

Employing a Three-Stage Data Mining Procedure to Develop Sizing System

¹R. Bagherzadeh, ¹M. Latifi and ²A.R. Faramarzi

¹ATMT Research Institute, Textile Engineering Dept.,
Amirkabir University of Technology, Tehran, Iran

²Faculty of Engineering, Golestan University, Gorgan, Iran

Abstract: In clothing engineering, sizing system plays an important role for manufacturing of clothing. This study aims to introduce a strong approach that it could be used for developing sizing systems by data mining techniques using anthropometric data. By using a new approach of Three-stage data mining procedure lower body figure type of Iranian male people determined. This approach included three phases. First of all, Factor analysis was employed to mine through the variable and to determine the main effective factors. Second, by using a two-step cluster analysis included both hierarchical and non-hierarchical clustering, cases sorted to the clusters according to the factor analysis results. Third, the decision tree analysis employed to extract the most significant classification rules of the body type based on the results of cluster and factor analysis. A sizing system developed for the Iranian male age between 16 and 22, based on the body type determined in the data mining procedure. The result showed that three body type and sizing system developed, have a good fit performance.

Key words: Sizing system • Data mining • Clothing industry • Anthropometric data

INTRODUCTION

Garment sizing systems were originally based on those developed by tailors in the late 18th century. Professional dressmakers and craftsmen were developed various sizing methods in the past years. They used unique techniques for measuring and fitting their customers. In the 1920s, the demand for the mass production of garments created the need for a standard sizing system. After that many researcher worked on developing sizing system by the different methods and data gathering approaches. It is proved that garment manufacturing is the highest value-added industry in the textile industry manufacturing chain [1]. Mass production by machines in this industry has replaced manual manufacturing, so the planning and quality control of production and inventory are very important for manufacturer. Moreover, this type of manufacturing has demand to certain standards and specifications. Furthermore each country has its own standard sizing systems for manufacturers to follow and fit in with the figure types of the local population.

A sizing system classifies a specific population into homogeneous subgroups based on some key body

dimensions [2]. Persons of the same subgroup have the same body shape characteristics and have the same garment size. Standard sizing systems can correctly predict manufacturing quantity and proportion of production, resulting more accurate production planning and control of materials [3, 4]. The standard sizing systems have been used as a communication tool among manufacturers, retailers and consumers. It can provide manufacturers with size specification, design development, pattern grading and market analysis. Manufacturers, basing their judgments on the information, can produce different type of garments with various allowances for specific market segmentation. Thus, establishing standard sizing systems are necessary and important. Many researchers worked on developing the sizing system by many approaches. They found very extensive data were made by using anthropometric data [2]. People have changed in body shape over time. Workman [5] demonstrated that the problem of ageing contributes to the observed changes in body shape and size, more than any other single factor, such as improved diet and longer life expectancy [6]. Sizing concerns will grow as the number of ageing consumers is expected to double by the year 2030. This presents a marketing

challenge for the clothing industry since poor sizing is the number one reason for returns and markdowns, resulting in substantial losses. Therefore, sizing systems have to be updated from time to time in order to ensure the correct fit of ready-to-wear apparel. Many countries have been undertaking sizing surveys in recent years. Since sizing practices vary from country to country, in 1968 Sweden originated the first official approach to the International Organization for Standardization (ISO) on the subject of sizing of clothing, it being in the interest of the general public that an international system be created. After lengthy discussions and many proposals, members of technical committee TC133 submitted documents relating to secondary body dimensions, their definitions and methods of measuring. This eventually resulted in the publication of ISO 8559 'Garment Construction and Anthropometric Surveys - Body Dimension', which is currently used as an international standard for all types of size survey [7]. The ISO is now a worldwide federation of national standards bodies from more than 140 countries. The latest version of international standard for clothes is found in ISO/TR 10652:1991. Further attempts done to establish a uniform size designation have been made by several organizations such as EN 13402 for the European size standard and so on.

