

## 데이터 마이닝에서 Cohen의 kappa를 이용한 분류정확도 측정

엄 용 환\*

# Assessing Classification Accuracy using Cohen's kappa in Data Mining

Yonghwan Um\*

### 요 약

본 논문에서는 데이터 마이닝에서 분류 작업을 실시할 때 그 분류정확도를 측정하기 위해 Cohen의 kappa 계수와 weighted kappa 계수를 제안하였다. kappa 계수는 우연에 의해 생기는 분류를 보정하여 분류정확도를 측정하며 명목척도와 순서척도의 데이터에 대해 사용된다. 특히 순서척도의 데이터에서는 오분류의 크기를 가중치에 의해 정량화하여 분류정확도를 측정하는 weighted kappa 계수가 더 유용하게 사용된다. weighted kappa 계수 계산을 위해서는 2가지 가중치(일차형 가중치, 이차형 가중치)를 사용하였다. 또한 실제 데이터인 지방간 데이터에 대해 kappa 계수와 weighted kappa 계수를 계산하여 비교하였다.

▶ Keywords : 분류정확도, 순서척도, weighted Cohen's kappa,

### Abstract

In this paper, Cohen's kappa and weighted kappa are applied to measuring classification accuracy when performing classification in data mining. Cohen's kappa compensates for classifications that may be due to chance and is used for the data with nominal or ordinal scales. Especially, for the ordinal data, weighted kappa which measures the classification accuracy by quantifying the classification errors as weights is used. We used two weights (linear weight, quadratic weight) for calculations of weighted kappa. Also for the calculation and comparison of kappa and weighted kappa we used a real data set, fat-liver data.

▶ Keywords : classification accuracy, ordinal data, weighted Cohen's kappa,

---

· 제1저자 : 엄용환

· 투고일 : 2012, 11, 19, 심사일 : 2012, 12, 1, 게재확정일 : 2012, 12, 18.

\* 성결대학교 산업경영공학부(Division of Industrial and Management Engineering, Sungkyul University)

## I. 서론

분류(Classification)는 데이터 마이닝의 여러 기능 중에 널리 사용되는 보편적인 작업이다. 분류작업은 정의된 분류기준에 의해 각각의 레코드를 특정 범주에 할당하고 이를 하나의 변수로 만들어 표현하게 된다. 은행에 대출 신청한 고객들의 신용상태를 우량, 불량으로 분류하는 것, 보험 청구 사례를 정상적인 청구와 사기 청구로 분류하는 것, 환자들의 증상을 악성, 양성으로 분류하는 것 등이 분류의 예이다. 분류의 정확성을 나타내기 위해서는 분류정확도(Classification Accuracy) 또는 분류일치도(Classification Agreement)의 통계수치가 사용되고 분류 결과는 종종 2차원의 오분류표(confusion matrix)로 제시된다. 이 표의 대각선 상에 놓인 것이 정확히 분류된 것이고 비 대각선 상에 있는 것은 잘못 분류된 것을 나타내며 이때 분류정확도는 전체 분류 대상 중에서 바르게 분류된 대상들의 비율(정분류율)로서 표현된다. 그러나 이 비율로서 분류정확도를 나타내는 것은 오해의 소지가 있으며, 오분류 결과는 값비싼 손실을 가져오므로 주의해야 한다. 특히 의학, 금융, 군사학 분야에 적용할 경우 오분류에 의한 손실은 치명적이라 할 수 있다[1][2].

