ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

KERNEL-BASED CLUSTERING APPROACH IN DEVELOPING APPAREL SIZE CHARTS

C. P. Vithanage*, T. S. S. Jayawardane, C. D. Thilakaratne, S. N. Niles

* Department of Textile & Clothing Technology, Faculty of Engineering, University of Moratuwa.
Department of Textile & Clothing Technology, Faculty of Engineering, University of Moratuwa.
Department of Statistics, Faculty of Science, University of Colombo.
Department of Textile & Clothing Technology, Faculty of Engineering, University of Moratuwa.

ABSTRACT

With the industry revolution, apparel products also become more sophisticated moving from the basic purpose of clothing to aesthetic appeal of the garment embracing the concepts garment fitting and fashion. Garment fitting is a key technical essential for comfortable wearing. In garment fitting, size refers to a set of specified values of body measurements, such that it will provide a means for garments perfectly fit to a person. With the advent of computer software and improved data mining techniques, researchers attempted new advances in formulation of size charts with a better fit. This article suggests a kernel-based clustering approach in developing an effective size chart for the pants of Sri Lankan females. A new kernel based approach "Global Kernel K- means clustering" was successfully deployed to cluster lower body anthropometric data of Sri Lankan females within the age range of 20-40 years. Through the proposed Kernel- based clustering method can effectively handle highly non-linear data in input space which is a key property of lower body anthropometric data and make it linearly separable in feature space without reduction in dimensions and also mathematically justified. Through this method promising results could be obtained and further clustering method was internally validated with kernel based Dunn's index. The level of fitness of the developed size chart was also evaluated with the aggregate loss of fit factor. The proposed method has strong implications to utilize globally in developing size charts.

KEYWORDS: Size charts, clustering, global kernel K-means, cluster validation.

INTRODUCTION

In parallel to the industry revolution apparel products also become more sophisticated moving from the basic purpose of clothing to aesthetic appeal of the garment embracing the concepts garment fitting and fashion. Garment fitting is a key technical essential for the comfortable wearing intact with the aesthetic appeal. In this context, size refers to a set of specified values of body measurements, such that it will provide a means for garments perfectly fit to a person. Fit of a garment depends on the closeness between garment measurements and the body measurements for which it is intended. The purpose of an apparel sizing system is to divide a varied population into homogeneous subgroups such that garments with set of standard sizes can be used by the entire population. The charts containing set of body measurements belonging to various sub groups of the population is termed as size charts.

Very few number of researchers have involved in the process of the development of size charts. However, with the introduction of online shopping, the importance of perfect size charts is globally recognized. Each country understood that they need their own size charts representing their population because researchers have found that human body shapes, proportions and measurements change significantly due to the geographical and demographical differences. With the development of computer software and improved data mining techniques, researchers attempted new advanced methods to formulate size charts with better fit. It was revealed that statistical methods [1], [2], cluster analysis methods combined with factor analysis [3],[4],[5], Classification and Regression Tree (CART) decision tree method [6],and Self Organizing maps [7] have been used to develop size charts.

Above mentioned approaches have intrinsic shortcomings in the way of application such as selection of key variables, reduction of variables, expectation of linear relationships between variables and expectation of normal distribution of variables in the process of developing size charts. However,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

anthropometric data do not follow exact normal distribution and also it is not linearly related to each other. Only two or three key variables were selected for analysis in above existing approaches and this variable selection may lead to a wrong decision which will produce size charts with imperfect fitting. Therefore, instead of considering only a few variables, if all relevant variables can be considered in the process, an effective size chart could be expected. Hence, such approach will inevitably capable of handling high dimensional non-linear data and this type of research has the highest potential to develop an effective size chart.

There are three revolutions in automated algorithms for pattern analysis[8]; detecting algorithm for linear relations in the 1960s, introduction of back propagation multilayer neural networks and decision trees in the mid 1980s and kernel-based learning methods which enabled the analysis of non linear relations efficiently, in the mid 1990s. The kernel based learning facilitates different methods of pattern recognition such as classification, correlations, rankings, clustering, principal component analysis on different types of data such as vectors, strings, images, text documents and so on [8].