Figure type plays a decisive role in a sizing system and contributes to the topic of fit. So to create a sizing system, different body types are first divided from population, based on dimensions, such as height or ratios between body measurements. A set of size categories is developed, each containing a range of sizes from small to large. The size range is generally evenly distributed from the smallest to the largest size in the most countries. For men's wear, the body length and drop value are the two main measurements characterizing the definition of figure type. Maier has compared existing sizing systems throughout the world [8]. In the UK, BSI identified three body heights; short (166 cm), regular (174 cm) and tall (182 cm). Each has two figure types; mature (drop ± 14 cm) and athletic (drop ± 17 cm) [2]. Emanuel, *et al.* [9] recommended the use of the difference in figure types as the classification of ready-to-wears and developed a set of procedures to formulate standard sizes for all figure types. In early times, the classification of figure types was based on body weight and stature. Later on, anthropometric dimensions were applied for classification. Based on this method, a linear structure was found in many of the commonly used sizing systems, such as KS K 0050 and JIS L 4002. This type of sizing system has the

advantages of easy grading and size labeling. But, the disadvantage is that the structural constraints in the linear system may result in a loose fit. Thus, some optimization methods have been proposed to generate a better fit sizing system, such as an integer programming approach [10] and a nonlinear programming approach [11]. For the development of sizing systems using optimization methods, the structure of the sizing systems tends to affect the predefined constraints and objectives. Tryfos [10] indicated that the probability of purchase depended on the distance between the sizing system of a garment and the real size of an individual. In order to optimize the number of sizes so as to minimize the distance, an integer programming approach was applied to choose the optimal sizes. Later on, McCulloch, *et al.* [11] constructed a sizing system by using a nonlinear optimization approach to maximize the quality of fit. Recently, Gupta, *et al.* [12] used a linear programming approach to classify the size groups. Using the optimization method has the advantages of generating a sizing system with an optimal fit, but the irregular distribution of the optimal sizes may increase the complexity in grading and the cost of production.

On the other hand, in recent years, data mining has been widely used in area of science and engineering. The application domain is quite broad and plausible in bioinformatics, genetics, medicine, education, electrical power engineering, marketing, production, human resource management, risk prediction, biomedical technology and health insurance. In the field of sizing system in clothing science, data mining techniques such as neural networks [13], cluster analysis [14], the decision tree approach [15] and two stage cluster analysis [16] have been used. Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait. Cluster analysis was used as an exploratory data analysis tool for classification. In the clothing a cluster which is typically grouped by the similarity of its members' body shape can be considered as a size category or a figure type. Moon and Nam [14] used the K-means cluster analysis method to classify the lower body shapes of elderly women and then used the control dimension and size interval to establish a lower garment sizing system. The pitfall of these methods is that it requires one to pre-assign the number of clusters to initialize the algorithm and it is usually subjectively determined by experts. Also by use the two stage clustering, it cannot define the rule of classification for categorizing the figure type. To

overcome these disadvantages, a three stage-based data mining procedure include cluster analysis, factor analysis and after that, decision tree algorithm, is proposed here to eliminate the requirement of subjective judgment and to improve the effectiveness of size classification. Also by using the decision tree algorithm, rule of classification can be derived for ever changing and it help for pre-assign number of figure type for other purpose.

Data Mining Procedure

Data Preparation: After the definition of industry problem, first stage of data mining, data preparation selected to increase the efficiency and ensure the accuracy of its analysis through the processing and transformation of the data. Before starting to mine the data, they must be examined and proceed with all missing data and outliers. By examining the data before the application of a multivariate technique, the researcher gains several critical insights into the characteristics of the data. In this study, we used an anthropometric database which was collected in a protective clothing producer company in Iran. All the missing or abnormal data were omitted and after that, anthropometric data of 1050 Iranian male with the age ranged from 16 to 22 years from the database were obtained.