오분류표에서 정분류율이 분류정확도로서 적합하지 않은 이유는 다음과 같다. 첫째로 오분류표에서 정분류된 결과들은 우연에 의해 발생한 것일 수 있기 때문에 이런 가능성을 배제하여 분류정확도를 측정할 수 있는 통계치가 필요한데 정분류율은 이 점을 충족시키지 못하고 있다는 것이다. 둘째로 정분류율을 사용하는데 있어서의 문제는 오분류된 결과들을 단순히 바르게 분류되지 못한 것으로서 모든 오류를 동일하게 취급한다는 것이다. 예를 들어 고객의 만족도(불만, 보통, 만족), 고객의 신용 등급(상, 중, 하), 지방간 환자의 간수치 정도(매우 심각, 심각, 정상) 등과 같이 목표변수의 범주가 순서척도의 데이터일 때 오분류된 결과는 그 의미에 차이가 있다고 말할 수 있다. 이를테면 신용 등급이 '상'인 고객을 '중'으로 분류하는 것보다 '하'로 분류하는 것은 그 오류의 정도가 크다고 할 수 있으며 이로 인한 손실도 커질 것이다. 따라서 이같은 이유로 정분류율을 분류정확도를 나타내는 통계치로 사용하는 것은 적절치 못하며, 이 점을 보완해 줄 수 있는 새로운 분류정확도가 필요한데, Cohen이 제안한 kappa 계수와 가중치가 부여된 Cohen의 kappa (Cohen's weighted kappa) 계수가 대표적인 통계수치라 할 수 있다.[3]-[5]. Cohen의 kappa는 우연히 분류가 일치할 가능성을 보정(chance-corrected)함으로써 분류정확도를 측정된 것이며,

그 값은 (실제 분류정확도-우연에 의해 기대되는 분류정확도)로 주어진다. Cohen's weighted kappa는 분류 범주가 순서화 되어 있을 때, 오분류 결과를 그 정도에 따라 오류를 정량화하여 분류정확도를 측정하는데 사용된다. 실제로는 범주  $i$ 에 속해 있는 대상을 범주  $j$ 에 분류함으로써 빚어지는 불일치 정도를 정량적으로 계산하면서 분류정확도를 산출한다. 또한 Cohen의 kappa는 chance-correctedness 뿐만 아니라 유의성 검증(test of significance)과 다변량 데이터로의 확장성 등 통계치와 분류정확도로서 갖추어야 할 성질들을 만족한다[6].

본 논문의 제 2장에서는 Cohen의 kappa 계수를 소개하고 제 3장에서는 오류를 정량화하는 방법과 더불어 weighted kappa 계수를 소개하며 제 4장에서는 실제 데이터에 대해 kappa 계수를 이용하여 측정된 분류정확도를 제시하였다. 이때 사용되는 분류기법에는 데이터 마이닝에서 널리 쓰이는 의사결정나무, 회귀분석, 신경망 등 다양한 방법이 있으나 본 논문에서는 의사결정나무와 회귀분석을 사용하였다.

## II. Cohen의 kappa 계수

Cohen의 kappa 계수는 심리학, 교육학, 사회학 분야에서 처음 사용된 개념으로서 2명 또는 그 이상의 평가자들이 동일한 대상을 관찰한 후 평가자들 사이의 일치 또는 불일치 정도를 나타내는 통계수치이다. 예를 들어 2명의 평가자(O1, O2)가 100명의 학생들을 관찰한 후 이들의 성격을 내향형(S1)과 외향형(S2)으로 분류한다고 가정하고 얻은 가상적인 결과가 표 1과 같다고 하자. 이 표 1에서 2명의 평가자는 87%  $((20+67)/100)$ 의 일치율을 보이고 13%의 불일치를 나타내고 있다. 그러나 이 87%의 일치율에는 평가자들의 분류가 우연히 일치할 수 있는 가능성이 포함되어 있기 때문에 정확한 일치율이라 볼 수 없다. 만일 평가자들이 학생들을 관찰하지 않고 단순히 동전을 던져 나오는 결과에(앞면이면 S1, 뒷면이면 S2) 따라 학생들을 분류한다고 하자. 이 때 얻어지는 4가지 결과((S1,S1), (S1,S2), (S2,S1), (S2,S2))의 각각은 1/4의 확률로 발생하므로 전체 100명 중에서 각 결과에 해당되는 학생 수는 근사적으로 25명이 될 것이므로 이로부터 계산되는 일치율은 50%  $(=(25+25)/100)$ 이다. 이 일치율은 순전히 우연에 의해 발생한 것이므로 실제의 일치율은 우연에 의한 일치율을 제거함으로써 계산되어야 하고 이 점이 반영된 Cohen의 kappa 계수는 다음과 같이 정의된다.

