THE USE OF K-MEANS AND KOHONEN SELF ORGANIZING MAPS TO CLASSIFY COTTON BALES

J I Mwasiagi[1,3], Huang XiuBao[1], Wang XinHou[1] and Chen Qing-Dong[2]

[1]Department of Textile Engineering, Donghua University
1882, Yan-an Xi Lu, Shanghai 200051, China
[2]Inspection Center of Industrial Products and Raw Materials, Department of Textiles Materials, B-412, No. 1208, Minsheng Lu, Pudong New Area Shanghai, 200135, China

Abstract

The use of High Volume Instrument (HVI) system has enabled fast and reliable measurements of cotton fiber characteristics thus producing high dimensional data. This calls for the use of clustering techniques to adequately interpret and utilize the data. Clustering techniques classify objects based on attributes into distinct classes (clusters). The HVI characteristics can be used to group cotton bales so that the within group variations are kept at a minimum. This will ensure that all the bales in a given group have the highest level of similarity hence help reduce lot to lot variations in the manufactured yarn. A bale classification model using K-means clustering technique and Kohonen self organizing maps (SOM) is discussed. The model is used to classify 2421 cotton bales whose HVI data containing 13 cotton attributes, was obtained from Shanghai inspection center of industrial products and raw materials. The model reduced the 2421x13 HVI high dimensional data into 18x13 grids, with a quantization error of 1.879 and a topographic error of 0.083, and resulted in the identification of 16 groups of cotton bales and one group of outliers. The outliers could be further subdivided into five subsets.

Keywords: *Cotton fibers, Kohonen self Organizing maps (SOM), k-means, clustering, Bale classification*

Introduction

The quality of cotton yarn is affected by cotton fiber quality characteristics such as fiber length, length uniformity, short fiber index, elongation, trash measurements, micronaire, strength, and color measurements. [Kothari,1999; Hequet and Ethridge,1999; Jackowski and Frydrych; 1999]. These fiber quality characteristics can be measured by a variety of instruments. [Chen, 1999; McCreight et al, 1999]. When the measurement of single fiber characteristics such as fiber strength is done using a stand alone strength measuring instrument, it will take a lot of time and manpower to measure many samples. Consequently only representative measurements were taken and used in cotton spinning before the introduction of the High Volume Instrument (HVI). The HVI system which incorporates all fiber measuring instruments into one testing unit was developed due to a concerted effort by researchers to provide an accurate, reliable and timely system which can measure important cotton quality characteristics speedily. In fact the HVI system can measure most of the known cotton quality characteristics in about forty seconds per sample [Ghorashi, 1999]. The high testing speed of the HVI system enables the spinning mill to test each individual bale of cotton. While HVI system provides a lot of data for cotton lint the utilization of the data has however been limited to a use of one or two important characteristics such as Spinning Consistency

[3] Author to whom any correspondence should be addressed

Index (SCI) and micronaire for bale selection [Kothari, 1999; Majumdar et al, 2004]. According to Kothari [Kothari, 1999] cotton bale selection can be done by first considering SCI followed by other HVI characteristics such as micronaire, fiber length or length uniformity depending on the spinning system. The aim of this procedure is to gain between mix long-term and short term consistency of all properties from all of the cotton bales available for yarn spinning. The bale selection procedure can be simplified and made more objective by using clustering techniques together with other statistical tools used for measuring central tendency and measure of dispersion for multivariate data. Clustering techniques group items into sets of similar objects based on given attributes [Jain and Dubes, 1988]. In this paper, the design of a bale classification model that uses Kohonen Self Organizing Maps (SOM) to visualize the high dimensional cotton lint HVI data, K-means technique to cluster the data is discussed.

Kohonen Self Organizing Maps

Kohonen Self Organizing Maps (SOM) whose architecture is given in figure 1, learn to recognize groups of similar input vectors in such a way that neurons physically close together in the neuron layer respond to similar input vectors [Kohonen, 1997].The SOM algorithm has been applied in many areas due to its ability to provide an effective platform for visualization of high-dimensional data [Vivekanandan and Doke, 2003; Vesanto and Alhoniemi, 2000].

