IJACT 16-2-6

# Predicting Discharge Rate of After-care patient using Hierarchy Analysis

Yong Gyu Jung[1], Hee-Wan Kim[2], Min Soo Kang[3,+]

[1,3] *Dept. of Medical IT Marketing, Eulji University, Korea*
*ygjung@eulji.ac.kr, mskang@eulji.ac.kr(Corresponding Author[+])*

[2] *Div. of Computer Science, Shamyook University, Korea*
*hwkim@syu.ac.kr*

## Abstract

*In the growing data saturated world, the question of "whether data can be used" has shifted to "can it be utilized effectively?" More data is being generated and utilized than ever before. As the collection of data increases, data mining techniques also must become more and more accurate. Thus, to ensure this data is effectively utilized, the analysis of the data must be efficient. Interpretation of results from the analysis of the data set presented, have their own on the basis it is possible to obtain the desired data. In the data mining method a decision tree, clustering, there is such a relationship has not yet been fully developed algorithm actually still impact of various factors. In this experiment, the classification method of data mining techniques is used with easy decision tree. Also, it is used special technology of one R and J48 classification technique in the decision tree. After selecting a rule that a small error in the "one rule" in one R classification, to create one of the rules of the prediction data, it is simple and accurate classification algorithm. To create a rule for the prediction, we make up a frequency table of each prediction of the goal. This is then displayed by creating rules with one R, state-of-the-art, classification algorithm while creating a simple rule to be interpreted by the researcher. While the following can be correctly classified the pattern specified in the classification J48, using the concept of a simple decision tree information theory for configuring information theory. To compare the one R algorithm, it can be analyzed error rate and accuracy. One R and J48 are generally frequently used two classifications...*

*Keywords: one R, J48, Hospital Inpatient, Discharges, County Residence, SPARCS*

## 1. Introduction

Recently, Smartphone has been popularized, young children from children as there are few people that use the 2g phone to grandparents many people are using Smart phones. Although sent by pawl 60 characters tight even send or previous character, due to recently been many data capacities of 3g and 4g plan everyone

cocoa talk SNS, it could not be a lot of data, was able to produce. Now, I became very many countless data that have been produced in a day. Therefore, I became a large number of these data so called big data. Big Data recently too are important, are utilized in various fields. Such in order to take advantage of big data, it is necessary first to use data mining techniques to the analysis that it is necessary to analyze the data. Data mining, by analyzing the relationships and characteristics of each attribute what data has, is intended to take the portion of data that they need. If there are fewer number of data, it might obtain a simple analytical results, hundreds, if there are thousands of data, it cannot all be obtained handmade results. With the development of data mining technique to apply the Mobile CRM design and implementation and information and computer technology, has come out full of data in various fields in recent years. The rapid development of information technology, to promote the automation of operations, is it possible to store, collect vast amounts of data, electronically the amount of collecting data is increased every year exponentially, is accumulated there. When such collected data is not processed in the form of information, if the amount of it does not have any meaning by itself clustering not enormous, experts simple statistical methods and questions and the data were analyzed through. These analytical methods, the efficiency decreases as the amount of data to be increased, brought to limit the information that can be obtained through the analysis.

## 2. Experimental Data

Data that was starting with the 2009 from SPARCS, a Hospital Inpatient Discharges by Patient County of Residence SPARCS Beginning 2009.csv data comparing the hospital patients and the autonomous regions of hospitalized patients in the patient's home that were investigated. In hospitals, it is possible to use a lot to have a data comparing hospital patients and those hospitalized patients. Data used in this experiment were used by selected 104 pieces of the total of 3654 pieces. Prior to data analysis, and visualization of the attributes of the data in the pre-treatment process, it is possible to view the attributes of the data. The experimental data, housing patients of residence (Patient County of Residence), gender (Patient Gender), number of hospital discharge patients (Number of Discharges), age (Patient Age Group), discharge year (Discharge Year), length of hospital stay (Average Length of Stay), 5 one of the numeric properties and discharge year, such as hospital rankings (Yearly Rank by Volume) in fiscal 2012, is a variable, resident of gender, female, it is a variable limited to male, residence name is a detailed description of each property to varchar attributes are shown in Table 1 below.

**Table 1. Properties of the experimental data**

| Attribute | Type | Value |
|---|---|---|
| Patient County of Residence | Nominal | {Bronx, Erie, Kings (Brooklyn), Monroe, Nassau, New York (Manhattan), Queens, Richmond (Staten Island), Suffolk, Westchester} |
| Patient Gender | Nominal | {Female, male, unknown} |
| Patient Age Group | Nominal | {13-17, 18-40, 41-64, 65-74, 75+} |
| Discharge Year | Nominal | {2012} |
| Number of Discharges | Numeric | Continuous from 1 to 64870 |
| Average Length of Stay | Numeric | Continuous from 1 to 17 |
| Yearly Rank by Volume | Numeric | Continuous from 1 to 10 |

The following is first load the prepared data Set in CSV or arff file. Click the upper left of the open file, and then load the data set that had been prepared. When loading the data, as shown in Figure 1, the middle portion of the left, the data attribute to the right of the middle portion out, and label number and frequency attributes come out. Then, in the lower right, it is possible to see the attributes of the data easily at a glance out is visualized in the graph screen.
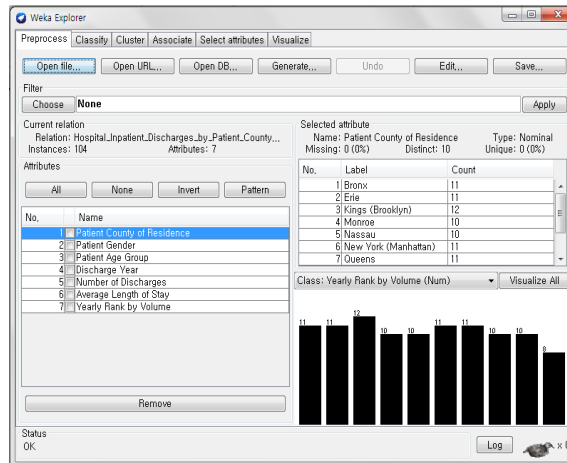


