면,

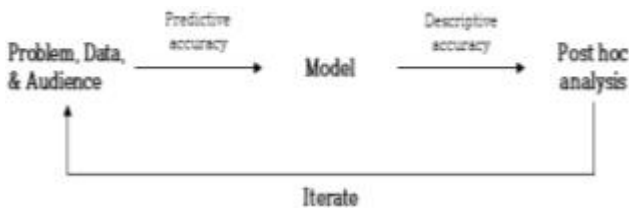
$$\Delta G(S) = G(S) - \frac{|S_l|}{|S|} G(S_l) - \frac{|S_r|}{|S|} G(S_r) \quad (7)$$

$$MAX_{x_i \in C_{i,j}} \Delta G(S) = G(S) \quad (8)$$

여기서,  $x_i$ 는  $i$ 번째 설명변수이며,  $C_{i,j}$ 는  $S$ 을  $S_l$ 과  $S_r$ 을 나누는  $j$ 번째 영역을 의미한다.

## 2.2 머신러닝 알고리즘의 해석력

머신러닝 알고리즘의 해석력은 지난 10년간 부분적으로 개별 척도를 제시하는 연구가 수행되어 오다가 최근 종합적인 차원에서 머신러닝의 해석력을 다루는 프레임워크를 제시하는 연구들이 이루어졌다. 이중 대표적인 연구로는 Murdoch et al.[22]의 연구와 Belle and Papantonis [3]의 연구가 있다. Murdoch et al.[22]은 머신 러닝 알고리즘의 해석력에 있어 예측적 정확도까지 포함하여 PDR (Predictive Descriptive Relevant) 프레임워크를 제시하고 해석력을 모델 기반 해석력과 모델 정립이후 해석력으로 나누어 설명하고 있다. 이를 도시하면 다음과 같다.



<Figure 1> PDR Framework[22]

Belle and Papantonis[3]는 해석력 용어 대신 ‘설명가능한’(Explainable)이라는 용어를 사용하여 설명가능한 머신러닝의 원칙과 실행 사례를 제시하고 있다. Belle and Papantonis[3]은 설명가능한 머신러닝 모델의 중요 특성으로, 정확성, 강건성, 개선성, 편향, 호환성 그리고 인간 이해가능성을 제시하고 있다.

이상의 6가지 핵심척도는 다양한 머신러닝 모델의 해석력이나 설명력을 평가하는 데에 사용된다. 가령, 신경망 기반의 모델과 같은 블랙 박스 모델은 인간 이해가능성이 매우 낮은 모델이라고 할 수 있다.

## 2.3 DT 알고리즘의 해석력

DT 알고리즘은 해석력이 뛰어난 분류 알고리즘이다 [20, 21]. 그러나 해석력이 뛰어난 분류 알고리즘은 DT 알고리즘 외에도 규칙 유도 알고리즘이나 로지스틱 회귀 모델 등이 있다. 따라서 이러한 여러 분류 알고리즘의 해석력을 비교하는 연구가 수행되고 있다[12]. 또한, DT 알고리즘의 나무 깊이나 전체 리프 노드의 개수 등에 따라 인간의 이해가능성이 어떻게 영향을 받는지를 연구한 논문도 최근에 등장하고 있다[21]. 한편 DT 알고리즘을 해석력 측면에서 개선시키고자 하는 연구도 최근 이루어

어지고 있다[11, 13, 17]. Kotsiantis[17]에서는 기존의 DT 알고리즘이 해석력 측면에서는 기본적인 두 가지 문제점이 있음을 보이고 있다. 첫째, DT 알고리즘의 분기기준이 분기 후 양쪽 노드의 동질성을 기계적으로 가중합하여 부모 노드와 자식 노드간 동질성 감소량을 최대화 하는 것은 어느 한쪽 노드에 동질적인 데이터들이 몰려 있는 것을 조기에 탐지하지 못할 가능성이 매우 크다는 것이다. 이를 해결하기 위해, Kotsiantis[17]은 한쪽 노드만을 최대화하는 분기기준을 제시하고 이러한 분기기준에 기반한 알고리즘이 DT의 해석력을 높일 수 있음을 입증하였다. 둘째, DT 알고리즘의 뿌리 노드의 분기기준은 이후 모든 리프 노드의 규칙에 모두 조건으로 들어가게 된다. 이는 해석력 측면에서 문제가 된다. 즉, 모든 규칙에 하나의 변수가 들어가는 것은 모든 규칙의 독립성이 성립할 여지가 전혀 없는 것이다. 목표변수의 클래스를 정하는 데에 있어 독립적인 규칙이 존재할 수 있는 다수의 사례에 비추어 볼 때, 이는 개선해야 할 문제임을 알 수 있다. 또한 뿌리 노드의 분기기준은 모든 규칙에 포함되므로 이의 안정성은 특히 문제가 된다. 제 3장에서 DT 알고리즘의 안정성 문제를 다룰 때 이러한 사항이 고려될 것이다.

## 3. DT 알고리즘의 해석력 평가척도

### 3.1 기존의 평가 척도

DT 알고리즘의 해석력은 DT 알고리즘에서 유도되는 규칙의 품질에 좌우된다. 규칙의 품질은 DT 알고리즘 외에도 규칙 유도 알고리즘이나 연관규칙등에서 모두 거론되는 주제이다.

규칙의 품질을 평가할 때는 규칙의 정확도까지 포함하므로 제시된 척도가 매우 다양하다[1].