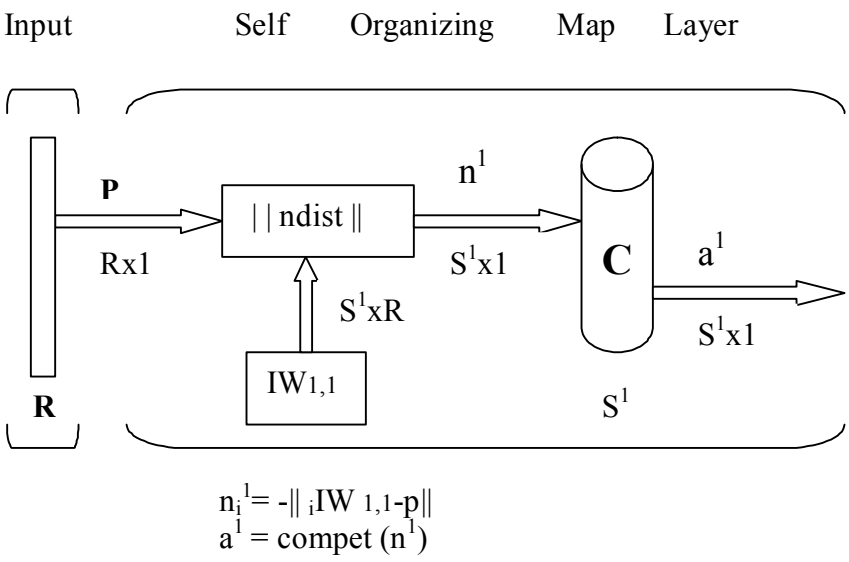

Fig. 1: The architecture of Kohonen Self Organizing Map

SOM learn both the distribution and topology of the input vectors they are trained on. The neurons in the layer of a SOM are arranged in physical positions according to given topology and distance functions [Demuth and Beale, 2005; Hagan et al, 2002]. SOM learning algorithm exploits two fundamental interactions between neurons, namely, competition and cooperation [Demuth et al, 2005; Ham and Kostanic, 2001]. Competition during learning refers to the characteristics of a neuron, where each input vector is compared with each weight vector of the weight matrix. The winning neuron, which is also referred to as the Best Matching Unit (BMU) is selected, based on the least distance between the vector and the neuron. There are a number of different ways for determining the distance between neurons. The most commonly used distance method is the

Euclidean distance. Cooperation learning on the other hand refers to the situation whereby all neurons located in a topological neighborhood of the winning neuron k have their weights updated in accordance with a neighborhood function [Jang and Mizutani, 1997] φ(j):

$$\varphi(j) = \exp\left(-\frac{\rho(j)^2}{2\sigma^2}\right) \qquad (1)$$

where $\rho(j) = \|\mathbf{V}(j,:) - \mathbf{V}(k,:)\|$ is the distance between the j^{th} and k^{th} neurons in the *l*-D (*l* =1,2 or 3) feature space, and σ^2 is a variance (or spread) of the Gaussian distribution.

When a vector is presented, the vectors of the winning neuron and its close neighbors will move towards it. Consequently after many presentations, the neighboring neurons will have learnt vectors similar to each other.

The quality of the SOM maps can be checked by using two factors: data representation accuracy and data set topology representation accuracy. Data representation accuracy can be measured using average distance between each data vector and its BMU. Data set topology representation accuracy can be expressed as the topographic error, which is the percentage of data vectors for which the first and second BMUs are not adjacent units.

K-means clustering algorithm

Clustering algorithms attempt to organize unlabeled feature vectors into clusters or natural groupings such that vectors within a cluster/group are more similar to each other than other vectors belonging to different clusters. K means is a partitioning nonhierarchical clustering method that constructs k (k fixed a *priori*) clusters for a given data set. The K-means algorithm defines k centers one for each cluster and hence has k groups. The grouping is done by minimizing the sum of squares of distances between the data members and the corresponding cluster centers. This is achieved by minimizing a squared error objective function;

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (2)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is the chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers. The algorithm is composed of the following steps;

(i) Place *k* points into the space represented by the objects that are being clustered. These points represent initial group centers.
(ii) Assign each object to the group whose center is closest.
(iii) When all objects have been assigned, recalculate the positions of the *k* centers.
(iv) Repeat the second and third steps until the centers no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The quality of the data partitioning using K-means clustering techniques can be analyzed by using cluster validity methods such as Davies Bouldin (DB) index and silhouette means. The DB index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation and can be computed as shown in equation 1 [Weisstein, 2006].

$$DB = \frac{1}{N_c}\sum_{j=1}^{N_c}\frac{\max_{j\neq k}(Sc(k)+Sc(j))}{dce(k,j)} \qquad (3)$$

where $Sc = \frac{\sum_i \| x_i - c_k \|}{N_k}$ and $dce = \| c_k - c_j \|$

In Eq. (1) *Sc* and *dce* denotes the centroid intra-cluster and inter-cluster distances respectively. The intra-cluster distance for a given cluster is obtained as the average of all pair wise distances from points in the cluster to the cluster centroid. The inter-cluster distance between two clusters is computed as the distance between their centroids. N_k is the number of objects belonging to cluster k, given that a total of N_c clusters are found to exist in the data. The DB index can be plotted for several numbers of clusters, whereby the cluster value with the lowest DB index will be deemed to be the optimum number of clusters for the data. The Silhouettes method on the other hand uses width plots. The silhouette width S_i for the i^{th} feature vector in a cluster q is defined by equation 2 [Jain and Dubes, 1988].

