ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH - TECHNOLOGY

APPLICATION OF DATA MINING METHODS IN ESTABLISHING SIZING SYSTEM FOR CLOTHING INDUSTRY

Elawad F Elfaki Elnour*, Akram H M Ali, Mohammed Elhafiz, Salah A latif Mohammed
Department of Textiles Engineering, College of Engineering & Industries Technology, Sudan University of Science and Technology, Khartoum, Sudan.

ABSTRACT

The purpose of this paper is to present the possible uses of data mining methods in establishing standard sizing system based on anthropometric body measurements variables. Sudanese army officers database was selected for establishing sizing system because of the need for producing army officers uniform (poshirt) for the whole body types (upper and lower).The anthropometric data was collected for 841 army officers from Sur Military Clothing Factory in Sudan. For each individual (13) anthropometric variables were involved resulting in a total of 10933 variables. The data mining methods applied to establish the standard sizing system were (WEKA 3.6.9 and SPSS). However, both methods were used for clustering and establishing sizing system by implementing the Simple K-means algorithm to determine the final cluster classification .Cluster analysis using chest and waist as a control anthropometric variables revealed a proposed new established sizing system with eight distinct clusters. The eight types of body shapes namely; XXS, XS, S, M, L, XL, XXL and XXXL respectively.

KEYWORDS: Data mining, anthropometric data, sizing system, clothing industry, clustering

INTRODUCTION

In the last twenty years there was an extraordinary expansion of computer accessible data about all kinds of human activities. The availability of these large volumes of data, and our limited capabilities to process them effectively, creates a strong need for new methodologies for extracting useful, task – oriented methodologies for deriving plausible knowledge from small and directly relevant data.

In order to automatically generate useful knowledge from a variety of data, and presented it in human oriented forms, a powerful tools is strongly needed. Researchers have been exploring ideas and methods in different areas as efforts to satisfy this need. Such areas include; data mining, text mining, machine learning, statistical data analysis, data visualization, and pattern recognition.

Fayyad, and Piatetsky – Shapiro, define, data mining as, "the process of extracting valid, previously unknown comprehensible information from large databases in order to improve and optimized business decisions [1]. In order to successfully implement data mining methods in design and manufacturing processes figure 1 illustrates the issues should be considered.

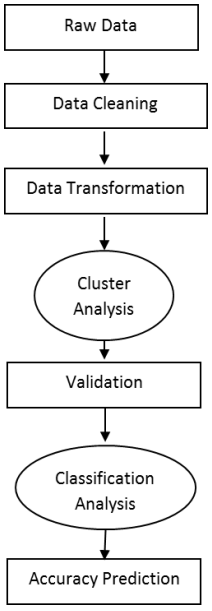


Figure1. Data Mining Processes

BACKGROUND

Data mining uses in apparel industry:

Data mining could be used effectively in apparel manufacturing produces products with highest added value in global textile manufacturing chain. In manufacturing apparel it is so important to produce

apparel with best design based on standard size chart to fit all the body types. In recent years, there have been more attention to develop new sizing systems based on data mining [2]. Applied a data mining techniques to develop industrial standards for adult females he applied two - stage clustering approach to generate a standard size chart (2). In another work he established systems for using a decision tree technique to determine the pants sizes of army soldiers [3].

In the 7th International Conference – TEXSCI held in Czech Republic, Jamal S. and Maryam S.E presented paper in which a new sizing chart based on Iranian male body size is developed using principle component analysis (PCA), and clustering approach and the effect of fitness of final sizing system chart is investigated[4].

In (2010) R. Bagherzadeh et.al introduced a three-stage data mining procedure for developing sizing system using anthropometric data for lower figure type of Iranian male people determined [5]. Result showed that the three body type and sizing system developed, have a good fit performance. Nor Saadah Zakaria et al 2008 conducted an anthropometric survey of Malaysian girls using data mining technique to explore anthropometric data towards the development of sizing system 33 different body dimensions were taken from each subject following the ISO 8559-1989 standard for body measurement [6, 7]. The whole data was analyzed using descriptive analysis of average, mean and standard deviation. Then the factor analysis method was used to explore data. Principle component analysis (PCA) technique was done to reduce the variables to similar factor components. Decision tree technique was used to validate cluster groups in order to convert these segmented groups into size tables.

