

Association Analysis of Parkinson's Disease using Apriori Algorithm

Yong Gyu Jung[†], Oh Jin Kim^{*}, Jae Kang Won^{**}

[†]Dept. of Medical IT Marketing, Eulji University

^{*}School of Medical Industry, Eulji University

^{**}Department of Computer Science, Kyonggi University, Corresponding Author

Abstract

Parkinson's disease is representative degenerative diseases of the nervous system, which is from deficiency of dopamine neurons to pass in which the gradual degeneration of the body. In this paper, open UCI repository data of Parkinson's patients is used for experiments. The classification based on correlation analysis is examined. In addition, the relationship between groups is differentiated by cluster analysis based on patients with Parkinson's disease by apriori algorithm and correlation analysis. It is used to find the properties that distinguish cluster analysis. Though the disease is the same in the basic structure, each group is compared as each gender group with the most distinctive part of the characteristics.

Key words : Apriori algorithm, Parkinson's patients, association analysis

1. INTRODUCTION

Parkinson's disease is representative degenerative diseases of the nervous system, which is from deficiency of dopamine neurons to pass in which the gradual degeneration of the body. [1] Physical condition in Parkinson's patients is appeared the fault of ability to adjust the movement of breathing, phonation, articulation affects. Limb muscle rigidity, resting progress (3-6Hz, move the disappearance), slow movement (bradykinesia), or the position of the characteristic symptoms of instability if there are two kinds of Parkinson's syndrome can be diagnosed. As the disease progresses in addition, a variety of movement disorders may occur. For example, micrographia, reduced facial expression, and the number of Blink, shake reduction in walk, foot dragging, hypophonia. [2]

In this paper, experimental data set is used as the UCI open repository data. Patients with Parkinson's disease through linkage analysis of the results obtained by the cluster analysis based on group characteristics were analyzed. And their associated methods of analysis to analyze the relationship see any linkage analysis using apriori algorithm that can understand the structure and association, it was constructed

on the basis of the relationship.

The research is differentiated for Parkinson's disease, while Parkinson's disease took most of the existing research focuses on the presence kept. With previous studies depending on the status of patients with symptoms of Parkinson's disease were able to determine only the presence. In this study, patients already suffering from Parkinson's disease has information about the characteristics of the patient to extract more granular classification of patients were studied.

2. RELATED RESEARCH

2.1 Association Rules

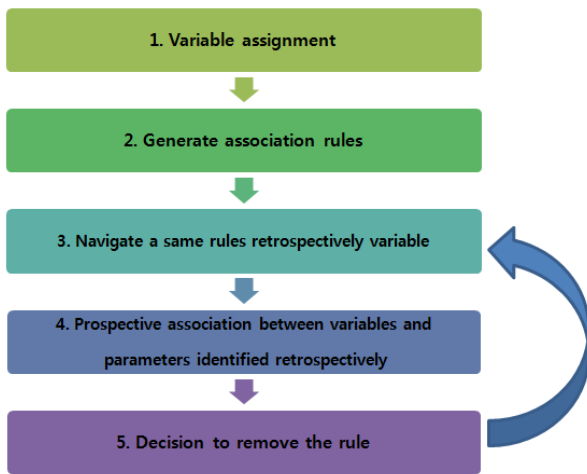
Association rule is simply the study to accompany any item with another item. The affinity analysis is called, for example, medical researchers with certain symptoms of illness. In addition, the analysis of original database is studied while customer transactions for the purchase of different items, as a way to determine the dependence associated with applying the rules like market basket analysis [3]. The steps to perform association rule is shown in Figure 1 below.

Manuscript received Jan. 26, 2012; revised Feb.13, 2012

^{**}Corresponding Author: formida@paran.com

Tel: +82-31-740-7178, Fax: +82-31-740-7251

Dept. of Computer Science, Kyonggi Univ., Korea



[Figure 1] The steps association rules

2.2 Apriori Algorithm

A transaction T is defined as a subset of the I, k, given a set of entries made into, if it contains all the items of X, a transaction T is a set of items support and confidence. Support and confidence are defined as items A and B respectively, if it satisfies the association rule is defined as follows:

Support (support): $Sup(A \rightarrow B) = P(A \cup B)$
 Reliability (confidence): $Conf(A \rightarrow B) = P(B|A) = \frac{P(A \cap B)}{P(A)}$
 Association rules $A \rightarrow B [Sup = s\%, Conf = c\%]$

Including the items A and B simultaneously, and the transaction s%, A and B of the transaction containing the same time. It means that the transaction containing the c%. In association rules the pre-specified minimum support threshold and minimum confidence threshold are calculated to satisfy both the rules with two items. In order to organize all the association rules among a set of items, first support and confidence from them are given the association rules based on the items. [4]

3. EXPERIMENTS

The Parkinson's UCI data set was used. Parkinson's data has 26 properties, as follows:

