

Wine Quality Assessment Using a Decision Tree with the Features Recommended by the Sequential Forward Selection

Seunghan Lee*, Kyungtae Kang**, Dong Kun Noh***

Abstract

Nowadays wine is increasingly enjoyed by a wider range of consumers, and wine certification and quality assessment are key elements in supporting the wine industry to develop new technologies for both wine making and selling processes. There have been many attempts to construct a more methodical approach to the assessment of wines, but most of them rely on objective decision rather than subjective judgement. In this paper, we propose a data mining approach to predict human wine taste preferences that is based on easily available analytical tests at the certification step. We used sequential forward selection and decision tree for this purpose. Experiments with the wine quality dataset from the UC Irvine Machine Learning Repository demonstrate the accuracies of 76.7% and 78.7% for red and white wines respectively.

▶ Keyword : Decision Tree, Wine Quality, Classification, Sequential Forward Selection

I. Introduction

와인이란 일반적으로 포도를 원료로 해서 만든 술이라고 할 수 있다. 넓은 의미에서 단맛을 지닌 과일주스에 효모를 넣어서 발효시킨 모든 양조주를 가리키기도 하지만, 대개는 순수한 포도만을 거두어 발효시킨 포도주를 의미한다. 와인국제사무국이 마련한 와인의 정의를 보면 신선한 포도송이나 과즙을 완전하게 또는 부분적으로 발효를 시켜 얻은 결과의 내용품을 와인이라고 한다. 현재 우리나라에 와인의 대중화로 와인 시장은 급속도로 성장하고 있다. 이러한 시장의 성장은 와인 산업의 발전을 가속화 하였고, 더불어 많은 소비자들이 와인의 품질에 대해 관심을 갖게 되었다 [1]. 그러나 최근 와인의 소비가 증가하면서 소비자들에게 친숙하게 인식되어 가고 있지만 품질에 대한 정

보는 아직 자세히 알 수가 없다. 이렇게 와인 시장이 꾸준히 성장함에 따라 소비자들은 각자에 맞는 욕구를 충족시키기 위해 정보 교류나 친목을 도모하기 위한 와인 동호회가 꾸준히 증가하고 있고 전문 교육기관을 통한 전문가 양성도 증가하고 있는 추세이다 [12]. 하지만 전반적으로 와인의 지식이 없는 상태에서 선택하는 경우 아무리 설명을 들어도 이해할 수 없는 경우가 있으며, 품질선택에 있어서 어려운 경우가 있다. 품질에 대해서 인터넷으로 와인의 라벨을 보고 정보를 찾게 되는데 이러한 경우 전문가가 아닌 일반사람에 의하여 부정확한 정보를 습득할 경우가 생길 수 있다. 일반적으로 와인의 분석은 전문가에 의하여 맛과 향으로 결정되기 때문에 이것은 객관적이기 보다 주관적이며 많은 시간과 비용이 들게 된다. 따라서 평가된

* First Author: Seunghan Lee, Corresponding Author: Dong Kun Noh

**Seunghan Lee(shlee10@hanyang.ac.kr), Dept. of Computer Science and Engineering, Hanyang University

***Kyungtae Kang(ktkang@hanyang.ac.kr), Dept. of Computer Science and Engineering, Hanyang University

***Dong Kun Noh(dnoh@ssu.ac.kr), Department of Smart Systems SW, Soongsil University

Received: 2016. 12. 06, Revised: 2017. 01. 26, Accepted: 2017. 02. 08.

