Int J Adv Manuf Technol (2005) 26: 669–674
DOI 10.1007/s00170-003-2032-0

ORIGINAL ARTICLE

Chih-Hung Hsu · Mao-Jiun J. Wang

Using decision tree-based data mining to establish a sizing system for the manufacture of garments

Received: 25 July 2003 / Accepted: 10 November 2003 / Published online: 1 December 2004

Abstract While data mining has been widely used in many fields, little research has been done on its applications to sizing systems for the manufacturing of garments. The goal of this study was to establish systems for using a decision tree technique to determine the pants sizes of army soldiers. We first defined the subject and constructed a large anthropometric database. Second, we prepared and analyzed data, performed factor analyses, and then extracted important sizing variables. Third, we used the decision tree technique to mine data in an effort to identify and classify significant patterns in the body shapes of soldiers. The use of decision tree-based data mining to establish sizing systems is advantageous because it can (1) allow for a wider coverage of body shapes with a fewer number of sizes, (2) generate regular sizing patterns and rules, and (3) provide manufacturers with reference points to facilitate production. The newly developed sizing system can provide garment manufacturers with size specifications, design development, pattern grading, and market analysis. Moreover, when production plans can be made more realistic, inventory costs due to mismatches can be minimized.

Keywords Anthropometric data · Data mining · Decision tree · Garment manufacturing · Sizing systems

1 Introduction

Among textile manufacturing industries, garment manufacturing produces products with the highest added value [1]. In the garment manufacturing industry, large-scale machine production has replaced manual production, which has greatly increased output and reduced manufacturing costs. Because certain standards and specifications must be followed in large-scale machine production, each country must have its own standard-sizing systems for manufacturers to follow.

The standard-sizing systems have been used as a communication tool among manufacturers, retailers, and consumers. The use of such a system can provide manufacturers with size specification, design development, pattern grading, and market analysis. Manufacturers can produce different types of garments with various allowances for specific market segmentation. Furthermore, standard-sizing systems can predict production numbers and proportion of sizes to be produced, resulting in accurate production planning and material control [2, 3].

Garment-sizing systems were originally based on those developed by tailors in the late 18th century. Before then, all garments were hand-made to order. Tailors measured the body dimensions of each customer, and then drew and cut patterns for each garment, for the specific customer. After many original patterns had been accumulated, the tailors discovered correlations between bodily dimensions, regardless of the individual differences. Tailors gradually developed these patterns into a system of garment storage, which could be used to make clothes for people with similar figures [4].

In 1959, Emanuel et al. determined a set of procedures for formulating standard sizes for all figure types [5]. According to this system, people of all figure types were first classified into one of four bodyweight groups. These bodyweight groups are further subdivided by height. People were, thus, divided into eight categories based on similar heights and weights. Other countries were using similar sizing systems with classifications based on two or three sizing variables [6]. The sizing variables for female garments are height, bust girth, and hip girth; the sizing variables most commonly used for male garments are typically height, chest girth, and waist girth [7, 8].

In 1998, McCulloch et al. proposed criteria by which sizing systems could be evaluated [9]. They stated that the sizing systems should ideally: a) cover the greatest number of people and b) require the fewest number of sizes. At times, these criteria conflict with each other. Depending on circumstances, one may take priority over another.

C.-H. Hsu · M.-J.J. Wang (✉)
Department of Industrial Engineering and Engineering Management,
National Tsing Hua University,
Hsinchu 300, Taiwan, R.O.C.
E-mail: mjwang@ie.nthu.edu.tw
Tel.: +886-3-5742655
Fax: +886-3-5737107

To minimize costs, manufacturers must be able to determine how to accomplish both using a sizing system. We propose to use decision tree-based data mining to best accomplish that goal. Data mining, a major step of knowledge discovery in database (KDD), has been successfully applied in many fields, including finance [10], health insurance [11], biomedicine [12], manufacturing [13], and e-commerce [14]. However, research on establishing sizing systems using data mining is lacking.

