International Journal of Advanced Culture Technology Vol.7 No.2 238-243 (2019)
DOI 10.17703/IJACT.2019.7.2.238

IJACT 19-6-29

Clustering for Home Healthcare Service Satisfaction using Parameter Selection

Jae Hong Lee[1], Hyo Sun Kim[2], Yong Gyu Jung[3], Byung Heon Cha[4*]

[1]CEO of Ubivelox, Dept. of {[2,3]Medical IT, [4]Biomedical laboratory science} Eulji University

E-mail: jhlee@ubivelox.com, {20180737, jabogy, ygjung}@eulji.ac.kr

Abstract

Recently, the importance of big data continues to be emphasized, and it is applied in various fields based on data mining techniques, which has a great influence on the health care industry. There are many healthcare industries, but only home health care is considered here. However, applying this to real problems does not always give perfect results, which is a problem. Therefore, data mining techniques are used to solve these problems, and the algorithms that affect performance are evaluated. This paper focuses on the effects of healthcare services on patient satisfaction and satisfaction. In order to use the CVParameterSelectin algorithm and the SMOreg algorithm of the classify method of data mining, it was evaluated based on the experiment and the verification of the results. In this paper, we analyzed the services of home health care institutions and the patient satisfaction analysis based on the name, address, service provided by the institution, mood of the patients, etc. In particular, we evaluated the results based on the results of cross validation using these two algorithms. However, the existence of variables that affect the outcome does not give a perfect result. We used the cluster analysis method of weka system to conduct the research of this paper.

Keywords *: Classification, Cross validation, Parameter selection,Home health care, Patient satisfaction*

1. Introduction

Recently, the importance of big data continues to be emphasized, and it is applied in various fields based on data mining techniques, which has a great influence on the health care industry. Now, with the development of information and communication technology, it is a rapidly changing era. The choice and focus for securing the competitiveness of the Big Data era, page 1) Due to the rapid development of information and communication technology, the Internet is changing the whole social field, and according to the wide variety of usage patterns, . Most Internet data has been repeatedly generated and destroyed except for limited use. Recently, interest in Internet data is rising due to the emergence of a key keyword called big data. In the past, efforts have been made to create value added data bases through data mining, business intelligence, and life logs. In the past, these efforts have been many, but the value of big data is getting bigger because the importance of securing and utilizing data is emphasized. In addition to the emphasis on big data emphasized in all areas of society, the scope of utilization of big data is also widening in the health care industry as well. This paper investigates the attributes and the patient satisfaction of the home health care organization, which is one of the healthcare industry using Big Data, to contribute to patient satisfaction.

Manuscript received: May 13, 2019/ revised: May 20, 2019 / Accepted: May 27, 2019
Corresponding Author: jabody@eulji.ac.kr
Tel:+82-10-4272-3094, Fax: +82-31-740-7114
Author's affiliation: Biomedical laboratory science, Eulji University

Home health care refers to providing medical care within a patient's home before a medical institution is admitted to the hospital and becomes ill or in a state where the patient is hospitalized. It can be seen that the importance of home healthcare institutions is emphasized in this age of increasingly aging society. In this paper, we use cluster analysis method of Weka system based on big data of health field. We also analyzed the attributes of these factors and evaluated the home healthcare service and patient satisfaction.

2. Research methods

2.1 Item set formation for clustering

For each item, the number of occurrences in the data set is given to one side of the table. These numbers are a set of items. The next step is to pair up a set of items to form a set of two items. Of course, it is not important to create a set containing the same properties with different values, since it can not happen in any real instance. It can be the case to find association rules with a minimum coverage. Therefore, we discard the set of items that contain instances smaller than two instances. It created a set of two items, some of which were recorded in the second column along with the number of occurrences. The next step is to create a set of three items, with the set having a large coverage. Next, there are six sets of four items, and there is no set of five items.

It can be explained how to efficiently create such a set of items. Once all the sets of items with the desired coverage are created, the next step is to convert them one by one into at least a set of rules or rules with a predefined minimum accuracy. Some item sets create one or more rules. Other sets do not generate any rules. Now, let's deal with an association rule generation algorithm that has a specific minimum coverage and accuracy in more detail. This algorithm is divided into nine phases. We create an item set with a certain minimum coverage and determine the rule with the minimum accuracy from each set of items.

The first step is to create a set of items of a single element with a given minimum coverage. Next, we increase the size of the set, such as the set of elemental items and the set of three items, one by one. The step-by-step action involves browsing the data set, counting the number of occurrences of items in each set of items, and then storing the set of items remaining in the hash table until the end. It creates these elementary item set candidates from a single elemental item set, then scans the data set and counts the coverage of each of these elemental item sets. Finally, candidate sets that do not meet the minimum coverage are removed from the hash table. These element candidate sets are simply composed of pairs of single elements. This set of elements does not satisfy the minimum coverage if two single element sets constituting itself do not satisfy the minimum coverage. This is a common part of a larger set of items. In other words, the triple subset satisfies the minimum coverage only if all three of its constituent elements have the minimum coverage, which is the same for the element set as well.