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \qquad (2)$$

In Eq. (2), a_i is the average distance from the i^{th} feature vector to all other feature vectors in the cluster k; b_i is the minimum average distance from the i^{th} feature vector to all the feature vectors in another cluster j (j =1,….,q; j $\neq$ q). When S_i is close to one, this implies that a_i is small with respect to b_i. This is an indication that the vector is correctly classified. If S_i is close to zero, this implies that a_i and b_i are approximately equal, thus making it unclear which cluster *i* should belong to, while negative S_i implies that *i* is assigned to the wrong cluster. For given k clusters, the overall average silhouette width is the average of the silhouette widths for all the feature vectors in the data set. The partition with the maximum overall average silhouette width is taken as the optimal partition.

As stated earlier, in k-means clustering, the value of *k* is unknown a *priori*. There is need therefore to use another algorithm to determine the value of *k*. SOM is good at pattern recognition, and hence can be used to determine the nature of clustering in a given data. This information can be used to determine the value of *k*, which can then be passed on to the K-means clustering algorithm, for the clustering of the data [Vesanto and Alhoniemi, 2000].

Statistical Techniques

There are many methods in statistical analysis which can be used to organize and summarize data. These can be studied in descriptive statistics. There are also many descriptive measures such as measure of tendency and measure of dispersion. Given a sample of data set for one attribute the measure of central tendency such as means, mode etc and the measure of variation such as standard

deviation and coefficient of variation (CV) can be used to describe the data. [Bluman, 2004; Weiss, 1997]. The results of the above analysis can be displayed using two dimensional graphs. In real life situation however there many occasions when the data to be analyzed is multivariate (i.e. it has more than one attribute). A good example of a multivariate data is HVI data, which may contain say 100 samples of cotton lint, samples each having over characteristics producing over 1000 data, having 100 rows and over ten columns. This type of data will be hard to visualize. Graphical display of data can only be done for at most three variables displayed in a three dimensional plots. For data with more than three variables it is hard to visualize their relationship. Boxplots (box-and –whisker diagram) together with principal components analysis technique can be used to study the characteristics of such high dimensional data. [Demuth et al, 2005]. A common characteristic of many multivariate data is that the variables within a given data set often move together. This may be partly due to the fact that several variables may be measuring the same driving principle governing the behavior of the system. Another common characteristic of data set with many variables is that in many systems there are only a few such driving forces. There are however an abundance of instrumentation which can be used to variables. Taking advantage of this redundancy of information, simplifies the problem by replacing a group of variables with a single new variable. The above principles are referred to as Principal components analysis. This method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other so there is no redundant information.

Materials and Methods

Materials

The HVI data used in this study was obtained from the Shanghai Inspection Center of Industrial Products and Materials (SICIPM). SICIPM is mandated to inspect all industrial products and material entering or leaving China, through Shanghai port. SICIPM has a well equipped laboratory for testing the properties of cotton fiber. The HVI data used in this study was selected from the SICIPM HVI data bank for cotton bales samples meant for importation into China through Shanghai port between October 2005 and July 2006. The cotton lint samples collected from the cotton bales were measured using the Uster HVI 1000 and 13 cotton lint characteristics were selected for this research work. The total data sample had 2424 cotton bales and 13 cotton HVI characteristics, namely micronaire (Mic), maturity (Mat), length (Len), elongation (Elg), strength (Str), Short Fiber Index (Sfi), length uniformity (Unf), Spinning Consistency Index (Sci), reflectance (Rd), yellowness (+b), trash cent (Trc), trash area (Tra) and trash grade (Trg).

Methods

The high dimensional HVI data containing 2421x13 inputs was first partitioned using K-means algorithm. DB index was run for all clusters partitioning starting from 2 to $\sqrt{2365}$ and the cluster with the lowest DB index was selected. Since DB index algorithm in the SOM toolbox starts out with random centers, the optimum cluster could vary from time to time depending on the random centers selected. The cluster validity using DB index was run for a 100 times and the least number of cluster was selected as the optimum number clusters and hence designated as k. The HVI data was then partitioned into k groups and the Coefficient of variation for each group was checked for

all the HVI attributes. Any group which was found to have a CV higher than that of the main group (MG) for any of its attributes was declared to have failed. The failed groups were further analyzed using other cluster validity methods and statistical techniques, such silhouette means and principal component analysis. All the algorithms were written in M-files using SOM and MATLAB toolboxes [Vesanto et al, 2000; The Mathworks, 2005]. All computations were performed on a Pentium IV computer with 512 MB RAM working under MS Windows XP. The bale classification model can be summarized as follows;

(i) Use SOM data visualization technique to get an idea of the nature of clustering within HVI data,

(ii) Partition the HVI data using K-means technique. The value of k should be obtained from (i) above, and

(iii) Check for data group compactness using other methods and techniques such as silhouette means, coefficient of variation and principal component analysis.