In March /April (2012) M. Martin Jeyasingh et al attempted to predict clothing insulation factors with the goal of understanding the computational character of learning, the data mining technique used was linear regression because it is quite effective in terms of high prediction rate [8]. Also, linear regression was able to discover the clothing isolation performance in a most efficient manner in comparison to all other learning algorithms experimented.

In another work more recently published online June 2012 in MECS http://www.mecs-press.org/ M. Martin Jeyasingh et al; conduct research for developing sizing systems by data mining techniques applied to Indian anthropometric dataset [9]. They proposed a new approach of two-stage data mining procedure for labeling the shirt types exclusively for Indian men. With the application of clustering technique on the original dataset they managed to categories the size label.

The sizing system classifies these clusters to a specific population into homogeneous subgroups based on some key body dimensions.

The result of that research, they have obtained classifications for men's shirt attributes based on clustering techniques. Table 1shows the uses of data mining in apparel industry.

This paper attempting to find standard size system for Sudanese army officers uniform (poshirt) sizes for Sudanese men by implementing data mining methods.

Table1. Application of data mining techniques in developing sizing system for clothes in different countries.

Country	Taiwan	Iran	Iran	Malay	India
Targeted size	Size System for Army Soldiers	Size system for pants Male:16-22	Size system for suit	Size system for female 7-12	Size system for shirt for men 25-66 years
Data Mining Techniques Applied	Classification Decision-tree	Quadratic average of difference Aggregate loss goodness of fit sizing Hierarchical &Clustering Classification Decision tree (CART)	Aggregate Loss of Fitness	Decision Tree Clustering	Error Improvement Regression Classification K-nearest neighbor 2(knn)+Random tree
Algorithms	Factor Loading Clustering Body Mass Index(BMI)	K-means finally Multivariate Analysis cluster SPSS(Equal-Variance Maximum like LiHood (EML) factor analysis	K-means Algorithm Clustering Non-Hierarchical Clustering approach	Principle Component Analysis(PCA) Factor Analysis exploration SPSS	X – means clustering missing data by multivariance techniques
Techniques	Decision Trees	Three Stages	PCA	Two Stage	Two stage
Year	2008	2010	2010	2010	2012

From this review the data mining methods used in the field of sizing system in clothing industry can be summarized as follows;
1) Neural network.
2) Cluster analysis.
3) The decision tree approach.
4) Two stage cluster analysis.
5) Three stage data mining procedure.
6) Two stage based data mining procedure include cluster analysis and classification algorithms.

MATERIAL

In this research, the anthropometric database for army officers obtained from (Sur Military Clothing Factory) located in Khartoum North (Industrial Area), in the Sudan used. Based on the experience and the advice of the experts the selected variables for the jacket were collar, chest, waist, length, across shoulder, sleeve and sleeve and cuff. For the trouser variables were; waist, hip circumference, thigh circumference, knee girth, foot, and trouser length. These measurements of anthropometric data followed the ISO 8559/1989 body measurement standard. The thirteen dimensions for (poshirt) measured by using a measuring tape (7).

METHODS

The choice of the software was not an easy job. From the review of earlier research, it was observed that there are several different mining software packages that had their own strengths and weakness. Based on this information it was decided to select two software packages that were suitable for establishing sizing system.

WEKA 3.6.9

In this research Waikato Environment of Knowledge (WEKA) was used. WEKA is software written in Java and runs on almost any plat form. It is a data mining system developed at the University of Waikato in New Zealand. WEKA is free software available under the GNU (General Public License). The WEKA work bench contains collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality.
The reasons for choosing these techniques were; WEKA 3.6.9 is more advanced version which has been implemented in Java with latest Windows7 operating in Intel Core ZQuad@2.83 Ghz and 2 GB memory, with a fairly simple microprocessor as software allows for the data to be imported into their software directly from Excel(8).

4.2 The Statistical Package for Social Sciences (SPSS) Version 18.0:

This package was the second method selected because it is a simple clustering method and shows optimal results. However all variables must be independent and normally distributed. Clusters were identified according to the different body types. It was employed for anthropometric data analysis. It reduces the large samples in same groups contains similar number. The co-efficient correlation in determining the relationships between the dimensions can be obtained. Finally the WEKA and the (SPSS) methods could be used to establish new sizing systems.