One of the most important data mining methods is the decision tree technique, which can be used to classify data according to rules deduced from input variables, to display the data in a tree-shaped form, and to explore the significance of influencing factors.

Different decision tree algorithms suit different types of data. Of the many available algorithms, the most appropriate method for processing continuous data, such as anthropometric data, is the classification and regression tree (CART) [15, 16]. CART puts pre-processed data into the tree root, and then, through a series of classifying processes, determines the best separating points, which it uses to form a tree-shaped structure. The mined data is split by classification rules according to the purity or impurity of the child nodes [17].

2 Subject definition for data mining

Because of the urgent need for accurate sizing systems for producing army uniforms, a large database was created based on 265 static anthropometric variables measured in a sample 610 soldiers in Taiwan, resulting in around 160 000 pieces of data. Because standard-sizing systems for soldiers' pants were outdated and incomplete, this study uses the CART data mining technique to identify systematic patterns in body dimensions. Based on these body dimension patterns, the bodies of Taiwanese soldiers can be classified into representative figure types, and standard-sizing systems can be established. This work will be beneficial to the production of military uniforms in Taiwan.

3 Data preparation and analysis for data mining

3.1 Data preparation

Before the data are mined, they must be examined and purified. In this study, samples containing missing or abnormal data were omitted. Twenty of the 610 samples had missing or abnormal data and were thus deleted, leaving a total of 590 valid samples for further processing. Because all body dimensions were measured in millimeters with decimals, they were converted into integers in centimeters, allowing comparison and calculation with commonly used international garment-sizing units. Additionally, two anthropometric variables, body weight and height, were converted into a new variable, the body mass index (BMI). The World Health Organization defines BMI as weight divided by height squared. It is used in medical science as a reference for judging obesity-related diseases and whether a figure type is standard or not. Taiwan's military only selects military personnel with a BMI between 16.5 and 32 [18]. In this study BMI served as the target variable of the decision tree and could be used to facilitate data mining.

3.2 Factor analysis

Not all of the 265 anthropometric variables were useful for establishing the sizing systems of pants; hence, the body dimensions commonly used in garment manufacturing are first considered. Domain experts were consulted to determine eight anthropometric variables that are strongly associated with garment production.

Using all eight anthropometric variables to establish sizing systems would be very difficult. To simplify, we attempted to extract the most important factors. Using Kaiser's eigenvalue criterion, we selected two factors with eigenvalues over one [19]. Factor loadings were calculated to determine the coefficients of correlation between the two factors and the anthropometric variables. Consequently, anthropometric variables with factor loadings of greater than 0.7 were found to be clustered in Factors 1 and 2, as shown in Table 1. Most of the variables that appeared in Factor 1 were girth-related variables and included waist girth, hip girth, thigh girth and weight. The variables appeared in Factor 2 were height-related variables and included outside leg length, crotch length, and height. Thus, Factor 1 was called the girth factor, and Factor 2 was called the height factor.

The results of the factor analysis presented in Table 1 reveal that, besides weight, the top three anthropometric variables most closely correlated with the girth factor were waist girth, hip girth, and thigh girth. Although hip girth was the anthropometric variable most closely correlated with the girth factor, waist girth is the most important variable in establishing sizing systems of pants in garment making [8, 20]. Therefore, waist girth was selected to represent the girth factor. Because the anthropometric variable that correlated most closely with height was outside leg length, it was selected to represent the height factor.