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1018) supervised by the IITP (Institute for Information & communications Technology Promotion)

와인의 품질에 대한 정보는 일반 소비자들이 쉽게 접근하기가 어렵다. 본 논문은 적은 데이터를 이용하여 높은 정확도를 나타낼 수 있는 장점을 가지는 의사결정나무를 이용한 와인품질 분류 기법을 제안한다. 우선 와인에 대한 각 특징과 와인의 품질 분석을 하기 위해서 11개의 특징(Feature)을 대상으로 정보이득치(Information Gain)를 측정하고 가장 큰 값을 가지는 특징을 시작으로 SFS (Sequential Forward Selection) 알고리즘을 적용한다. SFS 알고리즘은 특징들을 하나씩 새롭게 추가하면서 의사결정나무(Decision Tree)를 생성한다. 추가하는 특징들의 순서는 분류된 결과의 정확도에 근거하며, 이때 사용한 데이터는 표 1에 정리하였다. 최종적으로 생성된 의사결정나무는 향후 새로운 와인데이터를 입력 받았을 때 그 품질을 분류하는 용도로 사용이 되며, UC Irvine Machine Learning Repository에 정의된 데이터셋을 사용한 실험을 통해 구성된 의사나무트리틀 이용하여 Red wine과 White wine에 대해 각각 76.04%, 78.70%의 정확도로 품질을 분류할 수 있음을 확인하였다.

II. Related Works

현재 계속적으로 와인의 품질 평가에 대해서 데이터 마이닝 기법을 이용한 방법론이 제안되어 왔다. Cortez et al [1, 18]. 는 와인의 화학 분석에 Support vector machine, Multiple Regress, and 뉴런 네트워크 (Neural network)를 사용하여 와인의 화학적 분석을 측정하여 선호도를 예측하려고 시도 하였다. 또한 회귀접근 방식을 이용하여 주어진 입력 변수가 그 영역을 통해 변화 될 때 응답 변화를 측정하는 민감도 분석을 측정하였습니다. Shanmugathan et al [5]. 은 포도 수확량과 와인의 품질에 계절과 기후의 영향을 예측하였고, Chen et al [6]. Wineinformatics system의 계층적 클러스터링을 적용하여 맛과 와인의 특성을 추상화 하였다. 우리는 UC Irvine Machine Learning Repository [4] 에서 와인 퀄리티(Wine

quality) 데이터 셋을 사용하여 의사 결정 나무 [2], [3] 를 사용한 방식을 제안하고, Daniele Grifoni et al [19]는 기상 조건에 포도 재배를 하는 것에 따라 품질과 양을 결정하는 방식으로 예측하는 도구를 개발 하였다. 위에서 언급한 기존 관련연구와는 달리, 본 논문에서는 와인의 물리화학적 방식을 이용한 새로운 아이디어를 이용하여 와인 품질을 분류하였다.

III. Data Set

1. Wine Dataset

UC Irvine Machine Learning Repository의 와인 퀄리티 데이터 셋트는, 레드 와인 1599개의 인스턴스와 화이트 와인 4898개의 인스턴스를 포함한다. 각 인스턴스는 표 1과 같이 11개의 속성으로 이루어져 있다. 실험에서 사용된 데이터 세트는 세계 10대 와인 생산국 중 하나인 포르투갈의 Minho 지역의 와인 퀄리티 데이터이다. 데이터는 와인의 품질을 개선하고 마케팅 하는 공식 인증기관(CVRVV) [13] 에서 테스트 되었다.

2. Data Characteristic

표 1은 와인의 물리화학적 특성을 나타낸 것이다. 11개의 속성의 최소값과 최대값을 측정하고 각각의 평균을 측정한 값이다. 와인의 품질은 3명의 전문가의 블라인드 테스트를 통해 0(Very bad)에서 10(Excellent) 사이의 값으로 측정된 데이터를 이용하여 와인의 품질 3~5는 Bad, 6~9는 Good으로 사용하였다. 그림 1은 와인 퀄리티 데이터 세트의 물리화학적 데이터의 중요도를 표시한다. 레드 와인과 화이트 와인의 중요성분인 황산염(Sulphates)은 알코올 발효의 부산물로서 생성되거나

