Analysis of Anthropometric Data for Garment Sizing

Vivek Muskan

2017MT10755

Department of Mathematics, IIT Delhi

23rd November, 2020

TXQ - 301 Seminar Course Report

Professor In charge : Prof. Deepti Gupta

Department of Textile, IIT Delhi

Primary Citations : 1. Norsaadah Zakariaa,b , Wan Syazehan Ruznanc (2020)

2. D. Gupta and B.R. Gangadhar (2004)

Index

1. Objective

The objective of this report is to give a step by step method to obtain a standard size chart for a large population using Anthropometric data of a sample size based on the results obtained from various research papers mentioned in the reference section . Here is the key objectives for each part:

- Experiment design for collection of data in minimum cost
- Preprocessing and cleaning of data for further study
- Best steps for Exploratory Data Analysis for Anthropometric Data
- Deciding different class of size chart required for better outcomes
- Identifying better model among Decision tree and Random Forest

2. Introduction

Garments are manufactured in massive amounts using predefined size charts which allow for the reduction of production cost. It is, therefore, practically not possible to obtain a perfect fit between a piece of cloth and an individual customer (Kotha, 1995; Pine, 1993). The process of developing body size charts involves taking anthropometric body measurements of the sample population and its division into groups with similar anthropometric measurement for the purpose of garment manufacture (D. Gupta and B.R. Gangadhar, 2004).

There are basically three important steps in sizing system development:

1. Process of Collection of Anthropometric Data
2. Sizing Data Analysis
3. Sizing System Development

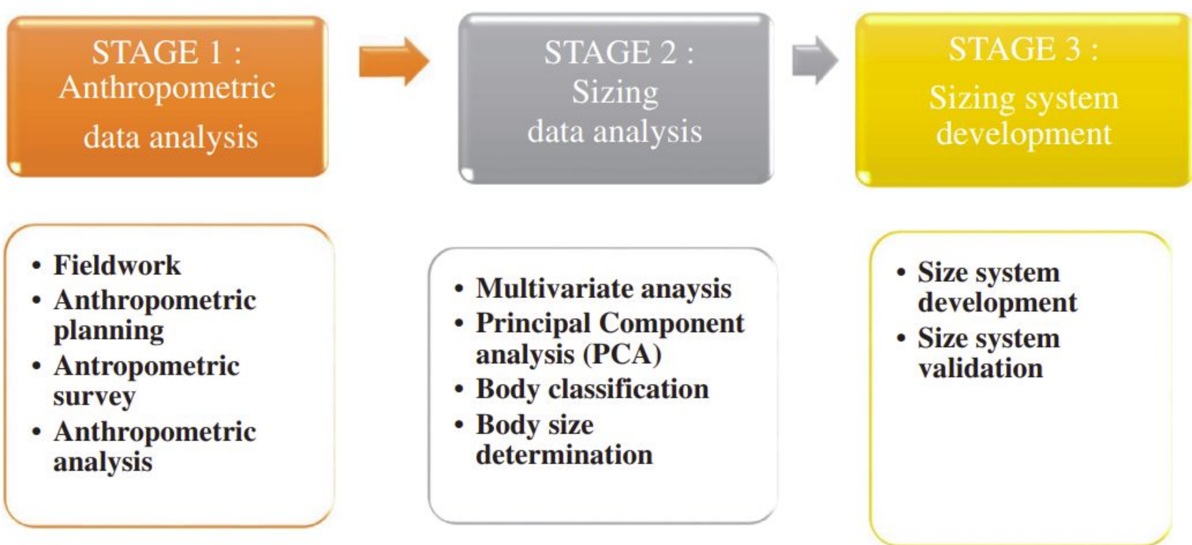

Fig 2.1 : Sizing system development in process (N Zakariaa,b , Wan S Ruznanc, 2020)

3. Data Collection and Preprocessing

3.1. Data Collection

The data collection for an anthropometric survey for clothing purposes is based on the standard ISO 8559/1989, which defines the terms used for each different body dimension. Under this ISO standard, there are 51 body dimensions to be measured for a garment sizing. These body dimensions are divided into three groups: vertical length, width, and girth, as shown in *Table 3.1.1.* Further these dimensions are divided into the upper and lower body. The 29 dimensions suitable for the upper body are marked with superscript *"a"*, while the other 20 dimensions are categorized as lower body dimensions.

Out of all these dimensions, we need not measure each and every one of them. In a study by Salusso (1982), it was found that 15 principal components of all these dimensions can be used to summarize trends in body form variation. First and second principle components, laterality (fullness) and linearity (length), respectively, were selected to describe body size and type. In another study by D Gupta (2004), bust for the upper body and hip for the lower body were identified as the critical dimensions affecting garment fit (D. Gupta and B.R. Gangadhar, 2004). In a paper from University of Minnesota, it was shown that six lower leg measurements - ankle, calf, and knee circumferences as well as knee-to ankle, knee to-calf, and ankle-to-calf lengths gave better performance with respect to other body dimensions for North American civilian population (using the CAESAR database) (Rachael Granberry, Julia Duvall, Lucy E. Dunne, Bradley Holschuh, 2017). Looking at the following results from the papers, I conclude that we can take the union of all these body dimensions for the purpose of creating an exhaustive dataset.