$$K = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

여기서  $P_o = P_{11} + P_{22} = \frac{n_{11}}{n} + \frac{n_{22}}{n}$  는 관찰되는 일치할 확률,

$P_e = P_{1.} \cdot P_{.1} + P_{2.} \cdot P_{.2} = \frac{n_{1.} \cdot n_{.1}}{n^2} + \frac{n_{2.} \cdot n_{.2}}{n^2}$  는 우연에 의해 일치할 확률이다. 따라서 식 (1)에 의해 표 1

의 데이터에서  $P_o = \frac{20}{100} + \frac{67}{100} = 0.87$ ,

$$P_e = \frac{24}{100} \times \frac{29}{100} + \frac{76}{100} \times \frac{71}{100} = 0.6092 \quad \text{이}$$

므로 일치율은 0.87이 아닌

$$K = \frac{0.87 - 0.6092}{1 - 0.6092} = 0.667 \quad \text{이다.}$$

표 1. 2×2 분류표  
Table 1. 2×2 classification table

|         |                         |                         |                         |
|---------|-------------------------|-------------------------|-------------------------|
| O2 \ O1 | S1                      | S2                      | 합                       |
| S1      | 20 (= n <sub>11</sub> ) | 4 (= n <sub>12</sub> )  | 24 (= n <sub>1.</sub> ) |
| S2      | 9 (= n <sub>21</sub> )  | 67 (= n <sub>22</sub> ) | 76 (= n <sub>2.</sub> ) |
| 합       | 29 (= n <sub>.1</sub> ) | 71 (= n <sub>.2</sub> ) | 100 (= n)               |

Cohen의 kappa 계수의 값은 -1과 1사이에 존재하며, 평가자들의 평가가 완전히 불일치 할 때에 -1이 되고 완전히 일치 할 때에 1이 된다. 그리고  $P_o = P_e$  일 때 kappa 계수는 0이 되고,  $P_o > P_e$  일 때는  $0 < K < 1$ ,  $P_o < P_e$  일 경우에는  $-1 < K < 0$  이다. Koch와 Landis는 kappa 계수의 값에 따른 일치도의 세기를 다음과 같이 표현하였다[7].

표 2. kappa 계수와 일치도 세기  
Table 2. Kappa and strength of agreement

| kappa 계수  | 일치도 세기      |
|-----------|-------------|
| <0.00     | 거의 일치하지 않음  |
| 0.00-0.20 | 일치정도가 매우 낮음 |
| 0.21-0.40 | 일치정도가 낮음    |
| 0.41-0.60 | 일치정도가 보통임   |
| 0.61-0.80 | 일치정도가 높음    |
| 0.81-1.00 | 거의 완전히 일치함  |

### III. Cohen의 가중치가 부여된 kappa 계수

만일 2명의 평가자가 1000명의 고객의 신용 상태를 (상, 중, 하)로 분류하여 표 3과 표 4를 얻었다 하자. 표 3과 표 4는 똑같이 1000명의 고객을 오분류 하고 있지만 표 3의 오분류 결과는 표 4의 오분류 결과보다 대각선으로부터 더 많이 벗어나 있기 때문에 더 심각한 오류라고 볼 수 있다. 그러나 이러한 차이가 Cohen의 kappa 계수를 계산하는 데는 전혀 반영되지 않으며 식 (1)에 의해 얻은 kappa 계수는 두 자료에서 똑같이 0.8486이다. 따라서 이러한 단점을 보완하기 위해 사용하는 것이 kappa 계수에 가중치를 부여된 Cohen의 weighted kappa 계수이며 식 (2)와 같이 쓸 수 있다.

$$K_w = \frac{\sum_{i=1}^I \sum_{j=1}^I w_{ij} P_{ij} - \sum_{i=1}^I \sum_{j=1}^I w_{ij} P_{i.} P_{.j}}{1 - \sum_{i=1}^I \sum_{j=1}^I w_{ij} P_{i.} P_{.j}} \quad (2)$$