RESULTS AND DISCUSSION

The Implementation of Data Mining Techniques:

Data mining is the process of extracting useful information from large data bases [10]. The increasing power of computer technology has increased data collection and storage. Automatic data processing has been aided by computer science, such as a neural network, decision trees, genetic algorithms, clustering, and support vector machines (machine learning). Data mining is the process of applying these methods for the prediction of uncovering hidden patterns [11]. Before starting to mine the data, the first step was to define the problems in clothing industry in Sudan related to this research.

Launching WEKA Explorer:

In this research the following steps were followed through the analysis of the problem using WEKA Explorer, preprocessing, classification, clustering, association, attribute selection, and visualization. Figure (2) shows these steps when using "cluster" tab at the top of WEKA Explorer window.
WEKA can be launched from c:\ program files directly, from the desktop selecting WEKA 3.4, shortcut 2 KB icon or from the windows task bar "start" → programs → WEKA 3.4. When WEKA Gui chooser, window appears on screen, one of the four options button of the window can be selected [12].
WEKA 3.6.9 could be launched by following the same steps mentioned for WEKA 3.4. When WEKA Gui chooser' windows appears on screen and one of the following steps could be selected. The steps one; (preprocessed, classify, cluster, associate, select attribute, and visualization) as shown in figure (3).

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

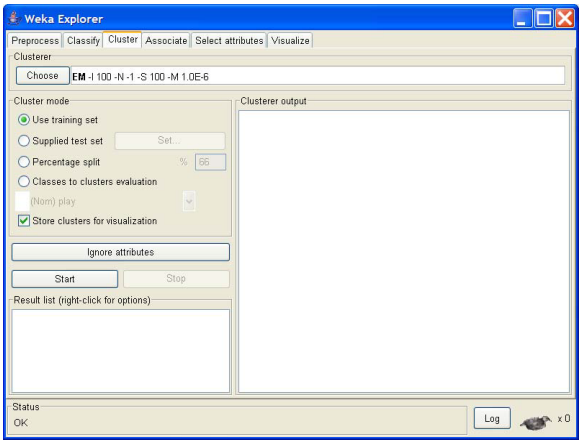


Figure2. Using 'Cluster' tab at the top of WEKA Explorer window

Cluster Analysis:

Cluster analysis is a powerful technique to divide heterogeneous data into groups. Clustering is the portioning of a dataset into subset (clusters), so that the data in each subset (ideally) share some common trait. Cluster analysis was used as an exploratory data analysis tool for classification. In the clothing cluster which is typically grouped by the similarity of its member's body shape can be considered as a size category or a figure type. The simple K-means method was implemented to determine the final cluster categorization. Simple K-means algorithm was used as the clustering approach.

The dataset in this research has different characteristics, such as the number of instances, the number of attributes, the number of classes, the number of records, and the percentage of classes' occurrences.

In this research the database was designed in (size.csv) file system to store the collected data. The data was formed according to the required format and structures. Further the database was converted to (size.csv) format to be processed in WEKA. The text file describes a list of instances that sharing a set of attributes. After processing the (size.csv) file in WEKA a list of all attributes, statistics, and other parameters can be utilized as shown in figure 3.

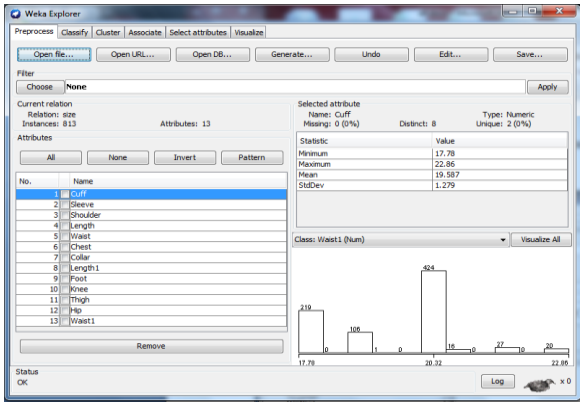


Figure3. Processed (size.csv) file in WEKA

Analysis of Data Processed in WEKA:

As mentioned before, the processed data in WEKA can be analyzed by using different data mining techniques such as, classification, clustering, association route mining, visualization algorithms etc; Figure 4. Shows the processed attributes visualized into a 2 dimensional graphical representation.

