Using Data Mining to Extract Sizing Knowledge for Promoting Manufacture

CHIH-HUNG HSU[1], BOR-SHONG LIU [2], SU-CHIN CHEN[3]
[1]Department of Industrial Engineering and Management, Hsiuping Institute of Technology, Taiwan, R.O.C.
[2]Department of Industrial Engineering and Management, St. John's University, Taiwan, R.O.C.
[3]Department of Fashion Imaging, Ming-Dad University, Taiwan, R.O.C.

Abstract: - Garment production is a high value-added industry in the global textile manufacturing chain. Standard sizing systems of garment are crucial issue, play an even important role for garment manufacturing industry. The extraction of knowledge from large database has been successfully applied in a number of advanced fields by data mining. However, little research has been done in the area of developing standard sizing systems of garment manufacture, using data mining. Focusing on the anthropometric data of adult males in Taiwan, the goal of this study was to develop standard sizing systems, using a novel cluster-based data mining cycle. Certain advantages may be observed when standard sizing systems are developed, using the data mining cycle. These include being able to cover a higher percentage of the population, using fewer sizes, and providing manufacturers with reference points to promote products, according to body type and distribution. Since the anthropometric database must be repeatedly updated, standard sizing systems may also be continuously renewed via application of the proposed data mining cycle. These standard sizing systems will remain continually beneficial for both production planning and reducing inventory costs, while facilitating the advanced garment manufacture.

Key-Words: - Data mining; Standard sizing systems; Garment manufacture

1 Introduction

Standard sizing systems of garment are very crucial issue, play an even important role for garment manufacturing industry. Garment industry designer seek to produce the best design that meets standard sizing systems. Many times companies require specification limits [1]. Standard sizing systems can correctly predict numbers of items and ratio of sizes to be produced, resulting in accurate inventory control and production planning [2, 3]. Owing to current variations in dimensions and body type among nations, thus, each nation requires its own standard sizing systems as mass production benchmarks for garment factories. Under traditional production procedures, Taiwan has never developed its own sizing systems for the local market, since it uses traditional order-oriented production modes with most sizing data borrowed or modified from the foreign sizes resulting in sizing confusion among garment markets and factories and an inability to control all the plausible sizing data, finally resulting in heavy stock burden to garment manufacturers. Apart from the fact that most overseas sizing data do not correspond to Taiwanese body types, domestic manufacturers have been inconsistent in their size classifications, so consumers must choose suitable garment by trial and error, resulting in enormous inconvenience, not to mention wasted time and money. Thus, the development of standard sizing systems for garment manufacturers and consumers is long overdue, and the establishment of standard sizing systems that conform to the body types is crucial to the garment manufacturing industry.

Human body types can be distinguished by taking various approaches. When classifying garment sizes, instead of targeting every consumer, manufacturers simply produce garment in several sizes. Although a greater number of sizes offers consumers a greater number of choices, this can cause difficulties for manufacturers, as far as production and stock are concerned. Therefore, it would be helpful to formulate standard sizing systems, which have the fewest number of sizes to fit the largest number of body types, for the majority of consumers [3].

The application domain of data mining is quite broad. However, little research has also been applied to standard sizing systems development using data mining. This study proposes the development of male standard sizing systems by using cluster-based data mining cycle. By applying the proposed approach, body types can be classified from a large anthropometric database. The standard sizing systems can then be developed to facilitate garment production.

2 Data Mining and Clustering

Berry and Linoff defined data mining as the analysis of huge amounts of data by automatic or semi-automatic means, in order to identify significant patterns or rules [4]. One of the most important data mining techniques is cluster analysis, which is an exploratory data analysis tool for solving classification problems. The objective of cluster analysis is to sort bservations into clusters, so that the degree of association is strong between members of the same clusters and weak between members of different clusters.