여기서  $P_i = \frac{n_{i.}}{n}$ ,  $P_j = \frac{n_{.j}}{n}$ 는 관찰된 주변확률(marginal probability)이고  $w_{ij}$ 는 i번째 행과 j번째 열의 해당 빈도에 대한 가중치로서 0과 1 사이의 값을 갖는다 ( $0 \leq w_{ij} \leq 1$ ). 평가자들의 평가 결과 (또는 분류 결과)가 완전히 일치할 때는  $w_{ij} = 1$ , 완전히 불일치 할 때는  $w_{ij} = 0$  이 되고 부분적으로 일치할 때는  $w_{ij}$  값이 0과 1 사이가 된다. 가장 널리 쓰이는 가중치는 일차형(linear)과 이차형(quadratic)이 있으며 다음의 식 (3), 식 (4)와 같다.

$$\text{일차형 가중치} : w_{ij} = 1 - \frac{|i-j|}{R-1} \quad (3)$$

$$\text{이차형 가중치} : w_{ij} = 1 - \frac{(i-j)^2}{(R-1)^2} \quad (4)$$

여기서  $R$ 은 순서형 범주의 갯수이다. 표 3, 표 4의 경우  $R=3$ 이므로 식 (3)과 식 (4)에 의해 얻을 수 있는 일차형 가중치의 값은 1(완전히 일치할 때), 0.5(한 개의 범주차이로 불일치 할 때), 0(두 개의 범주차이로 불일치 할 때)이며, 이차형 가중치의 값은 1(완전히 일치할 때), 0.75(한 개의 범주차이로 불일치 할 때), 0(두 개의 범주차이로 불일치 할 때) 이다. 범주의 수가 4( $R=4$ ) 일 때는 일차형 가중치의 값은 각각 1, 0.667, 0.333, 0이 되고 이차형 가중치의 값은 각각 1, 0.889, 0.5556, 0 이다. 완전히 일치(가중치=1)할 때와 완전히 불일치(가중치=0)할 때를 제외하고는 이차형 가중치가 일차형 가중치보다 크게 나타나고 있다. 오분류표에서 가중치는 대각선을 중심으로 대칭으로 나타나며 분류의 오류가 커질수록 가중치의 크기는 단조감소 (monotonically decreasing) 한다. 표 3과 표 4의 데이터에 계산한 kappa 계수와 weighted kappa 계수는 표 5와 같다. 표 5에서 보듯이 weighted kappa 계수의 값은 모두 0.7 이상으로 Koch와 Landis의 기준에 의하면 높은 일치도를 나타내며, 또한 weighted kappa 계수에 의해 표4에서의 일치도가 표3에서 보다 더 높음을 보여주고 있다.

표 3. 오분류표 1  
Table 3. Confusion matrix 1

|    |    | O1  |     |     |      |
|----|----|-----|-----|-----|------|
|    |    | 상   | 중   | 하   | 총합   |
| O2 | 상  | 300 | 0   | 50  | 350  |
|    | 중  | 0   | 300 | 0   | 300  |
|    | 하  | 50  | 0   | 300 | 350  |
|    | 총합 | 350 | 300 | 350 | 1000 |

표 4. 오분류표 2  
Table 4. Confusion matrix 2

|    |    | O1  |     |     |      |
|----|----|-----|-----|-----|------|
|    |    | 상   | 중   | 하   | 총합   |
| O2 | 상  | 300 | 50  | 0   | 350  |
|    | 중  | 50  | 300 | 0   | 350  |
|    | 하  | 0   | 0   | 300 | 300  |
|    | 총합 | 350 | 350 | 300 | 1000 |

표 5. kappa 계수와 가중치가 있는 kappa 계수의 비교  
Table 5. Comparison of kappa and weighted kappa

|           | 표3     | 표4     |
|-----------|--------|--------|
| $K$       | 0.8486 | 0.8486 |
| 일차형 $K_w$ | 0.7802 | 0.8857 |
| 이차형 $K_w$ | 0.7143 | 0.9227 |