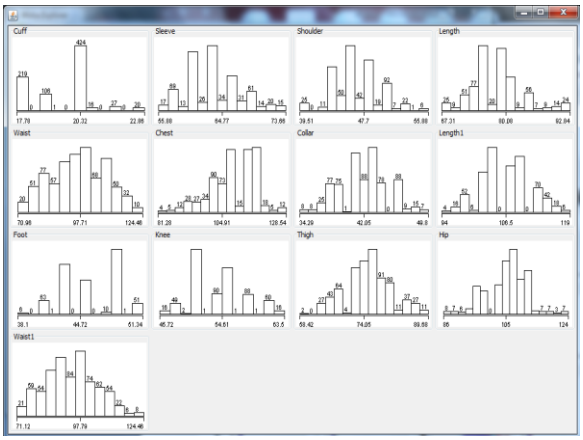


Figure4: Graphical visualization of processed attributes

Cluster 7, 8, 9:

These clusters include the attributes from 0 – 6, 0 – 7, and 0 – 8 respectively.

- First set the value in (num Clusters box) to 7 instead of default (2), 8 instead of 7 and finally 9 instead of 8. Therefore, the clustering scheme used was Simple K-Means with 7 clusters as an example. When set to run mode the following information will appear on the dialogue box.
- The relation name "size".
- Number of instances in the relation is 813.
- Number of attributes in the relation is 13.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

The 13 attributes used in clustering are shown in figure (5).

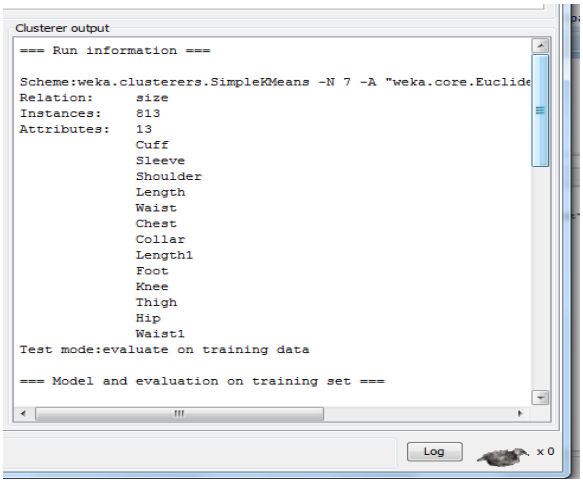


Figure6. The output from the run mode

WEKA Analysis and results:

From the results obtained from experimental work it is clear that WEKA 3.6.9 is quite effective in terms of clustering and visualization performance rate.

Table (2) the distribution of classes for cluster 8

Class category	No. of Records	Percentage of Class Occurrences%
XS	75	09
S	97	12
M	128	16
L	63	08
XL	161	19
2XL	64	08
3 XL	120	15
4 XL	105	13

The clustering model shows the centroid of each cluster and statistics on the number and percentage of instances assigned to different clusters. Cluster centroids are the mean vectors for each cluster; therefore, each dimension value and the centroid represent the mean value for that dimension in the cluster.

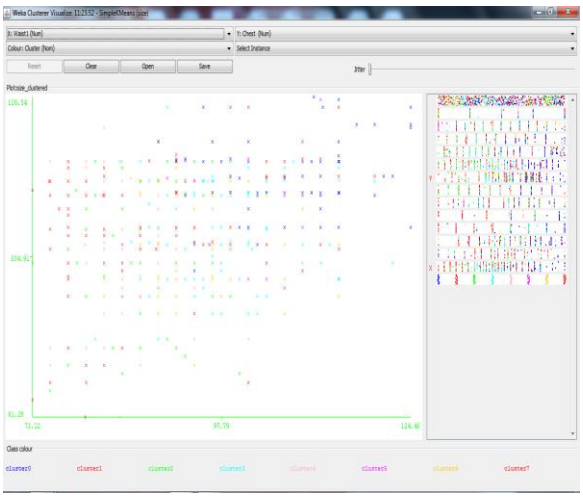


Figure 6. Visualization for cluster 8

Procedure for Establishing the New Sizing Systems:

Therefore, centroids can be used to characterize the clusters. Table 3 shows the statistical analysis for the size ranges from the raw data.