Table 1. Wine Quality Data Set

Attribute(6,497)	Red wine(1,599)			White wine(4,898)		
	min	max	mean	min	max	mean
Fixed acidity (g(tartaric acid)/dm ³)	4.6	15.9	8.3	3.8	14.2	6.9
Volatile acidity (g(tartaric acid)/dm ³)	0.1	1.6	0.5	0.1	1.1	0.3
Citric acid (g/dm ³)	0.0	1.0	0.3	0.0	1.7	0.3
Residual sugar (g/dm ³)	0.9	15.5	2.5	0.6	65.8	6.4
Chlorides (g(sodium chloride)/dm ³)	0.01	0.61	0.08	0.01	0.35	0.05
Free sulfur dioxide (mg/dm ³)	1	72	14	2	289	35
Total sulfur dioxide (mg/dm ³)	6	289	46	9	440	138
Density (g/dm ³)	0.990	1.004	0.996	0.987	1.039	0.994
pH	2.7	4.0	3.3	2.7	3.8	3.1
Sulphates (g(potassium sulphate)/dm ³)	0.3	2.0	0.7	0.2	1.1	0.5
Alcohol	8.4	14.9	10.4	8.0	14.2	10.4

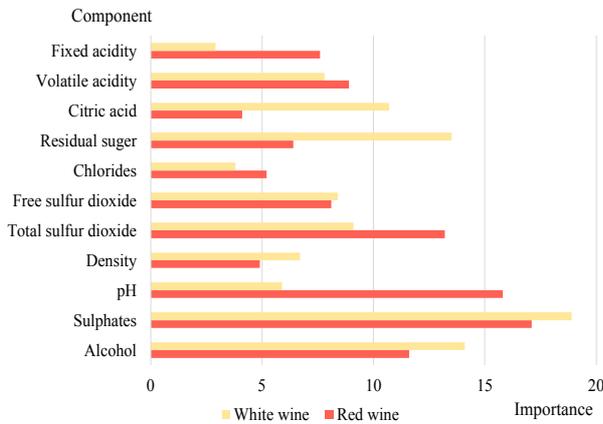


Fig. 1. Importance of physiochemical indicators

저장 중 공기와 접촉하여 비효소적 알코올 산화로 소량 생성되는 것이다. 대부분은 공기가 있는 상태에서 미생물적 산화로 생성되기도 한다. 와인의 종류에 따라 물리화학적 중요도가 달라진다. 각각의 레드 와인과 화이트 와인의 물리화학적 지표 중요도가 다른 것을 볼 수가 있는데 이것은 와인의 제조 과정에 따라서 달라지는 것을 알 수 있다. 알코올과 수소가 온 농도(pH)는 포도를 발효시킬 때 나오며, 밀도는 와인의 점도를 나타낸다. 또한 이산화황은 방부제, 살균제, 방부제 등의 역할로 발효에 도움이 되지 않는 효모를 죽여 발효가 효과적이고 위생적으로 진행되도록 돕는다. 이산화황은 와인을 만드는 과정 중 자연적으로 나오게 되는 성분이며, 잔류 설탕은 와인을 제조할 때 나오며 레드 와인보다 화이트 와인이 좀 더 단맛을 느낄 수가 있는데 화이트 와인은 껍질과 씨앗을 제거한 뒤 발효시키기 때문에 단맛이 좀 더 높은 것을 볼 수 있다. 구연산은 포도를 발효시키면 나오는 천연적은 식초이며, 고정 산도와 휘발성 산도는 와인의 포함된 산의 양을 나타낸다.

