Data Mining and Web-Based Children Shoe Suggestion System

P. Sudta, K. Kanchan, and Chantana Chantrapornchai

Department of Computing, Faculty of Science, Silpakorn University, Thailand, 73000
ctana@su.ac.th

Abstract. In this paper, we present a shoe suggestion system for children. The system stores the database about children's shoes from K1-K3 and P1-P3. The attributes considered are types of shoes, shoe sizes, brands. Given a type of children foot and age, the system suggested the type of shoes, brands, and sizes with the confidence level. The system uses the data mining technique for the classification and prediction.

Keywords: Shoe suggestion, data mining, decision tree, KNN, Neural net.

1 Introduction

For children, there are various kinds of attributes that are needed to consider for buying their shoes. The children shoes need to be comfort. The size needs to be suitable to them as well as the shape of the shoes. In this research, we are interested in shoe sizing and major attributes that affect the decision on buying shoes for Thai students. In particular, we would like to develop a prototype system that suggests proper shoes for small kids based on certain criteria.

Data mining is a popular technique used in classification and predictions. Many research work used it as a tool for decision support system. Palaniappan S. and Awang R. [7] developed a heart disease decision support system. It used naïve bayes, decision tree and neural net to model the classifications. It is developed as a web applications and used DMX query tool. C.Y. Ma, Frances, V Buontempo and X. Z Wang [6] used inductive data mining for historic data analysis. M.J. Aitkenhead [1] developed a co-evolution decision tree method. It combines an evolution method with the decision tree model for a better classification. Hudson S., and Ritchie B. [4] used cluster analysis for tourism in Alberta. They applied the model CRISP-DM. Jiao J. R et. al. [5] developed Kensei Mining which uses association rules to design user interface. Hsieh C., Huang S. [3] applied data mining technique to design new products. Apriori algorithm was used here. The attributes investigated are customer needs and product properties. Prassas G., et. al.[8] data mining for online shopping suggestion. Chen Y.L., Chen J.M., Tung C.W. [2] applied data mining to see the effect of shelf-space adjacency.

In our paper, we develop a children shoe suggestion system. The data mining engine is used. Particularly the decision tree technique is applied. We collect samples data from the neighborhood school from K1-K3 and P1-P3, each of which is 20-30

T.-h. Kim et al. (Eds.): UNESST 2011, CCIS 264, pp. 114–121, 2011.

students. About the students, we collect the student age, shoe shape, sex. About their shoes, we collect the shoe sizes, brands, and types. Assume the types are student shoes, sport shoes and leisure shoes. We design the database to save all the student data and shoe attributes. After the modeling for classification is developed, it is integrated to the web-based system.

2 Backgrounds and Data Preparations

Data mining technique is used often with large database, dataware house etc. It is mainly applied in decision support systems for modeling and prediction. There are several kinds of data mining: classification, clustering, association, sequencing etc.