Body Dimensions(cm)	Mean	Standard Error of Mean	Std. Deviation	Min	Max
Cuff	19.6	0.04	1.3	17.8	22.9
Sleeve	63.6	0.12	3.5	55.9	73.7
Shoulder	46.6	0.11	3	39.5	55.9
Length	78.2	0.07	4.8	67.3	92.8
Waist	96.5	0.41	11.6	71	124.5
Chest	109	0.3	8.5	81.3	128.5
Collar	41.9	0.1	2.9	34.3	49.8
Length 1	106.3	0.2	4.6	94	119
Foot	45.8	0.1	2.9	38.1	51.3
Knee	53.9	0.14	3.9	45.7	63.5
Thigh	75.2	0.2	5.7	58.4	89.7
Hip	105.2	0.21	6.0	86	124
Waist 1	94	0.37	10.7	71.1	124.5

Table3. Statistical Analysis for the Size Ranges from the Raw Data:

n = 813 (all values are in centimeters)

- Cluster 8 gives the size that represents these clusters and it shows the number of persons fall in these classes and the percentage from the total instances see Table 2.

On the other hand, in cluster 8 the percentage covered by the proposed sizing system was 96.7 and there was only one size figure (5XL) with no classes that was not represented. This may be due to the fewer number of persons in this size figure.

Therefore, cluster 8 seems to be the best sizing system that represents the data collected from Sur factory. This is because it covered nearly 96.7% of the data and it includes 8 figure types. The new established sizing systems consists of the following sizing figures, XXS, XS, S, M, XL, 2XL, 3XL, and 4XL respectively.

THE (SPSS) ANALYSIS AND RESULTS

The Statistical Package for the Social Sciences (SPSS) version 18.0 for windows was employed for an anthropometric data analysis. In this work data analysis were carried but by using K-means method to reduce the large samples in same groups contains similar number. The results were obtained for clusters 7 up to cluster 9. The iteration was 10 for all the clusters mentioned above. Descriptive statistics including Analysis of Variance (ANOVA) means square, standard error of mean, standard deviation, minimum and maximum values were calculated and utilized for the analysis and the determination of the correlations see table(4). The values were calculated in centimeters. All values of the standard deviation and others descriptive statistics are rounded to two decimals.

The results of the cluster samples were analyzed statistically using (SPSS) version 18.0 package. The analysis results (mean and standard deviation) of the sample are tabulated in table (3).

The development of the size chart was carried out by using values obtained from the statistical information of body dimensions. Winks (1997) states that, the mean value can be a convenient indication of obtaining central tendency (15). The mean values are the most widely used value for size steps. Table (4) shows the descriptive statistics for body dimension, and it is equivalent to the average size (mean) and also equivalent to size 12 of every size chart. Nine size steps approach was used to establish the new size chart as given in table (4). The nine size steps used as a base for the determination of the outliers. The values that were less than the smallest size and those higher than the biggest size were eliminated and classified as outliers.

Table 4 Nine Steps Size Ranges

	XXXS M-4STD	XXS M-3STD	XS M-2STD	S M-1STD	M Mean	L Mean+1 STD	XL M+2STD	XXL M+3STD	XXXL M+4STD
Cuff	14.4	15.7	17	18.3	19.6	20.9	22.2	23.5	24.8
Sleeve	49.6	53.1	56.6	60.1	63.6	67.1	70.6	74.1	77.6
Shoulder	34.6	37.6	40.6	43.6	46.6	49.6	52.6	55.6	58.6
Length	59	63.8	68.6	73.4	78.2	83	87.8	92.6	97.4
Waist	50.3	61.9	73.5	85.1	96.7	108.3	119.9	131.5	143.1
Chest	75	83.5	92	100.5	109	117.5	126	134.5	143
Collar	30.3	33.2	36.1	39	41.9	44.8	47.7	50.6	53.5
Length 1	87.9	92.5	97.1	101.7	106.3	110.9	115.5	120.1	125
Foot	36	38.9	40	42.9	45.8	48.7	51.6	54.5	57.4
Knee	38.3	42.2	46.1	50	53.9	57.8	61.7	65.6	69.5
Thigh	52.6	58.2	63.8	69.5	75.2	80.9	86.6	92.2	97.8
Hip	81.2	87.2	93.2	99.2	105.2	111.2	117.2	123.2	129.2
Waist 1	51.3	61.9	72.6	83.3	94	104.7	115.4	126.1	136.8

n = 813 all values are in centimeters

As shown in table (3), in order to obtain nine steps for nine categories of body size, (1STD), (2STD), (3STD) and (4STD) values were added to the mean and subtracted from the mean respectively. This was carried out in order to obtain four values that are higher and four values that are lower than the mean. According to Ashdown (1998) by subtracting (-1STD), (-2STD), (-3STD) and (-4STD) from the mean, the values obtained represents size4, 6, 8, and10 respectively [13]. When (1STD), (2STD), (3STD) and (4STD) are added to the mean, the values obtained represent sizes 14, 16, 18 and 20 respectively. The mean and standard deviation values

were all rounded up to 0.1 decimals. Values above 0.15 were rounded up to 0.2 cm, and values below 0.15 have been reduced to 0.1 cm.

