Association Rule Mining with Apriori Algorithm for Pediatric Foot Disorders

JUNGKYU CHOI[1], YONGGWAN WON[2], JUNG-JA KIM[3,4,*]

[1]Department of Healthcare Engineering, Graduate school
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
REPUBLIC OF KOREA

[2]School of Electronics and Computer Engineering
Chonnam National University
77 Yongbong-ro, Buk-gu, Gwangju, 500-757
REPUBLIC OF KOREA

[3]Division of Biomedical Engineering
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
REPUBLIC OF KOREA

[4]Research Center of Healthcare & Welfare Instrument for the Aged
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
REPUBLIC OF KOREA

Abstract: - An association rule mining has potential to discover important information of disorder in data. The purpose of this study was to show analysis process of the clinical data to acquire significant information effectively between the foot disorder groups and biomechanical parameters related to symptom by the association rule discovery. The first clinical health records of the total 279 pediatric patients diagnosed with the complex foot disorder, including pes planus basically, were used for study. The Apriori algorithm was applied to discover rule of the foot disorder groups. As the results, we were able to discover 8 rules of the complex foot disorder and confirm major information. In next study, another data mining methodology like neural network will be applied with a careful preprocessing for better analysis of the pediatric foot disorder from now on.

Key-Words: Data mining, Association rule, Apriori, Pediatric foot, Foot disorder

1 Introduction

In medical field, massive data sets like the electronic health records is created by fast development of hospital information technology. The clinical data includes quantitative data (e.g. laboratory values), qualitative data (e.g. text-based documents and demographics), and transactional data (e.g. record of medication delivery) [1]. When utilization of medical big data, value production of 330 billion dollars is expected every year on the US medical field. If effective treatment method by analysis data of diagnostic pattern, prognosis, cost, etc., direct effect of about 165 billion dollars is expected [2]. Data mining is the extraction technology of hidden and valuable information, and it has been recognized for a large amount of the clinical data by many study [3]. It is method to deal with large data and to discover meaningful knowledge with application of pattern recognition technology, statistics or mathematic algorithm [4]. An association rule discovery, is first introduced by Agrawal, Imieliński, and Swami, has potential to discover important information of disorder in data [5]-[7]. It is the most valid methodology to explore attributes pairs which contain useful knowledge in nonrestrictive patterns. In other words, rules from this method mean the relation between particular transaction and other transaction appeared progressively or simultaneously when specific transaction occurred. Accordingly, the advantage of

the association rule discovery is to be able to analyze a correlation of variables in data at the same time by searching items to set of rules.

According to previous studies, the data mining was applied for analysis of the clinical data. Kim presented that age, pathology scale, related disease, hospitalization period, respiratory failure and congestive heart failure were came out to be danger factors on death of pneumonia by using Apriori modeling algorithm in study for analysis of death factor on pneumonia patient [8]. Breault, Goodall and Fos utilized Classification and Regression Trees (CART) of data mining for analysis of diabetic data warehouse. They figured out that the most important variable associated with bad glycemic control was younger age, not the comorbidity index or whether patients had related disorders [9]. In addition, Duru used four decision tree algorithms for analysis postoperative status in the ovarian endometriosis patient under different conditions, and reported new meaningful information about recurrent ovarian endometriosis [10]. However, these data mining technology was not tried to utilize yet in terms of podiatric medicine. So far the earlier studies have verified significance only through a simple statistical method.

The lower limbs are the most important organ for movement action of human and basic activity of daily living [11]. Most of all, the foot have 26 bones, 33 joints, more than 100 muscles, tendons and ligaments, even if it is just 5% of the entire surface in the body. In addition, it works organically as a highly complex anatomical- and biomechanical structure in gait with supporting the whole weight and ground reaction force [12]. These pressure, weight and ground reaction force by push-off exercise, causes stress or soft tissue strain [13]. In this condition, there are close connections between form of the foot and disorder the lower limbs. However, complex correlation analysis is necessary than simple quantitative analysis between the foot disorder and symptom because the foot disorder appear with complex symptoms, and this symptoms are not commonly clear.

Accordingly, the purpose of this study was to show analysis process of the clinical data to acquire significant information effectively between the foot disorder groups and biomechanical parameters related to symptom by the association rule discovery.

2 Study Procedure

2.1 Subjects

The first clinical health records of the total 279 pediatric patients diagnosed with the complex foot disorder, including pes planus basically, were used for study. The data was collected from the Foot Clinic of Jeon-ju Pediatrics. 64 patients records with missing values were excluded, and the complex disorder groups over 5% of whole data were selected. Therefore, analysis data was composed of 174 patient records with five groups for the complex disorder.