Result and Discussion

Data Visualization

The visualization of the 2421 HVI data with SOM reduced the 2421x13 data into 18x13 grids, with a quantization error of 1.879 and topographic error of 0.083. The number of hits for each node of the SOM map is given in figure 2, which shows that there are multiple hits in 231 out of the 234 nodes of the visualization grids.

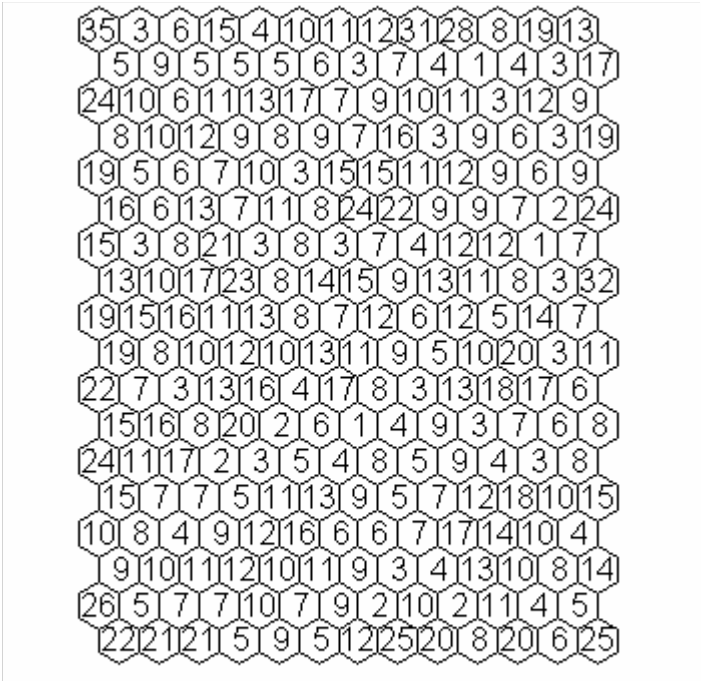

Fig. 2: Number of hits for the nodes of SOM map

Data Clustering

The partition of the HVI data was done using K-means clustering technique, and the occurrence of optimum cluster (figure 3) by using DB index as described in the methodology section. The optimum number of clusters (groups) as shown in figure 3 is 19. The means and the Coefficient of variation (CV) for the groups are given in Tables 1 and 2. The CVs for the 2nd, 7th and 17th group are higher than that of the main group (MG) for some attributes. These three classes are therefore considered not to be properly partitioned. The other 16 classes are considered to well partitioned, producing classes which are compact and different from one another.