그런데 해석력을 고려할 때 가장 기본적인 평가척도는 규칙의 커버리지와 동질성이다. An and Cercone[1]에서는 이 두 가지 척도의 가중합으로 규칙의 품질을 평가한다. 또한 DT 알고리즘의 경우는 해석력을 평가할 때 DT의 깊이가 중요한 평가요소가 된다[20]. 따라서 해석력 관점에서 DT 알고리즘을 평가하는 가장 기본적인 평가척도는 커버리지, 동질성 그리고 DT의 깊이이다. 그런데 동질성은 목표변수가 불균형 집합인 경우는 한계가 있으므로 동질성을 목표변수의 비율로 나눈 리프트를 추가하는 것이 바람직하다. 마지막으로 머신러닝 알고리즘의 해석력을 평가할 때, 사용되는 안정성 척도를 추가하고자 한다. 이 경우 DT 알고리즘의 해석력을 평가하는 요소는 다음 5가지 요소가 된다.

- (1) 커버리지 : 하나의 리프 노드에 속하는 데이터의 비율
- (2) 동질성 : 하나의 리프 노드에 속하는 데이터 중 목표 변수의 최빈 클래스에 속하는 데이터의 비율
- (3) 리프트 : 동질성을 최빈 클래스의 사전 비율로 나눈 값
- (4) DT의 깊이 : 루트 노드에서 리프 노드에 이르는 경로의 길이의 최댓값
- (5) 안정성 : 훈련 데이터의 변화에 따른 DT의 변화정도

예를 들어, 하나의 리프 노드에 데이터가 100개 있고, 이 중 목표변수가 1인 데이터가 90개, 목표변수가 0인 데이터가 10개 있다고 하자. 데이터 전체는 1000개이며 목표변수가 1인 비율은 30%라고 하자. 이 경우, 이 리프 노드의 커버리지는 10%이고, 동질성은 90%이며, 리프트는 0.9/0.3이므로 3이다.

위의 5가지 평가 척도를 2.2절의 머신러닝 모델의 해석력을 평가하는 6개 척도와 관련하여 설명하면 다음과 같다.

첫째, 정확성(correctness)은 정확도(accuracy)를 포함한, 보다 폭넓은 개념이다. DT에서 도출된 규칙이 적절한 의미를 갖고 신뢰할만한가를 뜻하는 척도이다. 따라서 커버리지와 동질성이 이 척도와 연관이 있다고 할 수 있다. 규칙이 커버하는 비율이 너무 작으면 규칙으로써의 적절한 의미를 잃고, 또한 동질성이 어느 정도 확보되어야 규칙의 적절성이 확보되기 때문이다. 둘째, 강건성은 안정성과 연관되는 개념이다. 강건하다는 것은 입력 요소의 작은 변화에도 규칙이 변화하지 않는다는 것을 의미하는데, 훈련데이터가 변해도 DT 변화가 적은 것을 안정성이 높다고 할 때, 이는 바로 강건성이 높다는 것을 의미하기 때문이다. 셋째, 개선성이나 편향 그리고 호환성은 DT 전체에 해당되는 척도이고 분기 기준에 따라 그 정도가 달라지는 척도가 아니다. 따라서 3개의 DT 알고리즘의 해석력을 비교하는 데에 사용될 척도는 아니라고 할 수 있다. 마지막으로 인간 이해가능성은 DT의 깊이에 해당한다고 할 수 있다. DT의 깊이가 낮은 곳에서 생성되는 리프 노드일수록 규칙의 조건의 개수가 줄어들고 이는 인간 이해가능성을 높이기 때문이다.

5가지 평가척도중 안정성을 제외한 4가지 평가척도의 계산은 간단하다. 리프 노드  $k$ 에 속하는 데이터의 집합을  $S_k$ 라 하면, 커버리지는  $\frac{|S_k|}{|S|}$ 가 된다. 리프 노드  $k$ 에 속하는 데이터 중  $y=1$ 의 비율과  $y=0$ 의 비율 중 높은 값이 동질성이다. 리프트나 DT의 깊이는 정의에 따라 간단히 계산된다.

안정성은 다른 4개의 척도처럼 간단하게 계산되지 않는다. 본 논문은 해석적 관점에서 안정성을 평가하는 척도를 새로이 제시하고자 한다.

### 3.2 해석력 관점에서 DT 알고리즘의 안정성

DT 알고리즘의 안정성 문제는 해석력 차원보다 정확도 차원에서 먼저 연구되기 시작했다[5]. 이 연구를 바탕으로 최근까지 DT 알고리즘의 안정성 연구는 주로 정확도 차원에서 안정성을 평가하고 이를 개선하는 알고리즘을 제시하고 있다. 해석력 차원에서 안정성을 다룬 논문은 Jacobucci[15]가 있다. 이 논문은 DT의 유사성을 평가하는 Briand의 논문[6]을 바탕으로 안정성을 다음과 같이 정의한다.

$$ST(DT_1, \dots, DT_N) = 1 - \frac{q}{N} \quad (9)$$

여기서,

- $DT_i$ 는  $i$ 번째 부스팅된 훈련 데이터로 생성된 DT
- $ST(DT_1, \dots, DT_N)$ 는 총  $N$ 번 부스팅된 훈련 데이터로 계산된 DT의 안정성
- $q$ 는  $N$ 개의 DT 중 서로 다른 형태의 DT의 개수

식 (9)는 분기 기준이 되는 변수뿐 아니라 분기되는 변수의 값이 다를 경우는 다른 패턴의 DT로 규정하므로 수치형 변수가 많을 경우에는 안정성이 매우 낮게 나오게 되는 단점을 가지고 있다. 따라서 수치형 변수가 분기 기준으로 사용될 경우, 분기하는 임계값만 다른 경우는 그 임계값간의 차이를 반영하는 방식으로 두 개의 분기 노드의 유사성이 정해져야 한다. 본 논문은 Jacobucci[15]의 DT간 유사도를 바탕으로 두 가지 방식으로 DT간 유사도를 정의하고자 한다.