2.2 Item sets

A consequent in the study was the foot disorder, and it was into encoding for more clear classification. Among them, the consequent in the analysis data was consisted of five complex disorder groups such as class A : D1 (Achilles tendinitis), D2 (Pes planus), class B : D2 (Pes planus), class C : D2 (Pes planus), D3 (Intoe gait), class D : D2 (Pes planus), D3 (Intoe gait), D5 (Genu valgum), class E : D2 (Pes planus), D5 (Genu valgum). The foot disorder commonly appeared complexly like this. An antecedent was the biomechanical parameters related to disorder, and it was 26 attributes at first. To check statue of the foot, the biomechanical parameters like the Resting Calcaneal Stance Position (RCSP), the tibia TransMalleolar Angle (TMA) or the knee Internal Malleolus Distance (IMD) were measured and recorded into a patient chart by a podiatrist, as shown in Fig. 1. Through the feature selection node, meaningless antecedents were excluded. The extracted 13 attributes were grouped for changing numeric type into nominal type based on Donatelli's clinical assessment of the foot by radiography and Choi, Kim, Won and Kim's result of decision tree model [14]-[15]. Finally, 7 antecedents were used for analysis, as follow Table 1.

Fig 1. Measurement of RCSP and patient charts

2.3 Data analysis

In the study, combination of the antecedent meant each of the foot disorder group. Therefore, 7 antecedents were inserted into Apriori algorithm of association rule mining for the consequent. Data analysis was performed by IBM SPSS statistics 18 (SPSS Inc., Chicago, IL, USA) and IBM SPSS Modeler 14.2 (SPSS Inc., Chicago, IL, USA). In the Apriori algorithm, a assocation rule (R) of the antecedent (X) cause the consequent item (Y), and this rule can be expressed as follows :

$$(R) : (X) \Rightarrow (Y) \quad (1)$$

For significance, a rule which satisfied with the minimum support and the minimum confidence set up by a user is only extracted. In this study, the Minimum Support (MS) was 5 %, and the Minimum Confidence (MC) was 85 %. The overall study process was shown in Fig. 2.

Table 1. Consequent Item

Class	Disorder
A	D1, D2
B	D2
C	D2, D3
D	D2, D3, D5
E	D2, D5

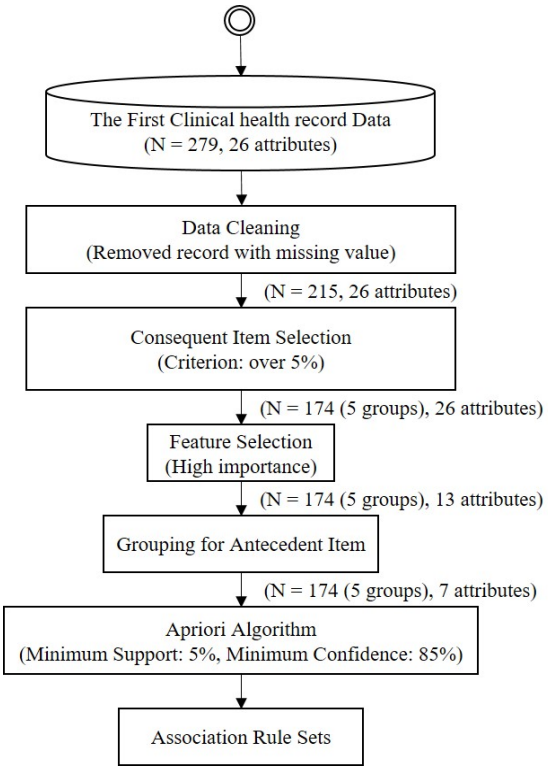

Fig 2. Study Process

3 Result

The first clinical health records of 174 pediatric patients with the complex foot disorder, including pes planus basically, were used for this study in the 279 whole data. The Apriori algorithm was applied to the experimental data to discover rule of the foot disorder groups.