<Table 1> attribute and descriptions

Attribute	Description
subject#	Integer that uniquely identifies each subject
age	Patient's age

sex	Patient's gender '0' - male, '1' - female
test_time	Time since recruitment into the trial. The integer part is the number of days since recruitment.
motor_UPDRS	Clinician's motor UPDRS score, linearly interpolated
total_UPDRS	Clinician's total UPDRS score, linearly interpolated
Jitter(%) Jitter(Abs) Jitter:RAP Jitter:PPQ5 Jitter:DDP	Several measures of variation in fundamental frequency
Shimmer Shimmer(dB) Shimmer:APQ3 Shimmer:APQ5 Shimmer:APQ11 Shimmer:DDA	Several measures of variation in amplitude
NHR HNR	Two measures of ratio of noise to tonal components in the voice
RPDE	A nonlinear dynamical complexity measure
DFA	Signal fractal scaling exponent
PPE	A nonlinear measure of fundamental frequency variation

In cases of this research, we apply the association rules the effective properties, which determined from the mainstream. It was deleted 85% of a property is concentrated upon them to show the same results.

First of each property (except sex) of the maximum and minimum values are adding dividing 4, which is a number of type of classification. Then divided by the minimum value and plus a value less than the values of things data type1 is to specify the value of two data type 1 below that value plus the value of dividing data type 2, Data type2 less than or equal to the value of the sum divided by the value of those of the data type 3, Data type 3 value plus the value divided by the value of Max than what data type 4 determined. To summarize this process, the formula is shown below.

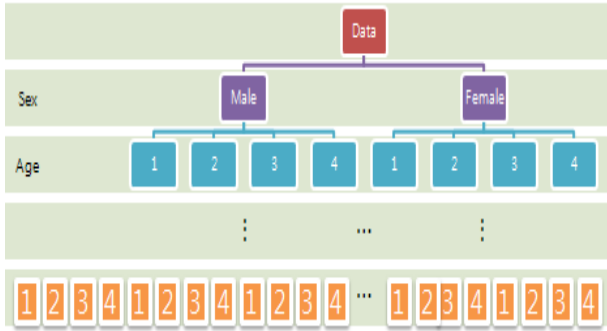
$$\begin{aligned}
 DataType1 &\leq Min + \frac{\{Max(x) - Min(x)\}}{4} \\
 DataType2 &\leq DataType1 + \frac{\{Max(x) - Min(x)\}}{4} \\
 DataType3 &\leq DataType2 + \frac{\{Max(x) - Min(x)\}}{4} \\
 DataType4 &\leq DataType3 + \frac{\{Max(x) - Min(x)\}}{4} = Max
 \end{aligned}$$

Each data was classified according to the attributes of each property determined by the data type. Classified information is shown in table 2 as below.

<Table 2> the data type of the property by DataType

Attribute	DataType
Sex	1(Male), 2(Female)
Age	1, 2, 3, 4
Test_time	1, 2, 3, 4
total_UPDRS	1, 2, 3, 4
HNR	1, 2, 3, 4
RPDE	1, 2, 3, 4
DFA	1, 2, 3, 4
PPE	1, 2, 3, 4

The apriori algorithm is used to classify data types. Using the Microsoft Excel, the results were analyzed. It is used apriori algorithm first and then classified the data by gender, age, aim to target communities. Tthe corresponding data were analyzed to split after the first selection as Figure 1 as below.



[Figure 1] Apriori classification algorithm.

The result of this data is summarized as below shown in Figure 2.

sex	age	test_Time	total_UPDRS	HNR	RPDE	DFA	PPE	결과
1	1							0
1	2							1277
1	3							1873
1	4							858
4	1							101
4	2							655
4	3							573
4	4							538
1	2	1						334
1	2	2						399
1	2	3						378
1	2	4						166
1	3	1						518
1	3	2						572
1	3	3						560
1	3	4						223
1	4	1						244
1	4	2						250
1	4	3						227
1	4	4						177

Figure 2: Examples of how the algorithm handles

4. EXPERIMENTAL RESULTS

In this paper, five kinds of data types were analyzed using the property. Experimental analysis of the output is shown below.

sex	age	test_time	total_UPDRS	HNR	RPDE	DFA	PPE	S_number
1	2			2	3		2	260
1	2			2		3	2	265
1	2		2	3			1	269
1	2			2	3	3		281
1	2		2	3	2			299
1	2			3	2		1	335
1	2				3	3	2	337
1	3			3	2	2		375
1	3			3	2		2	428
1	3			3	2		1	538
1	4		3	3		2		183
1	4			2	3		2	187
1	4			3	2	2		202
1	4			3		2	1	213
4	1	2	1	3				21
4	1			3	3		1	23
4	1	1			2		1	24
4	1		1		3		1	24
4	1	3	1				1	25
4	1	3		3			1	25
4	1		1	3	3			25
4	1	3	1	3				27
4	1				2	1	1	27
4	1		1		2	1		28
4	1			3	2	1		28
4	1	1	1		2			29
4	1	1		3			1	29
4	1	1	1				1	30
4	1				2	2	1	32
4	1	1	1	3				35
4	1			3		1	1	39
4	1		1	3		1		41

Figure 3: Experimental results

This analysis is used as the data from associated with each age that corresponds with the association of more than 20% in the table as below.