첫째, 해석력 관점에서 DT 알고리즘의 핵심적인 결과는 규칙들의 집합이다. 이 중에서도 커버리지가 가장 높은 규칙이 가장 의미가 있는 규칙이므로 이를 지배적인 규칙이라 정의하고, 이들간의 유사도로 DT 알고리즘의 안정성을 평가하고자 한다. 두 개의 규칙간의 유사도는 한 개의 분기기준간의 유사도를 바탕으로 정의된다. 하나의 분기기준이  $X_i$ 라고 하자.  $X_i$ 가 범주형일 경우는 두 개의 DT에서 발생한 두 분기기준이 임계치 범주가 같으면 유사도는 1이고 그렇지 않으면 0이다.  $X_i$ 가 수치형인 경우, 두 개의 DT에서 발생한 두 분기기준의 임계치가 각각  $\delta_1$ 과  $\delta_2$ 라고 하면, 분기기준  $x_i$ 의 유사도는 다음과 같이 정의된다.

$$ST(X_i) = 1 - \frac{|\delta_1 - \delta_2|}{range(X_i)} \quad (10)$$

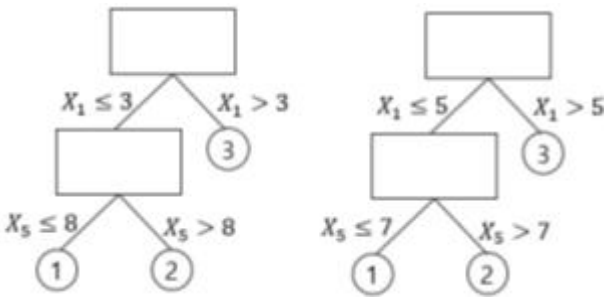
여기서,  $ST(X_i)$ 는 두 분기기준  $X_i$ 의 유사도이며,  $range(X_i)$ 는  $X_i$ 가 취하는 값의 범위이다. 이제 두 개의 DT에서

발생한 두 지배규칙  $R_{p,q}$ 의 유사도는 다음과 같이 정의된다. 지배규칙을 이루는 분기 기준 변수들의 집합을  $D_{p,q}$ 라 하자.

$$ST(R_{p,q}) = ST(\prod_{i \in D_{p,q}} X_i) = \prod_{i \in D_{p,q}} ST(X_i) \quad (11)$$

여기서, 두 지배규칙을 이루는 분기기준 변수가 하나라도 다르면, 유사도는 0으로 정의된다.

식 (10)의  $ST(X_i)$ 와 식 (11)의  $ST(R_{p,q})$ 를 예를 통해서 계산 과정을 간략히 보이기로 한다. <Figure 2>에 두 개의 DT가 나와 있다.



<Figure 2> Example Decision Trees

$DT_1$ 과  $DT_2$ 의 지배 규칙은 두 DT 모두 리프노드 ①에 대응한다고 하자.  $DT_1$ 과  $DT_2$ 의 지배 규칙의 유사도를 측정하기 위해 먼저  $ST(X_1)$ 과  $ST(X_5)$ 를 구해보자.  $X_1$ 의 범위는 10이며  $X_5$ 의 범위는 20이다. 이 경우,

$$ST(X_1) = 1 - \frac{|3-5|}{10} = 0.8$$

$$ST(X_5) = 1 - \frac{|8-7|}{20} = 0.95$$

따라서,

$$ST(R_{1,2}) = ST(X_1) \times ST(X_5) = 0.8 \times 0.95 = 0.76$$

N번 부스팅한 경우, N개의 DT가 형성되며 또한 N개의 지배규칙이 생성된다. 이들 간의 유사도를 계산하게 되는 조합의 가지 수는  ${}_n C_2$ 개다. 따라서, 지배규칙간 유사도를 바탕으로 안정성은 다음과 같이 정의 된다.

$$ST(R) = \frac{\sum_p \sum_{q < p} ST(R_{p,q})}{{}_n C_2} \quad (12)$$

둘째, 해석력 관점에서 DT 알고리즘의 모든 규칙에는 뿌리 노드의 분기기준이 들어간다. 따라서 뿌리 노드의 유사성이 높을수록 DT 알고리즘의 안정성이 높다고 할 수 있다. N번 부스팅을 시행하여 N개의 DT를 발생시켰

다면, N개의 뿌리 노드가 만들어진다. 이 경우 비교적 간단하게 N개의 뿌리노드의 동질성을 평가하는 방법은 분기의 임계값은 무시하고 분기 변수만을 판단기준으로 분기기준이 같은 변수이면 동질 노드로 보고 다른 변수이면 이질로 판단한다. 이 경우 지니지수로 간단하게 동질성을 계산할 수 있다. 가령 N개의 뿌리 노드가 분기기준이 같은 변수끼리 하나의 그룹으로 묶을 때,  $N_1, N_2, \dots, N_k$ 개의 그룹으로 나누어진다면, 지니지수는 다음과 같다.

$$ST(RN) = 1 - \sum_{k=1}^K p_k^2 \quad (13)$$

여기서,  $ST(RN)$ 은 뿌리 노드 RN의 안정성이고  $p_k$ 는  $\frac{N_k}{N}$ 이다.

