Data mining-based optimal assignment of apparel size for mass customization

Zhu-Jun Wang[1,2], Cheng Chi[3], Meng-Yun Zhang[3], Xian-Yi Zeng[3], Pascal Bruniaux[3], Jian-Ping Wang[1,*], Ying-Mei Xing[2], Shuo Xu[1]

[1]*College of Fashion and Design, Donghua University, Shanghai, China*
[2]*School of Textile and Garment, Anhui Polytechnic University, Wuhu, China*
[3]*GEMTEX Laboratory, Ecole Nationale Superieure des Arts et Industries Textiles, Roubaix, France*
*Corresponding author *E-mail address:* wangjp@dhu.edu.cn

INFO

ABSTRACT

In this study, we have explored and discussed the data mining-based solutions to apparel size assignment using an approach principle, K-means clustering, and support vector machine, respectively. A case of mass customization for men's pants in China with 200 adult males were employed to validate and evaluate the solutions. After anthropometric data acquisition and preprocessing, three key body dimensions were identified based on hierarchical clustering as well as their ranges and fit models. Sequentially, we calculated all the possible values of the distance between the target population and fit models by the enumeration algorithm. Afterward, we assigned the garment sizes for the target population using the above mentioned data mining approaches. Finally, the solution based on the support machine was considered as the optimal solution for the pant mass customization after being comprehensively assessed by the aggregate loss of fit, the number of poor fit, accommodation rate of ideal fit, and the number of garment size employed, since it employed only 48 sizes to reach the accommodation rate of target population up to 82%. The experimental results demonstrate that the present solution is a low-cost method for the size assignment by exploiting the potentials of existing sizing system, instead of creating new sizing systems, and also easy to be flexibly extended to any types of garments.

Keywords

1 Introduction

Among all facets associated with the purchase decisions of consumers, wearing comfort represents a fundamental and central concern, which is linked inextricably with a garment fit, which largely relies on the garment design and the relative size between body and clothing [1]. Appropriate garment fit can not only provide the adequate space allowance for body movements, but also give consumers a rewarding experience of self-expression and self-image [2]. Most recently, with the rapid development of mobile e-commerce, along with the convenient shopping way, the personalized demands for garments such as styles, colors, and fit have been dramatically enhanced. Enterprises have paid attention to mass customization (MC) in the clothing industry [3]. MC means enterprises are driven by personalized orders, offering individualized products by the industrial manufacturing process. Contributing to the advanced manufacturing technologies, i.e., 3D whole-body scanning, virtual reality, and apparel CAD/CAM, the garment enterprises can realize the customization using the full automation systems preliminarily. However, such a manner could not be made available and flexible under the circumstances of MC, i.e., small series production. Furthermore, the high costs of expensive systems with advanced technologies also hinder their applications in the garment MC. Therefore, it is a paramount issue of how to meet the needs of individual consumers, while getting the benefits of economy of scale should be concerned comprehensively [4].

Ordinarily, in the context of MC, garments are manufactured following the process illustrated in Fig. 1. First of all, the anthropometric data of target consumers are acquired after the personalized garment order is placed and signed. Before drafting the garment patterns for the target population, the body sizes and garment sizes are identified and determined based on the sizing system and the returned merchandising reports. Afterward, the products are produced by the industrial manufacturing process. In this way, the garment sizing system is the inextricable linkage of the development of personalized garments and the industrial manufacturing process. With the help of sizing system, the manufacturers could only utilize a limited number of patterns, while simultaneously catering to the fit requirements of the individual consumer in the industry. Besides, the correct assignment for apparel sizes is said to have the advantage of predicting sales of diverse sizes and determine production quantities, resulting in the accurate material cost control and manufacturing planning. However, the size assignment is still a challenging issue to resolve in the garment MC. Much of research have examined the utilization of advanced technologies, such as traditional step-wise methods, multivariable approaches, and optimization techniques, concerning the assignment of apparel sizes [5-11]. Nevertheless, plenty of the studies concentrated on establishing a new sizing system only, instead of exploiting the potentials of the existing sizing system. The effectiveness and feasibility of the new sizing system still need to be further verified. In practice, improving the existing one is more economical than developing an alternative. Moreover, the improved one can be more easily implemented by the operators in the enterprises.
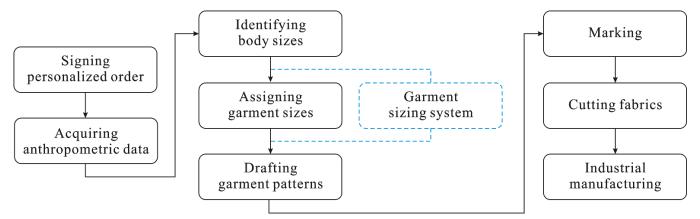