Length (vertical)	**Width (vertical)**	**Girth (horizontal)**
Height	1. [a]Shoulder length	Weight
1. [a]Under arm length	2. [a]Shoulder width	1. [a]Head girth
2. [a]Scye depth	3. [a]Back width	2. [a]Neck girth
3. [a]Neck shoulder point to breast point	4. [a]Upper arm length	3. [a]Neck-base girth
4. [a]Cervical to breast point	5. [a]Arm length	4. [a]Chest girth
5. [a]Neck shoulder to waist	6. [a]Seventh cervical to wrist length	5. [a]Bust girth
6. [a]Cervical to waist(front)	7. [a]Hand length	6. [a]Upper arm girth
7. [a]Cervical to waist(back)	8. Foot length	7. [a]Armscye girth
8. [a]Cervical height(sitting)		8. [a]Elbow girth
9. [a]Trunk length		9. [a]Wrist girth
10. [a]Body rise		10. [a]Hand girth
11. [a]Cervical to knee hollow		11. Waist girth
12. [a]Cervical height		12. Hip girth
13. Waist height		13. Thigh girth
14. Outside leg length		14. Mid-thigh girth
15. Waist to hips		15. Knee girth
16. Hip height		16. Lower knee girth
17. Crotch		17. Calf girth
18. Trunk circumference		18. Minimum leg girth
19. Thigh length		19. Ankle girth
20. Inside leg length/crotch		
Height		
21. Knee height		
22. Ankle height		

Lower body dimensions.
[a]Upper body dimensions.

Table 3.1.1 : Body dimensions by ISO8559/1989 (N Zakariaa,b , Wan S Ruznanc, 2020)

3.2 Data Preprocessing

Most of the features of the dataset is continuous as it is the length measurement of various kinds. Few features like Gender, Ethnicity or Country will be categorical data which should be label encoded (Scikit learn label encoder can be used in python).

How to deal with a data row having a missing column of Anthropometric data, if any?

- Instead of the usual way of replacing missing values with the mean of the column, we can use an alternate technique. Find the correlation of the missing row with all the dataset, then take the mean of that column of all the dataset which has correlation value greater than a critical value (~0.8).

Look at the size of your dataset. Typically the anthropometric data set has less than one thousand data points. In this case it is very costly to lose information based on taking average for missing values. The above procedure ensures that the person has missing body dimensions quite similar to those whose all other body dimensions match.

4. Data Analysis and Inference

4.1 Anthropometric Data Analysis

In this section we will mainly do two types of analysis: Correlation Matrix analysis and Principal Component Analysis. Correlation matrices help us to remove similar dimensions and PCA helps in reducing the dimensions of the dataset.

4.1.1 Correlation Matrix Analysis

A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses.

According to the results obtained in a paper , he correlation matrix is given below in *Table 4.1.1*

Dimension	Height	Nat waist	Hip 6	Bust
Height	1	0.104	0.204	0.132
Cervical height	*0.866*	0.139	0.224	0.159
Natural waist	0.104	1	*0.752*	*0.781*
Waist length_center front	0.172	0.224	0.138	0.242
Cervical _natural waist	0.246	0.33	0.277	0.39
Center back_natural waist	0.235	0.124	0.091	0.167
Artificial waist	0.139	*0.801*	*0.74*	*0.716*
Outer leg	*0.713*	0.08	0.185	0.11
Inner leg	*0.52*	0.101	0.173	0.07
Hip 4	0.185	*0.78*	*0.919*	*0.74*
Hip 6	0.204	*0.752*	1	*0.714*
Thigh	0.127	*0.541*	*0.611*	*0.562*
Knee	0.205	0.478	*0.55*	*0.507*
Ankle	0.179	0.201	0.243	0.218
Bust	0.132	*0.781*	*0.714*	1
Upper arm girth	0.073	0.49	0.472	0.467
Wrist	0.146	0.279	0.275	0.252
Neck_mid	0.18	*0.638*	*0.581*	*0.62*
Neck_neck	0.138	0.484	0.435	0.488
Arm length	0.48	0.109	0.136	0.125
Shoulder_shoulder	0.275	0.348	0.326	0.347
Outer leg-inner leg	0.116	−0.035	−0.013	0.031

Table 4.1.1 Correlation coefficients of key dimensions (D. Gupta and B.R. Gangadhar, 2004)

Here we should remove the dimensions having a high correlation but low variance.

4.1.2 Principal Component Analysis (PCA)

The aim of PCA is to project a high dimensional data into a low dimensional data preserving the variance or the information gain of the data. These dimensions constitute an orthonomal basis in which different individual dimensions of the data are linearly uncorrelated. Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.