The main features of the process of kernel methods can be summarized as follows;

(i) Data items are embedded into a vector space called the feature space.

(ii) Linear relations are sought among the images of the data items in the feature space.

(iii) The algorithms are implemented in such a way that the coordinates of the embedded points are not needed, only their pair wise inner products.

(iv) The pair wise inner products can be computed efficiently directly from the original data items using a kernel function [8].

However, mapping input data into high dimensional data may be very difficult with larger datasets and computationally expensive. Hence, without explicitly mapping data to high dimensional feature space through a mapping function, kernel can be used to implicitly fulfill it which is called a "kernel trick". Through this kernel function input data transform to a kernel matrix which is an n×n, (where n is the number of cases), symmetric, positive semi-definite matrix (where all Eigen values are essentially positive to be a valid kernel).

Although, this approaches were used in other areas such as image processing, text classification [9] , [10], object recognition[11], gene expression profile analysis[12], DNA and protein analysis [13], it is still new to the subject of size chart development.

The objective of this article is to suggest a kernel-based clustering approach which can handle high dimensional, non-linear data, in developing an effective size chart for the pants of Sri Lankan females. A new kernel based approach "Global Kernel K-means clustering" [14] which is used for MRI segmentation, was identified as a suitable approach in clustering anthropometric data. It was a combination of global k-means clustering method [15] and kernel k-means clustering method and it avoids the problems encountered with k-means clustering. This method has been tested using three different artificial datasets and MRI images segmentation and compare them with kernel k-means clustering. They have proved that global kernel k-means clustering is better than kernel k-means clustering in terms of clustering error [14].

Kernel k-means is capable of handling the high dimensional data[16] while classical k-means need to follow some variable reduction method because of computational reasons. This is a very strong reason in using kernel k-means for clustering anthropometric data because it has more variables which k-means can not handle efficiently. A historical development of clustering algorithms with a critical review presents in the next few sub sections for the benefit of the novel reader on clustering algorithms and it also provides a rationale for selection of the most appropriate algorithm.

K- means Clustering

K-means clustering is one of the most popular clustering method and it has less time complexity (which is O(n),where n is the number of cases)[17].

In this algorithm, number of clusters need to be finalized in advance and the objects are assigned to the nearest initial cluster centers which are arbitrary selected and these centers moves at each step in order to minimize the sum of squared distance from its objects to the cluster center. It iterates several times until the convergence occurs in a way that the intra-cluster squared distances is minimized while inter-cluster distance is maximized. This algorithm usually yields clusters approximately of equal size.

This algorithm has two major drawbacks namely the resulted clusters greatly depends on the initial positions of the cluster centers and it can only find linearly separable clusters [14]. Since, anthropometric data are not linearly separable, k-means approach will not work well and it was practically experienced in analysis of anthropometric data. For high dimensional data, the traditional distance measures can be ineffective and hence finding clusters using k-means approach can be unreliable[18].

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Global K- means Clustering

This algorithm has been proposed by [15] to fix the initialization problem in k-means clustering. Instead of selecting cluster center initially at arbitrary locations, global k-means works in an incremental way of finding new cluster centers. In other words, it starts with one cluster (k=1) and find the cluster center using k-means algorithm and it is the optimal position of the cluster center. Then it moves to two clusters (k=2) and the second best cluster center obtains after several iterations of k-means algorithm. Sequentially adding clusters in the above manner locates the cluster centers at the optimal positions in k-means algorithm [15].

Kernel K- means Clustering

According to Chitta et.al [19], "Kernel k-means is a nonlinear extension of the classical k-means algorithm. It replaces Euclidean distance function employed in the k-means algorithm with a non linear kernel distance function". According to Zhang and Rudnicky [20], "the incorporation of kernel functions enables the K-means algorithm to explore the inherent data pattern in the new space". Kernel k-means algorithms perform better and significantly more accurate than conventional k-means algorithms in unsupervised classification [20][21]. Different kernel functions are available such as Polynomial kernel, Gaussian (RBF) kernel, Sigmoidal kernel and the selection of relevant kernel function and its parameters depends heavily on the type of data itself[22].