Factor Analysis: In the first step of data mining approaches, we used factor analysis that provides direct insight into the interrelationships among variables or respondents and empirical support for addressing conceptual issues relating to the underlying structure of the data. It also plays an important complementary role with other multivariate techniques through both data summarization and data reduction. If the number of variables is too large or there is a need to better represent a smaller number of concepts, factor analysis can assist in selecting a representative subset of variables or even creating new variables as replacements for the original variables while still retaining their original character. From the data summarization perspective, factor analysis provides the researcher with a clear understanding of which variables may act in concert together and how many variables may actually be expected to have impacts in the analysis. In factor analysis, the factors are formed to maximize their explanation of the entire variable set, not to predict a dependent variable(s).

Here, two most important factors were selected though 8 anthropometrics data using Cattell's Scree Test. Eigen values, in *Cattell's Scree Test* obtained from a single analysis are plotted and an inflection point of the

Table 1: Results of factor analysis

	Factor 1	Factor 2
Body height	0.654	0.864*
Hip girth	0.775*	0.334
Crotch height	0.455	0.815*
Knee girth	0.764*	0.286
Waist girth	0.87*	0.454
Thigh girth	0.795*	0.487
Crotch length	0.687	0.802*
Weight	0.854*	0.215

*Factor loadings > 0.75

resulting curve (scree) is determined by visual inspection. The location of the inflection points indicates the number of factors to be extracted. Then two factors determined using calculated factor loading 0.75.

It could be concluded from the result of factor analysis (Table 1) that factor 1 has the most correlated variables that related to girth measurements include hip girth, knee girth, waist girth and weight (that is a exception). Also variables related to the height properties of the body, highly correlated with factor 2. Furthermore, in factor 1 and factor 2, waist girth and body height, respectively, are the most correlated and were transformed into a new variable, waist to height ratio (WHR) that used as the target variable of the decision tree.

Cluster Analysis: Second step of data mining approach was undertaken, using the cluster analysis, included both hierarchical and non-hierarchical clustering. EML (Equal-Variance Maximum Likelihood) method was selected to mine through the data to find the number of cluster to eliminate the subjective evaluation (by the experts) to determining the number of clusters. The EML method joins clusters to maximize the likelihood at each level of the hierarchy under the following assumptions.

- Multivariate normal mixture.
- Equal spherical covariance matrices.
- Unequal sampling probabilities.

The EML method is similar to Ward's minimum variance method but removes the bias toward equal-sized clusters. Practical experience has indicated that EML is somewhat biased toward unequal-sized clusters. With the difference between body height and waist girth, we determined the cluster numbers. In the second stage of cluster analysis, K-means method implemented to determine the final cluster classification. The result of two-stage cluster analysis is shown in Figure 1.

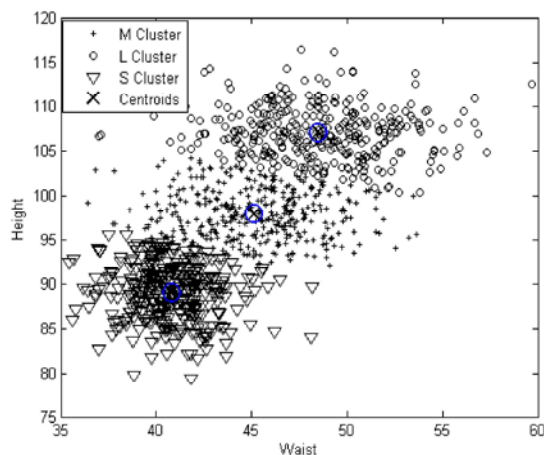


Fig. 1: Height versus waist girth for the three clusters.

This figure illustrates, for all samples, the distribution of height versus waist girth. It can be observed that there are three distinct clusters that named Small (S), Medium (M) and Large (L).

Cluster S with having 416 samples and cluster M and cluster L having 324 and 320 samples, respectively.

To investigate the accuracy of two stage of cluster analysis, data were examined with another multivariate analysis (MVA). With using General Linear Model in SPSS software, data was investigated to determine whether significant differences existed among the difference cluster S, M and L. The results of the MVA (with P-value <0.05) indicated that for the three clusters, there is a significant differences in the waist girth, body height and hip girth variables. After determining the significant difference between variables, Duncan Post hoc test was used to detect the means of variables differences among body types. From the result of Duncan test concluded that there is a significant difference between the mean of body height, waist girth and hip girth for the three clusters. Therefore, analyzing of two stages of data mining approaches indicated that, there are three distinct clusters through anthropometric data of lower body that used to garment manufacturing.

