DEVELOPMENT OF A NEW SIZING SYSTEM BASED ON DATA MINING APPROACHES

Jamal SHAHRABI[1], Maryam Salehi ESFANDARANI[2]
[1] *Amirkabir University of Technology, Faculty of Industrial Engineering, Tehran , Iran*
[2] *Amirkabir University of Technology, Faculty of Textile Engineering, Tehran, Iran*
msalehi_87@aut.ac.ir

Abstract
Due to wide differences in race, nutrition and climate people who live in different countries have their own body size also most of current sizing systems are out-dated, so there is an urgent need to develop a new sizing system. The main goal in this work is develop a new suit sizing system based on up-dated data with using data mining techniques, to improve the final quality and reduce the waste of fabric. This paper aims to investigate effect of data reduction on the final fitness of the sizing chart. Principle component analysis is applied to reduce the sizing variables, and non-hierarchical clustering approach is used to segment the heterogeneous population to more homogeneous one, the aggregate loss of fitness is used to evaluate the resulted sizing chart. The results show that, when principle component analysis reduces the 10 sizing variables to 2 main components, the final fitness for the resulted sizing chart is the best. These two main components are height and circumference components. The hierarchical clustering approach could effectively group all body type to 7 clusters. The resulted sizing chart could be used as a reference for suit manufactures. Due to growing rate of globalization, the final results will be useful for the companies that want to connect to global business chain.
Key words: Sizing System, Data Mining, Principle Component Analysis, Clustering

1. Introduction

Data mining is the science of extracting useful information from large data bases [1], due to wide availability of huge amount of data and the urgent need for turning such data into useful information and knowledge, data mining has attracted great attention in the information industry and in society in the recent years [2]. Data mining could help the companies to increase their costumers and reduce the costs. It could be used effectively in textile field such as textile commerce, production plan and apparel design [3-5]. Apparel manufacturing produces products with highest added value in global textile manufacturing chain [6]. In manufacturing apparel it is so important to produce apparel with best design based on standard size chart to fit all body types [7]. In recent years, there have been more attention to develop new sizing systems based on data mining, [8,9] .Hsu, applied a data mining techniques to develop industrial standards for adult females, he applied two-stage clustering approach to generate a standard size chart [6], in another work he established systems for using a decision tree technique to determine the pants sizes of army soldiers. The newly developed sizing system can provide garment manufacturers with size specifications [10]. In this work a new sizing chart based on Iranian male body size is developed using principle component analysis and clustering approach and the effect of dimension reduction on the fitness of final sizing chart is investigated. In most recent researches regard to the large number of variable, the factor analysis is done, but dimension reduction maybe cause loosing useful data so in this research a new sizing system is developed and the effect of the number of components on the resulted sizing chart fitness is analyzed.

1. Translate the business problem into datamining problem

Due to differences of race, age, gender, social class and even the occupation, the human body shape and type diversify in many aspects. People who live in different countries have their own body size, also most of current sizing systems are out-dated, thus there is an urgent need for each country to develop, a new sizing chart based on up-dated data. The resulted sizing system could be useful for the garment companies that want to produce garment with the best fitting, and also for international companies that want be prosper in the global business chain.

2. Collecting and preparing the data set

With consulting domain experts, ten sizing variables that are necessary for producing suits are selected. The selected variables are coat height, armhole girth, sleeves height, waist circumference, chest circumference, stomach circumference, thigh circumference, hip circumference, knee girth, trouser height, then a large data set consist of 600 body size is collected from a garment company in Iran. Missing data replaced with new value by series mean algorithm, the out layer data consider data that doesn't limited to ±3σ from the mean [11], and are omitted.

3. Data Reduction Process

Principle component analysis (PCA) is used as the variable reduction method, PCA reduces a set of observed variables into a smaller set of new variables called principle components. Data reduction maybe cause losing some useful data that effects on the fitness of the final sizing chart, thus in this research the effect of principle components number on the fitting of the final suit sizing chart for all body types is investigated.

The selected variables for producing suits are ten, so the data reduction procedure is run for 7 components to 2 components. The critical value to retaining the components is chosen 1% as table 1 shows, components 8,9 and 10 are omitted and only the 7 components are important enough to be considered for the next steps. These components are highlighted in table 1.

Table1. Initial eigen values for all components

component	Initial eigen values		
	total	% of variance	Cumulative %
1	6.668	66.676	66.676
2	1.807	18.072	84.749
3	.467	4.668	89.417
4	.458	4.575	93.992
5	.176	1.763	95.755
6	.114	1.138	96.893
7	.106	1.062	97.956
8	.088	.876	98.832
9	.074	.740	99.572
10	.043	.428	100.000

4. Clustering Approach

In this work cluster analysis is used to develop new sizing chart based on principle components extracted in the last part. Cluster analysis is a powerful technique to divide hetrogenous data into groups. K-mean algorithm is used as the clustering approach. This is prototype-based clustering technique that attempts to find a user-specified number of clusters [12]. k certain number of clusters fixed as a priori . k centroids are defined for each cluster. Each sample is then assigned to the closest centroid, and each collection of samples that assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the samples assigned to clusters, this action is repeated until the centroids don't change. Clustering approach is done on training data set consist of 460 samples (80% of total population).