2.2 Parameter selection for cross validation

This method is a class that performs parameter selection by cross validation for a certain class. This classifier can be optimized through any number of parameters, and only the default classification of one nested option can not optimize the direct option.

Table 1. General Characteristics of Subjects

Subjects	Characteristics
CVParameters	Set the planning parameters that can be set by cross validation. The format of each string should be param_char LOWER_BOUND UPPER_BOUND number_of_steps To retrieve the parameter -P from 1 to 10 in units of 1
Debug & Folds	Console In fact, set the category to output additional information numFolds.
	Gets the number of wrinkles used for cross validation.

3. Experiment

3.1 Experimental data

The home health care institution refers to an institution that provides medical care within a patient's home before a medical institution is admitted to a hospital and becomes unwell. These services include visiting the patient's home on time, prescribing medication on time, explaining the medication to the patient, and treating the patient physically. The data used are healthcare.arff on services, patient satisfaction of home healthcare institutions investigated in the United States. The data used in the experiments consisted of 12 numeric attributes and 6 string attributes, and a total of 1,164 data items.

Relation: healthcare

No.	Nursing_Care String	Physical_Therapy String	Occupational_Therapy String	Speech_Pathology String
1	TRUE	TRUE	TRUE	TRUE
2	TRUE	TRUE	TRUE	TRUE
3	TRUE	TRUE	TRUE	FALSE
4	TRUE	TRUE	TRUE	TRUE
5	TRUE	TRUE	TRUE	TRUE
6	TRUE	TRUE	TRUE	TRUE
7	TRUE	TRUE	TRUE	TRUE
8	TRUE	TRUE	TRUE	TRUE
9	TRUE	TRUE	TRUE	TRUE
10	TRUE	TRUE	TRUE	TRUE
11	TRUE	TRUE	TRUE	TRUE
12	TRUE	TRUE	TRUE	TRUE
13	TRUE	TRUE	TRUE	TRUE
14	TRUE	TRUE	TRUE	TRUE
15	TRUE	TRUE	TRUE	TRUE
16	TRUE	TRUE	TRUE	TRUE
17	TRUE	TRUE	TRUE	TRUE
18	TRUE	TRUE	TRUE	TRUE
19	TRUE	TRUE	TRUE	TRUE
20	TRUE	TRUE	TRUE	TRUE
21	TRUE	TRUE	TRUE	TRUE
22	TRUE	TRUE	TRUE	TRUE
23	TRUE	TRUE	TRUE	TRUE
24	TRUE	TRUE	TRUE	TRUE

Figure 1. A part of experimental data

3.2 Preprocess

The numeric attribute is an integer from 0 to 100. The attributes from timely_manner to check_pain indicate the service execution rate of the home healthcare institution. The patients_bathing and patients_hospital attributes are closer to 0, and the patient's dissatisfaction, . It did not seem to matter whether the string attribute, TRUE or FALSE, is a type of service provided by a home healthcare organization. Data was evaluated and visualized with six numeric attributes, excluding the string attribute, for each instance. The data used are healthcare.arff on services, patient satisfaction of home healthcare institutions investigated in the United States.

3.3 Experimental result

We used the CVParameterSelection and SMOreg described above as the clustering algorithms for the numeric properties based on healthcare.arff data. Figure shows a visualization of all string and numeric attributes. There are two types of string attributes: TRUE and FALSE. The left is the number of TRUEs for the string attribute, and the right is the number of FALSEs. The numeric attribute displays the minimum to maximum values for each attribute and shows the mean and variance values.

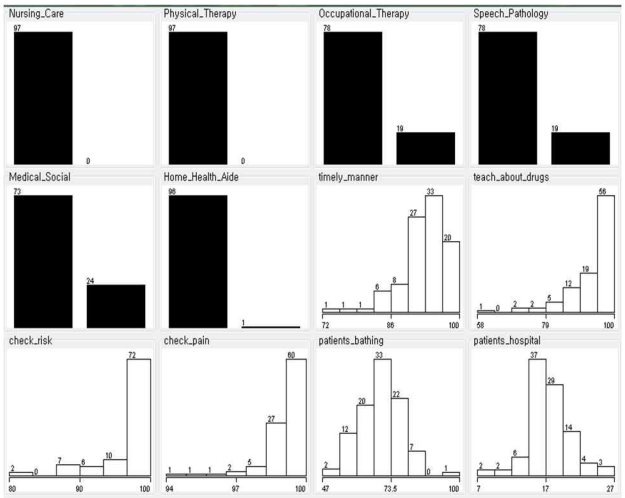

Figure 2. Experimental result

4. Discussion

Figure 3 shows the CVParameter Selection output through experiment. There are 97 instances and 12 attributes. The names of six numeric attributes have been identified and created. The classifier errors is visualized like the result as shown below.

```
=== Run information ===

Scheme:weka.classifiers.meta.CVParameterSelection -X 10 -S 1 -W weka.classifiers.rules.ZeroR
Relation:     healthcare-weka.filters.unsupervised.attribute.Remove-R1-6
Instances:    97
Attributes:   6
              timely_manner
              teach_about_drugs
              check_risk
              check_pain
              patients_bathing
              patients_hospital
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Cross-validated Parameter selection.
Classifier: weka.classifiers.rules.ZeroR
Classifier Options:

ZeroR predicts class value: 17.65979381443299

Time taken to build model: 0 seconds

=== Cross-validation ===
```