Cluster analysis seeks to maximize between-group variances and minimize within-group variances, including both hierarchical and non-hierarchical methods [5]. Agglomerative hierarchical algorithms are commonly used with hierarchical methods, to calculate the distance between observations; the two nearest observations are combined into a cluster. This procedure continues until all observations are appointed in one cluster. Ward's minimum variance is an important agglomerative hierarchical algorithm method, as the smallest increase in total within-group variance has the highest priority of combination. Ward's method is distinct from all the other methods because it uses an analysis of variance approach to evaluate the distances between clusters. This method is regarded as very efficient. On the other hand, the most widely used method for non-hierarchical algorithms is the K-means method. The K-means method is known as a partitional method since the user must first predefine the number of clusters after which the algorithm partitions the data iteratively until a solution is found. Some researches have proposed a feasible solution for clustering by integrating the hierarchical method with the non-hierarchical method [6]. Thus, this study has integrated Ward's minimum variance method with the K-means method to come up with a cluster-based data mining cycle. A two-stage cluster-based data mining was proposed, in order to mine the patterns of anthropometric data for developing standard sizing systems.

3 The Data Mining Procedure

This study uses data mining cycle to establish the standard sizing systems with each respective step closely involved. The data mining cycle involves a series of activities, from defining the subject to evaluating and applying the results. The previous steps can be served as the baseline reference for the next step, and the steps for developing the standard sizing systems are described below.

3.1 Defining the problem for data mining

Owing to incomplete and outdated sizing systems for adult males in Taiwan, a large anthropometric database was created, based on 52 anthropometric variables, measured in each of 976 males, according to the definition of the ISO 8559; this resulted in 50,752 pieces of data [7]. The intent of this study was to explore and analyze a huge amount of data, by employing a cluster-based data mining cycle, so as to identify systematic patterns within body dimensions. Based on these patterns, the lower body types of Taiwanese adult males may be classified and standard sizing systems can be developed for use by manufacturers and consumers.

3.2 Data preparation and analysis

The data was processed, and analyzed, in order to enhance the efficiency and ensure the accuracy of the results [8]. Before mining the data, it had to be checked and processed, with all abnormal or missing data being separated out. As a result, of the 976 samples of adult males, 18, which had missing or abnormal data, were deleted; this left a total of 958 valid samples.

Not all of the 52 anthropometric variables were suitable for use in developing the standard sizing systems; therefore, in coordination with the judgment of domain experts, as well as international standards, this study identified 16 variables [9].

To use all of the 16 anthropometric variables, as a basis for developing standard sizing systems, would make things too complicated; therefore, the more important factors were identified first. Based on the results of the Bartlett's test ($p < 0.01$) and Kaiser-Meyer-Olkin measure of sampling adequacy (0.85), these 16 dimensions were all suitable for factor analysis. Factor analysis gave the eigenvalues of the 16 anthropometric variables. In accordance with Kaiser's eigenvalue criterion, two factors whose eigenvalues were greater than 1 were selected. Consequently, anthropometric variables, with factor loadings of greater than 0.5, were found to be clustered within factors 1 and 2. The major variables concentrated within factor 1 were waist girth, hip girth, thigh girth, mid-thigh girth, knee girth, calf girth, ankle girth, total crotch length and body rise; those in factor 2 included hip height, knee height, ankle height, waist to hips, outside leg length, thigh length and inside leg length. Therefore, two important factors were determined, with factor 1 being named the girth factor and factor 2, the height factor. The result is shown in Table 1.