유사도와 달리 지니지수는 0에 가까울수록 동질성이 높아지며, 따라서 안정성이 높아진다. 이 경우는 임계값이 무시되는 단점이 있으므로, 임계값을 고려할 경우, 식 (10)을 토대로 뿌리 노드 분기기준의 안정성을 측정할 수 있다. 따라서 뿌리 노드의 안정성은 식 (13)과 식 (10), 두 가지 방식에 의해 측정된다.

#### 4. 세 개의 DT 알고리즘의 해석력 비교실험

세 개의 DT 알고리즘의 해석력을 비교하는 실험을 수행하기 위해 수집된 데이터의 특성이 <Table 1>에 나와 있다. <Table 1>의 속성(Attributes)에서 N은 수치형 변수(Numerical variable)을 의미하고 C는 범주형 변수(Categorical variable)을 의미한다. <Table 1>에 나타난 데이터는 UCI DATASET 7개, Kaggle DATASET 3개로 총 10개이다. 훈련 데이터와 검증 데이터를 8:2로 분리하여 실험을 30번 반복 시행했다. 예측적 정확도를 비교하는 실험이 아니므로, 가지치기는 가장 간단한 방식인 종료 조건으로 대체되었다. 본 예제 실험에서 종료 조건은 다음과 같다. 하나의 노드에서 분기되는 두 개의 자식노드의 데이터 개수가 각각 전체 데이터 개수의 3% 미만이면 분기하지 않고, 그 노드는 리프노드가 된다.

이들 데이터에 대해 세 개의 DT 알고리즘을 적용한 결과의 해석력을 비교하는 작업은 두 가지로 나누어진다. 안정성을 제외한 나머지 해석력 평가척도는 서로 긴밀하게 얽혀 있다. 즉, DT의 깊이가 낮은 리프 노드에서 동질성(리프트)이나 커버리지가 높은 리프 노드가 많은 것이 해석력 차원에서 바람직한 것이기 때문이다. 따라서 안정성을 제외한 나머지 4가지 평가 척도는 동시에 비교가 이루어질 것이다. 다음 DT 알고리즘의 안정성은 별도로 비교분석이 이루어질 것이다.

<Table 1> Description of Datasets

Data information				Terminating condition
Datasets	Instances	Attributes	Classes	Minimum number of samples at leaf node (trainset * 0.03)
abalone	4177	8(N:8, C:0)	3	100
australian	690	14(N:14, C:0)	2	17
Bank-Personal-Loan	5000	12(N:12, C:0)	2	120
breast-cancer-wisconsin	699	9(N:8, C:1)	2	17
Churn	5000	19(N:15, C:4)	2	120
ionosphere	351	34(N:34, C:0)	2	8
pima-indians-diabetes	768	8(N:8, C:0)	2	18
sportscar-choice-long	6000	8(N:5, C:3)	2	144
tic-tac-toe-endgame	958	9(N:0, C:9)	2	23
yeast	1484	9(N:8, C:1)	9	36