IV. Classification

1. Decision Tree

의사결정나무 분석은 학습의 정확도 보다 분석과정의 설명이 필요한 경우에 더 유용하게 사용된다. 일반적으로 의사결정나무는 의사 결정 규칙(Decision rule)을 도표화 하여 관심 대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측 나무 구조에 의해서 표현되기 때문에 판별 분석(Discriminant Analysis), 회귀분석(Regression Analysis), 신경망(Neural Networks)등과 같은 방법들에 비해 연구자가 분석과정을 쉽게 이해하고 설명할 수 있다는 점을 가지고 있다. 의사결정나무는 분류 또는 예측 목적으로 하는 어떤 경우에도 사용 될 수 있으나 분석의 정확도 보다는 분석과정의 설명이 필요한 경우에 더 유용하게 사용 된다. 의사결정나무의 장점은 결과가 나오게 된 과정에 대해 명확하며 속도가 빠르다. 또한 어떤 변수들이 분류에 중요한 영향을 미치는지 알 수 있다. 하지만 연속성 변수의 처리에 한계가 있고, 정확한 모형을 만들기 위해서는 모든 변수가 가질 수 있는 경우를 모두 표현할 수 있는 충분한 표본 수가 있어야 하는 단점도 가지고 있다. 의사결정나무 중 대표적인 알고리즘이 ID3이다. ID3는 엔트로피(Entropy) 기반의 이득률에 의한 동질성 측도를 이용하여 분류를 한다. 그리고 속성 선택 측도는 정보이익(Information Gain)을 이용한다. 그리하여 ID3의 단점들을 보완하여 나온 알고리즘이 C4.5 [7] 이다. ID3 알고리즘은 모든 속성이 명목 형 속성일 경우에만 처리가 가능하였다. 그러나 실세계의 데이터들은 대부분 수치 형 속성들을 포함하고 있다. ID3 알고리즘에 수치 형 속성을 처리할 수 있도록 확장하는 것이 필요하다. C4.5는 의사결정나무를 형성하기 위하여 처음 수행하는 작업이 분할 정복이다. 입력되는 훈련 집합이