Table 1. Factor loadings (un-rotated)

	Factor 1	Factor 2
Waist girth	−0.831*	0.363
Hip girth	−0.903*	0.265
Thigh girth	−0.823*	0.395
Outside leg length	−0.490	−0.819*
Crotch height	−0.419	−0.811*
Total crotch length	−0.631	0.262
Body height	−0.586	−0.758*
Weight	−0.937*	0.162
Variance explained	4.219	2.357
Total proportion	0.527	0.295

* Marked loadings are > 0.7

3.3 CART analysis

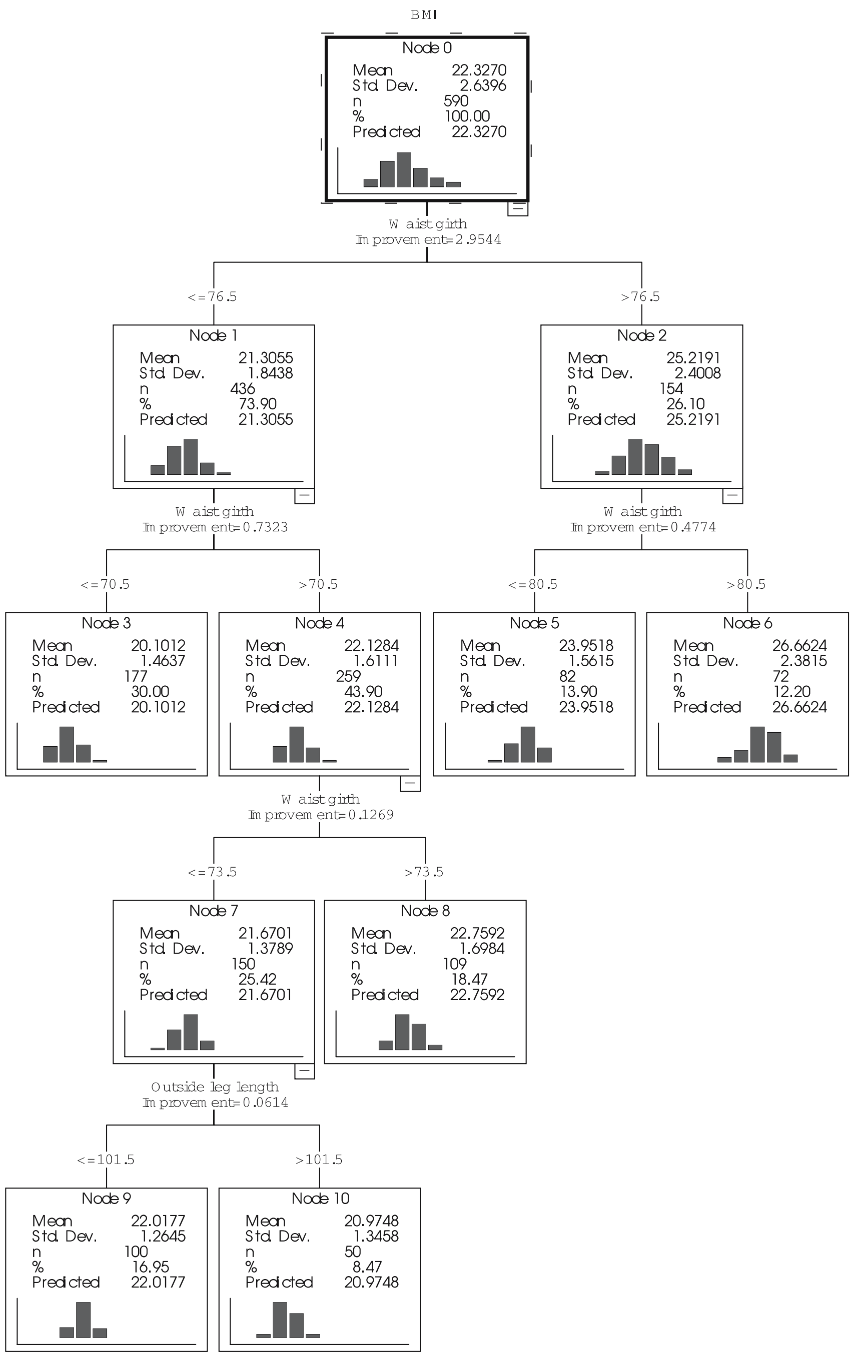
Once waist girth and outside leg length were selected by factor analysis to be the most important sizing variables, the CART decision tree technique was used to mine data. This study takes BMI as the target variable, with waist girth and outside leg length being predictors used to classify the target variable. The following stopping rules were set as follow.