Fig. 1   Flowchart of garment mass customization

Data Mining (DM) is an interdisciplinary technique focusing on exploiting useful knowledge from a large dataset, involving the application of algorithms that explore the data, develop models and discover the useful knowledge that supports decision making, which has been viewed as an efficient tool of solving

complicated engineering problems [12]. Therefore, an approach of optimizing sizing assignment toward MC was proposed based on the DM technique in this study, to fulfill the individualized fit requirement for consumers with few quantities of sizes.

2　Method

2.1　Research scheme

Fig.2 presents the research process in this study. Research work can be split into five sequential procedures, i.e., an anthropometric data acquisition, data preprocessing, body size identification, apparel size assignment, and evaluation. We mainly took into account the following factors affecting the size assignment: the target population, garment, standard sizing system, manufacturer as well as the market feedback.

Step 1: The anthropometric data of the target population were measured according to the desired garment style. Both the manual and automatic machine measurement are available.

Step 2: The data noises are inevitable during the process of data acquisition. To minimize the impact of data noises, we preprocessed the original data obtained in the previous step using the detection of outliers, descriptive analysis, and test of normality.

Step 3: The hierarchical algorithm identified key body dimensions of the target population.

Step 4: According to the target population and the relevant standard sizing system, we determined the intervals, fit models, and size range. The assignment of apparel sizes was executed using approaching principle, *K*-means clustering, and support vector machine, respectively.

Step 5: The optimal solution to size assignment was found by evaluating the experimental results eventually.
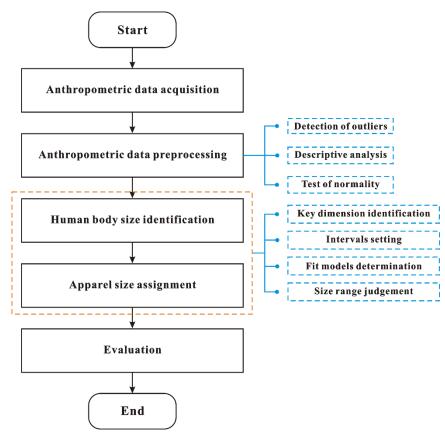

Fig. 2　Research scheme

2.2 Applied algorithms

2.2.1 Hierarchical clustering algorithm

To reflect how the body dimensions are merged or divided evidently, the hierarchical clustering algorithm was employed to identify the key body dimensions after anthropometric data of the target population were captured and preprocessed.

2.2.2 Enumeration algorithm

Enumeration algorithm was utilized by listing all the possible values while computing the distance between the body dimensions of target population and the dimensions of the fit models.

2.2.3 K-means clustering algorithm

K-means algorithm is a popular partitioning approach. The classic clustering procedure was as follows: (a) Choose K as the number of clusters; (b) Initialize the cluster centroid vector $\{c_1, c_2, \ldots, c_k\}$; (c) For each new object vector, calculate the distance between the new object vector and cluster centroid vector; (d) Assign the new object to its closest cluster centroid, and re-compute the centroid of the closest cluster with the new object; (e) If the position of any cluster centroid changed, return to (c), otherwise, stop.