<Table 2> Performance of 3 Splitting Criteria

Dataset	Measure	Performance of 3 splitting criteria																	
		D1			D2			D3			D4			D5			Means		
		ENT	GINI	GR	ENT	GINI	GR	ENT	GINI	GR	ENT	GINI	GR	ENT	GINI	GR	ENT	GINI	GR
abalone	L	-	-	2.3603	2.1890	2.4184	2.1874	1.5331	1.4733	1.3199	1.4596	1.4010	1.4128	1.5564	1.6498	1.7870	1.6845	1.7356	1.8135
	H	-	-	0.7605	0.7156	0.7782	0.7102	0.5140	0.5025	0.4562	0.5008	0.4779	0.4897	0.5221	0.5518	0.5829	0.5631	0.5776	0.5999
	C	-	-	0.2378	0.3068	0.2700	0.2945	0.4220	0.4011	0.4554	0.6774	0.6313	0.5629	0.8401	0.7621	0.6758	-	-	-
australian	L	1.6726	1.6639	1.7007	2.0071	1.9090	2.0111	1.7223	1.8343	1.7834	1.9109	1.8063	1.8499	1.5281	1.4751	1.5827	1.7682	1.7377	1.7856
	H	0.9313	0.9346	0.9349	0.9644	0.9470	0.9650	0.8099	0.8420	0.8329	0.8573	0.8139	0.8446	0.7428	0.7408	0.7491	0.8611	0.8557	0.8653
	C	0.4776	0.4791	0.4701	0.6353	0.6418	0.6551	0.7373	0.7768	0.7583	0.8071	0.8527	0.8199	0.9054	0.9474	0.8814	-	-	-
bank-personal-loan	L	1.0962	1.0826	1.3999	0.9919	5.5185	5.3233	5.6273	-	1.0015	-	-	5.5532	-	-	-	2.5718	3.3006	3.3195
	H	0.9914	0.9791	0.9488	0.8971	0.9264	0.9381	0.8588	-	0.9045	-	-	0.8535	-	-	-	0.9158	0.9528	0.9112
	C	0.7434	0.8061	0.7684	0.8997	0.9999	0.9871	1	-	0.9310	-	-	0.9990	-	-	-	-	-	-
breast-cancer-wisconsin	L	-	-	2.1416	2.0284	1.9899	1.7673	1.8843	1.7835	1.8259	1.7113	1.7145	1.8256	1.6214	1.6808	1.7013	1.8114	1.7922	1.8523
	H	-	-	0.9806	0.9484	0.9278	0.8563	0.9231	0.8823	0.8934	0.8679	0.8628	0.9059	0.8092	0.8354	0.8478	0.8872	0.8771	0.8968
	C	-	-	0.0836	0.3182	0.3331	0.1539	0.5184	0.6087	0.4156	0.7279	0.8348	0.5999	0.8911	0.9521	0.7794	-	-	-
churn	L	-	-	5.4938	1.8048	3.0019	4.3572	3.1851	3.4325	1.7106	3.4123	3.1790	2.4853	2.4249	2.4551	2.7928	2.7068	3.0171	3.3679
	H	-	-	0.7755	0.8645	0.6878	0.6151	0.8288	0.8240	0.5155	0.7512	0.8312	0.7426	0.7457	0.7421	0.8440	0.7976	0.7713	0.6935
	C	-	-	0.0232	0.0359	0.0623	0.0521	0.7386	0.2416	0.0885	0.9469	0.8602	0.2839	0.9994	0.9998	0.8890	-	-	-
ionosphere	L	2.7664	2.7312	2.7883	-	1.454	2.7907	2.2397	2.1281	2.7753	2.1507	2.1248	2.6147	1.7594	1.6729	1.8230	2.2291	2.0222	2.5584
	H	0.9976	0.9695	1.0	-	0.95	1.0	0.9625	0.8408	0.9952	0.9261	0.9559	0.9636	0.8464	0.8203	0.8092	0.9332	0.9073	0.9536
	C	0.1943	0.2129	0.1902	-	0.8594	0.2461	0.6069	0.5322	0.2798	0.8348	0.8805	0.3757	0.9568	0.9929	0.7982	-	-	-
pima-indians-diabetes	L	2.3955	2.3648	2.4539	1.6496	1.5457	1.9950	1.5454	1.6471	1.7019	1.4996	1.6071	1.5984	1.5695	1.6917	1.6353	1.7319	1.7713	1.8765
	H	0.831	0.8187	0.8624	0.9403	0.9132	0.8158	0.8728	0.8274	0.8316	0.8039	0.7898	0.8933	0.7959	0.7601	0.8618	0.8488	0.8218	0.8530
	C	0.145	0.161	0.1125	0.306	0.3257	0.1906	0.5525	0.5313	0.3156	0.68	0.6841	0.4219	0.8083	0.8092	0.5199	-	-	-
sportscar-choice-long	L	1.2546	1.2565	1.2546	1.0701	1.2186	-	1.3450	1.4209	1.3216	1.9228	1.3664	2.0469	1.3459	1.3051	1.3350	1.3877	1.3135	1.4895
	H	0.837	0.8375	0.837	0.717	0.7348	-	0.6866	0.7295	0.6889	0.6748	0.6011	0.6812	0.5653	0.562	0.5644	0.6961	0.6930	0.6929
	C	0.3305	0.3319	0.3305	0.664	0.553	-	0.7054	0.7682	0.7004	0.8561	0.9161	0.8511	0.9294	0.9563	0.9301	-	-	-
tic-tac-toe-endgame	L	-	-	-	-	-	-	2.0548	2.0547	1.9568	1.7767	1.7893	1.8082	1.8045	1.8006	1.7813	1.8786	1.8815	1.8488
	H	-	-	-	-	-	-	0.9363	0.9362	0.9155	0.8449	0.8416	0.8553	0.8266	0.8269	0.8222	0.8692	0.8682	0.8643
	C	-	-	-	-	-	-	0.3313	0.3313	0.3327	0.7141	0.7155	0.6995	0.9949	1	0.9856	-	-	-
yeast	L	-	-	-	10.774	9.9433	13.997	9.1461	3.3507	4.0720	2.9861	11.473	3.1111	1.9923	2.1853	2.6246	6.2247	6.7395	5.9512
	H	-	-	-	0.549	0.5822	0.5382	0.4513	0.6316	0.5332	0.6198	0.4509	0.5183	0.5591	0.6419	0.6313	0.5448	0.5767	0.5553
	C	-	-	-	0.2209	0.2322	0.1949	0.3116	0.4007	0.2709	0.4812	0.4927	0.348	0.656	0.6742	0.4495	-	-	-

L, H and C mean Lift, Homogeneity and Coverage respectively. Boldface means the best performance of 3 splitting criteria.

#### 4.1 DT의 깊이에 따른 3가지 분기기준의 동질성, 커버리지 그리고 리프트 비교

<Table 2>에 각 데이터마다 30번 부스팅하여 DT 알고리즘을 적용한 결과의 평균값이 나와 있다.

여기서 엔트로피를 사용한 분기기준은 약어로 ENT로 표시되었고, 마찬가지로 GINI, GR은 각각 지니수와 게인비율을 사용한 분기기준을 의미한다. 일단 깊이 1에서 보면 10개 중 3개의 데이터에서 GR 분기기준에 따른 DT 알고리즘만 리프노드가 생성됨을 알 수 있다. 2개의 데이터에서는 모든 DT 알고리즘에서 리프 노드가 형성되지 않았다. 5개의 데이터를 보면 동질성이나 리프트는 3개의 데이터에서 GR이 우수하고 하나는 ENT가 나머지 하나는 3가지가 거의 같은 결과를 보인다. 커버리지 측면에서는 5개중 4개의 데이터에서 GINI가 우수하고 나머지 하나에서 ENT가 우수한 결과를 보임을 알 수 있다.