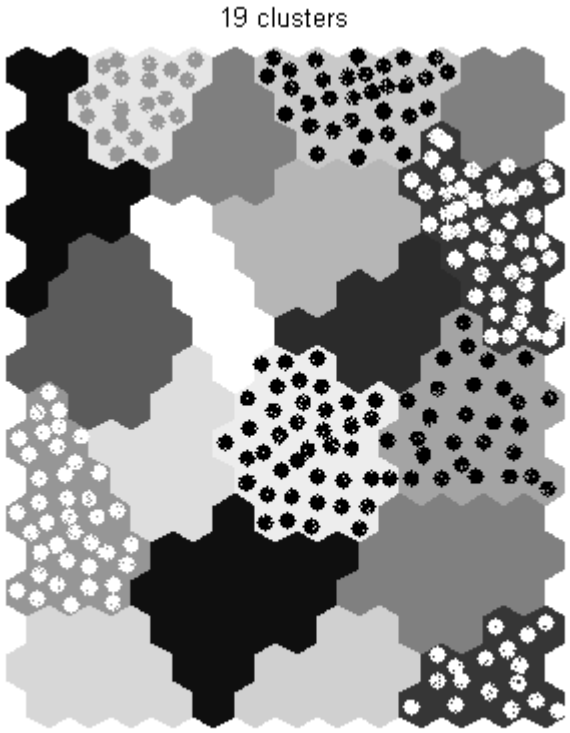

Fig. 3: Optimum number of clustering for the HVI data

Table 1: Means for the 19 groups of cotton bales

Group No.	Sci	Len	Unf	Str	Elg	Sfi	Mic	Mat	Rd	+b	Trc	Tra	Trg	No. of Data
1	99	1.05	78.59	24.04	5.79	12.71	4.36	0.86	77.70	9.14	21.94	0.28	2.56	34
2	108	1.05	79.44	25.30	5.94	12.39	4.38	0.86	78.96	10.37	24.12	0.34	2.93	57
3	110	1.07	79.78	25.01	5.88	11.80	4.28	0.86	79.46	9.23	13.35	0.17	1.58	84
4	110	1.07	79.79	25.01	5.89	11.81	4.28	0.86	79.49	9.23	13.31	0.16	1.55	83
5	111	1.07	80.18	25.93	5.79	11.68	4.44	0.87	75.38	8.62	34.66	0.43	3.54	71
6	116	1.07	80.27	26.16	6.05	11.18	4.31	0.86	79.19	9.69	22.00	0.26	2.47	123
7	117	1.09	80.86	26.44	5.90	11.11	4.28	0.86	74.28	9.03	52.20	0.66	4.68	25
8	122	1.09	81.49	27.27	6.11	10.51	4.37	0.87	74.98	8.96	47.47	0.55	4.20	55
9	123	1.09	81.42	27.18	6.01	10.34	4.42	0.87	77.81	9.40	30.56	0.33	3.06	125
10	124	1.09	80.82	26.70	6.07	10.70	4.18	0.86	80.89	9.63	11.71	0.16	1.56	186
11	129	1.10	81.57	27.34	6.04	10.21	4.15	0.86	80.36	9.70	19.75	0.25	2.32	310
12	132	1.11	81.49	27.77	5.94	10.55	3.72	0.84	76.51	8.97	45.21	0.48	3.82	82
13	136	1.12	80.91	27.58	6.12	10.98	3.16	0.82	78.60	8.98	40.27	0.39	3.16	49
14	137	1.12	82.78	28.62	6.03	9.50	4.25	0.87	78.27	9.32	33.26	0.43	3.57	176
15	138	1.13	82.45	28.07	6.17	9.55	4.10	0.86	81.15	9.18	8.93	0.12	1.22	306
16	142	1.14	82.66	28.56	5.98	9.45	3.98	0.86	80.64	9.54	18.51	0.26	2.32	279
17	149	1.17	84.04	29.48	6.80	8.78	4.45	0.89	81.46	9.89	12.76	0.19	1.95	62
18	151	1.17	83.86	29.57	6.13	8.68	4.16	0.87	81.48	9.48	11.07	0.15	1.49	205
19	162	1.19	84.83	31.70	6.15	8.46	4.17	0.88	80.80	9.54	19.03	0.26	2.39	109

Table 2: CV for the groups of cotton bales

Group No.	Sci	Len	Unf	Str	Elg	Sfi	Mic	Mat	Rd	+b	Trc	Tra	Trg
1	3.67	2.76	1.50	5.69	8.65	10.41	10.29	3.04	3.44	12.29	31.32	50.20	45.65
2	16.45	5.13	2.77	9.87	12.44	18.52	12.86	3.80	5.59	16.04	43.53	83.85	56.41
3	5.45	3.19	1.50	6.24	8.49	10.13	12.30	3.16	2.69	11.70	34.37	59.40	54.44
4	5.48	3.13	1.50	6.28	8.50	10.14	12.37	3.18	2.69	11.77	34.59	56.50	53.09
5	3.58	3.31	1.57	5.81	8.24	10.15	11.85	3.21	3.47	12.13	13.21	43.42	27.61
6	2.89	3.43	1.60	5.61	9.07	9.67	12.98	3.35	3.39	11.53	14.56	47.19	42.33
7	12.50	4.25	2.55	10.52	6.17	17.47	13.88	4.19	5.41	9.56	41.41	30.44	19.79
8	2.98	3.16	1.54	4.64	8.25	9.28	13.04	3.38	3.79	12.04	7.00	41.78	24.20
9	2.88	3.03	1.49	5.44	8.11	10.75	11.94	3.19	3.81	11.53	9.63	36.36	24.59
10	2.73	3.53	1.50	5.95	9.66	10.27	11.83	3.19	2.81	11.49	26.20	57.21	53.38
11	2.90	3.20	1.58	5.36	9.70	10.71	13.04	3.25	3.09	10.92	17.43	65.09	46.51
12	5.10	3.86	1.78	6.56	8.64	9.74	12.34	2.88	3.88	10.23	15.06	29.52	21.70
13	8.59	3.09	1.77	5.41	6.39	9.51	10.33	1.92	3.65	10.10	46.79	58.70	40.06
14	5.00	3.60	1.43	6.12	9.19	9.55	13.02	3.26	3.95	12.10	16.48	49.27	28.22
15	3.02	3.21	1.31	5.52	9.32	8.78	9.81	2.59	3.07	11.58	35.15	73.38	50.13
16	2.78	3.41	1.67	6.15	9.11	9.59	12.92	3.28	2.85	12.21	16.59	68.71	45.65
17	12.18	6.52	2.42	11.91	15.33	15.45	15.16	4.46	2.87	16.43	60.64	100.92	64.70
18	2.35	2.94	1.41	6.63	9.41	8.87	10.28	2.84	2.93	11.57	43.48	60.55	51.94
19	4.57	3.68	1.77	8.01	8.34	8.82	11.33	2.94	3.12	11.58	37.92	57.64	47.34
MG	11.71	4.78	2.36	8.55	9.73	14.70	13.14	3.39	3.96	12.31	58.70	74.12	55.48

