Indian Journal of Fibre & Textile Research
Vol. 47, September 2022, pp. 356-361
DOI: 10.56042/ijftr.v47i3.53130

Classification of ring-spun yarns using cluster analysis

Elham Naghashzargar[1,a] & Haleh sadat Nekoee Zahraei[2]
[1]Department of Textile Engineering, Faculty of Engineering, University of Bonab, Bonab 5551395133, Iran
[2]Biostatistics Unit, Department of Public Health, University of Liège, Belgium

Received 2 August 2021; accepted 5 January 2022

The aim of this study is to classify ring-spun yarns according to their unevenness, imperfections, and hairiness parameters using cluster analysis. The mentioned features of ring-spun yarns are measured for five different ranks. Five ranks of ring-spun yarns including compact and conventional as well as combed and carded types are chosen and produced. In the modeling section, the model-based clustering method was applied as a strong method based on the distribution of each variable. In order to select the best fit and to find out the final clustering, bayesian information criterion (BIC) is applied. According to the results of modeling, five ranks of selected ring-spun yarns are classified in four clusters and the acceptable agreement is measured according to Cohen's kappa method. The highest value for Kappa represents a high agreement to match between the clustering result and the real rank.

Keywords: Clustering validation, Model-based clustering, Ring-spun yarn, Statistical analysis

1 Introduction

Quality control of ring-spun yarns is one of the important aspects for spinners and for further processors in textiles. The classification of ring-spun yams is an important need in assessing textile quality for commercial purposes, since it provides a useful means of expressing yarn standards in the market. Traditionally, it is done by a comparison of measurements of various quality-related parameters with their values as recommended in a standard. With the rapid development of computer facilities, more and numerical classification methods for the quantitative estimate of subjective properties have been used in various fields[1].

So, the classification of ring-spun yarn according to its quality aspects, which is the main purpose of this study, can be an important attribute in establishing quality control of yarns. The unevenness (CVm), imperfections and hairiness parameters have become some measurable yarn parameters[2]. Nikolić *et al.*[3] studied the influence of differences in compacting systems on yarn quality, and to compare the compact and conventional yarns produced. The quality parameters of cotton/polyester rotor yarn blends with various contents of the components and different length distributions of the cotton fibres were analyzed by Cyniak *et al.*[4]. As a result, the quality parameters changed depending on the percentage of cotton fibre in the blend. Also, Canoglu & Yukseloglu[5] investigated the hairiness of ring-spun polyester/viscose blends, which are commonly used in the textile industry. The blend ratios and fibre locations were scanned on the scanning electron microscope. The observations were concluded both on the hairiness and pilling values, depending on the blend proportions.

One of the most challenging and difficult issues in machine learning, especially due to its unsupervised nature, is clustering. The unsupervised nature of the subject implies that its structural characteristics are unknown. The main goal of clustering is that the similar objects are grouped into one cluster, while dissimilar objects are grouped into different clusters. Clustering involves identifying similar characteristics in data sets that are always hidden in nature; dividing them into groups referred to as clusters[6].

When the data is unlabelled, a clustering algorithm is used. This is because unsupervised learning methods have to be used in order to analyze the data. Similarity indices are calculated between individual data points and the data points are segmented into groups. The number of clusters is a direct input parameter in most clustering algorithms. There are a lot of aspects to consider when choosing a suitable clustering algorithm for a particular dataset[7]. So, the most important problems in clustering are choosing

[a]Corresponding author.
E-mail: e.naghashzargar@ubonab.ac.ir

the number of components. As part of model-based clustering, one of the commonly used methods for determining the number of clusters is BIC (bayesian information criterion), which is used in this study[8,9]. BIC is a likelihood criterion that penalizes the number of parameters in a model. BIC can be used to select a model by maximizing the posterior probability of various candidate models[10].