As a result, total 8 rules which were satisfied with the minimum support and the minimum confidence were generated, as follows : Rule 1 : KneeIMD was abnormal and TibiaTMA was abnormal ⇒ Class D (MS : 14.451%, MC : 92%), Rule 2 : KneeIMD was abnormal, TibiaTMA was abnormal and RCSP was 1 group ⇒ Class D (MS : 12.139%, MC : 90.476%), Rule 3 : KneeIMD was abnormal, TibiaTMA was abnormal and TalarDeclination was 2 group ⇒ Class D (MS : 12.717%, MC : 90.909%) Rule 4 : Talocalcaneal was 3 g roup, CuboidAbduction was 2 g roup and Intermetatarsal was 2 group ⇒ Class B (MS : 9.249%, MC : 93.75%), Rule 5 : CuboidAbduction was 1 group, TibiaTMA was abnormal and RCSP was 1 group ⇒ Class C (MS : 16.763%, MC : 86.207%), Rule 6 : KneeIMD was abnormal, TibiaTMA was abnormal, RCSP was 1 group and TalarDeclination was 2 group ⇒ Class D (MS : 10.405%, MC : 88.889%), Rule 7 : Intermetatarsal was 1 group, TibiaTMA was abnormal, RCSP was 1 group and TalarDeclination was 2 group ⇒ Class C (MS : 12.717%, MC : 86.364%), Rule 8 : CuboidAbduction was 1 group, TibiaTMA was abnormal, RCSP was 1 group and TalarDeclination was 2 group ⇒ Class C (MS : 15.607%, MC : 88.889%), as shown in Fig. 3. In case of class A and class E, reliable rules were not discovered.

4 Conclusion

The purpose of this study was to perform analysis process of the clinical data to acquire significant information between the foot disorder groups and biomechanical parameters related to symptom by the association rule discovery. The first clinical health records of 174 pediatric patients with the complex foot disorder were analyzed for this study, in the 279 whole data. Through preprocessing like grouping, the experimental data was composed of a consequent, five classes, and 7 antecedent. The Apriori algorithm was applied to the experimental data to discover rule of the foot disorder groups.

As the results, we were able to discover 8 rules of the complex foot disorder and confirm major

Table 2. Antecedent Item

Variable	Type	Value	Description
RCSP	Nominal	1: below -6°, 2: over -5°, 3: asymmetry	Resting calcaneal stance position angle
Tibia TMA	Flag	1: abnoraml, 0: normal	Angle of the tibia transmalleolar
Knee IMD	Flag	1: abnoraml, 0: normal	The knee internal malleolus distance
Talocalcaneal	Nominal	1: below 28°, 2: over 27°, 3: asymmetry	Angle between the talus and the calcaneus
Cuboidabduction	Nominal	0: normal, 1: below -1°, 2: over 6°, 3: asymmetry	Angle of the cuboid abduction
Intermetatarsal	Nominal	0: normal, 1: below -7°, 2: over 9°, 3: asymmetry	Angle of the metatarsus primus adductus
Talardeclination	Nominal	1: below 20°, 2: over 22°, 3: asymmetry	Angle of the talus declination

Consequent	Antecedent	Rule ID	Support %	Confidence %
Class = D	KneeIMD TibiaTMA	1	14.451	92.0
Class = D	KneeIMD TibiaTMA RCSP = 1.0	2	12.139	90.476
Class = D	KneeIMD TibiaTMA TalarDeclination = 2.0	3	12.717	90.909
Class = B	TaloCalcaneal = 3.0 CuboidAbduction = 2.0 Intermetatarsal = 2.0	4	9.249	93.75
Class = C	CuboidAbduction = 1.0 TibiaTMA RCSP = 1.0	5	16.763	86.207
Class = D	KneeIMD TibiaTMA RCSP = 1.0 TalarDeclination = 2.0	6	10.405	88.889
Class = C	Intermetatarsal = 1.0 TibiaTMA RCSP = 1.0 TalarDeclination = 2.0	7	12.717	86.364
Class = C	CuboidAbduction = 1.0 TibiaTMA RCSP = 1.0 TalarDeclination = 2.0	8	15.607	88.889

Fig 3. Result of Association Rules

information. In case of the class D, pes planus, intoe gait and genu valgum, there was possibility to occur with high probability if the knee internal malleolus distance and the tibia transmalleolar angle were abnormal, RCSP was below -6° and the talar declination angle was over 22°. The class C, pes planus and intoe gait, was likely to appear highly in condition of that the tibia transmalleolar angle was abnormal, RCSP was below -6°, the talar declination angle was over 22° and the cuboidabduction angle was below -1°, or the intermetatarsal angle was below -7°. As the result of class C and D, especially, we were able to know that key factor which decided genu valgum was the knee internal malleolus distance. In case of the class B, pes planus, there is likely to appear highly if the talocalcaneal angle was asymmetry between both feet, the cuboid abduction angle was over 6° and the intermetatarsal angle was over 9°. Generally, the pes planus was diagnosed by the resting calcaneal stance position angle [16]. All of classes had pes planus basically because the highest value of RCSP was just -4° in the experimental data. However, the result of the class B showed a new rule with other parameters, instead of RCSP. In other words, it meant that RCSP was a factor for diagnosis of pes planus-nothing more and nothing less.