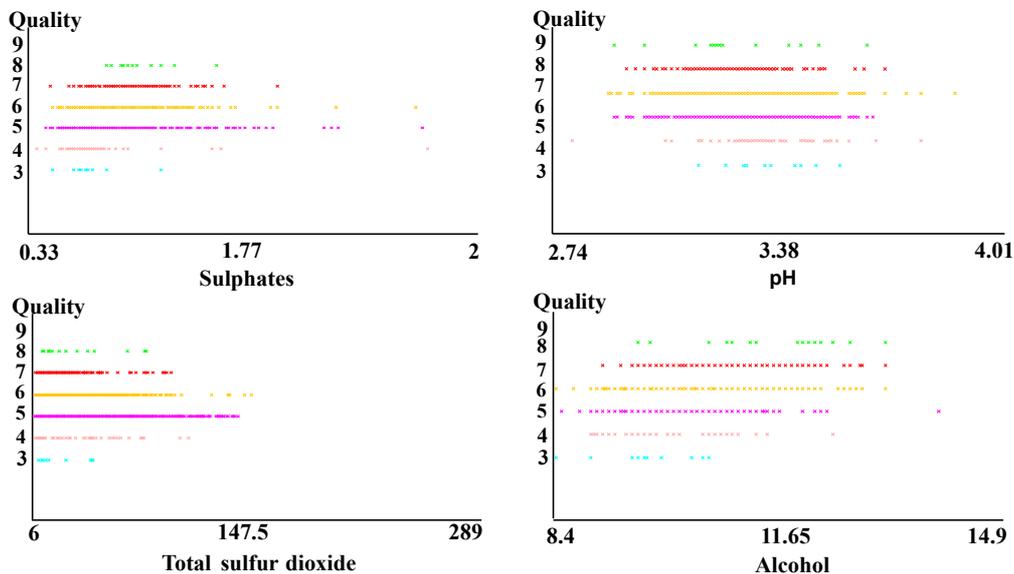


Fig. 2. Effect of key physiochemical indicators on the quality of red wines

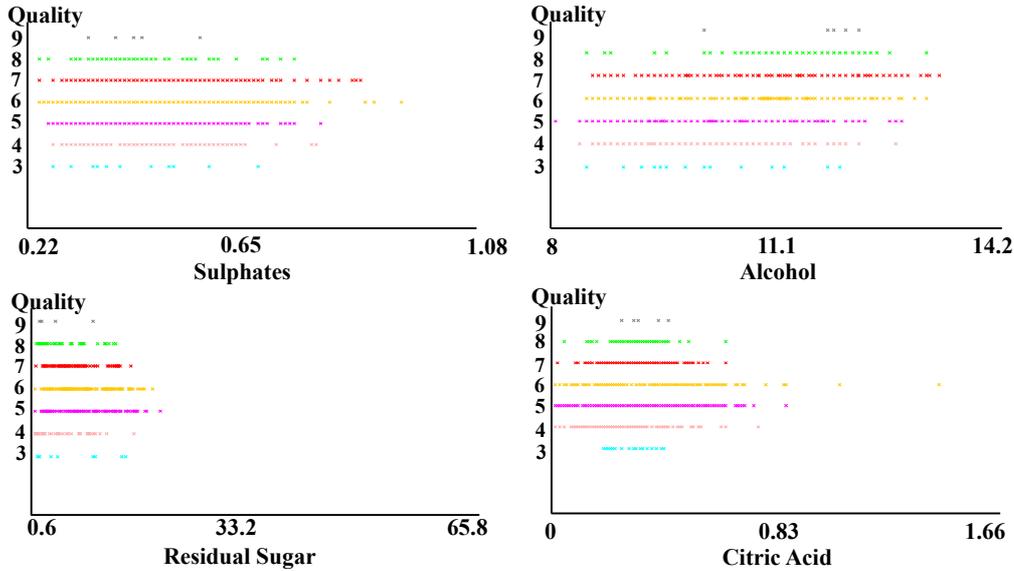


Fig. 3. Effect of key physiochemical indicators on the quality of white wines.

성공적으로 분할 되도록 모든 하부 집합에 하나의 클래스가 속하는 경우들로 구성될 때까지 나무를 형성한다. 정보이득비율이 노드를 분리하는 기준으로 사용한다. 표 2와 같이 의사결정나무의 특징을 볼 수 있으며, 표 3은 의사결정나무의 구성요소를 보여주고 있다.

Table 2. Characteristic of Decision Tree

Advantages	<ul style="list-style-type: none"> • Easy to understand model • It is possible to use all variables • Automatically exclude variables that do not affect classification in model building • Do not need assumptions such as linearity or normality
Disadvantages	<ul style="list-style-type: none"> • The ability to process data is less than that of neural networks or other statistical techniques • It is important to the size of the sample used to construct the model
Characteristic	<ul style="list-style-type: none"> • It is more useful than other techniques, and it can easily tell the basis of classification or prediction

Table 3. Component of the decision tree

Root Node	• The first node at the beginning of the tree structure.
Child Node	• Two or more nodes that are separated from one node
Parent Node	• Upper node of child node
Terminal Node	• A node located at the bottom end of each tree root.
Internal Node	• The node between the root node and the end node
Branch	• A series of nodes from root node to end node

2. Decision Tree Purity

우리는 C4.5 알고리즘 예측모델로 의사결정나무를 구성하여 와인의 물리화학적 특징으로 측정하였고, 취향 선호도를 예측하는 데 사용하였다. C4.5 알고리즘의 속성이 높고 낮은 값에 따라서 가장 영향력 있는 속성을 선택한 뒤 세분화하여 나무를 구성하게 된다. 세분이 완료된 후 모든 인스턴스는 하위 노드에 걸쳐 분산된다. C4.5에서 가장 영향력 있는 속성을 선택하게 되는데 이것은 Information Gain을 측정하여 선택하게 되는데 Entropy [8], [9] 를 고려하여서 측정할 수가 있다.

$$Entropy(S) = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

S는 주어진 식의 집합을 나타내고 있으며 p_i 는 클래스 요소의 비율을 나타내고 있다. 엔트로피가 0 일 경우 모든 인스턴스는 똑 같은 클래스로 속해져 있고 엔트로피가 1일 경우 모든 클래스가 다르게 속하게 된다.

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

속성 A의 모든 가능한 값의 집합 A를 사용하여 구분된 S의 Entropy에 대한 기대값을 나타낸다. 따라서 Gain(S,A)는 속성 A의 값을 사용하여 Entropy 감소량에 대한 기대치를 의미하는 것이다.