Figure 3. Output of CV Parameter Selection

The below figure shows the patient satisfaction with the services of the home healthcare institution. The X axis is patienrs_bathing and the Y axis is patients_hospital. If it is satisfied with bathing, we could find to have

fewer hospital admissions. The total number of string and numeric attributes add up to 12, but only the numeric attribute is handled and the cluster number is set to 6 in the cluster option.

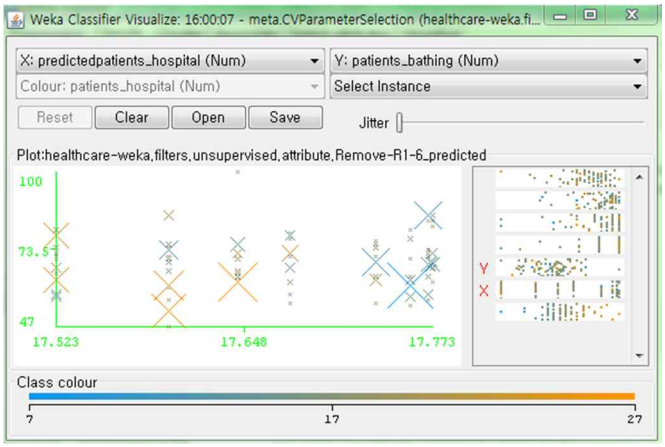

Figure 4. Visualization for CV Parameter Selection

5. Conclusion

Recently, the importance of big data has been emphasized, and the securing and utilization of information has been emphasized. Based on this vast amount of data, extracting useful information is called data mining, which we can use in various fields. In the previous study, we measured the satisfaction of patients receiving services based on US home health care agency service data. Here, the home health agency's service means that the medical staff stops at the patient's home and provides the service to the patient. The six numeric attributes of timely_manner, teach_about_drugs, check_risk, check_pain, patients_bathing and patients_hospital are provided by the home healthcare organization. The service type Nursing_Care, Physical_Therapy, Occupational_Therapy, Speech_Pathology, Medical_Social, and Home_Health_Aide. The service provided by each institution is divided into TRUE and FALSE, which concludes that it does not affect patient satisfaction. Home health care institutions, such as when they come in time, the right to prescribe the drug attributes, such as a percentage of 100 gave a rating. Patients' satisfaction was related to their satisfaction with the bath provided by the home healthcare organization and the degree of hospitalization of the patient. The higher the satisfaction of bathing, the less the percentage of hospitalized patients, which means the satisfaction of home health care services.

In an age of increasing aging, the services of these home healthcare organizations will become more and more important. In addition, the development of data systems and information and communication technologies are expected to help the healthcare industry.

In this paper, we evaluate clustering algorithms in more detail, and in the future, we will study how to apply multiple models in a variety of ways.

5. Acknowledgment

The research and writing of this article were made possible by funding from the “Development on the BigData Analysis platform technology for IoT Home appliance product of small and medium sized home appliance company (No,20000195)”

References

[1] Alwadi, Mohammad, Girija Chetty, and Mohammad Yamin. "A Virtual Sensor Network Framework for Vehicle Quality Evaluation." 12th INDIACom: 2018 5th International Conference on Computing for Sustainable Global Development. BVICAM, 2018.

[2] Castelli, Mauro, Luca Manzoni, and Aleš Popovič. "An artificial intelligence system to predict quality of service in banking organizations." Computational intelligence and neuroscience 2016 (2016).

[3] Doukas, Charalampos N., and Ilias Maglogiannis. "Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components." IEEE Transactions on Information Technology in Biomedicine 15.2 (2010): 277-289.

[4] Doukas, Charalampos, and Ilias Maglogiannis. "Advanced classification and rules-based evaluation of motion, visual and biosignal data for patient fall incident detection." International Journal on Artificial Intelligence Tools 19.02 (2010): 175-191.

[5] Kolesnikova, Olga, and Alexander Gelbukh. "Binary and Multi-class Classification of Lexical Functions in Spanish Verb-Noun Collocations." Mexican International Conference on Artificial Intelligence. Springer, Cham, 2017.

[6] Rani, A. Swarupa, and S. Jyothi. "Performance analysis of classification algorithms under different datasets." 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2016.

[7] Selvakuberan, K., M. Indradevi, and R. Rajaram. "Combined Feature Selection and classification−A novel approach for the categorization of web pages." Journal of Information and Computing Science 3.2 (2008): 083-089.

[8] Sung, Sheng-Feng, et al. "Developing a stroke severity index based on administrative data was feasible using data mining techniques." Journal of clinical epidemiology 68.11 (2015): 1292-1300.