Table 1. Factor analysis results

	Factor 1	Factor 2
Waist girth	−0.817*	0.389
Hip girth	−0.899*	0.318
Thigh girth	−0.753*	0.445
Mid-thigh girth	−0.726*	0.364
Knee girth	−0.769*	0.163
Calf girth	−0.789*	0.339
Ankle girth	−0.689*	0.245
Total crotch length	−0.669*	0.264
Body rise	−0.523*	−0.247
Hip height	−0.519	−0.829*
Knee height	−0.258	−0.635*
Ankle height	−0.198	−0.612*
Waist to hips	−0.224	−0.653*
Outside leg length	−0.483	−0.883*
Thigh length	−0.323	−0.647*
Inside leg length	−0.428	−0.823*
Variance explained	7.539	3.243
Total proportion	−0.492	0.335

* Factor loading > 0 .5

3.3 Data mining by cluster analysis

Through factor analysis, the girth factor and the height factor were found to be the most important factors in garment making. Subsequently, data mining was undertaken, using a two-stage cluster analysis, which included both hierarchical and non-hierarchical clustering. Ward's minimum variance method was integrated with the K-means method, to mine the patterns of anthropometric data, for developing standard sizing systems. Ward's minimum variance method was used to determine the initial clustering information for the K-means method, while the K-means method determined the final clusters.

In the first hierarchical clustering, this study analysed the factor scores of the girth factor and the height factor to decide the cluster numbers, using Ward's minimum variance method. A tree diagram, shown in Figure 1, presents the results. As shown, a total of 958 males were grouped from three to six obvious clusters; thus, these cluster numbers were chosen for the next stage of processing.

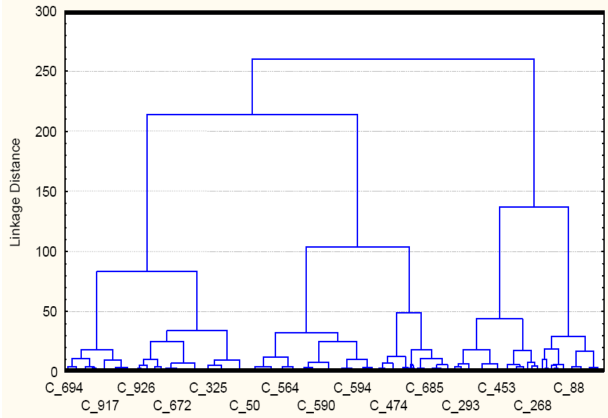

Fig. 1. The tree diagram for the cluster analysis

In the second non-hierarchical clustering, this study discovered that five clusters are the best for clustered result by the K-means method iteratively. Thus, this study analyzed all of the anthropometric data belonging to the 958 males, to decide the numbers and body types for each of the clusters by the K-means method. A total of 256 males were grouped into cluster 1 (body type C), with 52 males being grouped into cluster 2 (body type D), with 322 males being grouped into cluster 3 (body type B), with 198 males being grouped into cluster 4 (body type A), and 127 males into cluster 5 (body type Y). The result is shown in Table 1.

Table 1. Definitions of five body types

Clusters	5	4	3	1	2
Numbers	127	198	322	256	52
Girth variables	Smaller	Small	Medium	Large	Larger
Height variables	—*	—*	—*	—*	—*
Body types	Y	A	B	C	D

To gain a better insight into the differences between the five clusters (five body types) resulting from the two-stage cluster-based data mining, a line plot was drawn of the averages of the five body types and the 16 anthropometric variables. As can be seen in Figure 2, the five body types bear marked differences in the girth anthropometric variables, displaying a trend of body type D > body type C > body type B > body type A > body type Y. The height anthropometric variables did not have significant differences. Generally speaking, as far as garment manufacturing is concerned, the height variables of the lower body were similar within each body type.