3. Sequential Forward Selection

Sequential Forward Selection(SFS)[17] 알고리즘을 통해 기존의 특징 중 최적의 조합을 도출하는 방식이다. 특징 추출하는 방법으로 주성분분석(Principal Component Analysis)[20, 21]이 보통 사용되지만 특징을 선택하는 방법으로는 SFS 또는 SFBS(Sequential Floating Forward Selection)등의 방법들을 자주 사용하고 있다. SFS는 가장 간단한 Greedy Search Algorithm이다[15,16]. 초기에 공집합으로 시작해 첫 단계 에서 모든 특징 중에 가장 높은 정확성을 갖는 하나의 특징을 선택하여 분류작업을 수행하고 다음단계에서 두 번째 정확성을 갖는 특징이 추가되어 두개의 특징을 조합해 분류작업이 수행된다[14]. 이렇게 각 단계에서 특징들을 하나씩 추가하는 작업을 반복해 나가면서 가장 좋은 정확도를 나타내는 특징부터 순차적으로 나열되는 새로운 특징이 선정되어지고 이 특징을 이용한 분류작업의 정확성은 어느 한 점에서 더 이상 크게 향상되지 않게 된다. 그림 4는 SFS 알고리즘 순서도를 보여준다. 처음에는 비어있는 집합에서 시작하여 한 개씩 각각의 새로운 특징을 추가하고 평가하여 최적의 적합도를 보여주는 특징을 선택하고, 반복하여 한 개씩 새로운 특징을 추가하는 방법이다.

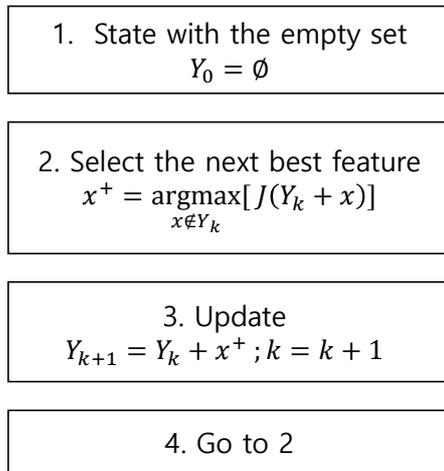


Fig. 4. Sequential Forward Selection algorithm flowchart

V. Experiment

1. Experimental results of Decision Tree using Sequential Forward Selection

Sequential Forward Selection를 사용하여 각 특징을 하나씩 추가하고 의사결정나무의 5-fold Cross Validation 사용하여 측정하였다. 표 5, 6은 레드와인과 화이트와인의 각 특징을 추가함에 따라서 각각의 정확도를 나타내었다. 그리고 표

7, 8은 SFS알고리즘을 사용하지 않는 전체적인 특징과 SFS 알고리즘을 사용하여 가장 높은 정확도를 나타내는 특징을 나누어 TP, FP, Precision, Recall, F-Measure, Accuracy를 측정하였다.

Table 4. Feature Selection

Feature
1. Fixed acidity
2. Volatile acidity
3. Citric acid
4. Residual sugar
5. Chlorides
6. Free sulfur dioxide
7. Total sulfur dioxide
8. Density
9. pH
10. Sulphates
11. Alcohol

Table 5. Red Wine Quality Feature Selection using Sequential Forward Selection

Feature	Accuracy(%)
11	69.8561
11,2	73.3583
11,2,7	73.6710
11,2,7,9	74.3589
11,2,7,9,10	75.2970
11,2,7,9,10,3	76.0475
11,2,7,9,10,3,8	76.0475
11,2,7,9,10,3,8,4	75.7973
11,2,7,9,10,3,8,4,5	75.1719
11,2,7,9,10,3,8,4,5,6	75.4846
11,2,7,9,10,3,8,4,5,6,1	74.6716

Table 6. White Wine Quality Feature Selection using Sequential Forward Selection

Feature	Accuracy(%)
11	69.0690
11,2	74.5610
11,2,9	75.9085
11,2,9,6,3	77.0518
11,2,9,6,3,5	77.5622
11,2,9,6,3,5,8	77.8481
11,2,9,6,3,5,8,1	78.1339
11,2,9,6,3,5,8,1,4	78.6035
11,2,9,6,3,5,8,1,4,7	78.7055
11,2,9,6,3,5,8,1,4,7,10	78.4810