- The greatest depth of the tree extends to the fourth level beneath the root node.
- The minimum number of samples in the parent node is 100, and the minimum number of samples in the child node is 40.

Figure 1 presents the results of CART analysis. The root node was split according to waist girth, resulting in the first level. A total of 436 samples with waist girth values were smaller than or equal to 76.5 cm were grouped into Node 1, and 154 samples with waist girths greater than 76.5 cm were grouped into Node 2.

Node 1 was then split according to waist girth, yielding the second level of the tree. Grouped into Node 3 were 177 samples

Fig. 1. CART branching



with waist girths of less than or equal to 70.5 cm, and grouped into Node 4 were the other 259 samples, whose waist girths exceeded 70.5 cm. Node 2 was similarly split. Eighty-two Node 2 samples with waist girths of less than or equal to 80.5 cm were grouped into Node 5; the other 72 Node 2 samples with waist girths of over than 80.5 cm were grouped into Node 6.

The first three levels were split according to waist girths. From the fourth level, outside leg length was used for splitting. Eventually, only the four nodes generated at the second level were chosen to represent the four figure types leaving us with a reduced number of figure types.

Figure 2 plots a distribution graph of waist girth as the X-axis against hip girth on the Y-axis to show the distribution of all of figure types. Hip girth is also an important variable for sizing male pants in many countries and it significantly correlates with the girth factor. This study, as represented in Table 2, identifies the figure type of 177 samples with smaller waist girth and hip girth as Y; the figure type of 259, the majority, as A; the figure type of 72 samples with larger waist girth and hip girth as C; and the figure type of the remaining 82 samples as B.

A line graph was plotted to yield better insight into the differences among the four CART-classified figure types. As shown in Fig. 3, the four figure types showed differences in waist girth, hip girth, and thigh girth. The four figure types also followed a $C > B > A > Y$ trend. Analysis of Variance (ANOVA) results and Duncan's multiple range tests further confirmed distinct differences among the four figure types.

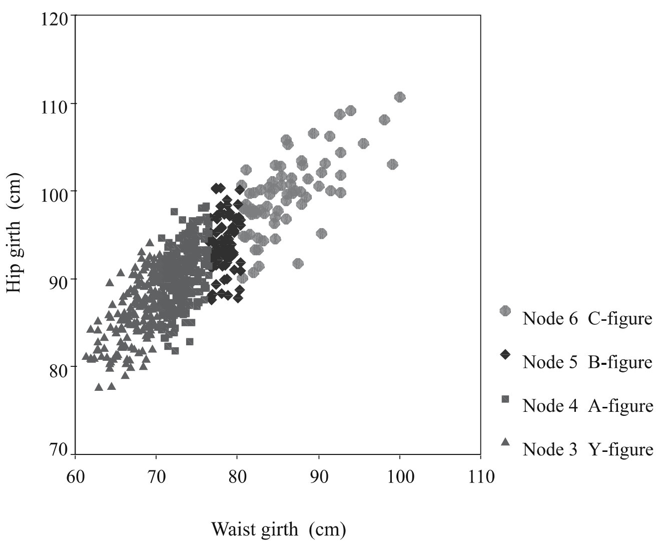


Fig. 2. The scatter plot of waist girth vs. hip girth for four figure types

Table 2. Definitions of the four figure types

Nodes	Figure types	Classified rules	Numbers
Node 3	Y	Waist girth $\leqq$ 70.5 cm	177
Node 4	A	70.5 cm < Waist girth $\leqq$ 76.5 cm	259
Node 5	B	76.5 cm < Waist girth $\leqq$ 80.5 cm	82
Node 6	C	80.5 cm < Waist girth	72