5. Evaluation the effect of components number on the final sizing chart fitness

To compare the sizing charts with different components number, the aggregate loss of fit is calculated for each sizing system. Aggregate loss of fit, represents how well the sizing system

performs in fitting the population. Consulting with domain experts shows that chest circumference, waist circumference and hip circumference are most important dimensions for determining size, thus the aggregate loss of fit for these dimensions are determined. To determine the aggregate loss of fit, all individual distance from their assigned size are averaged over the validation data [8], using equation 1.

$$loss\ of\ fit = \frac{\sqrt{\sum_{i=1}^{117}\left(a_{i1}-a'_{j1}\right)^2+\sum_{i=1}^{117}\left(a_{i2}-a'_{j2}\right)^2+\sum_{i=1}^{117}\left(a_{i3}-a'_{j3}\right)^2}}{n_j} \quad (1)$$

In this equation, a_{i1} , a_{i2} , a_{i3} are the value of chest, waist and hip circumference in population for each person, and a'_{i1} , a'_{i2} , a'_{i3} are the value of assigned chest, waist and hip circumference, for this person in the sizing chart, n_j is the number of members in each clusters. Aggregate loss of fitness is calculated for validation data consist of 114 samples (20 percent of population).

Table 2. loss of fitness for all sets of components

Loss of fitness for each cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Total
7 components	2.66	2.26	4.26	3.67	3.67	7.84	5.81	30.17
6 components	2.75	4.6	7.32	5.02	4.51	4.26	3.76	32.2
5 components	4.24	2.56	5.98	6.97	1.86	3	4.17	28.78
4 components	4.33	3.55	4.21	2.42	4	3	4.2	25.71
3 components	2.72	2.92	4.51	3.07	4.71	4.13	3.8	25.82
2 components	2.8	4	2.7	0.1	4.3	3.2	2.8	19.9

The loss of fitness for all sets of components are shown in table 2. As the total loss of fitness is less when two components are considered, the sizing chart based on two components is proposed as a new sizing chart for Iranian male suits (table 3). In this article, our means of circumference is half of girth which is usual among the suit producers.

Table 3. The new sizing chart based on two components

Sizing variable(cm)	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Coat height	76	82	70.5	56	77.5	73	81
Shoulders distance	50	54	50	34	56	51	60.5
Sleeve height	65	68	55	48	58.5	55.5	60.5
Chest circumference	49	60	47	36	53.5	55	62.5
Stomach circumference	47	55	45	36	52.5	56	64.5
Trouser height	111	111	100	78	103	96	104.5

Waist circumference	41	54	39	30	46	49.5	60.5
Hip circumference	52	63	47	40	58.5	59.5	70
Thigh girth	36	41	32	28	37.5	37.5	43.5
Knee girth	27	32	26	23	31.5	29.5	33.5

7. Conclusion:

Globalization has altered the competitive dynamics of nation, firms and industries, so there is an urgent need for all companies to change, their designing based on up-dated data. Thus in this research a new sizing system based on up-dated data is developed using a series of data mining techniques. Principle component analysis is used for data reduction. The effect of the number of the components on the fitness of final sizing chart is analyzed, K-mean algorithm is used to change the hetrogenous population to a homogenous population, all components are grouped to 7 clusters. For measuring the fitness of each sizing chart, aggregate loss of fitness is calculated, the results show that the sizing chart created by 2 components has the least aggregate loss of fitness, so it could be useful as the refrence for designing suits. These two components are girth and height components, these results are in good agreement with the last researches. The final sizing chart could be used as a reference for suits manufactures, it can help them to produce more fitted garments, also it will decreases waste of fabrics, in mass production it affects on the final suit price. Producing garments with lower price and better quality will be so attractive for garment manufactures.

References:

1. D.Hand, H.Mannila, P.Smyth, *Principle of Datamining*, The MIT Press, Massachusetts LondonEngland, 2001
2. J.Han, M.Kamber, *Data Mining Handbook*, Elsevier, University of Illinois at Urbana-Champaign, 2006
3. Y.Shin, Y.Kim, E.Yi Kim, *Image and Vision Computing*, Vol .28, pp.526–537, 2010
4. H.F.Lin, C.H.Hsu, M,J.Wang ‘’ An application of data mining technique in engineering to facilitate production management of garments’’, Proceedings of the 11th WSEAS International Conference on Computers, 2007
5. A.D.Toni, A.Meneghetti, *Int. J. Production Economics*, Vol.65, pp.17-32, 2000
6. C.Hsu, *Expert Systems with Applications*, Vol.36, pp.4185–4191, 2009
7. C. L.Regan, D.H.Kincade, G.Sheldon, *Clothing and Textiles Research Journal*, Vol.16, pp.36–46, 1998
8. M.J.Chung, H.F.Lin, M.J.Wang, *International Journal of Industrial Ergonomics*, Vol.37, pp.707–716, 2007
9. C.Hsu, *Journal of the Chinese Institute of Industrial Engineers*, Vol. 24, No.3, pp.245-251, 2007
10. C.Hsu. J.Wang, *Int J Adv Manuf Technol*, Vol.26, pp.669–674, 2005
11. J. Y.Moon, Y.J.Nam, “A study the elderly women’s lower figure type classification and lower garment sizing systems”, *In Proceedings of international ergonomics association 2003 conference, Korea*
12. Pang-Ning Tan, “ Introduction to data mining”, Addison-wesley companion book site, 2006