Global Kernel K -means Clustering

This algorithm combines the advantages of both global k-means and kernel k-means. Therefore it avoids both limitations local minima problem and linearly separable clusters of k-means and also produces a final partition which is independent of the cluster initialization [14].This method identifies linear clusters in high dimensional feature space and it becomes nonlinear clusters in input data space.

Kernel-based Dunn's Index

The Dunn's index [23] measures the ratio between the smallest inter-cluster distance and the largest intra-cluster distance and hence the maximum value of Dunn index is corresponding to the optimal number of clusters. Dunn's Index was extended to kernel Dunn Index as follows [24];

$$kDI(K) = \min_{1 \le i \le K} \left\{ \min_{1 \le j \le K} \left\{ \frac{\delta^K(C_i, C_j)}{max_{1 \le k \le K}\{\Delta^K(C_k)\}} \right\} \right\} \quad \text{----- (1)}$$

where $\Delta^{\kappa}(C_{\kappa})$ is the largest intra-cluster separation of cluster k in the feature space, $\delta^{\kappa}(C_i, C_j)$ is the minimum of kernel based distance between cluster i and cluster j in the feature space.

The kernel distance is given by [25],

$$D_K^2(\{p\},\{q\}) = K(p,p) + K(q,q) - 2K(p,q) = 2(1 - K(p,q)) \quad \text{---- (2)}$$

where D is the kernel distance between two points *p, q* and *K(p,q)* is the cross similarity of *p* and *q*.

METHODOLOGY

A sample of 1067 females, aged 20-40 years, were selected using convenient sampling technique covering all provinces in Sri Lanka and including different professions such as females in Army, Navy, Air force, university students, employed females, non-employed females etc. Thirteen lower body measurements (waist, hip, thigh, mid thigh, knee, mid calf, ankle, inseam, outseam, crutchlength, hip height, knee height, ankle height) were collected using a measuring tape following manual method [26]. In order to minimize the measurement error, same person was employed in obtaining the measurements of each subject.

Before applying clustering method to the data set, different lower body shapes need to be identified to make the size chart more successful and Waist-to-hip ratio (waist / Hip) is a good standard measure for sorting lower body [27] [28]. Hence, three limits of WHR were identified to sort the lower body shape in to three groups as follows;

WHR ≤ 0.7 - Small WHR ; 0.71 ≤ WHR ≤0.79 - Medium WHR ; WHR ≥ 0.8 - Large WHR and further analysis were done on these three datasets separately.

For analysis, Matlab version 7.7 software and SPSS version 16 were used. Different kernel functions, polynomial and Gaussian (RBF) kernel functions, were used with different parameters in order to map data in input space into high dimensional feature space. This results in an (n×n) square kernel matrix and it was the input for Global kernel *k*-means clustering algorithm[14].

For global *k*-means clustering algorithm number of clusters is a priory knowledge. several cluster numbers were tried with this algorithm and the best number of clusters are obtained with kernel- based Dunn's Index which is used as a technique for cluster validation. In developing size charts from those clustered data, size intervals for the key measurements need to be decided. Out of these 13 variables, waist, hip and inseam length were identified as key variables based on Pearson correlation coefficients and the same was also endorsed in literature. Hence, size interval for waist, hip and inseam were decided based on measurement range and international size standards [29].

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Resulted size charts were validated using Aggregate loss of fit factor which is given by,

$$ALF = \frac{\sum_{i=1}^{n}\sqrt{(a_{i1}-b_{i1})^2+(a_{i2}-b_{i2})^2+(a_{i3}-b_{i3})^2}}{n} \quad \text{---- (3)}$$

where a_1, a_2 and a_3 are the assigned values for the key variables of individual, b_1, b_2, and b_3 are the actual values of the key variables of individual and n is the number of members in each clusters[2].

RESULTS AND DISCUSSION

The dataset was basically categorized into three groups as described above according to the body shapes. Accordingly, 124 samples for Small WHR, 629 samples for Medium WHR and 315 samples for Large WHR were yeild.

