Decision Support Systems 57 (2014) 22–33

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Looking for representative fit models for apparel sizing

G. Vinué [a,*], T. León [a], S. Alemany [b], G. Ayala [a]

[a] Department of Statistics and O.R., Avda, Vicent Andrés Estellés, 1. 46100-Burjasot, University of Valencia, Valencia, Spain
[b] Biomechanics Institute of Valencia, Universidad Politécnica de Valencia, Valencia, Spain

ARTICLE INFO

Article history:
Received 15 February 2013
Received in revised form 23 July 2013
Accepted 23 July 2013
Available online 1 August 2013

Keywords:
HIPAM
Hierarchical tree
Partitioning around medoids
Fit models
Mean split silhouette
INCA statistic

ABSTRACT

This paper is concerned with the generation of optimal fit models for use in apparel design. Representative fit models or prototypes are important for defining a meaningful sizing system. However, there is no agreement among apparel manufacturers and each one has their own prototypes and size charts i.e. there is a lack of standard sizes in garments from different apparel manufacturers.

We propose two algorithms based on a new hierarchical partitioning around medoids clustering method originally developed for gene expression data. We are concerned with a different application; therefore, the dissimilarity between the objects has to be different and must be designed to deal with anthropometric features. Furthermore, one of the algorithms incorporates a different rule to split the clusters, which, in our case, provides better results. Our procedures not only make it possible to obtain optimal prototypes, but also to detect outliers. These outliers should be removed before defining prototypes so that the companies' market share can be optimized.

All the analyses are performed using the anthropometric database obtained from a survey of the Spanish female population.

1. Introduction

Clothing fit is the major aim in the design and manufacture of clothing to ensure user comfort and better appearance. It is a complex issue influenced by the customer's anthropometry, fashion and social trends [16,30]. The development of ready-to-wear (RTW) clothes involves estimating the body measurements of the target population to provide patterns based on a basic size, sizing charts and grading parameters. However, most apparel manufacturers create and adjust their own size charts. They usually use sales studies, returned merchandising reports and small customer surveys and, based on this information, create their size charts i.e. they use a trial and error methodology [12]. This methodology cannot be used when a company explores a (completely) new market. Additionally, each company has its own sizing system, so garments with the same size-label fit differently. This results in a poor fit, unsold garments and a less competitive business. Many people return bought clothes because they are not satisfied [2,10]. A major challenge is to develop an efficient method of sales forecasting to predict the future demand of a product and to avoid unsold stock [17,21,44,45]. Apparel firms would gain a competitive advantage if they could enhance their organizational capabilities. The use of new information technology (IT) should play a critical role to that end [32]. In particular, Radio-Frequency IDentification (RFID) allows firms to track and trace products and could minimize inventory shrinkage [56].

In customer surveys, certain models, called the fit model and/or prototypes, represent the basic size [3]. The fit model represents the body dimensions selected by each company to define the proportional relationships needed to achieve the fit the company has determined [52]. Most apparel companies develop their own sizing systems by using a different fit model which covers their whole target market [51]. In other words, apparel companies only aim to fit one body type, generating base patterns and grade rules from the measurements and proportions of their fit model [5]. However, there might be many body types within a target market and this single idealized fit model may not adequately describe the various shapes of the other body types within each size [4, page 133]. Furthermore, there is little information to help choose a fit model whose body size and shape are consistent with the target market [5]. Our paper proposes a methodology to identify representative fit models. A good fit model is the basis for defining an accurate sizing system.

The use of anthropometric databases to enhance apparel fitting has been mainly aimed at defining new sizing systems. Three types of approaches can be distinguished for creating a sizing system: traditional step-wise sizing, multivariate approaches and optimization methods. Traditional methods select two independent dimensions (named key or control and secondary dimension, respectively) and choose an inter-size interval to determine the optimal number of

* Corresponding author. Tel.: +34 963543988.
E-mail addresses: guillermo.vinue@uv.es (G. Vinué), teresa.leon@uv.es (T. León), sandra.alemany@ibv.upv.es (S. Alemany), guillermo.ayala@uv.es (G. Ayala).

0167-9236/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2013.07.007

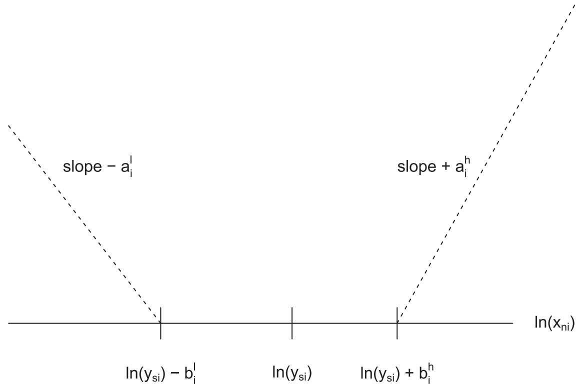

Fig. 1. Marginal dissimilarity proposed in [35].

sizes that accommodates the highest percentage of population. This approach is too simplistic because it is not possible to cover the different body types of the population, perhaps due to the fact that other relevant anthropometric dimensions are not considered.