<Table 3> experimental result using threshold

Sex	Age_Type	Number	Standard Number
Male	1	0	0
Male	2	1277	256
Male	3	1873	375
Male	4	858	172
FeMale	1	101	21
FeMale	2	655	131
FeMale	3	573	115
FeMale	4	538	108

In this table, statistics based on each person's age, type of property is greater than or equal to the number of data to analyze the following emerges.

<Table 4> numbers and types of attributes

Sex	Age	Test_time	Total_UPDRS	HNR	RPDE	PPE	%
1(M)	2~4		2or3	2or3	2or3	1or2	27
4(F)	1~4	1~3	1or2	3	2or3	1	42

According to this analysis table, age and gender are the most affected factors as known. Based on these results, the data shows the differences between the values of plates from gender discrimination in the direction reanalyzed. First, the gender differences seen in the best thing about a type of male age 48-year that it does not exist in patients with Parkinson's disease, reflecting the male over the age of 49-year was based only. When compared to both these conditions are applied differently to each side, plus the conditions for the analysis concluded.

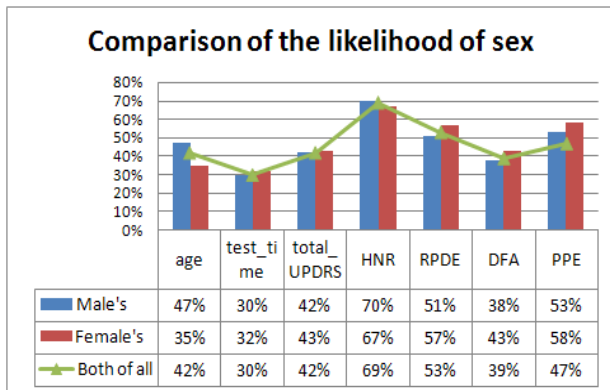


Figure 4 Likelihood compared by genders

As a result did not change in the additional condition, It makes up each group not only by the inclusion in any condition.

5. CONCLUSIONS

In this paper, apriori algorithm and correlation analysis were used to find the properties that distinguish cluster analysis. Though the disease is the same in the basic structure, each group is compared as each gender group with the most distinctive part of the characteristics. For the analysis of survey results because it was like most of the context, but each has different characteristics in terms of the instrument was capable of analyzing. In other words, It was an

opportunity to compare and quantify, even confined to the relationship including other additional outliers.

REFERENCES

- [1] N. H. Kim, "Rotor fault detection system for inverter driven induction motor using current signals and an Encoder," *Journal of the Institute of Webcasting, Internet and Telecommunication(IWIT)*, Vol. 10, No. 5, pp. 128-135, Oct. 2010.
- [2] F. A. Huliehel, F. C. Lee, and B. H. Cho, "Small-signal modeling of the single-phase boost high power factor converter with constant frequency control," in *Proc. PESC*, pp. 475-482, 2011.
- [3] T. J. E. Miller, *Reactive Power Control in Electric Systems*, John Willey&Sons, chap. 2, 2009
- [4] Myway labs -PSIM- <http://www.myway-labs.co.jp/psim>, May 13th 2011.
- [5] S. K. Dwivedi, "Power Quality Improvements and Sensor Reductions in Permanent Magnet Synchronous Drives," *PhD. Thesis*, IIT Delhi, 2006.
- [6] Yong-Gyu Jung, Seung-Ho Lee and Ho Joong Sung, Effective Diagnostic Method Of Breast Cancer Data Using Decision Tree, *Journal of IWIT* (2010), Vol.10 No. 5 pp.57-62
- [7] I.C Kim, Y.G Jung, Using Bayesian Network to analyze Medical Data, LNAI2734, Springer-Verlag (2003), pp.317-327
- [8] Yong Gyu Jung, Ki Young Lee and Myung Jae Lim, Discharge Decision for Post-Operative Patients, *Proceedings of ICHIT* (2010), pp.195-199
- [9] Shmueli, G., Patel, N. R., and Bruce, P., *Data Mining for Business Intelligence* (2009)
- [10] Ian H. Witten and Eibe Frank, *Data Mining*, Addison Wesley (2005), pp.315-333



Yong Gyu Jung received the B.S. in physics Education from Seoul National University in 1981. And then he got the M.S. and ph.D. degree of Computer Science from Yonsei and Kyonggi University in 1994 and 2003, respectively. Since 1999, he has been a Faculty of Department of Medical IT marketing in Eulji University. His research interests are in the areas of

medical information analysis and international standards including e-Business. He is a Member of the UN/ECE/CEFACT and ISO/TC154 standard organization.



Oh Jin Kim was born in Uijeongbu, Korea, in 1989 and has been studied in Dept. of Medical Computer Science of Eulji University in Korea. Her interest area is information analysis and system implementation in medical and hospital.



Jae Kang Won received the B.S. in biology from Kangreung University in 1999. And then he got the M.S. and ph.D. degree of Computer Science from Kyonggi University in 2002 and 2009 respectively. His research interests are in the areas of workflow, BPM and data mining in medical and

other industries.