Decision Tree Analysis: In third stage of procedure, the decision tree was used to mine data based on the results of cluster analysis. There has been increasing interest in the use of the Classification and Regression Tree (CART) technique in the similar field such as identifying the significant rules of body type for application in medical, sport and so on. The most important property of CART that caused to use them is ability to reduce the complexity of the decision tree,

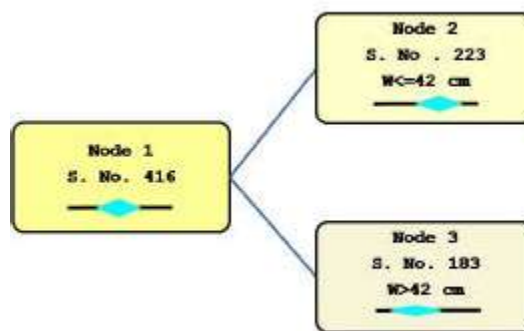


Fig. 2: Graphical view of the CART for the S cluster

including its depth and the number of child nodes, while enhancing the accuracy of classification. Each splitting reduces the impurity of the child nodes, until the samples in every child node have the same purity [17]. Consequently in this approach, we used the CART to gain the rules of classification of the body type.

As we mentioned before, it could be concluded from the Table 1 that waist girth is most correlated with factor 1. Furthermore, in garment manufacturing waist girth is the most important variable of pants production [18]. Therefore, waist girth was selected to represent the factor 1.

To start a decision tree process we need to define some variable as “predictor” and a variable as “target variable”. In this study we take the waist-to- height ratio (WHR) as the target variable, with waist girth and hip girth being predictors used to classify the target variable. Also such as any classification and optimization approaches we need to define the stopping rules as follow:

- When the amounts of purity of the Childs have no significant improve in comparison to their parents.
- When the minimum number of samples in the parent node is 4, or /and the minimum number of samples in the child node is 2.
- The greatest depth of the tree extends to the three levels after the root node.

Because the long tree shape of analyzing in this step, we take the cluster 1 (S Cluster) as an example and Figure 2, presents the results of decision tree analysis for this cluster with the defined variables.

In the first dividing point in the decision tree, the root node was divided to two nodes according to the waist girth variable. From total of 416 samples in the parent node, whose waist girths were smaller than or equal to

Table 2: The classification rules extracted for the different body type resulted from CART

Cluster type	Body type	Rule of classification	No. in each node
S Cluster	SWL	Waist girth greater than 42 cm (W42)	183
	SWS	Waist girth smaller than or equal to 42 cm (W42 cm)	223
M Cluster	MWL	Waist girth greater than 45 cm (45W)	157
	MWS	Waist girth smaller than or equal to 45 cm (45W)	167
L Cluster	LWL	Waist girth greater than 49cm (W49cm)	153
	LWS	Waist girth smaller than or equal to 49cm (W49cm)	167

42 cm were grouped into node 2 (223 samples) and 183 samples with waist girths greater than 42 cm were classified into the node 3. So, at the first level of process only the two child nodes generated by the rule of waist girth. With continuing the algorithm, more new child nodes generated according to the waist girth and hip girth in the dividing point of parent nodes. Although, more depth in the algorithms, more precise in classification rule, but in garment manufacturing producing such a width range of product according to the many classification rules extracted through the decision tree approach is not appropriated and we used the first level of division. Table 2 shows the classification rules for the all cluster body type extracted through the CART.