#### IV. 실제 데이터에서의 분류정확도

Cohen의 kappa 계수를 이용하여 분류정확도를 산출하기 위해 실제 데이터인 지방간 자료를 이용하였다[8]. 이 자료는 국내 모 병원에서 검진 받은 수검자들을 대상으로 지방간 상태, 그리고 이와 관련된 변수들을 조사한 것이다. 이 자료는 총 511명의 수검자로부터 얻은 17변수(연령, 성별, 음주 정도, 혈압, 체질량 지수, 콜레스테롤 수치, 간기능 검사 수치, 중성지방 수치, 혈색소, 지방간 상태 등)로 구성되고 목표변수인 '지방간 상태'는 0(정상), 1(약), 2(심각), 3(매우 심각)의 순서 척도의 값을 갖는다. 지방간 발병에 영향을 주는 주된 요인들을 파악하고, 지방간 환자들에 대한 특성을 이해하기 위해 데이터 마이닝에서 분류 기능에 널리 쓰이는 의사결

정나무 기법과 로지스틱 회귀분석을 사용하였다. 결과적으로 지방간 발병에 큰 영향을 주는 변수들은 체질량 지수, 나이, 간기능 검사 수치 순으로 나타났으며, 총 17개 변인들을 이용해 수검자들의 지방간 상태를 예측하였다. 표 6과 표 7은 의사결정나무와 회귀분석에 의해 각각 수검자들의 지방간 상태를 예측한 오분류표이고 이로부터 계산된 kappa 계수와 정분류율은 표 8과 같다. 의사결정나무와 로지스틱 회귀분석에 의한 정분류율은 각각 0.7886, 0.7378이고 kappa 계수는 각각 0.4386와 0.4122 이다. 즉 우연히 일치하는 분류 결과가 나올 가능성을 보장하지 않을 때, 실제보다 과대하게 분류 정확도를 추정하게 됨을 알 수 있다. 또한 일차형 weighted kappa 계수는 각각 0.4941와 0.5363, 이차형 weighted kappa 계수는 각각 0.5621과 0.5056을 나타내어 Koch와 Landis 기준에 의하면 둘 다 보통 정도의 분류정확성을 나타낸다고 할 수 있다. 의사결정나무는 로지스틱 회귀분석보다 더 큰 값의 kappa 계수, 일차형 weighted kappa 계수와 이차형 weighted kappa 계수를 보여주어 더 정확하게 분류 기능을 수행하였다.

표 6. 의사결정나무에 의한 오분류표  
Table 6. Confusion matrix by decision tree

|                  |    | 예측 결과 |    |    |    |     |
|------------------|----|-------|----|----|----|-----|
|                  |    | 0     | 1  | 2  | 3  | 총합  |
| 실<br>제<br>상<br>태 | 0  | 354   | 10 | 2  | 0  | 366 |
|                  | 1  | 47    | 29 | 5  | 2  | 83  |
|                  | 2  | 28    | 3  | 15 | 4  | 50  |
|                  | 3  | 2     | 3  | 2  | 5  | 12  |
|                  | 총합 | 431   | 45 | 24 | 11 | 511 |

표 7. 로지스틱 회귀분석에 의한 오분류표  
Table 7. Confusion matrix by logistic regression

|                  |    | 예측 결과 |    |    |    |     |
|------------------|----|-------|----|----|----|-----|
|                  |    | 0     | 1  | 2  | 3  | 총합  |
| 실<br>제<br>상<br>태 | 0  | 350   | 11 | 5  | 0  | 366 |
|                  | 1  | 0     | 12 | 18 | 53 | 83  |
|                  | 2  | 26    | 11 | 12 | 1  | 50  |
|                  | 3  | 4     | 1  | 4  | 3  | 12  |
|                  | 총합 | 380   | 35 | 39 | 57 | 511 |

표 8. 지방간 자료에 대한 분류정확도  
Table 8. Classification accuracy for fat liver data

|           | 의사결정나무 | 회귀분석   |
|-----------|--------|--------|
| 정분류율      | 0.7886 | 0.7378 |
| $K$       | 0.4386 | 0.4122 |
| 일차형 $K_w$ | 0.4941 | 0.4363 |
| 이차형 $K_w$ | 0.5621 | 0.5056 |