2.2.4 Support vector machine method

Due to the distinct advantages, i.e., an excellent generation performance, fast convergence, and robustness for noise, support vector machine (SVM) have been diffusely applied in solving nonlinear, high-dimensional pattern recognition, classification, and regression problems. SVM was introduced to assign garment size in this work.

3 Test

In this section, we adopt a case of men's pants to interpret the implementation of the present solution to the garment size assignment.

3.1 Data acquisition

Owing to the advantages of taking automatic body measurements in a manner of non-contact precisely and expeditiously, we procured the anthropometric data of 200 adult males aged from 18 to 55 using the Vitus Bodyscan shown in Fig. 3.

Fig. 3　Vitus Bodyscan

According to the characteristics of the garment type and the relationship with garment pattern blocks, we extracted nineteen lower body measurements from the whole body datasets. They were: stature, waist height, abdomen height, hip height, knee height, inseam length, waist girth, abdomen girth, hip girth, thigh root girth, mid-thigh girth, crotch length, knee girth, waist width, hip width, waist depth, abdomen depth, mid-hip depth, and hip depth. Fig. 4 demonstrated, how the body dimensions were measured.
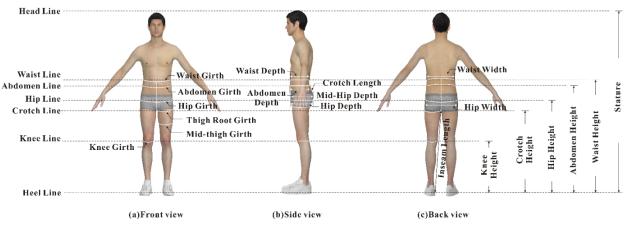

Fig. 4　Diagram of lower body measurements

3.2　Data preprocessing

3.2.1　Detection of outliers

In the statistical analysis, the outliers are a set of data to be an observation or subset of data that appears to be inconsistent with the remainder of that set of data. Outliers in a set of data will influence the modeling precision. Thus, the original data were conduct outlier analysis by the method of 3σ-rule. Once the outliers were detected, the data will be re-measured and executed outlier analysis again. If they were still outliers, it could be inferred that they were particular dimensions. We maintained all the particular dimensions in this study.

3.2.2　Descriptive analysis

Descriptive analysis is a simple, straightforward, and practical approach to mirroring the data features. The features of the experimental data were presented in Table 1.

Table 1. Description of the anthropometric dimensions

Measurements [cm]	Sample number	Minimum [cm]	Maximum [cm]	Mean [cm]	Standard deviation [cm]
Stature	200	150.85	183.21	166.04	5.32
Waist height	200	93.42	116.41	102.89	4.09
Abdomen height	200	83.50	102.76	92.28	4.13
Hip height	200	72.30	88.12	80.06	3.40
Knee height	200	39.28	49.72	44.34	2.16
Inseam length	200	64.45	83.34	74.55	3.79
Waist girth	200	61.21	101.90	79.75	8.87
Abdomen girth	200	63.10	101.95	82.41	7.50
Hip girth	200	76.57	105.25	90.69	5.04
Mid-thigh girth	200	40.10	58.00	49.03	3.48
Crotch length	200	65.50	90.30	77.50	4.67
Knee girth	200	31.80	41.50	36.17	1.92
Thigh root girth	200	42.20	62.90	51.67	3.87
Waist width	200	21.25	32.46	26.64	2.37
Hip width	200	28.00	35.70	31.81	1.55
Waist depth	200	14.10	29.35	20.73	3.11
Abdomen depth	200	14.70	27.90	20.84	2.77
Hip depth	200	15.40	26.35	20.63	2.28
Mid-hip depth	200	15.82	25.87	20.95	2.08