Based on the eigenvalues corresponding to each principal component we can decide about the no of principal components we should select so that we don't lose much information. According to Hair et al. (1998) the first few components should extract at least 50% of the variance to prove the usefulness of PCA technique. The study shows that female samples between 13 and 17, with less percentage value, are observed, at 54%. Thus this result indicates variance of >50% in the first component. This indicates that PCA technique is an effective method for this study to obtain a parsimonious solution in describing the variations of body shapes in this sample group.

4.2 Sizing Data Analysis

After the Correlation matrix and principal component analysis we are left with a minimal dataset containing a high amount of data.

4.2.1 Clustering

Now the next task is to decide the no of class in the standard size chart. For this purpose we have unsupervised learning techniques of clustering tools like K-Means Clustering, Gaussian Mixture Models and Fuzzy C-means Clustering. Since the size of the data set is very less so it rules out the use of Gaussian Mixture Model, which is usually used over a large population. We should use K- means clustering technique to with a Grid search for the hyper-parameter K.

According to results obtained in the paper by (Norsaadah Zakariaa,b , Wan Syazehan Ruznanc, 2020) it is found that three and four are the most ideal because the cluster groups were distinct from each other.

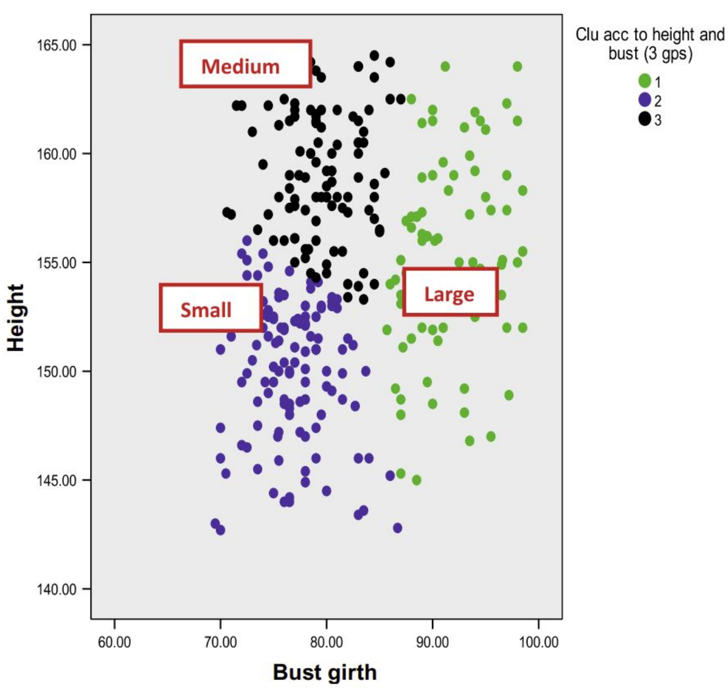

Fig 4.2.1 illustrated the distribution of height (y-axis) versus bust girth (x-axis) for all three clusters. The profile of each cluster is defined as this: Cluster 1 with 134 samples, tall stature with medium to big bust; Cluster 2 with 194 samples, average height with small to medium bust; and Cluster 3 with 197 samples, short with small bust.

(Norsaadah Zakariaa,b , Wan Syazehan Ruznanc, 2020)

4.2.2 Decision Tree vs Random Forest

Random forest classifier is just an ensemble technique of decision trees which performs better than Decision trees for large dataset. Since we have very little dimension as well as dataset so using random forest might increase computational cost much with respect to increase in performance.

According to a paper by (Chih-Hung Hsu · Mao-Jiun J. Wang, 2004), 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

We can set up a grid search or random search to tune hyper-parameters like *max_leaf and min_split* etc.

5. Conclusion

Based on the results that we saw, we can drastically reduce the cost of data collection of Anthropometric measurement by choosing the correct set of attributes. Correlation Matrix, PCA and Decision tree classifier gives great results on such data due to high similarity between measured data and low size of dataset. Further improvement can be made by taking these measurement into 3D scanner as sourced in paper by (Slavenka, Maja, Darko, 2012)

6. References and Bibliography

1. Norsaadah Zakariaa,b , Wan Syazehan Ruznanc (2020), “Developing apparel sizing system using anthropometric data: Body size and shape analysis, key dimensions, and data segmentation”
2. D. Gupta and B.R. Gangadhar (2004), “A statistical model for developing body size charts for garments”, Indian Institute of Delhi
3. Kotha, S. (1995), “Mass customization: implementing the emerging paradigm for competitive advantage”, Strategic Management International, Vol. 16, pp. 21-42.
4. Salusso, C.J. (1982), “A method for classifying adult female body form variation in relation to the US Standard for apparel sizing”, Doctoral Dissertation, University of Minnesota
5. Rachael Granberry, Julia Duvall, Lucy E. Dunne, Bradley Holschuh (2017), “An Analysis of Anthropometric Geometric Variability of the Lower Leg for the Fit & Function of Advanced Functional Garments”, University of Minnesota
6. JF Hair, WC Black, BJ Babin, RE Anderson, RL Tatham Prentice hall 5 (3), 207-219
7. Chih-Hung Hsu · Mao-Jiun J. Wang, (2004)
8. Slavenka, Maja, Darko, (2012), University of Zagreb, Zagreb, Croatia