The Re-combined Subgroup (RSG)

The visualization and the hits for the re-combined sub group (RSG), made of the members of 2nd, 7th and 17th groups is given in figure 4, which shows that the level of grouping is very low with

only five out of the 66 nodes having hits of more than 5, and about 44% of all the nodes have one (1) or zero hits.

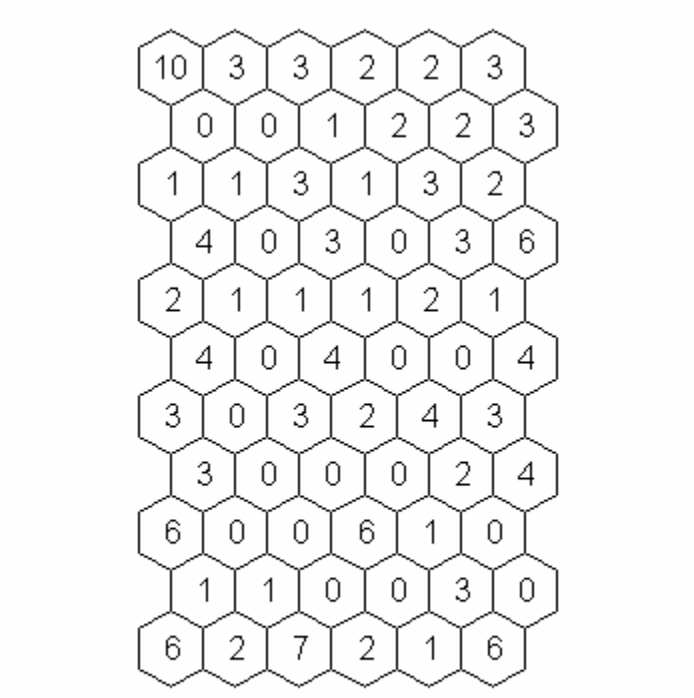

Fig. 4: Visualization for thr re-combined 2, 7 and 17 groups

The silhouette means for the RSG is given in figure 5. The optimum clustering is 3, which was the initial number of groups in RSG before re-combining. Checking out the principal components of the attributes given in Table 3, shows that 97. 87% of the variance is accounted for by the first two principal components.

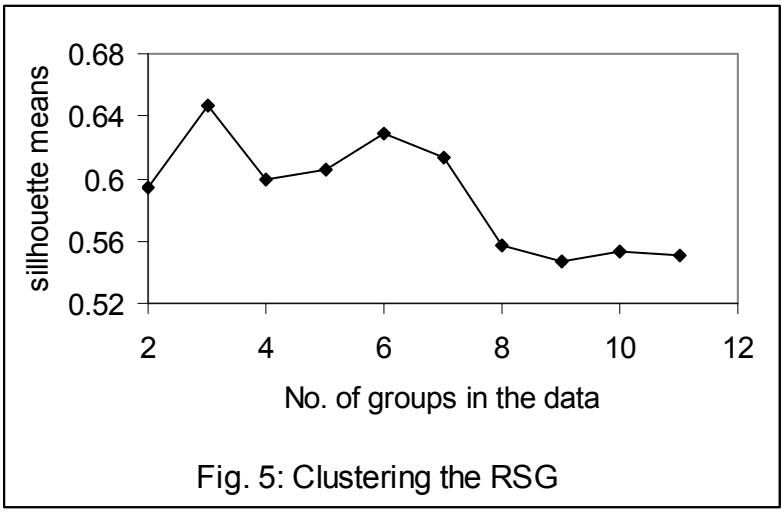

Fig. 5: Clustering the RSG

Table 3: The Principal Components of RSG

Principal Component	Cumulative sum of variance
1	70.3390
2	97.8709
3	98.7609
4	99.2784
5	99.5244
6	99.7281
7	99.8697
8	99.9521
9	99.9967
10	99.9995
11	100.0000
12	100.0000
13	100.0000

A scatter plot of the scores of the first two principal components (figure 6) shows that there are three distinct regions. These regions marked A, B and C are widely scattered but region B and C seem to be more scattered than A. B and C can be subdivided into two groups each. Thus RSG will now have a total of five subgroups.