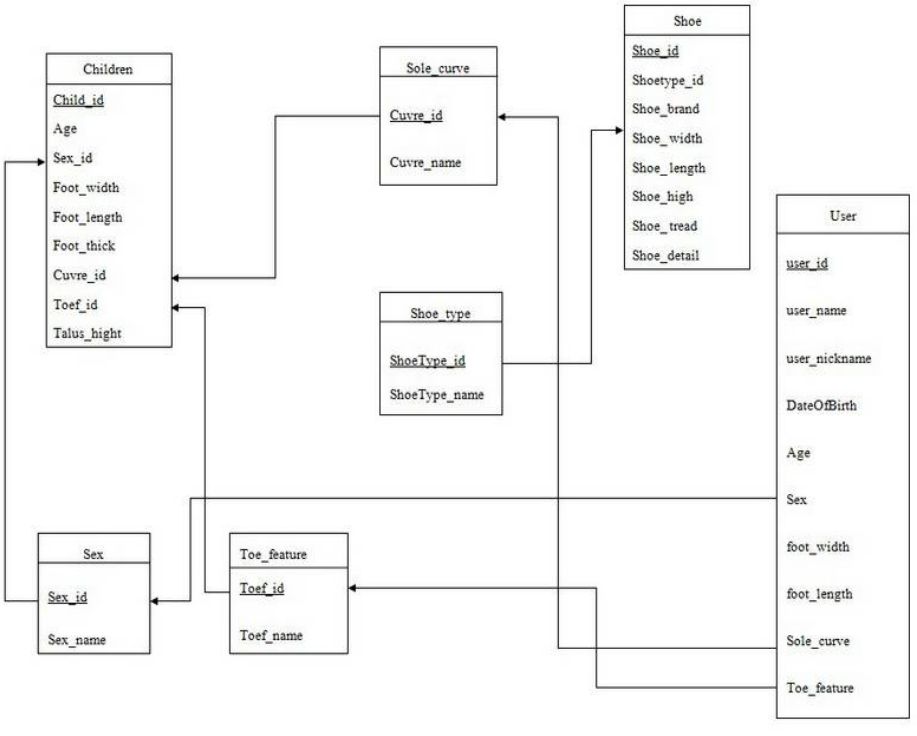

Fig. 1. E-R diagram of the collected data

Typically, for classification, the classifier model is needed. Data are divided into training set and test set. The training data is used to create the model. Then the test set applied for checking the model correctness. Until satisfied, the model is trained and adjusted by training data. Common techniques used in classification are decision tree, neural network, naïve bayes, etc.

In this work, we collect the data from one of the public kindergarden school. The school is a small size. There are 20 boys each in K1, K2 and K3 respectively. There are 30 boys in P1, P2, and P3 respectively. There are 20 girls each in K1, K2, and K3

respectively. Also, there are 30 girls each in P1,P2 and P3 respectively. There are 3 types of shoe types such as sport shoes, student shoes and leisure shoes. We collect 120 pairs of student shoes dividing into 60 boys' and 60 girl's, 70 pairs of sport shoes from 35 boys and 35 girls, 50 pairs of leisure shoes from 25 boys and 25 girls.

We design a form for collecting these data from students. The attributes that are specified are student information such as name, sex, age, foot information such as sole curve, toes characteristics, thickess, height. We also draw the foot figures for left and right sides for references. The shoe sizes for each type are collected as well. The sample form is shown in appendix. From the data collected, we design a E-R diagram which corresponds to the database table for the application in Figure 1. Then these data are divided into training set and test set as in Figure 2.

In Figure 2, the attributed used to train are age, sex, foot width, foot length, foot thickness, sold curve, foot type, toe type, and length upto the angle. The last two columns are the brands and shoe sizes.

3 Data Analysis

From the above data collection, we analyze the attributes of the feet as following.

-Foot length. The students for these ages mostly have the same foot length. There are very few samples that have a distinct foot size.

-Sole curve. Usually, there are three types of sole curve. However, since these are very small children, the sole curves are in only flat level and small curve.

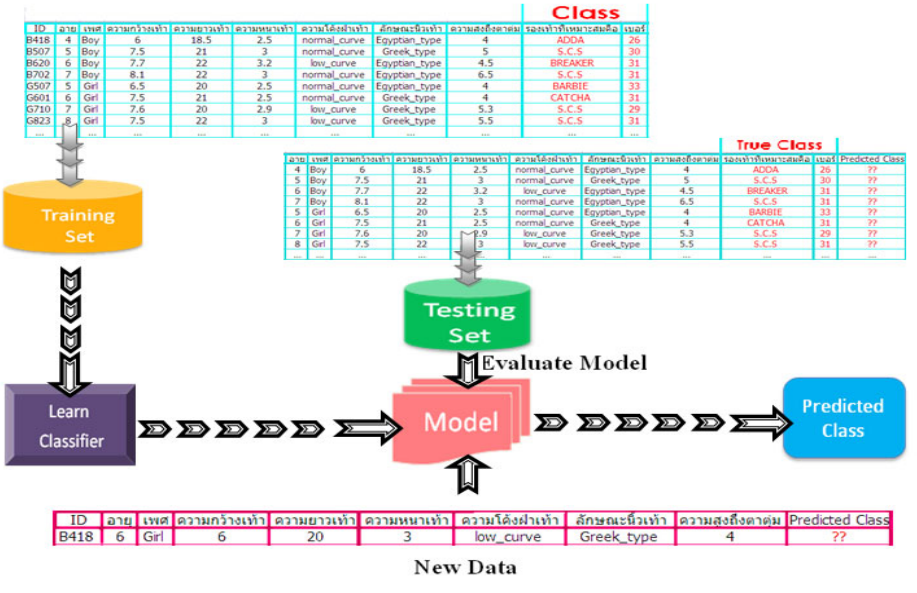

Fig. 2. Training the model

-Toe feature. Typically, there are three types: Egyptian style (where left one in the right side is the longest one and the others are shorter in the decreasing order), Greek style (where the second one is the longest) and the square style (where every finger is of the same height). From our data, most of the students are in the Greek style, then Egyptian style. Then the for the student shoe type, it can be matched with all kinds of feet. For the sport shoes, by inspecting the head of the shoes, it may be matched only Greek style feet. For the leisure shoes, by inspecting the head of the shoes, it may also be matched only Greek style feet.