Polynomial kernel function and Gaussian (RBF) kernel function were tested with different parameter values and finally RBF kernel function was selected as the most suitable one as polynomial kernel function did not generate a valid kernel capable of successful clustering. The sigma value which produced a kernel matrix with highest variance was taken as the optimum value for sigma. Therefore, different sigma values were used to create kernel matrices and then checked for its variances and selected the sigma value which gave the highest variance. Accordingly, sigma value was selected as 11 for small WHR dataset and 13 for the other two datasets. Optimum number of clusters based on kernel based Dunn's Index for small WHR category was two and for other two categories were three(Table 3.1).

Table 3.1 Number of Clusters and relevant KDI for 3 datasets

	Cluster No.	KDI
Small WHR	**2**	**0.1040**
	3	0.1014
	4	0.0961
Medium WHR	2	0.0339
	3	**0.0466**
	4	0.0415
Large WHR	2	0.0636
	3	**0.0727**
	4	0.0523

After finalizing the number of clusters, using Global kernel k-means clustering approach, three datasets were clustered separately. In developing size charts from these clustered datasets, 4 cm interval was maintained for waist and hip[26] while 6 cm interval for inseam height which resulted three inseam groups; short (68 cm/26.5"), regular (74 cm/29") and long (80 cm/31.5") for all clusters. After grouping the key variables as explained above, the mean values of the secondary measurements in each group were calculated and tabulated accordingly.

For example, Table 3.2 shows the size chart for small WHR category and Table3.3 shows the adjusted size chart where the measurements were adjusted maintaining a constant gap which facilitate production process.

Table 3.2 Size chart of small WHR category (waist/ hip ≤ 0.7)

Inseam	Short 68 cm			Regular 74 cm						Long 80 cm		
Waist	54	54	58	54	54	58	58	62	66	58	62	66
Hip	78	82	86	78	82	86	90	90	98	86	90	94
Thigh	42.2	45.6	48.4	42.4	45.3	47.8	50.7	50.6	56.6	47.8	51.4	53.6
Mid thigh	36.4	38.4	40.3	36.1	37.5	39.6	41.8	41.5	45.8	39.4	42.2	44.5
Knee	30	30.6	31.8	30.4	30.8	31.8	31.6	31.7	32.5	31.2	31.4	31.7
Mid calf	26	26.5	27.8	26.1	26.4	26.8	27.7	27.9	28.6	27.2	27.7	28.3
Ankle	20	20.5	21	20.3	20.6	21	21.4	21.5	22.2	20.5	22.1	22.4
Out seam	94.5	95.4	95.3	99.7	100	102.2	103.1	102	101	106.8	106.6	107.4
Crutch length	60.2	62.4	63.8	60.6	62.1	66.5	68.2	68	70.1	66	67.8	69.7
Hip height	16.1	16.4	16.5	17.4	17.1	18.6	18.9	18.6	19.1	19.3	19.4	19.8
Knee height	43.4	42.6	43.1	44.9	45.6	44.6	44.6	45.1	44	47.8	47.6	47.9
Ankle height	5.8	6.3	6.4	6.5	6.6	6.8	6.3	6.6	6.6	6.8	7.4	7.2
Percentage	7.48	6.54	7.48	6.54	3.74	14.02	3.74	20.56	9.35	5.61	7.48	4.67

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Table 3.3 Adjusted size chart for small WHR category (waist/ hip ≤ 0.7)

Inseam	Short 68cm			Regular 74 cm						Long 80 cm		
Waist	54		58	54		58		62	66	58	62	66
Hip	78	82	86	78	82	86	90	90	98	86	90	94
Thigh	42	45	48	42	45	48	51	51	57	48	51	54
Mid thigh	36	38	40	36	38	40	42	42	46	40	42	44
Knee	30	31	32	30	31	32	33	33	35	32	33	34
Mid calf	26	27	28	26	27	28	29	29	31	28	29	30
Ankle	20	20.5	21	20	20.5	21	21.5	21.5	22.5	21	21.5	22
Out seam	95	95	95	100	100	102	102	102	102	107	107	107
Crutch length	60	62	64	60	62	66	68	68	70	66	68	70
Hip height	16	16	16	17	17	19	19	19	19	19.5	19.5	19.5
Knee height	43	43	43	45	45	45	45	45	45	48	48	48
Ankle height	6	6	6	6.5	6.5	6.5	6.5	6.5	6.5	7	7	7
percentage	7.48	6.54	7.48	6.54	3.74	14.02	3.74	20.56	9.35	5.61	7.48	4.67