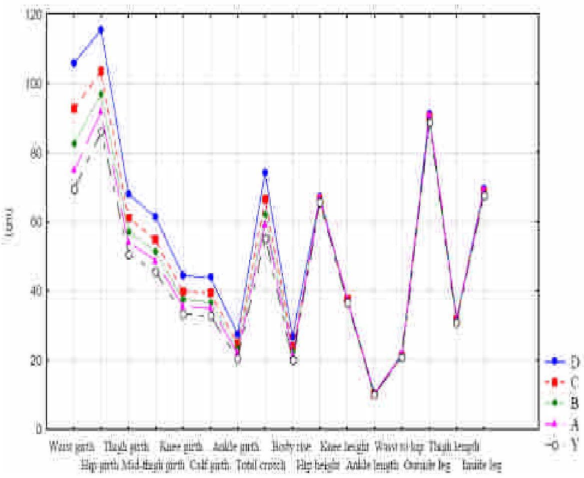

Fig. 2. The line plot of the means of anthropometric variables for each cluster

Therefore, according to the ISO/TR10652 [9], this study defined the body type, formed by cluster 2, with larger girth anthropometric variables, as type D; the body type, formed by cluster 3, with medium girth anthropometric variables, were defined as type B; the body type, formed by cluster 5, with smaller girth anthropometric variables, were defined as type Y; and the other body types, formed by cluster 1 and 4, were defined as type C and A. This definition of the five body types, used in this study, is shown in Table 1.

These five body types were produced by cluster-based data mining. Because hip girth, waist girth and outside leg length are the most important anthropometric variables in male garment manufacturing throughout the world, we drew a distribution graph of all five body types, with hip girth as the X-axis and waist girth as the Y-axis. By studying the distribution in detail, and coordinating our findings with the judgment of the domain experts, the standard sizing systems for the five body types were developed.

We take, as an example, the body type B, on the distribution graph. As most countries use 4 to 6cm as the interval for girth [7], and after coordinating this with the experts' judgment, as well as following the principle of "covering as many people as possible with the least number of sizes", this study set five sizes – 78cm, 82cm, 86cm, 90cm and 94cm - as representative hip girth sizes, and set three sizes - 96cm, 100cm and 104cm - as representative waist girth sizes, for body type B. The setting of sizes for the other body types was carried out in the same manner

Some samples were not included in determining the standard sizing systems for the five body types -- the D body type for example. Among them, the sample with 138cm hip girth and 110cm waist girth is at the far end of the scale for the standard sizing systems. The sample was eliminated, as it was felt to be unwise to add another group of sizes for such measurements, as it would increase costs. In the end, these 22 samples were still strictly excluded. Table 2 shows the body size distribution for all five body types. Out of the 958 samples, only 22 samples were excluded. Therefore, the coverage of the proposed "hip girth and waist girth" standard sizing systems was 97.7%. When the three sizing variables, hip girth, waist girth and outside leg length, were taken into account, the total coverage of was 95.6%.

Table 2. Body size distribution with five body types

			Outside leg length (cm)					
Body types	Waist girth	Hip girth	86	90	94	98	102	%
Y	88	70	86	90				2.5
		74	86	90	94			2.7
		78		90	94			1.1
	92	70			94	98		2.6
		74		90	94	98		3.3
A	92	78	86	90	94	98		5.7
		82		90	94	98		3.1
	96	74		90	94	98		4.3
		78	86	90	94	98		6.4
		82		90	94	98	102	6.3
	100	74			94	98		1.2
		78				98		2.6
B	96	86	86	90	94	98		3.8
		90		90	94	98		2.1
	100	82	86		94	98	102	4.7
		86	86	90	94	98	102	6.9
		90		90	94	98	102	6.6
		94			94			1.3
	104	78			94	98		1.2
		82			94	98		1.8
		86			94	98	102	2.1
		90				98	102	2.2
C	104	94		90	94	98	102	3.8
		98			94	98	102	2.4
		102		90	94	98		1.2
	108	90			94	98		2.3
		94			94	98		3.1
		98			94	98	102	2.9
		102			94	98	102	1.5
	112	98				98	102	2.1
		102					102	0.8
D	114	108				98		2.3
		114			94			0.9
	120	108			94	98	102	1.3
		114				98	102	0.9

3.4 Evaluation and Application of Results

Five body types were obtained using a cluster-based, data mining cycle, and male standard sizing systems were developed, according to the distribution conditions of the clusters, as well as the opinions of the domain experts. The newly standard sizing systems were found to have the following characteristics.