3.2.3 Test of normality

The normality test was utilized to check if the data sample deviates from the Gaussian distribution. In this study, all the measurements were tested for the normality. From Fig. 5, the data of waist girth agreed with the normal distribution. After being tested, all the measurements used in this study were following the normal distribution.
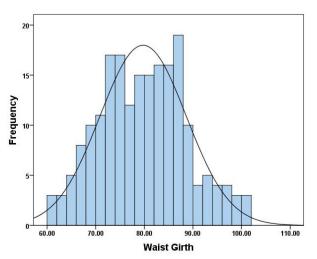

Fig. 5   Histogram of waist girth

3.3   Preparation for garment size assignment
3.3.1   Key body dimension identification

Key body dimension (KBD) denotes the essential dimensions influencing the garment fit [13]. Since the hierarchical clustering algorithm has the advantages of ease of handling any attributes type and interpreting the consequence, we applied it to identify the KBD in this study. The dendrogram (see Fig. 6)

presented the clustering results and order for each anthropometric measurements. All of dimensions could be split into three blocks. Stature, waist girth, and hip girth were the first dimension clustered in each block. Therefore, they were selected for the KBD in this study.
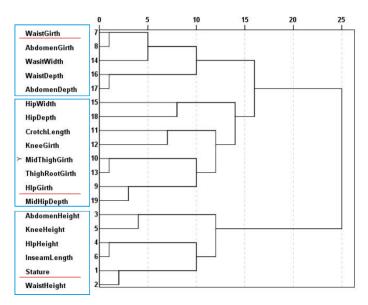

Fig. 6 Dendrogram of hierarchical clustering

3.3.2 Intervals setting

Considering the specific target population, characterized by its region, age, and gender, and the required garment represented, represented by its style and type, the standard sizing systems for garments - Men (GB/T 1335.1-2008) in China was taken into account. According to this sizing system, the intervals for the statue, waist girth, and hip girth were set to 5 cm, 2 cm, and 1.6 cm, respectively.

3.3.3 Fit models determination

Fit models refer to the basic sizes, which are selected by the companies to cover their whole target market [11]. According to the range and the intervals of KBD, we determined the fit models by the sequential steps:

1) Setting the grade of each KBD

Stature: 155, 160, 165, 170, 175, 180, and 185;

Waist girth: 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, and 102;

Hip girth: 77.2, 78.8, 80.4, 82, 83.6, 85.2, 86.8, 88.4, 90, 91.6, 93.2, 94.8, 96.4, 98, 99.6, 101.2, 102.8, 104.4, and 106.

2) Assigning the fit models

Since there were 7, 21, and 19 grades in the stature, waist girth, and hip girth, respectively, 2793 fit models were determined in total.

3.4   Garment size assignment

In this section, we assigned the garment sizes in the light of approaching principle, *K*-Means clustering, and support vector machine method, respectively.

3.4.1 Application of approaching principle

The enumeration algorithm was introduced to calculate the distances between the target population and fit models. Moreover, the distance was computed by the formula (1). For each instance, the closest fit model filtered out according to the distances was chosen as the optimal garment size. In this phase, we arranged 159 sizes for the target population.

$$D_{ij}^{Euclidean} = \|p_i - f_j^{model}\|, i = 1,2,\cdots,200, j = 1,2,\cdots,2793 \tag{1}$$

3.4.2 Application of K-means clustering

Before clustering, the initial centroids should be determined. We set the initial centroids following the distribution of key body dimensions in the standard sizing system. For instance, in the standard sizing system GB/T 1335-2009, if the statue equals to 155 cm, the range of waist girth (WG) is from 62 cm to 74 cm. Since the interval of WG is 2 cm, there are six grades for WG except for 68 cm. For hip girth (HG), the range is set in [77.2, 85.2], and the interval is 1.6 cm. It could be inferred that there are six grades for HG. Consequentially, there exist 36 sizes for the statue of 155 cm. For other grades of stature, the number of sizes was computed through the same process. In total, we chose 669 sizes as the initial centroids. Afterward, the target population was classified using *K*-means clustering with the fixed initial centroids. Finally, 112 sizes were assigned.