Several multivariate methodologies have been proposed for defining a sizing system. Principal component analysis (PCA) has been widely used as a dimensionality reduction technique using the first two principal components to generate the bivariate distribution in which to define the sizing chart [11,22–24,33,41]. Partitioning clustering methods, especially the k-means algorithm, have been used to classify the target population into different morphologies by using every anthropometric measurement available as an input [13,37,55]. Other alternatives combining data mining and decision trees have also been proposed [7,25].

The first reference using an optimization was [46], where an integer programming method was used to determine the number of sizes in order to optimize garment sales. A modification of Tryfos' proposal was introduced in [35], where the objective function is related with quality of fit (instead of sales) and a nonlinear optimization technique was used.

Good accommodation percentages are achieved using multivariate techniques or optimization. However, it is difficult to translate a selection method to the consumer using straightforward labeling.

None of the aforementioned approaches aims to identify fit models. As far as we know, no statistical method has been developed with this goal in mind. On the other hand, body models as a representation of part of the population pose a classic problem in ergonomics for designing products and working environments. Percentiles have been widely used in this context. However, these quantities are not additive and provide a univariate accommodation [36,39,54]. Nowadays, the most commonly used method is based on PCA: first, only the first two principal components are selected. Then, a normality ellipse (or circle if the data are standardized) covering a given percentage of the population (usually 95%) is calculated. Finally, eight subjects are identified at the intersections with the major axes and at the so-called octant points. They are estimates of extreme patterns of the data [8,19,20,26,40,54]. The limitations of this procedure are described in [18]. We are faced with the usual drawback of using PCA and the variation not considered may correspond to people with a difficult accommodation [18]. Recently, a new alternative based on archetypal analysis has been proposed [14]. Nevertheless, all these approaches used in ergonomics are not useful for obtaining fit models because they look for boundary cases and the fit models must be central individuals.

Table 2
Summary statistics for the variables considered.

Measurement (cm)	Minimum	First quantile	Median	Mean	Third quantile	Maximum
Neck to ground length	116.4	132.9	136.8	137	140.8	161.9
Bust circumference	73	87.4	93.3	95.02	100.7	145.7
Chest circumference	45.91	90.78	96.37	97.92	103.7	150.30
Waist circumference	58.60	75.6	83.10	84.98	92.40	167.6
Hip circumference	72.8	98.3	103.3	104.9	109.9	170.8

The final evaluation of fit in the apparel development process needs models representing the target population to test every new design before the production phase. These models are the dress form, the human fit model and the virtual fit model. Of these three, the role of the human fit model is the most important. Companies aim to improve the quality of fit by scanning their fit models and deriving dress forms from those scans [4,43]. The current practice in apparel fit analysis is based on using expert panels [6]. An expert panel is an experienced work team that judges the garment fit. As far as the apparel industry is concerned, fit analysis is carried out using a live fit model.

Research has recently been carried out to check the reliability of using virtual 3D scan models instead of fit models to improve garment fit [6,9]. These virtual 3D models come from scanned live fit models. They can be used in a similar way to fit models but have certain advantages. Emerging technology for body scanning offers many benefits in different areas of apparel design and manufacturing [6,42]. Indeed, the tailoring procedure followed by designers requires the scanning of real persons to generate 3D clothes from 2D patterns [34,48]. In this way, a representative fit model of the target population, whether a live fit model or a 3D scan model, is critical for improving the garment fit. The major aim of fit analysis is to define fit models for different sizes [9].

In this article we propose a new approach for defining optimal prototypes for apparel design. We introduce two classification algorithms based on the hierarchical partitioning around medoids (HIPAM) clustering algorithm [49], modified to deal with anthropometric data. The outputs of both algorithms include a set of representative subjects or medoids taken from the original data set which constitute our prototypes. The HIPAM algorithm is a divisive hierarchical clustering method based on the PAM algorithm. Hierarchical clustering methods seek to build a hierarchy of clusters. In particular, divisive methods begin with one cluster and split this group recursively. Usually, but not always, in the last step there are as many groups as observations. Partitioning clustering methods classify the objects into k clusters, where k is usually fixed in advance (k can also be data-adaptively selected). The Partitioning Around Medoids (PAM) algorithm [29] is a robust partitioning algorithm which seeks to find k representative subjects (also known as medoids) from the data set in such a way that the sum of the within-cluster dissimilarities is minimized.

HIPAM starts with one large cluster and, at each level, a given cluster is partitioned using PAM. It splits the parent cluster at a given node of the classification tree by taking into account a cluster structure measure. The number of child clusters is obtained by maximizing the silhouette

Table 1
Aggregation weights associated with the anthropometric variables.

Chest circumference	Neck to ground length	Waist circumference	Hip circumference	Bust circumference
0.42805	0.24580	0.12430	0.10180	0.10005