표 5는 레드와인의 Information Gain 값이 높은 특징부터 차례대로 추가하면서 측정한 값이다. 레드와인의 6번째, 7번째 특징을 넣은 결과 값과 비교하여 8번째부터 더 이상 향상되지 않고 정확도가 감소하는 것을 볼 수가 있다. 모든 특징

들을 사용하지 않더라도 더 좋은 결과를 나올 수 있다는 것을 알 수가 있다. 표 6은 화이트와인의 데이터를 측정된 값이다.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F\text{-Measure} = \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

표 7과 8은 레드와인 데이터 셋과 화이트와인 데이터 셋을 순차 전향 선택 알고리즘 사용하여 높은 정확도를 보여주는 결과 값과 알고리즘을 사용하지 않고 11개의 특징을 모두 사용하여 의사결정나무로 분류한 결과 값이다. 그 결과 전체적으로 향상된 것을 볼 수가 있었다.

Table 7. Red Wine Quality Compare with Sequential Forward Selection and Non Sequential Forward Selection

Non Sequential Forward Selection	
TP Rate	0.736
FP Rate	0.241
Precision	0.778
Recall	0.736
F-Measure	0.756
Accuracy	74.6717%

Sequential Forward Selection	
TP Rate	0.774
FP Rate	0.255
Precision	0.777
Recall	0.774
F-Measure	0.776
Accuracy	76.0475%

Table 8. White Wine Quality Compare with Sequential Forward Selection and Non Sequential Forward Selection

Non Sequential Forward Selection	
TP Rate	0.847
FP Rate	0.339
Precision	0.832
Recall	0.847
F-Measure	0.840
Accuracy	78.4810%

Sequential Forward Selection	
TP Rate	0.844
FP Rate	0.326
Precision	0.837
Recall	0.844
F-Measure	0.841
Accuracy	78.7056%

VI. Conclusions

본 논문은 순차 전향 선택 알고리즘을 제안하여 특징을 분석한 뒤 의사결정나무를 사용하여 특징벡터를 분류하였다. 기존의 순차 전향 선택을 사용하지 않고 의사 결정 나무만을 사용하여 분류한 결과보다 순차 전향 선택과 의사 결정 나무를 사용한 결과가 모든 특징을 사용하지 않고 일부의 특징을 사용하여도 분류 결과가 향상 되는 것을 볼 수 있었다. 11개를 모두 사용하지 않고도 순차 전향 선택 알고리즘을 사용하여 좀 더 효율적인 방식으로 결과를 도출할 수 있었다. 표 7,8에서 볼 수 있듯이 기존의 순차 전향 선택 알고리즘을 사용한 것과 사용하지 않는 것을 UC Irvine Machine Learning Repository의 Wine Quality dataset을 이용하여 분석을 하였다. 제안된 알고리즘 기반으로 향후 다양한 Wine Quality dataset에 대해서 실험하고 특징분석을 다양하게 함으로써 더 좋은 성능을 실험할 수 있을 것이다. 이러한 결과가 소비자들에게 와인을 선택하는데 더 나은 기회를 제공할 수 있음을 시사하며 좀 더 좋은 방향으로 나아갈 수 있도록 연구를 계속할 것이다.

REFERENCES

- [1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," Decision Support Systems, vol. 47, no. 4, pp. 547-553, Nov. 2009
- [2] A. Abdelhalim and I. Traore, "A new method for learning decision trees from rules," IEEE International Conference on Machine Learning and Applications, pp. 693-698, Dec. 2009.
- [3] M. A. Hussain, M. K. Rao, and A. M. Mahmood, "An optimized approach to generate simplified decision trees," IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-5, Dec. 2013
- [4] A. Asuncion and D. Newman, UC Irvine Machine Learning Repository, [Online] Available: <http://archive.ics.uci.edu/ml/index.html>.
- [5] S. Shanmuganathan, P. Sallis, and A. Narayanan, "Data mining techniques for modelling seasonal climate effects on grapevine yield and wine quality," IEEE International Conference on Computational Intelligence, Communication Systems and Networks, pp. 82-89, July 2010.
- [6] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, "Wineinformatics: applying data mining on wine sensory