DT의 깊이가 2부터는 깊이 1과 같이 GR만 리프 노드가 생성되는 현상은 보이지 않는다. 또한 모든 평가척도에서 우월한 분기기준이 존재하지 않으며, 특정 척도에서도 전체 10가지 데이터에 대해서 전체적으로 우수한 분기기준이 깊이 1에서와 같이 분명하지 않다. 따라서 나머지 깊이에서는 이들을 평균한 값을 토대로 평가척도별 분기기준의 성능을 비교해 보고자 한다. 깊이 1부터 깊이 5까지의 평가척도별로 각 분기기준의 성능을 비교하면, 첫째 리프트의 경우 GR이 매우 우수한 성능을 보인다. 10개의 데이터 중 8개에서 GR이 가장 우수한 성능을 보였다. 리프트와 유사한 개념인 동질성에서는 리프트만큼은 아니지만 GR이 5개의 데이터에서 가장 우수한 결과를 보였고, ENT는 3개의 데이터, 그리고 GINI는 2개의 데이터에서 우수한 결과를 보인다. 커버리지는 평균값은 의미가 없고 깊이 5까지의 누적 커버리지가 의미가 있으므로 깊이 5에서의 누적 커버리지를 보면 모든 데이터가 깊이 5이전 리프 노드로 수렴된 한 개 데이터를 제외한 9개의 데이터중 8개의 데이터에서 GINI가 가장 좋은 결과를 보인다. 이중 4개의 데이터에서는 GINI와 ENT간의 차이가 경미함을 알 수 있다. 특징적인 것은 9개의 데이터중 6개의 데이터에서 GR은 나머지 분기기준에 비추어 커버리지가 20%에서 30% 떨어지는 점을 보인다는 것이다. 이는 GR의 경우 매우 많은 규칙들이 깊이 6 이상에서 형성됨을 보여, 해석력이 떨어짐을 알 수 있다. 이상을 종합하면, GR은 리프트와 동질성 측면에서 우수한 성능을 보이거나 커버리지에서는 매우 낮은 성능을 보이고, GINI와 ENT는 그 반대의 특징을 보임을 알 수 있다.

#### 4.2 3가지 분기기준의 안정성 비교

안정성 비교를 위해 먼저 지배 규칙의 안정성을 비교해

<Table 3> Stability of Dominant Rules in 3 Splitting Criteria

Dataset	classes	ENT	GINI	GR
abalone	3	0.7418	0.8673	0.4947
australian	2	1	1	1
bank-personal-loan	2	0.9858	0.9920	0.8472
breast-cancer-wisconsin	2	0.6960	0.2414	0.2736
churn	2	0.8487	0.9270	0.5319
ionosphere	2	0.3495	0.2046	0.0252
pima-indians-diabetes	2	0.3565	0.4735	0.0776
sportscar-choice-long	2	0.9333	0.5471	1
tic-tac-toe-endgame	2	0.2437	0.2437	0.2437
yeast	9	0.1498	0.0320	0
mean	-	<b>0.6305</b>	<b>0.5529</b>	<b>0.4494</b>

보자. <Table 3>에 지배 규칙의 유사성이 세 가지 분기기준에 대해 나와 있다. <Table 3>을 보면 3가지 분기기준에 대해 모두 안정성이 1로 나온 ‘australian’ 데이터를 제외한 9개의 데이터중 4개의 데이터에서 ENT의 안정성이 가장 높게 나왔다. 나머지 4개 데이터에서는 GINI가 가장 높게 나왔다. 그런데 ENT가 높은 데이터에서는 GINI와의 차이가 GINI가 높은 데이터에서의 GINI와 ENT와의 차이보다 커서 10개 데이터의 평균값을 비교 기준으로 하면 ENT가 지배 규칙의 안정성 측면에서 가장 좋은 결과를 보인다.

다음으로 뿌리 노드의 안정성을 비교해 보자.

뿌리 노드의 안전성은 두 가지 방식으로 측정하고자 한다. 첫째, 식 (10)을 이용하여 뿌리 노드의 분기기준의 변수와 임계값을 토대로 뿌리 노드 분기기준의 유사도를 측정하는 것이다. 둘째, 임계값을 고려하지 않고, 분기 변수만을 고려하여 식 (13)을 사용한 지니계수를 구하여 뿌리 노드의 동질성을 측정하는 것이다. <Table 4>에 식(10)을 이용한 측정 결과가 나와있다. 모든 분기기준이 유사도 1을 보여준 두 개의 데이터를 제외한 8개의 데이터를 보면 유사도가 가장 높은 분기기준이 데이터별로 고르게 나누어져 있는 것을 볼 수 있다. 즉, 뿌리 노드의 유사성 측면에서 안정성이 타 분기기준에 비해 우월한 분기기준이 보이지 않는다. 그러나 ENT분기기준은 유사도가 타 분기기준에 비해 낮은 경우 그 차이가 크지 않다. 결과적으로 10개 데이터의 유사도의 평균값을 보면 ENT가 가장 좋은 결과를 보이고 있다. <Table 5>에 식 (1)을 사용한 측정 결과가 나와있다. 임계값의 차이를 고려하지 않으므로 5개 데이터에서 지니계수가 모든 분기기준에 대하여 0으로 나왔다. 나머지 데이터의 경우, ENT는 ‘yeast’와 ‘breast-cancer-wisconsin’에서 가장 동질성이 높았고, GINI는 ‘abalone’ 그리고 GR는 ‘churn’, ‘ionosphere’와 ‘breast-cancer-wisconsin’에서 가장 동질성이 높은 결과를 보였다. 그러나 10개 데이터의 평균을 보면 ENT가 가장 좋은 결과를 보인다. 따라서 지배 규칙의 유사성이나 뿌리 노드의 유사성 두 가지를 모두 고려하면 ENT가 가장 안정성이 우수한 분기기준이라고 할 수 있다.