"Aggregate loss of fit" factor for the size chart for Small WHR category was 5.8 cm which is slightly deviated from the ideal value of 4.4 cm as found in the literature. Therefore, it can be justified that the resulted size chart is accurate enough to avoid the fit problems. For currently available size charts, aggregate loss of fit factor is many times of the ideal value and hence it drastically deviated from the ideal value. So it can expect that the developed size chart using the proposed method based on global k-means clustering algorithm with Dunn's Index cluster validation will success in solving the fit problems of pants in Sri Lankan females.

CONCLUSION

Despite several methods ranging from statistical methods to data mining approaches used by the past researchers in developing size charts could not successfully address the apparel fit problems. Global kernel k-means clustering method, which was used for MRI segmentation, was successfully deployed to cluster the Sri Lankan females' lower body anthropometric data for developing size charts for female pants. The proposed kernel- based clustering method can handle high dimensional, non-linear data in input space effectively and make it linearly separable in feature space.

The number of clusters which is a priori knowledge to the algorithm as well as the parameters of transforming the data space to feature space was also optimized through Dunn's Index, a clustering validation technique. The effectiveness of the developed size chart was justified with calculation of aggregate loss of fit factor, which is an acceptable index found in literature to evaluate the level of fit.

Since the proposed algorithm can be run on a modern personal computer with a larger set of data using Matlab and Minitab softwares, it is possible to further refine the size chart for better fit with large sample size of anthropometric data. In addition, the rationale behind the proposed method of developing the size chart is mathematically justified in addition to providing a measure to assess the effectiveness of fitting and hence the proposed method is independent of the set of data. So the method has strong implications to generalize globally in developing size charts with the anthropometric data in other countries.

ACKNOWLEDGEMENTS

This work was supported by the University Grant Commission of Sri Lanka (UGC/ICD/RG 2013)

REFERENCES

1. Beazley, A. (1999). Size and Fit: The development of size charts for clothing; part 3. Journal of Fashion Marketing and Management; Vol 3, No. 1 , 66 -84.
2. Gupta, D., & Gangadhar, B. (2004). A Statistical Model for Developing Body Size Charts for Garments. International Journal of Clothing Science and Technology,Vol 16 Iss:5 , pp458-469.
3. Hsu, C. (2009). data mining to improve industrial standards and enhance production and marketing:An empirical study in apparel industry. Expert systems with Applications, Vol36(3) , 4185-4191.
4. Shahrabi, J. ,. (2010). Development of a new Sizing System Based on Data Mining