한편 의사결정나무와 로지스틱 회귀분석의 분류 결과를 비교하기 위해 통계적 가설검정을 실시하였다. 이 비교를 위해 통상적으로 대응표본 t검정(paired t-test)을 가장 널리 사용하나 이 검정은 제1종 오류(Type I error)를 부풀리는 결함을 갖고 있기 때문에 제1종 오류가 크지 않은 McNemar 검정을 이용하였다. McNemar 검정에 의해 각 기법으로부터 얻은 분류정확도에 차이가 있는지를 가설검정 하였으며 가설은 다음과 같다. (이 때 사용한 분류정확도는 kappa 계수나 weighted kappa 계수가 아닌 오분류된 비율을 사용하였다.)

- $H_0$  : 두 방법에 따른 오분류율에 차이가 없다.
- $H_1$  : 두 방법에 따른 오분류율에 차이가 있다.

McNemar 검정은 귀무가설 ( $H_0$ ) 하에서 기대빈도의 분포를 관찰빈도와 비교하는  $\chi^2$ 을 이용한 적합성 검정에 기초하고 있다. 만일 표 9와 같이 관찰빈도가 주어지면 귀무가설 ( $H_0$ ) 하에서 기대빈도는 표 10과 같이 되므로 McNemar 검정의 검정통계량 식 (5)와 같이 정의된다(9).

표 9. 분할표(관찰빈도)  
Table 9. Contingency table(observed counts)

|          |          |
|----------|----------|
| $n_{xx}$ | $n_{xo}$ |
| $n_{ox}$ | $n_{oo}$ |

여기서

- $n_{xx}$  : 두 방법 모두에 의해 오분류된 수검자 수
- $n_{xo}$  : 의사결정나무에 의해 오분류 됐으나 로지스틱 회귀분석에 의해 옳게 분류된 수검자 수
- $n_{ox}$  : 의사결정나무에 의해 옳게 분류 됐으나 로지스틱 회귀분석에 의해 오분류된 수검자 수

$n_{00}$  : 두 방법 모두에 의해 옳게 분류된 수검자 수이다.

표 10. 분할표(기대빈도)  
Table 9. Contingency table(expected counts)

|                       |                       |
|-----------------------|-----------------------|
| $n_{xx}$              | $(n_{x0} + n_{ox})/2$ |
| $(n_{ox} + n_{x0})/2$ | $n_{00}$              |

$$\frac{(n_{ox} - n_{x0} - 1)^2}{n_{ox} + n_{x0}} \quad (5)$$

귀무가설 ( $H_0$ ) 하에서는 의사결정나무와 로지스틱 회귀 분석 모두 동일한 오분류율을 가지므로 귀무가설을  $H_0: n_{x0} = n_{ox}$  로 쓸 수도 있다.

식 (5)의 통계량은 근사적으로 자유도 1인  $\chi^2$  분포를 따르고 이 통계량의 값이  $\chi^2_{1,0.05} = 3.841$  보다 클 때  $H_0$  을 기각한다.

지방간 데이터의 511명의 수검자에 대해 의사결정나무와 로지스틱 회귀분석 기법에 의한 분류결과로부터 표 9의 분할표와 같이 4가지 경우로 분류하여  $n_{xx} = 87, n_{x0} = 21, n_{ox} = 47, n_{00} = 356$  을 얻었으며 식 (5)에 의해 계산된 McNemar 검정통계량의 값은 9.1912 이다. 이 값은  $\chi^2_{1,0.05} = 3.941$  보다 크고  $H_0$  을 기각하므로 ( $p$  값 = 0.002) 의사결정나무와 로지스틱 회귀분석에 의한 분류 결과의 오분류율에는 차이가 있다고 말할 수 있다. 또한  $n_{ox} = 47 > n_{x0} = 21$  이므로 로지스틱 회귀분석에 의한 오분류율이 의사결정나무에 의한 오분류율보다 크다고 할 수 있으며 의사결정나무 기법이 로지스틱 회귀분석보다 더 정확하다고 말할 수 있다. 이것은 표 8에서 의사결정나무 기법에서 더 큰 kappa 계수와 weighted kappa 계수를 얻은 것과 일치하는 결과이다.