<Table 4> Stability of Root Node Splitting Variables in 3 Splitting Criteria

Dataset	classes	ENT	GINI	GR
abalone	3	0.7532	0.8673	0.6133
australian	2	1	1	1
bank-personal-loan	2	0.9858	0.9920	0.9484
breast-cancer-wisconsin	2	0.9134	0.7673	0.9078
churn	2	0.9105	0.9270	0.9975
ionosphere	2	0.8504	0.7794	0.8602
pima-indians-diabetes	2	0.9584	0.9440	0.9743
sportscar-choice-long	2	1	0.7977	1
tic-tac-toe-endgame	2	1	1	1
yeast	9	1	1	0.7607
mean	-	0.9721	0.9360	0.9614

<Table 5> Gini Index of Root Nodes in 3 Splitting Criteria

Dataset	classes	ENT	GINI	GR
abalone	3	0.2311	0.1244	0.3578
australian	2	0	0	0
bank-personal-loan	2	0	0	0
breast-cancer-wisconsin	2	0.0644	0.1800	0.0644
churn	2	0.0644	0.0644	0
ionosphere	2	0.1267	0.1800	0.1244
pima-indians-diabetes	2	0	0	0
sportscar-choice-long	2	0	0	0
tic-tac-toe-endgame	2	0	0	0
yeast	9	0	0	0.2311
mean	-	<b>0.0487</b>	<b>0.0549</b>	<b>0.0778</b>

## 5. 결론 및 추후 연구방향

DT 알고리즘에서 가장 널리 사용되는 3가지 분기기준의 해석력을 평가하는 척도로 본 논문은 리프트, 동질성, 커버리지, DT의 깊이, 그리고 안정성이라는 5가지 척도를 제시하였다. 5개의 평가척도중 안정성은 별도로 측정되어야 하는 개념으로 DT 알고리즘의 안정성을 평가하기 위해, 본 논문은 지배규칙의 유사도와 뿌리 노드의 유사도라는 두 개의 개념과 측정 방법을 제시하였다. UCI와 Kaggle에서 수집한 10개의 데이터를 토대로 3개의 분기기준에 따른 DT 알고리즘의 해석력을 비교한 결과는 다음과 같다.

첫째, 깊이가 1로 가장 단순한 규칙을 생성하는 경우는 GR 분기기준에 따른 DT 알고리즘이 가장 좋은 결과를 보였다.

둘째, 깊이가 1부터 깊이 5까지 종합한 결과, GR 분기기준에 따른 DT 알고리즘은 리프트와 동질성 측면에서 우

수한 성능을 보이나 커버리지에서는 매우 낮은 성능을 보이고, GINI와 ENT 분기기준에 따른 DT 알고리즘은 그 반대의 특징을 보임을 알 수 있다.

셋째, 지배 규칙의 유사성이나 뿌리 노드의 유사성 두 가지를 모두 고려하면 ENT 분기기준에 따른 DT 알고리즘이 가장 안정성이 우수한 결과를 보였다.

이러한 결과를 보면, GR 분기기준에 따른 DT 알고리즘은 안정성을 개선시키는 방법[10]과 함께 사용해야 한다는 것을 알 수 있다. 또한 데이터의 일부분이라도 동질성이 높은 규칙을 찾고 싶으면 GR을 사용하고 비교적 데이터의 많은 부분을 특정 규칙으로 설명하고자 할 때는 ENT나 GINI 분기기준에 따른 DT 알고리즘을 사용하는 것이 바람직하다고 할 수 있다.

본 논문은 가장 널리 사용되는 분기기준에 따른 DT 알고리즘의 해석력을 비교하였으나, 비교의 대상을 규칙 유도 알고리즘 등으로 넓히는 것이 향후 중요한 연구과제가 될 것이다.

## Acknowledgement

This study was supported by the Research Program funded by the SeoulTech (Seoul National University of Science and Technology).

## References

- [1] An, A. and Cercone, N., Rule Quality Measures for Rule Induction Systems, *Computational Intelligence*, 2001, Vol. 17, No. 3, pp. 409-424.
- [2] Baesens, B., Mues, C., De Backer, M., and Vanthienen, J., Building Intelligent Credit Scoring Systems Using Decision Tables, *In : Enterprise Information Systems V*, 2004, pp. 131-137.
- [3] Belle, V. and Papantonis, I., Principles and Practice of Explainable Machine Learning, *arXiv preprint arXiv: 2009.11698*, 2020.
- [4] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Monterey, CA, USA : Wadsworth and Brooks, 1984.
- [5] Breiman, L., Heuristics of Instability and Stabilization in Model Selection, *The Annals of Statistics*, 1996, Vol. 24, No. 6, pp. 2350-2383.
- [6] Briand, B., Ducharme, G.R., Parache, V., and Mercat-Rommens, C., A Similarity Measure to Assess the Stability of Classification Trees, *Computational Statistics and Data Analysis*, 2009, Vol. 53, No. 4, pp. 1208-1217.