A massive amount of medical data is available now with hospital system, and an intelligent analysis method is necessary to get meaningful information effectively in complex data [17]. However, these efforts were not attempted yet in the podiatric medicine field, and the pattern analysis of foot disorder in pediatric patient was particularly insufficient. Through this study, we were able to get useful information with the result that we tried to discover rule of the podiatric foot disorder. However, rules of two classes were not confirmed. This result would be judged by reason of that the foot disorder had commonly complex symptoms and the symptoms were not clear. This being so, more detailed data preprocessing like grouping is very important in the foot clinical data. In next study, another data mining methodology like neural network will be applied with a careful preprocessing for better analysis of the pediatric foot disorder from now on.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A04016782)

References:

[1] T. B. Murdoch, and A. S. Detsky, "The inevitable application of big data to health care. Jama, *Jama*, vol. 309, no. 13, 2013, pp. 1351-1352.

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity, Mckinsey Global Institute, June 2011.

[3] A. R. Islam, and T. S. Chung, An Improved Frequent Pattern Tree Based Association Rule Mining Technique. Information Science and Applications (ICISA), *2011 International Conference on. IEEE*, 2011, pp. 1-8.

[4] T. L. Lincoln, and P. W. Suen, Common rotational variations in children, *Journal of the American Academy of Orthopedic Surgeons*, vol. 11, no. 5, 2003, pp. 312-320.

[5] R. Agrawal, T. Imieliński, and A. Swami, Mining association rules between sets of items in large databases, *ACM SIGMOD Record*, vol. 22, no. 2, 1993, pp. 207-216.

[6] T. Imamura, S. Matsumoto, Y. Kanaqawa, B. Tajima, S. Matsuya, M. Furue, and H. Oyama, A technique for identifying three diagnostic findings using association analysis, *Medical & Biological Engineering & Computing*, vol. 45, no. 1, 2006, pp. 51-59.

[7] W. T. Hwang, and D. S. Kim, Improved association rule mining by modified trimming, *Journal of the Institute of Electronics Engineers of Korea*, vol. 45, no. 3, 2008, pp. 15-21.

[8] Y. M. Kim, A study on analysis of factors on in-hospital mortality for community-acquired pneumonia, *Journal of the Korean Data & Information Science Society*, vol. 22, n o. 3, 2011, pp. 389-400.

[9] J. L. Breault, C. R. Goodall, and P. J. Fos, Data mining a d iabetic data warehouse, *Artificial Intelligence in Medicine*, vol. 26, no. 1, 2002, pp. 37-54.

[10] N. Duru, An application of apriori algorithm on a diabetic database, *Knowledge-Based Intelligent Information and Engineering Systems*, Springer Berlin Heidelberg, 2005, p. 398-404.

[11] Y. J. Ko, and H. W. Kim. Diagnosis and Conservative Treatment of Common Foot Diseases, *Journal of the Korean Medical Association*, vol 47, 2004, pp. 247-257.

[12] H. C. Kim, Management of Foot and Ankle Disorders, *Journal of the Korean Medical Association*, vol. 48, 2005, pp. 663-671.

[13] D. J. Lott, M. K. Hastings, P. K. Commean, K. E. Smith, and M. J. Muller, Effect of footwear and orthotic devices on stress reduction and soft tissue strain of the neuropathic foot, *Clinical Biomechanics*, vol. 22, 2007, pp. 352-359.

[14] R. A. Donatelli, The Biomechanics of the foot and ankle, The F. A. Davis Company, Philadelphia, PA, 1995.

[15] J. Choi, H. I. Kim, Y. Won, and J. J. Kim, Pattern Analysis of Pediatric Foot Disorders Using Decision Tree, Recent Researches in Applied Computer Science, 2015, pp. 9-14.

[16] W. H. Lee, and S. W. Lee, A study of the Relationship between Normal Adults Resting Calcaneal Stance Position and Postural Sway, *J Korean Soc Phys Ther*, vol. 11, 2004, pp. 5-17.

[17] W. Raghupathi, Data mining in health care, *Healthcare Informatics: Improving Efficiency and Productivity*, 2010, pp. 211-223.