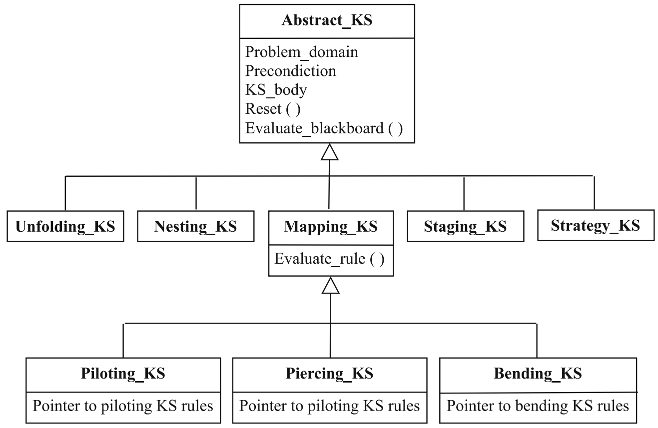


Fig. 3. The four figure types and the corresponding anthropometric variables

4 Establishing sizing systems

Waist girth, hip girth, and outside leg length are the most commonly used variables for sizing male pants. This section includes relevant scatter plots of waist girth on the X-axis against hip girth

Hip girth (cm)	61	62	63	64	65	66	67	68	69	70	Total
Total	1	4	7	10	9	24	20	26	36	40	177
96											0
95											0
94									1		1
93								1		1	2
92								1	4	4	9
91						1	1	1	2	5	10
90							2	3	1	7	13
89				1			2	2	5	1	11
88						2	2	3	4	6	17
87					1	1	2	4	3	6	17
86				2	1	2	4		6	2	17
85			1			7	3	2	3	6	22
84		1	1	1	2	2		6	4	1	18
83			1	1		2			2		6
82			2	1		4	2	1		1	11
81	1	3	1	1	4	1	1	1	1		14
80				3		1	1	1			6
79						1					1
78			1		1						2
Waist girth (cm)	61	62	63	64	65	66	67	68	69	70	Total

Fig. 4. The scatter plot for Y figure

on the Y-axis, at intervals of 4 cm. The waist girths were classified as follows.

- *Y* figure, 64 cm and 68 cm;
- *A* figure, 72 cm and 76 cm;
- *B* figure, 80 cm; and
- *C* figure, 84 cm, 88 cm, 92 cm, 96 cm and 100 cm.

The sizing systems for the four figure types were established using the methods of Emanuel [5].

Figure 4 shows a scatter plot for the *Y* figure of hip girth on the Y-axis against waist girth on the X-axis. The waist girths ranged from 61 cm to 70 cm. Since most countries use 4 cm as the interval for waist girth [21], the 64 cm and 68 cm waist girths were used for the *Y* figure. Another size of 70 cm was added because more people have a *Y* figure.

Figure 4 shows a scatter plot of 177 samples of the *Y* figure. The scatter plots of the other types of figures are similar to that of the *Y* figure type. Table 3 shows complete pants sizing systems for all four-figure types. Samples less than 0.3% were deleted in each size group. Out of the 590 samples, only 12 samples were excluded; therefore, the proposed "waist girth and hip girth" sizing systems covered 97.97% of the sample population.