The concept of clustering has become increasingly popular among diverse fields with heterogeneity. Recently, the classification of textile products using statistically clustering methods is more attractively applied by investigators in this field, which is often considered helpful for data exploration. A discriminant function based on the properties of polyester, lyocell/viscose, and treated polyester was oblained by Kiruthika and Chandrasekaran[11] in order to classify 105 samples of textile fabrics. K-means clustering was applied to all three categories of fabrics that resulted in three clusters. These clusters were subjected to discriminant analysis that again yields a 100% correct classification, indicating that the clusters are well-separated. Hamdi *et al.*[12] introduced an algorithm for automatic and unsupervised detection of defects in patterned fabrics. A standard deviation filter and K-means clustering algorithm were used to detect defective blocks in pattern fabric images. A thorough analysis of the k-means algorithm using printed fabric patterns as input data was presented by Pan *et al.*[13]. Analysis of experimental results demonstrated that the complexity in clustering has reduced to its minimal with good application specific and good time yield. In addition, a textile grading method that comprises image processing and machine learning classification was proposed as a replacement for visual textile grading[14].

In view of above, in this investigation, an application of the clustering method especially Model-based clustering using Bayesian information criterion technique has been studied to classify ring-spun yarns, which can be used in yarn quality control. The early features of textile are measured for five different ranks. The features that we consider in this study are CVm (%), thin (-50%), thick (+50%), nep (+200%), IPI and hairiness. Even though, we know the clustering group in this study, we attempt to determine the new grouping among all samples using heterogeneous features. Cluster analysis is one of the popular statistical methods that classified samples into several subgroups. Using cluster analysis, we investigate whether samples could be divided into separate clusters such that samples in each cluster present similar differences between two groups and have different differences in separate clusters. Also, the second aim of this study is to observe the agreement of the real rank of samples and the new grouping according to cluster analysis.

2 Materials and Methods

2.1 Raw Materials and Sample Preparation

Cotton and polyester were selected as the yarn components in this investigation. The properties of polyester fibre used in this investigation are summarized in Table 1. In addition, some of the HVI results of the cotton used are also shown in table.

Yarns were produced from 100% cotton and polyester/cotton with a blend ratio 65/35. Five different ring-spun yarns with count of 30 Ne and 40 Ne were spun. The unevenness (CVm), thin (-50%), thick (+50%), nep (+200%), IPI and hairiness parameters of different types of yarn in conventional and compact system were analyzed and compared. Also, to achieve better classification of different types of yarn, combed and carded yarns were chosen. All samples were produced with the same raw materials as mentioned in Table 1. During yarn spinning, the same roving was fed in the same position to eliminate any variation between samples. Finally, 5 ranks of ring-spun yarns, each rank 30 samples, were prepared. So, the total number of prepared samples was 150. Different types of selected yarns with their related features are summarized in Table 2.

2.2 Uster Tests

Uster testing of yarn to measure the irregularity before knitting and weaving may reduce the faulty fabric production, faulty dyeing which may minimize the production hassle in different stages as well as customer rejection of end product. Unevenness, thin (-50%), thick (+50%), nep (+200%), IPI and hairiness of spun yarn are the major indicator of yarn quality measurement for the textile technologist in ring spinning industries, and such types of yarn qualities can be confirmed by uster tester[15]. The unevenness (CVm), imperfections and hairiness parameters of the

Table 1 — Properties of cotton and polyester fibres

Fibre property	Polyester	Cotton
Linear density, dtex	1.66	1.7
Staple length, mm	38	31.2
Tenacity, cN/tex	43.24	32.89
Fibre elongation, %	25.61	7.1

Table 2 — Yarn ranks with related features

Yarn rank	Materials	Type	Yarn count, Ne	Twist level, turns/m
Rank 1	Polyester/Cotton (65/35%)	Combed/ conventional	40	850
Rank 2	Cotton (100%)	Combed /compact	30	780
Rank 3	Cotton (100%)	Combed /compact	40	890
Rank 4	Cotton (100%)	Combed/ conventional	30	790
Rank 5	Cotton (100%)	Carded/ conventional	30	730

Table 3 — Physical properties of yarns measured by uster tester

Yarn rank	Average values of physical properties					
	Unevenness CVm, %	Thin (-50%)	Thick (+50%)	Nep (+200%)	IPI	Hairiness (H)
Rank 1	13.04(0.31)	5.3 (6.03)	20.9(5.38)	17.8(3.49)	43.9(11.94)	4.31 (0.27)
Rank 2	11.13 (0.52)	0.3(0.78)	13.9 (10.45)	30.3 (16.29)	44.3(14.99)	3.85 (0.42)
Rank 3	11.49 (0.70)	0.07(0.36)	24.3 (15.01)	44.1 (16.53)	68.4 (30.18)	4.24 (0.57)
Rank 4	10.87 (0.23)	0.03 (0.18)	10.3 (2.99)	18.9 (4.41)	29.3 (8.32)	4.75 (0.62)
Rank 5	14.07 (0.43)	4.4 (3.41)	91 (36.44)	134 (63.27)	229.5 (96.17)	6.03 (0.67)