From our data, we still have problems of insufficient data for children. Most likely, children in the same area and nationality, have the same style of feet. Also, the available shoes are very similar to each other. Then, the foot styles are not influence the classification so much.

4 Results

We use Weka (http://www.cs.waikato.ac.nz/ml/weka/) tool to create the decision tree. Figure 3 is an example result for student shoes. In this example, 1) foot length is the most important attribute. Then it is sole curve (small curve) and sex (female). We obtain CATCHA_31(6.0) which is the brand CATCHA, size 31, and there are 6 students on this leaf. In 2), also, foot length is the most important attribute. Then it is sole curve (small curve) and sex (male). The shoes is SCS brand whose size is 33 and there are 5 students in this leaf.

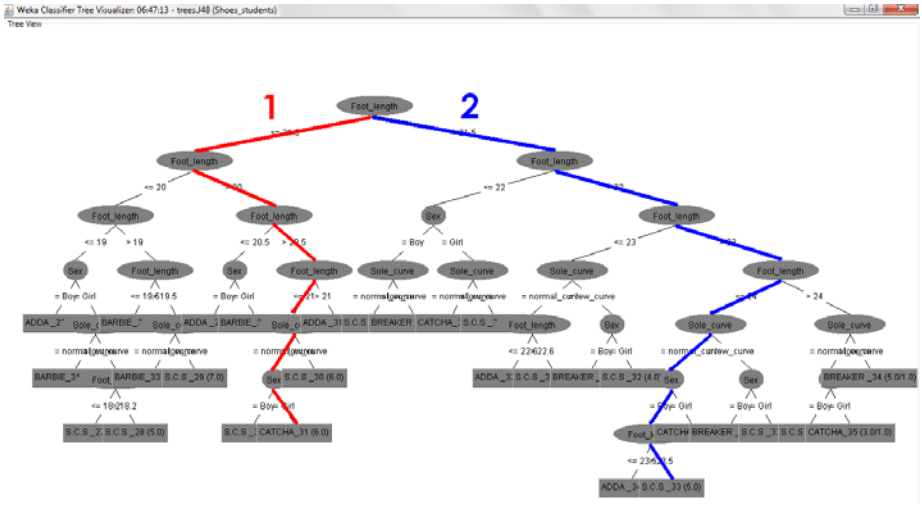

Fig. 3. Sample tree results for student shoe classification

In Figure 4, it is the sample results from sport shoes. In 1), only the foot length is used. In 2), the attributes are foot length, sex (male). In Figure 5, it is the tree for leisure shoes. Same here for 1) and 2), the foot length is the major attribute for decision.