Table 3. The sizing systems for the pants of army soldiers with four figure types

Figure types	Waist girth	Hip girth	Outside leg length (cm) 92	96	100	104	108	%
Y	**64**	**80**		96				0.7
		84		96	100	104		2.5
		88		96	100	104		0.6
	68	**82**		96	100	104		3.1
		86	92	96	100	104	108	5.4
		90	92	96	100	104	108	3.9
		94			100	104	108	0.9
	70	**84**		96	100	104	108	1.6
		88		96	100	104	108	6.2
		92		96	100	104	108	5.1
		96				104		0.4
A	**72**	**86**		96	100	104		3.8
		90	92	96	100	104	108	6.4
		94		96	100	104	108	6.8
		98					108	0.4
	76	**88**	92	96	100	104	108	5.6
		92		96	100	104	108	11.5
		96		96	100	104	108	8.6
		100		96	100	104	108	1.2
B	**80**	**90**	92	96	100	104		1.6
		94		96	100	104	108	6.9
		98		96	100	104	108	4.5
		102			100	104	108	0.9
C	**84**	**94**	92	96	100	104	108	0.9
		98			100	104	108	3.1
		102		96	100	104	108	1.4
	88	**98**		96	100	104	108	1.1
		102		96	100	104		1.7
		106				104	108	1.1
	92	**102**			100	104		0.7
		106			100	104		0.4
	96	**106**			100		108	0.4
		110			100			0.4
	100	**110**				104	108	0.4

Table 4. The simplified sizing systems for the pants of army soldiers

Figure types	Waist girth	Outside leg length (cm) 92	96	100	104	108	%
Y	**64**		64S	64M	64L		3.8
	68	68XS	68S	68M	68L	68XL	13.3
	70		70S	70M	70L	70XL	13.3
A	**72**	72XS	72S	72M	72L	72XL	17.4
	76	76XS	76S	76M	76L	76XL	26.9
B	**80**	80XS	80S	80M	80L	80XL	13.9
C	**84**	84XS	84S	84M	84L	84XL	5.4
	88		88S	88M	88L	88XL	3.9
	92			92M	92L		1.1
	96			96M		96XL	0.7
	100				100L	100XL	0.3

When three sizing variables, waist girth, hip girth, and outside leg length, were taken into account, the total coverage of the military garment-sizing systems was 94.58%, as shown in Table 3. Furthermore, this study could generate simple sizing systems from Table 3 according to the those used by previous soldiers' pants sizing systems, as shown in Table 4.

5 Discussion and application of results

The four figure types were obtained by applying the classification rules based on the data mining decision tree technique. The newly developed sizing systems exhibit the following three characteristics.

(1) High coverage rates and few sizing groups

As stated above, the coverage of the sizing systems with the two sizing variables (waist girth and hip girth) is 97.97%. Even if outside leg length is added to each figure type, the total coverage rate remains a high 94.58%.

This study generated simple sizing systems based on the format of the previous soldiers' sizing systems. Forty-two groups were found in the simple sizing systems, which was fewer than the previously used systems. Additionally, the newly developed sizing systems use the army's labeling method: 76 M means that the waist girth is 76 cm (figure *A*) with an outside leg length of "M" (100 cm). Thus, the soldiers' body dimensions can be obtained using easy-to-understand garment manufacturing sizing systems.

(2) Regular patterns and rules

The sizing systems, like sizing systems in other countries, have regular patterns and rules. Four-centimeter intervals are used for both waist girth and hip girth. Waist girths are 64 cm and 68 cm (*Y* figure), 72 cm and 76 cm (*A* figure), 80 cm (*B* figure) and 84 cm, 88 cm, 92 cm, 96 cm and 100 cm (*C* figure). Another size of 70 cm was added because more people have a *Y* figure.

The first corresponding hip girth for each size group is presented in order. They are 80 cm, 82 cm and 84 cm (*Y* figure), 86 cm and 88 cm (*A* figure), 90 cm (*B* figure), and 94 cm, 98 cm, 102 cm, 106 cm and 110 cm (*C* figure). The regularity for hip girth classification is also obvious. Table 3 shows a clear regular pattern and rule.