- reviews processed by the computational wine wheel," IEEE International Conference on Data Mining Workshop, pp. 142–149, Dec. 2014.
- [7] J. R. Quinlan, "Improved use of continuous attributes in C4.5," *Journal of Artificial Intelligence Research* 4, vol. 4, no. 1, pp. 77–90, Jan. 1996.
- [8] I. M. Mitchell, *Machine Learning*, McGraw-Hill International Editions, 1997
- [9] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann; Second Edition, 2005.
- [10] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," In *Proc. International Joint Conference on Artificial Intelligence*, pp. 1137–1143, Aug. 1995.
- [11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, June 1997.
- [12] WineKorea, [Online] Available: <http://www.winekorea.asia/>
- [13] A Comissão de Viticultura da Região dos Vinhos Verdes, [Online] Available: <http://www.vinhoverde.pt/>
- [14] C.M Lee and S.S Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 2, pp. 293–303, May 2005
- [15] Chu Weibo, Zhu Bin B., Xue Feng, Guan Xiaohong, and Cai Zhongmin, "Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs," *IEEE International Conference on Communications*, pp. 1990–1994 June 2013.
- [16] Ladha, L., Deepa, T., "FEATURE SELECTION METHODS AND ALGORITHMS," *International Journal on Computer Science & Engineering*, Vol. 3, no. 5, pp. 1787–1797, May 2011
- [17] Dimitrios Ververidis and Constantine Kotropoulos, "Sequential forward feature selection with low computational cost" *European Signal Processing Conference*, pp. 1–4, Sept. 2005
- [18] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Using Data Mining for Wine Quality Assessment," *International Conference on Discovery Science* pp. 66–79, Oct. 2009
- [19] Daniele Grifoni, Marco Mancini, Giampiero Maracchi, Simone Orlandini and Gaetano Zipoli, "Analysis of Italian

Wine Quality Using Freely Available Meteorological Information," *American Journal of Enology and Viticulture*, pp. 339–346, Sep. 2006

- [20] Sungmain Myoung and Chang-Hwan Oh, "Pattern Recognition for Typification of Whiskies and Brandies in the Volatile Components using Gas Chromatographic Data" *Journal of the Korea Society of Computer and Information*, Vol. 21, no. 5 pp. 167–175, May 2016
- [21] Seok-Woo Jang, Moon-Haeng Hun and Gye-Young Kim, "Effective Handwriting Verification through DTW and PCA" *Journal of the Korea Society of Computer and Information*, Vol. 14, no. 7 pp. 25–32, July. 2009

Authors



Seunghan Lee received the B.S. degrees in computer science and engineering from Inje University, Korea, in 2014. he currently joined the Department of Computer Science and Engineering, Hanyang University, Korea.

His research interests include artificial intelligence and machine learning.



Kyungtae Kang received the B.S. degree in computer science and engineering and the M.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 1999, 2001, and 2007, respectively.

From 2008 to 2010, he was a Postdoctoral Research Associate with the University of Illinois at Urbana-Champaign, Urbana, IL, USA. In 2011, he joined the Department of Computer Science and Engineering, Hanyang University, Korea, where he is currently an Associate Professor. His research interests are primarily in systems, such as operating systems, wireless systems, distributed systems, real-time embedded systems, and the interdisciplinary area of cyber-physical systems.



Dong Kun Noh received BS, MS, and PhD degrees in EECS from Seoul National University in 2000, 2002, and 2007, respectively. He has been in University of Illinois at Urbana-Champaign as a postdoctoral researcher from 2007 to 2010.

He is currently an associate professor in department of smart systems software at Soongsil University in Korea. His research interests include cyber-physical system and mobile communication.