So, by the using of two-stage cluster analysis, three distinct clusters (S, M and L) were found. Body clusters could be classified to the sub-clusters according to the hidden rules of the data by the CART result. These results not only capable to use in the clothing industry such as sizing system, but also could be used in the other fields like physiology, medical, human ecology, sports and so on. Therefore in this research we presented a well-defined figure type classification that could be continue to the classification according to the industrial (application) demands. The S cluster figure type with waist girth smaller than or equal to 42 cm defined as type SWS and the others in the S cluster, defined as type SWL; The M cluster figure type that waist girth of them is smaller than or equal to 45 cm named as type MWS and the M cluster figure type, which waist

girth is greater than 45 cm labeled as type MWL; the L cluster figure type formed with waist girth smaller than or equal to 49 cm classified as type LWS and the L cluster with waist girth greater than 49 cm classified as type LWL.

DEVELOPING THE SIZING SYSTEM

As in international standards such as ISO/TC and those mentioned in the literature review, use of control dimensions and size interval can effectively facilitate to recognize the parameters for developing sizing systems [19]. In this research the control dimensions were height and waist girth. The size interval for height was 5 cm and for the waist girth, it was 4 cm. first of all, we categorized each figure type into several size subgroups based on the control dimensions and size interval and second, according to the control dimensions and size intervals, sizing system for Iranian male pants developed. Table 3, illustrates the sizing system for the Iranian male with age between 16 -22.

The size table for the pants of Iranian male with age between 16 -22 has showed in the Table 4. Each size label presents 6 reference dimensions for pants.

In the sizing system, definition of the size label is a critical issue that it determines quickly locating the right garment size for further consideration for customers as an interface. In this study we used a new definition size label, for instance SWS36-85, that it means this size of pants is suitable for the small figure type with waist girth between 36 and 40 cm and height between 85 and 95 cm.

Table 3: The sizing system for the Iranian male with age between 16 -22

Height (cm)					
110			LWS44-110	LWL48-110	LWL48-110
105			MWS44-105	MWL48-105	LWL48-105
100		MWS40-100	MWS44-100	MWL48-100	MWL52-100
95	SWS36-95	SWS40-95	SWL44-95	MWL48-95	
90	SWS36-90	SWS40-90	SWL44-90		
85	SWS36-85	SWS40-85			
	36	40	44	48	52
					56
Waist girth (cm)					

Table 4: The size table for the pants of Iranian male with age between 16 -22

Size Label	Control Dimension (cm)				Reference Dimension (cm)			
	Waist Girth	Height	Hip girth	Crotch height	Knee girth	Thigh girth	Crotch length	Weight (kg)
SWS36-85	36	85	42.3	24.2	22.4	25.0	21.4	38.5
SWS40-90	40	90	46.2	25.4	25.7	27.9	22.6	42.6
SWL44-95	44	95	50.3	26.8	29.6	30.6	24.0	45.0
MWL48-100	48	100	52.6	30.7	33.5	35.0	26.8	49.4
LWL48-105	52	105	63.8	33.6	36.4	38.6	30.6	54.0
LWL56-110	56	110	69.3	38.4	36.2	44.7	36.1	62.7

To evaluate the approach that we used to identify the significant rules for classifying the body type and as an application, developed sizing system, we used aggregate loss as a measure of goodness of fit sizing system to the whole population. When all individuals' distances from their assigned size are averaged over the whole population, an aggregate measure is created that represents how well the sizing system performs in fitting the population. In such a good sizing system the average distance of individuals from their size is as low as possible. Therefore, on average, the system provides the best fit. In this study according to the body measurements that used to create the sizing system, the quadratic average of log differences was chosen as the distance measure in calculation of aggregate loss. The advantage of this simple measure is that both positive and negative differences increase the distance. Therefore the fit is worse both when the garment is too big and when the garment is too small along a particular dimension [20].

Based on the results of factor analysis and the critical dimensions in clothing design, we used waist girth, hip girth and crotch height to calculate the aggregate loss for the pants. The aggregate loss of the young male age between 16- 22, was 6.54 cm, that is similar to the results reported by [21], (6.77 cm for the lower garment). Thus, the performance for the size system developed is well-good fit and is satisfactory.