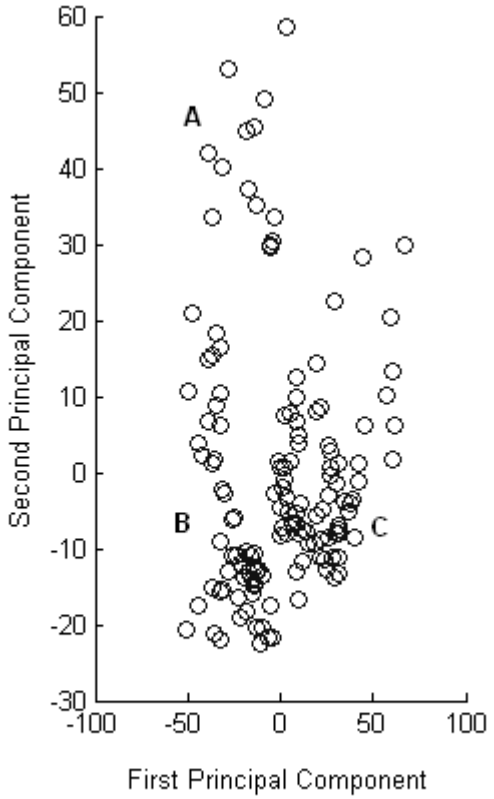

Fig. 6: Principal Component scatter plot for RSG

The means and CVs of the five groups of RSG are given in tables 4 and 5. These subgroups still do not show a sense of being more compact than the MG, since all the groups still have attributes with CVs which are higher than the original MG. Further subdivision of RSG will produces very small subgroups having less than ten members, so it can be concluded that RSG is actually a collection of outliers which can be subdivided into five sets as shown in table 4.

Table 4: Means for the 5 groups in RSG

	GROUP					
Attribute	RSG 1	RSG 2	RSG 3	RSG 4	RSG 5	MG
Length	1.042	1.118	1.109	1.212	1.047	1.115
Uniformity	78.96	82.65	82.19	85.23	78.46	81.88
Strength	24.33	27.82	28.63	31.22	24.62	27.70
Elongation	6.03	6.41	6.03	6.88	5.78	6.06
Sfi	12.88	9.69	10.18	7.99	12.85	10.12
Micronaire	4.27	4.42	4.53	4.43	4.45	4.16
Maturity	0.86	0.88	0.88	0.89	0.86	0.86
Rd	78.78	81.43	73.66	82.26	73.36	79.70
+b	10.52	9.52	8.96	10.49	9.23	9.43
Trash cent	17.60	17.88	66.77	11.38	46.12	20.70
Trash area	0.35	0.23	0.67	0.20	0.53	0.26
Trash grade	2.88	2.29	4.77	2.03	3.94	2.32
Sci	103	135	128	163	96	132
No. of data	40	42	11	32	17	2421

Table 5: CVs for the 5 groups in RSG

	GROUP					
Attribute	RSG 1	RSG 2	RSG 3	RSG 4	RSG 5	MG
Length	4.973	4.000	3.965	5.773	3.430	4.782
Uniformity	2.195	1.569	1.997	2.114	1.929	2.357
Strength	7.710	6.745	5.017	12.376	7.646	8.554
Elongation	13.591	17.887	5.012	12.326	5.834	9.728
Sfi	16.253	11.800	13.595	11.478	16.260	14.701
Micronaire	12.906	18.229	8.412	9.583	15.093	13.143
Maturity	3.888	4.844	2.453	3.491	4.422	3.390
Rd	4.180	3.451	4.584	2.264	5.303	3.962
+b	17.278	12.704	9.995	16.453	11.891	12.310
Trash cent	30.301	46.072	14.700	75.553	19.563	58.704
Trash area	101.123	60.234	15.450	123.681	41.707	74.116
Trash grade	68.773	49.870	12.070	72.275	29.475	55.484
Sci	9.279	5.367	8.387	8.641	5.483	11.709

Conclusion

A cotton bale classification model has been proposed and used to classify 2421 cotton bales. The model reduced the 2421x13 HVI high dimensional data into 18x13 grids, with a quantization error of 1.879 and a topographic error of 0.083, and initially identified 19 groups in the data. Three of the groups containing a total of 144 data failed the compactness test. The final classification of the 2421 cotton bales contains 16 groups of cotton bales having a total of 2277 bales and five sets of 144 bales containing outliers.