- [7] Buntine, W. and Niblett, T., A Further Comparison of Splitting Rules for Decision-Tree Induction, *Machine Learning*, 1992, Vol. 8, No. 1, pp. 75-85.
- [8] Cano, A., Zafra, A., and Ventura, S., An Interpretable Classification Rule Mining Algorithm, *Information Sciences*, 2013, Vol. 240, pp. 1-20.
- [9] Chandra, B., Kothari, R., and Paul, P., A New Node Splitting Measure for Decision Tree Construction, *Pattern Recognition*, 2010, Vol. 43, No. 8, pp. 2725-2731.
- [10] Dannegger, F., Tree Stability Diagnostics and Some Remedies for Instability, *Statistics in Medicine*, 2000, Vol. 19, No. 4, pp. 475-491.
- [11] Freitas, A.A., Comprehensible Classification Models : A Position Paper, *ACM SIGKDD Explorations Newsletter*, 2014, Vol. 15, No. 1, pp. 1-10.
- [12] Garcia, S., Fernandez, A., and Herrera, F., Enhancing the Effectiveness and Interpretability of Decision Tree and Rule Induction Classifiers with Evolutionary Training Set Selection over Imbalanced Problems, *Applied Soft Computing*, 2009, Vol. 9, No. 4, pp. 1304-1314.
- [13] Goldstein, A. and Buja, A., Penalized Split Criteria for Interpretable Trees, *arXiv preprint arXiv:1310.5677*, 2013.
- [14] Harris, E., Information Gain Versus Gain Ratio : A Study of Split Method Biases, *The MITRE Corp., McLean, VI, USA, Tech. Rep.*, 2001.
- [15] Jacobucci, R., Decision Tree Stability and its Effect on Interpretation, Retrieved from [osf.io/m5p2v](https://osf.io/m5p2v), 2018.
- [16] Jaworski, M., Duda, P., and Rutkowski, L., New Splitting Criteria for Decision Trees in Stationary Data Streams, *IEEE Trans. Neural Netw. Learn. Syst.*, 2018, Vol. 29, No. 6, pp. 2516-2529.
- [17] Kotsiantis, S.B., Supervised Machine Learning : A Review of Classification Techniques, *Informatika*, 2007, Vol. 31, No. 3, pp. 249-268.
- [18] Leiva, R.G., Anta, A.F., Mancuso, V., and Casari, P., A Novel Hyperparameter-Free Approach to Decision Tree Construction that Avoids Overfitting by Design, *IEEE Access*, 2019, Vol. 7, pp. 99978-99987.
- [19] Li, R.-H. and Belford, G.G., Instability of Decision Tree Classification Algorithms, *In Proceedings of the 8<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, 2002, pp. 570-575.
- [20] Lustrek, M., Gams, M., and Martincic-Ipsic, S., What Makes Classification Trees Comprehensible?, *Expert Syst. Appl.*, 2016, Vol. 62, pp. 333-346.
- [21] Martens, D., Vanthienen, J., Verbeke, W., and Baesens, B., Performance of Classification Models from a User Perspective, *Decision Support Systems*, 2011, Vol. 51, No. 4, pp. 782-793.
- [22] Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B., Interpretable Machine Learning : Definitions, Methods, and Applications, *arXiv e-prints, p. arXiv:1901.04592*, 2019.
- [23] Norouzi, M., Collins, M., Johnson, M.A., Fleet, D.J., and Kohli, P., Efficient Non-Greedy Optimization of Decision Trees, *Proceedings of the 28<sup>th</sup> International Conference on Neural Information Processing*, 2015, pp. 1729-1737.
- [24] Pazzani, M.J., Mani, S., and Shankle, W.R., Acceptance of Rules Generated by Machine Learning Among Medical Experts, *Methods of Information in Medicine*, 2001, Vol. 40, No. 5, pp. 380-385.
- [25] Quinlan, J.R., *C4.5 : Programs for Machine Learning*, San Francisco, CA, USA : Morgan Kaufmann, 1993.
- [26] Quinlan, J.R., Induction of Decision Trees, *Machine Learning*, 1986, Vol. 1, No. 1, pp. 81-106.
- [27] Verbeke, W., Martens, D., Mues, C., and Baesens, B., Building Comprehensible Customer Churn Prediction Models with Advanced Rule Induction Techniques, *Expert Systems with Applications*, 2011, Vol. 38, No. 3, pp. 2354-2364.
- [28] Wang, Y. and Xia, S.-T., Unifying Attribute Splitting Criteria of Decision Trees by Tsallis Entropy, *in Proc. IEEE Int. Conf. Acoust., Speech Signal Process(ICASSP)*, 2017, pp. 2507-2511.
- [29] Wang, Y., Xia, S.-T., and Wu, J., A Less-Greedy Two-Term Tsallis Entropy Information Metric Approach for Decision Tree Classification, *Knowledge-Based Systems*, 2017, Vol. 120, pp. 34-42.
- [30] Zhao, Y. and Zhang, Y., Comparison of Decision Tree Methods for Finding Active Objects, *Advances in Space Research*, 2008, Vol. 41, No. 12, pp. 1955-1959.

#### ORCID

Jung-Sik Hong | <http://orcid.org/0000-0003-0712-9934>

Geun-Seong Hwang | <http://orcid.org/0000-0003-4462-0482>