**Figure 1. Up the data on the screen**

In addition, of the attributes in the left center, properties required in addition to the attributes of the data he or she desired, can be removed using the remove button. In this study, select the one R to the analysis by using the one R using the rules for us to close. For the analysis of the following algorithm in the test option in Figure 1 it is to compare in succession to one another, Folds and me to select Cross-validation gives about 30. Then, by selecting from among the attributes of either (Nom) yourself in the middle section to analyze the reference of any attribute, pressing following the start button, the analysis of the data is started.

## 3. Experimental Results

After the analysis is finished, willing to progress as well J48 techniques. In other respects, willing to close by selecting the J48 from the trees in the Choose items. The next, when the classification attributes of the data me selected in J48 in Figure 2 press the start button, is the analysis of data. In this experiment, on the basis of data for analyzing the number of discharges who depending on the age, first by using the one R classification method rules, and tried to classify large. The properties of the data analysis, I went given the fold value of Cross Validation to 30. It is a simple structure than the hospital's number in accordance with the reason residence that was me organize the age data in the process of data pre-processing. The Confusion Matrix is the result of analysis can be seen in Figure 2.

```
 a  b  c  d  e   <-- classified as
20  0  0  0  0 |  a = 13-17
 2  1  4 11  4 |  b = 18-40
 2  6  6  5  3 |  c = 41-64
 1  3  0 15  2 |  d = 65-74
 1  3  4  7  4 |  e = 75+
```

**Figure 2. Hospital's number Confusion Matrix in accordance with the age**

The following is the result of the analysis in the classification method of J48. Figure 3 is a representation of the number of discharges who in accordance with the age J48 classified Confusion Matrix.

```
 a  b  c  d  e   <-- classified as
19  0  0  1  0 |  a = 13-17
 0  0 10  9  3 |  b = 18-40
 1 10  5  1  5 |  c = 41-64
 0 10  0  5  6 |  d = 65-74
 0  4  6  7  2 |  e = 75+
```

**Figure 3. Hospital's number Confusion Matrix in accordance with the age**

To see the test results to date, the number of discharged who, according to age clearly also was easy reference is classified, the matrix tree was also more easy, it was more accurate. Therefore, in the evaluation of results, try more of the evaluation of several that have a matrix of age. The next, he saw the results of the tree and the matrix with additional Visualize margin curves. Visualize margin curve is a little characteristic is significantly appear than J48 algorithm.
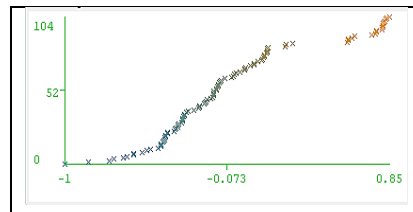


**Figure 4. Visualize margin curve of one R**

## 4. Conclusion

Recently, rather than to use what data, it must have a question of whether that data can be how effectively utilized. Even more data generated day in situations that are utilized by moving stored, as so many data increases, data mining techniques have also become more and more accurate. That's why if it is possible to effectively utilize the weka such data analysis program, it will be able to more easily analyze the data. Analyzes the results of the analysis data in the following that were analyzed, have their own on the basis it is possible to obtain the desired data.

In the data mining method, a decision tree, clustering, there is such a relationship has not yet been fully developed algorithm actually still impact of various factors.

In this experiment, the classification method of many data mining techniques, to Ogido blows with easy decision tree or decision tree in the decision tree. Also, I was using a special technology one R and J48 classification technique in the decision tree. After selecting a rule that a small error in the "one rule" in one R classification, to create one of the rules of the prediction data, it is simple and accurate classification algorithm. To create a rule for the prediction, we make up a frequency table of each prediction of the goal. It is displayed by creating precise rules than one R classification state-of-the-art classification algorithm while creating a simple rule to interpret human. While the following can be correctly classified the pattern specified in the classification J48, using the concept of a simple decision tree information theory for configuring information theory disorder also based on the entropy. And it is trying to compare the one R algorithm and error rate and accuracy come to know more the algorithm of J48 through this experiment. One R and J48 are generally frequently used two classifications.

## References

[1]  KHARADE, Kalpana; THAKKAR, Ms Rupal. Promoting ICT enhanced constructivist teaching practices among pre-service teachers: A case study.International Journal of Scientific and Research Publications, 2012, 2: 1-7.

[2]  DOGRA, Ashish Kumar; WALA, Tanuj. A Comparative Study of Selected Classification Algorithms of Data Mining. 2015.

[3]  SAXENA, Ritika. Educational Data Mining: Performance Evaluation of Decision Tree and Clustering Techniques Using WEKA Platform. International Journal of Computer Science and Business Informatics, IJCSBI. ORG, 2015, 15.2.

[4]  SATHYA, R.; GEETHA, M. Kalaiselvi. Vision Based Traffic Personnel Hand Gesture Recognition Using Tree Based Classifiers. In: Computational Intelligence in Data Mining-Volume 2. Springer India, 2015. p. 187-200.

[5]  WAGH, Sharmila, et al. Effective Framework of J48 Algorithm using Semi-Supervised Approach for Intrusion Detection. International Journal of Computer Applications, 2014, 94.12.