To get the new established sizing charts from the nine steps, the values of the 5XL figure size were omitted based on the results of the outlier from cluster 8, where there were no classes represented in this figure size. See table (4). Therefore, the new established sizing systems chart which consists of 8 figure size is given in table (5).

Table5. The Proposed New Established Size System

	XXS M-3STD	XS M-2STD	S M-1STD	M Mean	L Mean+1STD	XL M+2STD	XXL M+3STD	XXXL M+4STD
Cuff	15.7	17	18.3	19.6	20.9	22.2	23.5	24.8
Sleeve	53.1	56.6	60.1	63.6	67.1	70.6	74.1	77.6
Shoulder	37.6	40.6	43.6	46.6	49.6	52.6	55.6	58.6
Length	63.8	68.6	73.4	78.2	83	87.8	92.6	97.4
Waist	61.9	73.5	85.1	96.7	108.3	119.9	131.5	143.1
Chest	83.5	92	100.5	109	117.5	126	134.5	143
Collar	33.2	36.1	39	41.9	44.8	47.7	50.6	53.5
Length 1	92.5	97.1	101.7	106.3	110.9	115.5	120.1	125
Foot	38.9	40	42.9	45.8	48.7	51.6	54.5	57.4
Knee	42.2	46.1	50	53.9	57.8	61.7	65.6	69.5
Thigh	58.2	63.8	69.5	75.2	80.9	86.6	92.2	97.8
Hip	87.2	93.2	99.2	105.2	111.2	117.2	123.2	129.2
Waist 1	61.9	72.6	83.3	94	104.7	115.4	126.1	136.8

n = 813 all values are in centimeters

EVALUATION OF THE NEW ESTABLISHED SIZING SYSTEMS

As stated by the International Standards Organization (ISO), and those others Organizations mentioned in the literature; "the use of control dimensions and size interval can effectively facilitate to recognize the parameters for developing sizing systems" (Winks 1997) [14].

After the nine figure sizes were classified by the WEKA and SPSS soft wares, the new established size system of the eight figures size were determined and the results are given in table (5). Figure (7) shows the relevant scatter plots of chest on the X-axis verse the waist on the Y-axis and the interval was 4 cm to demonstrate the distribution of all figures type. It has been reported that, Cooklin (1992) the chest is the most important anthropometric variable most in establishing sizing systems in the field of garment making [15]. The waist is also an important variable for sizing male garments in many countries. Figure (8) illustrates the differences between the eight types for the new established sizing systems. The figure was plotted as a line graph to yield a better insight into the differences between the new established sizing systems. The eight figures types are exhibited by clear differences in chest, waist and hip. The eight figure types also follows the order; 2XS, SX, S, M, L, XL, 2XL, and 3XL. In addition, figure (8) show significant differences between the eight figures types mentioned above.

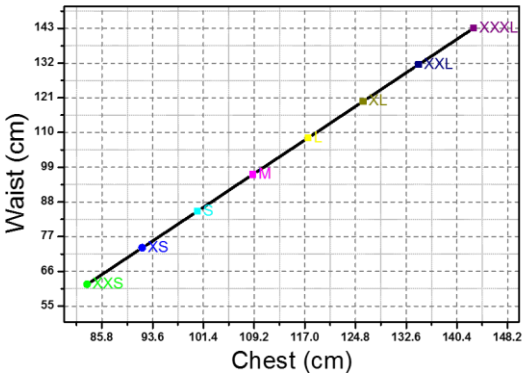
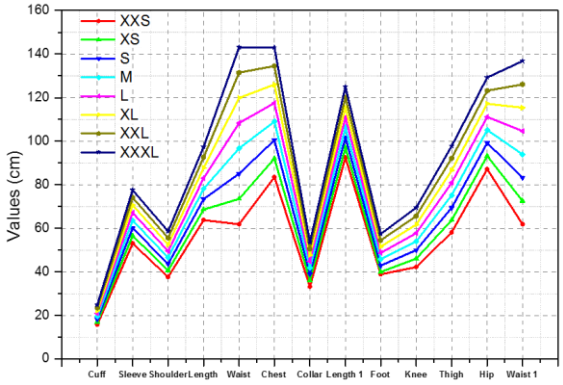