3.4.3 Application of support vector machine

Based on the result of section 3.4.2, we classified the garment sizes by the support vector machine (SVM). The parameters of penalty factor *C* and spread factor *q* have a significant impact on the performance of SVM for class. We searched the optimal combination of *C* and *q* using the grim search technique. Finally, we employed 48 sizes, when the parameters of *C* and *q* equaled to 4.351 and 0.023.

4 Results

4.1.1 Indexes for evaluating the effects of size assignment

(1) The aggregate loss of fit

The aggregate loss of fit stands for the averaged distance between the body dimensions of instances and dimensions of the selected garment size, calculated by the formula (2), which has been considered as a general criterion for evaluating the sizing selection [14]. Hence, the aggregate loss was employed in this work. A lower aggregate loss indicates that the shorter distance between the body and the selected size, in which case the garment is considered to have better fit. An optimal solution for garment size selection would have the lowest value of aggregate loss.

$$D_{ave} = \frac{\sum \|d_i - c_i\|}{n} \tag{2}$$

where: d_i represents the body dimensions of the instance, c_i stands for the dimensions of the selected garment size; *n* refers to the number of instances.

(2) Number of poor fit and the accommodation rate of ideal fit

The ideal aggregate loss was about 3.58 cm presented by Gupta [13]. Based on this criterion, the case, in which the distance between the body and the selected size exceeded 3.58 was regarded to have a poor fit. Otherwise, the case, in which the distance did not exceed 3.58 was viewed as the one, which has an ideal fit. The accommodation rate of the ideal fit was calculated by the formula (3).

$$R_{Acc} = \frac{n_{ideal}}{N} \tag{3}$$

where: n_{ideal} represents the number of individuals with the ideal fit; N refers to the total amount of the individuals.

4.1.2 Evaluation of apparel size assignment

Table 2 demonstrates the results of apparel size assignment based on the approaching principle, *K*-means clustering, and support vector machine, respectively. Among the three methods, the aggregate loss of fit and the poor fit number of the solution using approaching principle was the lowest, while the accommodate rate of ideal fit was the highest reaching 100%. All of these indicated that the method using approaching principle could reach the optimal garment fit comprehensively. However, this method required the largest

number of garment sizes, meaning that the cost of the method was the highest. From the aggregate loss of fit and accommodation rate of ideal fit, SVM was inferior to approaching principle slightly. Nevertheless, SVM only employed approximately one-third of the sizes used by the approaching principle to obtain the accommodate rate up to 82%. Therefore, it could draw to the conclusion that the method based on SVM was the optimal solution for the garment size assignment, compromising the comprehensive garment fit and cost.

Table 2. Performance comparison of apparel size assignment using different approaches.

Methods	Aggregate loss of fit [cm]	Accommodation rate of ideal fit [%]	Number of poor fit	Number of size employed
Approaching principle	1.58	100	0	159
K-means clustering	4.08	56	88	112
Support vector machine	3.07	82	32	48

5 Conclusions

In this study, solutions for the apparel size assignment toward the mass customization have been discussed and probed based on differing data mining techniques, involving approaching principle, *K*-means, and support vector machine (SVM). The solutions were assessed by the same case of men's pants in China with 200 adult males. Before assigning the garment size, key body dimensions (KBD) were determined using the hierarchical clustering algorithm as well as the distribution features of the KBD, after anthropometric data acquisition and preprocessing. Hereafter, the fit models were identified under the standard sizing system. Two thousand seven hundred ninety-three fit models were determined concerning the standard sizing system initially. Afterward, the garment sizes of the target population were assigned using the approaching principle, *K*-means, and SVM methods, respectively. And then, the solutions were evaluated by the aggregate loss of fit, the number of poor fit, accommodation rate of ideal fit, and the number of garment size employed. Ultimately, the solution based on SVM was justified as the optimal one, which could cover more populations with fewer sizes compared with another solutions. The SVM-based solution for garment sizes has the merits as follows: (1) a low-cost solution for the size assignment by exploiting the potentials of existing sizing system, without creating new sizing systems; (2) Easy of understanding and handling by the operators in the company, based on their existing experience and expertise; (3) Flexible feasibility to other types of garments. Besides, the solution proposed could be improved and extended in the future. For the customers, a sizing selection approach can be developed to support their buying decisions. For the enterprises, a garment pattern recommendation system can be put forth to promote the accuracy and efficiency of patternmaking in the context of MC. Indeed, as the advent of the big data era, it is significant to apply artificial intelligence techniques to improve the accuracy and efficiency due to the complicated relationship between human body types and size combinations in the future work.