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Approaches. 7th International Conference-TEXSCI .
5. Chung, M., Lin, H., & Wang, M. (2007). The development of sizing systems for Taiwanese elementery and high school students. International Journal of Industrial Ergonomics Vol 37,issue8 , pp 707- 716.
6. Lin, H., Hsu, C., & Wang, M. (2007). An application of data mining technique in Engineering to facilitate production management of garments. 11th WSEAS international Conference on Computers. Greece: Agios Nikolaos.
7. Doustaneh, A., Gorji, M., & Varsei, M. (2010). Using Self Organization Method to Establish Nonlinear Sizing System. World Applied Sciences Journal 9 (12) , 1359-1364.
8. Taylor, J., & Cristianini, N. (2004). Kernel Methods for Pattern Analysis. NewYork: Cambridge University Press.
9. Lodhi, H., Saunders, C., Taylor, J., Cristianini, N., & Watkins, C. (2002). Text Classification using String Kernels. The Journal of Machine Learning Research vol 2 , 419-444.
10. Methasate, I., & Theeramunkong, T. (2007). Experiments on Kernel Tree Support Vector Machines for Text Categorization. In Z. Zhou, & Q. Yang, Advances in Knowledge Discovery and Data Mining Vol 4426 (pp. 720-727). Berlin: Springer-verlag.
11. Wang, B., Xiong, H., Jiang, X., & Ling, F. (2012). Semi-Supervised Object Recognition Using Structure Kernel. Image Processing (ICIP), 2012 19th IEEE International Conference on (pp. 2157 - 2160). IEEE.
12. Liu, Z., Chen, D., & Bensmail, H. (2005). Gene Expression Data Classification With Kernel Principal Component Analysis. Journal of Biomedicine and Biotechnology Vol 2005(2) , 155-159.
13. Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T., & Müller, K. (2000). Engineering support vector machine kernels that recognizetranslation initiation sites in DNA. Bioinformatics, vol. 16 , 799–807.
14. Tzortzis, G., & Likas, A. (2009). The Global Kernel K-Means Algorithm for Clustering in Feature Space. IEEE Transactions on Neural Networks, Vol20(7) , 1181-1194.
15. Likas, A., Vlassisb, N., & Verbeekb, J. (2003). The global k-means clustering algorithm. The Journal of Pattern Recognition Society ,Vol 36 , 451-461.
16. Pochet, N., Ojeda, F., De Smet, F., De Bie, T., Suykens, J., & De Moor, B. (2007). Kernel Clustering for Knowledge Discovery in Clinical Microarray Data. In G. Valls, J. Alvarez, & M. Ramon, Kernel Methods In Bioengineering, Signal and Image Processing (pp. 64-92). Hershey: Idea Group Publishing.
17. Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. Journal of Intelligent Information Systems, Vol 17(2/3) , 107–145.
18. Han, J., Kamber, M., & Pei, J. (2012). Data mining Concepts and Techniques. Waltham: Morgan Kaufmann publishers.
19. R. Chitta, R. Jin, T. C. Havens, and A. K. Jain, "Approximate Kernel k-means: solution to Large Scale Kernel Clustering", KDD, San Diego, CA, August 21-24, 2011.
20. Zhang, R., & Rudnicky, A. (2002). A large scale clustering scheme for kernel k-means. The 16th International Conference on Pattern Recognition (pp. 289-292). IEEE.
21. Kim, D., Lee, K., Lee, D., & Lee, K. (2005). Evaluation of the performance of clustering algorithms in kernel-induced feature space. Pattern Recognition vol 38 , 607 – 611.
22. Cristianini, N., Taylor, J., & Saunders, C. (2007). Kernel Methods: A Paradigm for Pattern Analysis. In G. Valls, J. Alvarez, & M. (. Ramon, Kernel Methods in Bioengineering, Signal and Image Processing (pp. 1-41). Hershey: Idea Group Publishing.
23. Dunn, J. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics Vol 3(3) , 32-57.
24. Fa, R., Nandi, A., & Jamous, B. (2012). Development and Evaluation of Kernel-Based Clustering Validity. 20th European Signal Processing Conference (pp. 634-638). Bucharest: EURASIP.
25. Phillips, J., & Venkatasubramanian, S. (2011). A Gentle Introduction to the kernel Distance. Computing Research Repository-CORR Vol. abs/1103.1 .
26. Beazley, A. (1998). Size and fit: Formulation of body measurement tables and sizing systems; part2. Journal of Fashion Marketing and Management Volume 2 Number 3 , 260-284.
27. Singh, D. (2002). Female Mate Value at a Glance: Relationship of Waist-to-Hip Ratio to Health, Fecundity and Attractiveness.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Neuroendocrinology Letters,Suppl.4,Vol23 , 81-91.
28. Sugiyama, L. (2005). Physical Attractiveness in Adaptionist Perspective. Handbook of Evolutionary Psychology (pp. 292-343). Hoboken: John Wiley
29. Beazley, A. (1997). Size and Fit: Procedures in undertaking a survey of body measurements. Journal of Fashion Marketing and Management, Vol.2 (1) , 55-85.