The coverage rate of the standard sizing systems, including the three main anthropometric variables, was 95.6%. This study generated simple body size distribution with hip girth and outside leg length based on the format of the previous system. Forty-three groups were found in the simple body size distribution, which were fewer than the previously used system. In practice, manufacturers hope to work with as few sizes as possible, as too many sizes can result in too much inventory, which can encumber cash flow.

To obtain detailed production information, manufacturers may refer to Table 2. Taking the 100cm waist girth of body type B as an example, it can be seen that this waist girth of 102cm, matches up with four waist girths, of 82cm, 86cm, 90cm and 94cm; this means that garment with four different hip girths can be produced for one waist girth of 100cm. Of course, outside leg length must also be taken into account when planning the production of garment of a certain size. The newly standard sizing systems can supply basic sizes with pattern makers who design garment patterns. The percentage of males, within each particular body type and size were also recorded in the body size distribution; this may result in more accurate production planning and materials control for specific markets. The standard sizing systems established by this study are a basis for garment manufacturing specifications; the data obtained from standard sizing systems are actual human measurement values, and are highly flexible when used. Manufacturers can make different types of garment with various allowances, by referencing these newly standard sizing systems.

These standard sizing systems used the size labels as a reference [9]. For example, 96B86 means that waist girth is 96cm, body type is B, and outside leg length is 86cm. In this way the detailed body dimensions of a male can be easily determined, using these easy-to-understand sizing systems. Moreover, it is very convenient for men to be able to find suitable clothes within a short time. The size labels can also be used as a communication tool among pattern makers, manufacturers, retailers and consumers. Of course, no size chart is suitable for everyone. If a man finds it difficult to find a suitable size, he would be well advised to try the size closest to her own, with slightly larger being preferable.

4 Conclusion

Garment production is a high value-added manufacturing process, so accurate standard sizing systems are critical aspect for garment manufacturing industry. This study applied a cluster-based data mining cycle, using anthropometric data, to develop standard sizing systems for adult males in Taiwan. The advantages observed were as follows. (1) The total coverage rate of the standard sizing systems reached 95.6%, with the total number of size groups being only 43, fewer than the number of groups in previous used system. The standard sizing systems also used simple and easy-to-understand size labels to describe body dimensions, enabling consumers to quickly find suitable clothes. (2) These standard sizing systems also provided the percentage of males within each size group, for every body type, as well as the distribution of body types; this allows manufacturers access to reference points, facilitating garment production.

References:

[1] Regan CL, Kincade DH, Sheldon G, Applicability of Engineering Design Process Theory in the Apparel Design Process, *Clothing and textiles Research Journal*, 16, 1998, pp.36-46.

[2] Hsu KM, Jing SH, The chances of Taiwan apparel industry, *Journal of the China Textile Institute*, 9, 1999, pp.1-6.

[3] Burns LD, Bryant NO, *The Business of Fashion: Designing, Marketing, and Manufacturing*, Fairchild, 2000.

[4] Berry M, Linoff G, Data Mining Techniques: for Marketing, Sales, and Customer Support, Wiley, New York, 1997.

[5] Giudici P, Applied Data Mining: Statistical Methods for Business and Industry, Wiley, England, 2003.

[6] Kuo RJ, Ho LM, Hu CM, Cluster analysis in industrial market segmentation through artificial neural network. *Computer and Industrial Engineering*, 42, 2002, pp.391-399.

[7] ISO, Garment Construction and Anthropometric Surveys – Body Dimensions. International standard, ISO 8559, 1989, pp.1-9.

[8] Pyle D, Data Preparation for Data Mining, Morgan Kaufmann, California, 1999.

[9] ISO, Standard Sizing Systems for Clothes. Technical Report, ISO/TR10652, 1991, 19-25.