Figure7. Scatter plot of chest verse waist for the proposed new established size system



Anthropometric Variables

Figure8. The distinct anthropometric variables Between the proposed new established size systems

CONCLUSION

In this study, data mining methods (WEKA and SPSS) were applied in order to establish a sizing system for Sudanese army officer's uniform (poshirt). The study used database obtained from Sur Military Clothing Factory in Sudan. Anthropometric data for (841) officers was used. For each individual (13) anthropometric variables were involved resulting in a total of 10933 variables. Large amounts of data were

analyzed by applying the (WEKA and SPSS) methods, and identify systematic patterns in bodily dimensions. Based on these patterns, the representative figure types of Sudanese army officers were clustered and classified, and then standard sizing systems was established. The army officers' database was selected for establishing sizing systems because of the urgent need for accurate sizing systems for producing army officer's uniforms (poshirt).

The WEKA and SPSS methods were used for clustering and establishing sizing system by implementing simple K-means algorithm to determine the final cluster classification.

The study also showed that within the army officers aged 16-60 years there exist eight types of body shapes namely; XXS, XS, S, M, L, XL, XXL and XXXL respectively.
The percentage of army officers who fall in a certain figure type and sizes can serve as a good reference to indicate the quantity of garments to be produced for specific market. Thus a realistic plan for producing male army officer's uniforms can be established.

REFERENCES

[1] Fayyad, U.M. Piatetsky- Shapiro, G. Smyth, P. Uhturusamy R.(Eds)(1996b). Advances in Knowledge Discovery and Data Mining.AAAI Press, San Mateo, CA.
[2] Hsu, C. H, Lin. H F and Wang, M.I.J (2007) Developing Female Size Charts for facilitating garments Production by using data mining. Journal of the Chinese institute of industrial engineers.Vol 24, Issue3, 245-251.
[3] Hai-Fen Lin,Chih-Hung Hsu, Mao-Jiunj. Wang,Yu-Cheng Lin. An Application of Data Mining Technique in Developing Sizing System for Army Soldiers in Tai Wan, WESAS TRANSACTION on Computers, ,Issue 4, Volume 7,April 2008
[4] Jamal et al 2010, Development of a New Sizing System Based on Data Mining Approaches, 7th International Conference. TEXSCI 2010, September 6 – 8, Liberec, CZECH REPUBLIC.
[5] R. Bagherzadeh, M. Latis and A.R Faramarzi, 2010, employing a Three- Stage Data Mining Procedure to Develop Sizing System, Worled Applied Sciences Journal 8(8):923-929.
[6] Norsaadah Z, Jamil S M, Nasir T, Young Y Y and Bee Wah, Using Data Mining Techniqu to Explor Anthropometric Data Toward the Development of Sizing System
[7] ISO 5859, 1989. Garment Constriction and Anthropometric Surveys- Body Dimensions, International Organization for Standardization
[8] M. Martin Jeyasingh, Kumaravel, Appavoo, 2012. Mining the shirt sizing for Indian Men By Clustered Classification I.J. Information Technology and Computer science, 2012, 6, 12-17
[9] M. Martin Jeyasingh, Kumaravel, I.J. Information Technology and Computer Science 2012, 6, 12-17 Published online June 2012 in MECS (http://www.mecs-press.org/)
[10] D. Hand, H. Mannila, P. Smyth, Principle of data Mining. The MTT Press, Massachusetts. London. England. 2001
[11] Liang Xun, 2006, Data Mining: Algorithms and Application. Beijing University Press: p.p. 22-42
[12] R. Kirkby, WEKA Explorer User Guide for Version 3-3-4, University of Weikato, 2002.
[13]Ashdown, P.S. (1998) An investigation of the structure of sizing systems: 'A comparison of three multidimensional optimised sizing system generated from anthropometric data with ASTM standard D55855 – 94'. Journal of Clothing Science and Technology. Vol. 10, No. 5, pp. 324 – 341.
[14] Winks, J. M., Clothing Sizes: International Standardization, Redwood, U.K. 1997.
[15] Cooklin, G., Pattern Grading for Men's Clothes, Blackwell, Oxford. 1992.