CONCLUSION

This study proposed a three-stage data mining procedure applied to classifying the human body type (lower body). This well design approach has many advantages of capability to use for applications such as clothing industry, physiology, medical treatments, sport and so on. To approve this approach, we developed a sizing system for pants of Iranian males and showed that there is a good fit performance. Also, the rule of classification of the lower body types

extracted can help the garment manufacturers to understand better the customer's body shape characteristics. Moreover, the proposed approach in this study can provide an effective procedure of extraction of the significant rules to categorize human body type for many applications.

REFERENCES

1. Chang, C.F., 1999. The model analysis of female body size measurement from 18 to 22, J. Hwa Gang Textile, 6: 86-94.
2. Fan, J., W. Yu and H. Lawrance, 2004. Clothing appearance and fit: Science and technology, Woodhead Publishing Limited, Cambridge, England.
3. Tung, Y.M. and S.S. Soong, 1994. The demand side analysis for Taiwan domestic apparel market, J. the China Textile Institute, 4: 375-380.
4. Hsu, K.M. and S.H. Jing, 1999. The chances of Taiwan apparel industry, J. the China Textile Institute, 9: 1-6.
5. Workman, J.E., 1991. Body measurement specification for fit models as a factor in apparel size variation, Cloth Text Res. J., 10(1): 31-36.
6. LaBat, K.L. and M.R. Delong, 1990. Body cathexis and satisfaction with fit of apparel, Cloth Text Res. J., 8(2): 42-48.
7. ISO 8559, 1989. Garment Construction and Anthropometric Surveys - Body Dimensions, International Organization for Standardization.
8. Maier, A., 2000. International size chart comparison for men, Knitting Technol., 22(1): 32-36.
9. Emanuel, I., M.Alexander, E. Churchill and B. Truett, 1959. A height-weight sizing system for flight clothing", Wright Air Development Center, WADC Technical Report 56-365, Wright-Patterson Air Force Base, Ohio.
10. Tryfos, P., 1986. An integer programming approach to the apparel sizing problem, J. the Operational Research Society, 37(10): 1001-1006.

11. McCulloch, C.E., B. Paal and S.A. Ashdown, 1998. An optimal approach to apparel sizing, *J. the Operational Res. Society*, 49: 492-499.
12. Gupta, D., N. Garg, K. Arora and N. Priyadarshini, 2006. Developing body measurement charts for garments manufacture based on a linear programming approach, *J. Textile and Apparel Technology and Management*, 5(1): 1-13.
13. She, F.H., L.X. Kong, S. Nahavandi and A.Z. Kouzani, 2002. Intelligent animal fiber classification with artificial neural networks, *Textile Research J.*, 72(7): 594-600.
14. Moon, J.Y. and Y.N. Nam, 2003. A study the elderly women's lower body type classification and lower garment sizing systems, *Proceedings of International Ergonomics Association Conference*.
15. Hsu, C.H. and M.J. Wang, 2005. Using decision tree based data mining to establish a sizing system for the manufacture of garments, *International J. Advanced Manufacturing Technol.*, 26(5& 6): 669-674.
16. Meng, J.C., L. Hai and J.J.W. Mao, 2007. The development of sizing systems for Taiwanese elementary- and high-school students, *International J. Industrial Ergonomics*, 37: 707-716.
17. Moisen, G.G. and T.S. Frescino, 2002. Comparing five modeling techniques for predicting forest characteristics, *Ecological Modeling*, 157: 209-225.
18. Jongsuk, C.Y. and C.R. Jasper, 1993. Garment-sizing systems: an international comparison. *Int. J. Clothing Sci. Technol.*, 5(5): 28-37.
19. Winks, J.M., 1997. Clothing sizes: international standardization, Redwood, U.K.
20. Susan, P.A., 1998. A comparison of three multidimensional optimized sizing systems generated from anthropometric data with the ASTM standard D5585-94, *Int. J. Clothing Sci. Technol.*, 10(5): 324-34.
21. Gupta, D. and B.R. Gangadhar, 2004. A statistical model for developing body size charts for garments, *International J. Clothing Sci. and Technol.*, 16(5): 458-469.