Acknowledgement

The financial assistance received from University Research Fund (URF), Moi University, Eldoret and China Scholarship council (CSC), Beijing is hereby acknowledged.

Disclaimer

Mention of a trademark, warranty, proprietary product or vendor does not constitute a guarantee by the authors and does not imply approval or recommendation of the product to the exclusion of others that may be suitable.

References

Allan G. Bluman, "Elementary Statistics", McGraw Hill, New York, 5th Edition, chapter 3 (2004)

Abhijit Majumdar, Prabal Kumar Majumdar and Bijon Sarkar, "Selecting Cotton Bales By Spinning Consistency Index And Micronaire Using Artificial Neural Networks", Autex Research Journal, vol 4, No1 www.autexrj.org/No1-2004/0092.pdf (2004)

Anil k. Jain and Richard C. Dubes, "Algorithms for Clustering Data", Prentice-Hall Englewood Cliffs, 92- 97 (1998)

Dan J. McCreight, Ralph W. Feil, James H. Booterbaugh and Everwtt E. Backe, "Short Staple Yan Manufacturing", Carolina Academic press, Durham, North Carolina, 439-456 (1977)

Eric Hequet and Dean Ethridge, "Effect of Cotton Fiber Length Distribution on Yarn Quality", Proc. of the Beltwide Cotton Conference, vol 2, 1507-1514 (1999)

Eric W. Weisstein, K-Means Clustering Algorithm (From MathWorld- A Wolfram Web Resource. http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html) (2006)

Fredric M. Ham and Ivica Kostanic, "Principles of Neurocomputing for Science and Engineering", China Machine Press, Beijing 169-172(2001)

Hossein M. Ghorashi, "New Generation of High Volume Instruments", Proc. of the 58th International Cotton Advisory Committee (Charleston), 24-27 (1999)

Howard Demuth and Mark Beale, "Neural Network Toolbox User Guide", 9th printing, The MathWorks, Natick chapters 7 and 9 (2005)

Howard Demuth, Mark Beale and Martin Hagan, "Neural Network Toolbox User Guide", version 4 (Natick) chapters 5, 7 and 8 (2005)

J.S.R Jang and E. Mizutani, "Unsupervised Learning and Other Neural Networks, Neuro-Fuzzy and Soft Computing", Prentice Hall, NJ, 301-310 (1997)

Juha Vesanto and Esa Alhoniemi, "Clustering of the Self-Organizing Map", IEEE Transactions on Neural Networks, vol 11 (No. 3), 586-600 (2000)

Juha Vesanto, Johan Himberg, Esa Alhoniemi and Juha ParhanKangas, "SOM Toolbox for Matlab 5", Helsinki University of Technology www.cis.hut.fi/projects/somtoolbox (2000)

Martin T. Hagan, Howard B. Demuth and Mark Beale, "Neural Network Design", China Machine Press, Beijing, chapter 14 (2002)

MV Vivekanandan and SS Doke, "Pattern recognition in cotton data using Kohonen neural networks", The Indian Textile Journal, vol 113 (No.1), 21-25 (2003)

Neil A. Weiss, "Introductory Statistics" Addison-Wesley Publishing Company, Inc., New York, 4th Edition, pp. 41-178(1997)

Stephen Chen, "Matching Cotton Fiber Properties with Textile Properties", Proc. of the 58th International Cotton Advisory Committee (Charleston), 5-7 (1999)

Sueli A. Mingoti and Joab O. Lima, "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms", European Journal of Operational Research www.elsevier.com/locate/ejor (Article in Press) (2005)

Tadeusz Jackowski and I. Frydrych, "What I learnt in Fiber Quality: Practical Experience", Proc. of the 58th International Cotton Advisory Committee (Charleston), 7-19 (1999)

Teuvo Kohonen, "Self-Organizing Maps", 2nd Edition, Springer-Verlag Berlin/Heidelberg, 85-144 (1997)

The MathWorks, "Matlab 7.1 R 14 SP3", Natick, USA (2005)

The MathWorks, "Silhouettes", Statistical toolbox 5.3 http://www.mathworks.com 2006)

V.K. Kothari Progress, "Textiles: Science & Technology Vol: 1 Testing & Quality Management" IAFL Publications, New Delhi, 53-91 (1999)