References

[1] Kilinc-Balci F.S. 4 - How consumers perceive comfort in apparel. In Improving Comfort in Clothing; G. Song, Editor; Woodhead Publishing: Oxford, Great Britain, 2011; pp. 97-113.

[2] Tangchaiburana, S.; KW. Techametheekul. Development model of web design element for clothing e-commerce based on the concept of mass customization. *KJSS*, 2017, 38(3), pp. 242-250. DOI:10.1016/j.kjss.2016.07.007

[3] Jost, P.-J.; T. Süsser. Company-customer interaction in mass customization. *IJPE*, 2020, 220: pp. 107454. DOI:10.1016/j.ijpe.2019.07.027.

[4] Song, X.; L. Tang. Customization-Massing-New Angle of View to Implement C2B Development in China. AJIBM, 2017, 7(6): pp. 735-740. DOI: 10.4236/ajibm.2017.76052.

[5] Salusso-Deonier, C.J., et al., A Multivariate Method of Classif in Bod Form Variation for Sizing Women's Apparel. CTRJ, 1985, 4(1): pp. 38-45. DOI: 10.1177/0887302X8500400106.

[6] Hsu, C.-H.; M.-J.J. Wang. Using decision tree-based data mining to establish a sizing system for the manufacture of garments. IJAM, 2005. 26(5): pp. 669-674. DOI: 10.1007/s00170-003-2032-0.

[7] Zheng, R.; W. Yu; J. Fan. Development of a new chinese bra sizing system based on breast anthropometric measurements. IJIE, 2007, 37(8): pp. 697-705. DOI: 10.1016/j.ergon.2007.05.008.

[8] Hsu, C.-H., Developing Accurate Industrial Standards to Facilitate Production in Apparel Manufacturing Based on Anthropometric Data. HFM, 2009, 19(3): pp. 199-211. DOI:10.1002/hfm.v19:3.

[9] Bagherzadeh, R.; M. Latifi; A.-R. Faramarzi. Employing a Three-Stage Data Mining Procedure to Develop Sizing System. WASJ, 2010, 8(8): pp. 923-929.

[10] Esfandarani, S. M.; Shahrabi J. Developing a new suit sizing system using data optimization techniques. *IJCST*, 2012, 24(1): pp. 27-35. DOI: 10.1108/09556221211194327.

[11] Vinué, G., et al. Looking for representative fit models for apparel sizing. *DSS*, 2014, 57: pp. 22-33. DOI: 10.1016/j.dss.2013.07.007.

[12] Maimon, O.; L. Rokach, Eds. *Data Mining and Knowledge Discovery Handbook*, 2nd ed. 2010.

[13] Zakaria, N. 4-Body shape analysis and identification of key dimensions for apparel sizing systems, In *Anthropometry, Apparel Sizing and Design*, D. Gupta, N. Zakaria, Editors; Woodhead Publishing: Oxford, Great Britain, 2014; pp. 95-119.

[14] Gupta, D. A statistical model for developing body size charts for garments. *IJCST*, 2004, 16(5): pp. 458-469. DOI: 10.1108/09556220410555641.