(3) Manufacturing reference points to facilitate production

Garment manufacturers who need to consult detailed production information, can refer to Table 3. The *B* figure for example, with a representative waist girth of 80 cm, matches hip girths of 90 cm, 94 cm, 98 cm, and 102 cm. Therefore, pants with four different hip girths can be produced for soldiers with a waist girth of 80 cm. Of course, outside leg length should be taken into account when the production of pants of a certain size is planned. The percentages people who have certain figure types and sizes are also recorded in the sizing systems and may serve as good reference points for the number of pants to be produced. With this information, a realistic plan for producing army soldiers' uniforms can be established.

6 Conclusion

A standard sizing system is for a necessity the garment manufacturing industry to develop effective production planning. This study applies decision tree-based data mining to develop systems for sizing pants of soldiers in Taiwan. The advantages to the resulting system are as follows:

- The total coverage of the sizing systems is relatively high at 94.58%.
- The sizing systems based on the decision tree show regular patterns and rules.
- The total number of size groups is only 42, much fewer than the number in the previous sizing systems. They also outlined the percentages of people for every figure type in each size group.

These advantages provide manufacturers with detailed reference points for garment production. This newly developed sizing system can accurately predict the requirements for different sizes of uniforms, and thus generate a realistic production plan, thereby minimizing unnecessary inventory costs due to sizing mismatches between number of soldiers and produced garments.

References

1. Chang CF (1999) The model analysis of female body size measurement from 18 to 22. J Hwa Gang Textile 6(1):86–94
2. Hsu KM, Jing SH (1999) The chances of Taiwan apparel industry. J China Textile Inst 9(2):1–6
3. Tung YM, Soong SS (1994) The demand side analysis for Taiwan domestic apparel market. J China Textile Inst 4(5):375–380
4. Burns LD, Bryant NO (2000) The Business of fashion: designing, marketing, and manufacturing. Fairchild, USA
5. Emanuel I, Alexander M, Churchill E, Truett B (1959) A height-weight sizing system for flight clothing. WADCTR 56-365, Aero Med Lab, Ohio, USA
6. Cooklin G (1992) Pattern grading for women's clothes. Blackwell, Oxford
7. Ashdown SP (1998) An investigation of the structure of sizing Systems. Int Journal of Clothing Sci Technol 10(5): 324–341
8. Jongsuk CY, Jasper CR (1993) Garment-sizing systems: an international comparison. Int J Clothing Sci Technol 5(5):28–37
9. McCulloch CE, Paal B, Ashdown SP (1998) An optimal approach to apparel sizing. J Oper Res Society 49(5):492–499
10. Gerritsen R (1999) Assessing loan risks: a data-mining case study. IT Professional 1(6):16–21
11. Chas YM, Ho SH, Cho KW, Lee DH, Ji SH (2001) Data mining approach to policy analysis in a health insurance domain. Int J Med Informatics 62:103–111
12. Maddour M, Elloumi M (2002) A data mining approach based on machine learning techniques to classify biological sequences. Knowledge-Based Syst 15(4):217–223
13. Feng CX, Wang X (2002) Development of empirical models for surface roughness prediction in finish turning. Int J Adv Manuf Technol 20(5):348–356
14. Ying WT, Lee SP (2002) Data mining using classification techniques in query processing strategies. Computer Systems and Applications, ACS/IEEE International Conference, pp 200–202
15. Apte C, Weiss S (1997) Data mining with decision trees and decision rules. Future Gener Comput Syst 13:197–210
16. Breiman L, Friedman JH, Olshen RA, et al. (1998) Classification and regression tree. Chapman & Hall, Florida
17. Berry M, Linoff G (1997) Data mining techniques: for marketing, sales, and customer support. Wiley, New York
18. Ministry of National Defense (2002) Defense report for Republic of China: recruit procedures. Taiwan
19. Chen SY (2000) Multivariate analysis. Hwa-Tai, Taipei
20. Cooklin G (1992) Pattern grading for men's clothes. Blackwell, Oxford
21. Winks JM (1997) Clothing sizes: international standardization. Redwood, U.K.