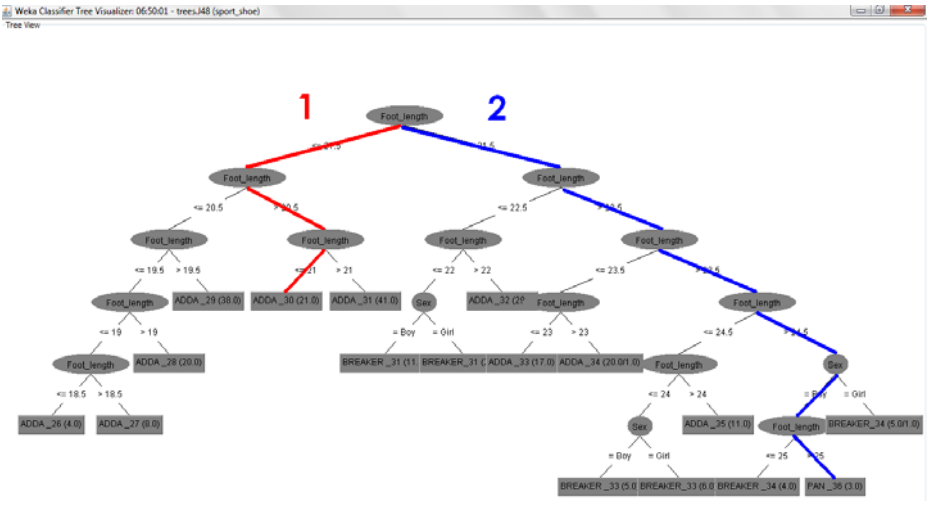

Fig. 4. Sample tree results for sport shoe classification

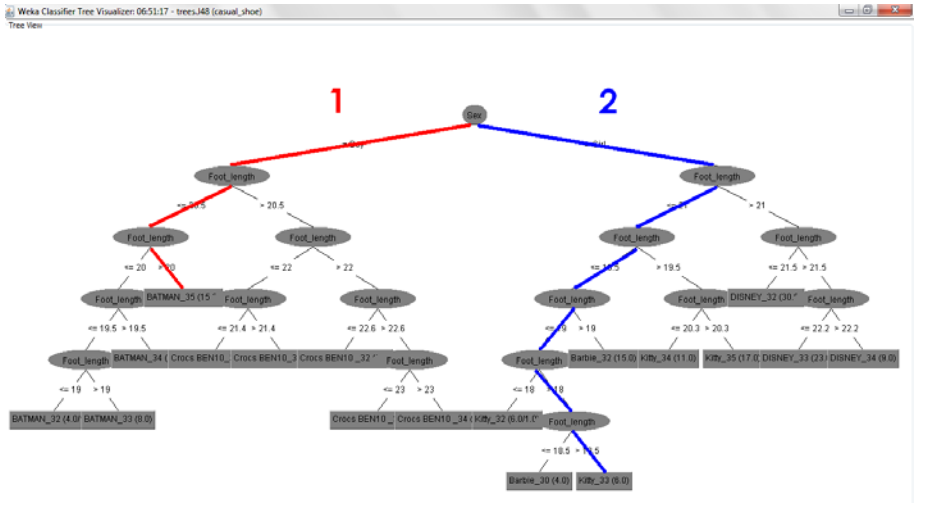

Fig. 5. Sample tree results for leisure shoe classification

For the given data, we test the accuracy results for different models: KNN, Neural net and decision tree. These are shown in Table 1. It is seen that decision tree performs best among the three models. We suspect this is because the mentioned reason in Section 4.

Table 1. Accuracy for different models

Type of shoes	Accuracy		
	NN	KNN	Decision tree
Student shoes	53%	56%	82%
Sport shoes	76%	54%	98%
Leisure shoes	80%	55%	99%

We develop the web application for suggesting the shoes as well. Sample user interface is shown in Figure 6 and Figure 7. In Figure 6, the user enters his profile about name, date of birth, sex. Then, in Figure 7, the user enters their feet information for all about foot attributes that were mentioned. Then, the suggested shoes are displayed for each type along with the accuracy in Figure 8.

Fig. 6. User profile

Fig. 7. User foot information

Fig. 8. Suggested shoes with accuracy

5 Conclusion

In this paper, we use a data mining to help suggest shoes for children. We are interested in factors that make the children select appropriate shoes for their comfort. The studied attributes include foot length, width, sole curve, foot style, as well as user profiles. However, the collected data may not be sufficient to train the model to see the affect of the shoe selections. Currently, the length of the shoes is the major concern. Due to this, decision tree performs well among the three approaches. In reality, sole curve, width, thickness are important to the comfort as well. To reflect this, data about the shoes need to be collected more intensively and in more detailed. This will be investigated further.

We also develop a web application that integrates the model. The user can enter their foot information and the system will suggest the proper shoes with accuracy.

References

[1] Aitkenhead, M.J.: A co-evolving decision tree classification method. Journal Expert Systems with Applications: An International Journal 34(1) (January 2008), doi:10.1016/j.eswa.2006.08.008

[2] Chen, Y.L., Chen, J.M., Tung, C.W.: A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Journal Decision Support Systems 42(3) (December 2006), doi:10.1016/j.dss.2005.12.004

[3] Hsieh, C., Huang, S.: Mining product maps for new product development Expert Systems with Applications 34(1), 50–62 (2008)

[4] Hudson, S., Ritchie, B.: Understanding the domestic market using cluster analysis. Journal of Vacation Marketing 8(3), 263–276 (2002), doi:10.1177/135676670200800305
[5] Jiao, J.R., Zhang, Y., Helander M.G.: A Kansei mining system for affective design. Expert Syst. Appl., 658–673 (2006)
[6] Ma, C.Y., Buontempo, F.V., Wang, Z.: Inductive Data Mining: Automatic Generation of Decision Trees from Data for QSAR Modeling and Process Historical Data Analysis. In: 18th European Symposium on Computer Aided Process Engineering, vol. 25(Idm), pp. 581–586. Elsevier (2008)
[7] Palaniappan, S., Awang, R.: Web-based Heart Disease Decision Support System Using Data Mining Classification Modeling Techniques. In: Proceedings of iiWAS 2007, pp. 157–167 (2007)
[8] Prassas, G., et al.: Recommender System for Online Shopping Based on Past Customer Behaviour. In: e-Everything: e-Commerce, e-Government, e-Household, e-Democracy: The 14th Bled Electronic Commerce Conference, Bled, Slovenia, June 25–26 (2001)

Appendix