### V. 결론

본 논문에서는 분류 기법에 의해 얻은 분류결과의 정확성을 가장 타당하게 측정해주는 분류정확도를 제안하는 것을 목적으로 한다. 이를 위하여 Cohen의 kappa 계수를 제안하였

으며 이 Kappa 계수는 명목척도와 순서척도의 데이터에 대해 사용되고, 기존의 정분류율이 할 수 없었던 우연히 발생할 수 있는 분류일치를 보정함으로써 분류정확도를 계산한다는 장점을 갖고 있다. 특히 순서척도의 데이터의 경우 분류 일치 정도에 따라 차등적으로 가중치를 부여하여 kappa 계수를 측정해 주는 weighted kappa 계수를 사용하였다.

예시적으로 사용한 지방간 데이터의 경우, 의사결정나무와 로지스틱 회귀분석의 두가지 기법으로 얻은 분류 결과에서는 의사결정나무 기법이 로지스틱 회귀분석보다 더 큰 값의 kappa 계수나 weighted kappa 계수를 제시하였고 McNemar 검정에 의해 의사결정나무가 더 정확하게 분류 기능을 수행함을 통계적으로 입증하였다.

결과적으로 Cohen의 kappa 계수와 weighted kappa 계수는 분류정확도를 측정하는데 있어 단순히 빠르게 분류된 빈도를 세는 것보다 더 유용하다고 할 수 있다.

향후에는 분류 기법들을 비교하고 분류정확도를 측정하기 위한 보다 효과적인 통계치와 가중치에 대한 연구가 요구된다. 특히 순서척도의 데이터의 경우에는 일차형 weight, 이차형 weight를 포함하여 오분류 크기를 세밀히 나타낼 수 있는 가중치 연구가 필요할 것이다.

### 참고문헌

- [1] J. Hollmen, M. Skubacz, and M. Taniguchi, "Input Dependent Misclassification costs for cost-sensitive classifiers". In N. Ebechen & C. Brebbia (Eds.), Data Mining II - Proceedings of the Second International Conference on Data Mining (pp. 495-503) Cambridge, MA: MIT Press, 2000.
- [2] D. D. Margineantu, and T. G. Ditterich, "Bootstrap Methods for the Cost-sensitive Evaluation of Classifiers", In: Proceedings of the Seventh International Conference on Machine Learning, Morgan Kaufmann (pp. 582-590), 2000
- [3] J. Cohen, "A Coefficient of Agreement for Nominal Scales", Educational and Psychological Measurement, 20, 37-46, 1960.
- [4] J. Cohen, "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit", Psychological

- Bulletin, 70, 213-220, 1968.
- [5] J. L. Fleiss, "Statistical Methods for Rates and Proportions (2nd ed.)", New York: Wiley, 1981.
  - [6] J. J. Barko, and W. T. Carpenter, "On the Methods and Theory of Reliability", The Journal of Nervous and Mental Disease, 163, 307-317, 1976.
  - [7] J. R. Landis, and G. G. Koch, "Measurement of Observer Agreement for Categorical Data", Biometrics, 33, 159-174, 1977.
  - [8] H. S. Bae, et. al. "SAS Enterprise Miner, Data Mining", Kyowoo Co. 2004.
  - [9] B. S. Everitt, "The Analysis of Contingency Tables", London: Chapman and Hall, 1977.

### 저자 소개



#### 엄 용 환

1990 : Emory University,  
Department of  
Biostatistics, 통계학석사.

1995 : University of Florida,  
Department of Statistics,  
통계학박사.

현 재 : 성결대학교  
산업경영공학부 교수

관심분야 : 통계추론, 데이터마이닝.

Email : uyh@sungkyul.ac.kr