Values in parantheses indicate standard deviation.

yarns were measured by an Uster Tester 3. The mass or weight variation per unit length of yarn is defined as unevenness or irregularity. The coefficient of variation of the mass (CVm%) is the calculation method to determine the variation of mass for yarns. Hairiness can be identified as fibres protruding from the main body of the yarn. Also, an Uster Tester 3 measured hairiness and calculated as the total length in centimeters of all hairs within one centimeter of yarn. The tests were performed at 400 m/min speed for 1000 m yarns. Mainly two replications were carried out for the tests and each replication consisted of seven separate measurements of different cops. The mean of the overall replications of physical properties of the yarns are shown in Table 3. All samples were kept in standard testing condition for 24 h before testing. The yarn tests were carried out at 20 ± 2 °C and 65% ± 2 RH conditions after the samples reaching equilibrium.

2.3 Modeling Framework

All of the variables collected for samples, such as CVm%, thin (-50%), thick (+50%), nep (+200%), IPI and hairiness were considered as potential variables for cluster analysis to discrimination of samples into subgroups. The traditional method for cluster analysis, K-means, and hierarchical clustering are heuristic and randomly initialized without certain indices for determination number of clusters which have to apply different distance measures for each variable in the dataset. Therefore, in this study, model-based clustering was applied as a strong method based on distributions of each variable that is the number of clusters as a mixture of distribution[16].

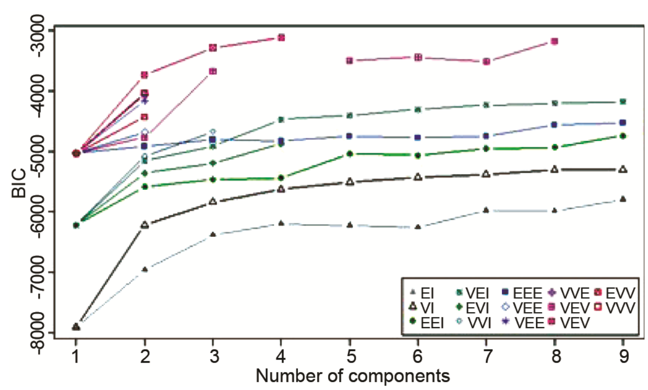

Fig. 1 — BIC values for fitted models with different clusters (criterion was maximized by the VEV model with four clusters)

In model based clustering, multivariate continuous variables follow finite mixture model of multivariate Gaussian distribution, using following relationship:

$$y_i \sim \sum_{k=1}^{K} \eta_k \phi(y_i|\mu_k, \Sigma_k) \quad \ldots (1)$$

where $\phi(y_i|\mu_k, \Sigma_k)$ is the probability distribution function of the multivariate Gaussian distribution with mean μ_k and variance-covariance matrix Σ_k; and K defines the cluster size with $\mu_k > 1$ and $\sum_{k=1}^{K} \eta_k = 1$. The covariance matrix describes the volume, shape, and orientation of the clusters. Therefore, the decomposition of the variance-covariance matrix of the kth clusters is given as:

$$\Sigma_k = \lambda_k \mathrm{D}_k \mathrm{A}_k D_k^t \quad \ldots (2)$$

where λ_k is the volume; D_k corresponds to orientation and A_k is for shape. The various combinations of these three parameters are listed in Fig. 1. These

Table 4 — Descriptive statistics for clustering results

Clusters	Variables					
	Unevenness CVm, %	Thin (-50%)	Thick (+50%)	Nep (+200%)	IPI	Hairiness (H)
Cluster 1	12.93	3.04	19.42	17.42	39.87	4.31
(n=24)	(0.21)	(0.91)	(4.32)	(3.77)	(6.50)	(0.29)
Cluster 2	10.91	0.05	10.88	23.24	34.17	4.19
(n=58)	(0.29)	(0.22)	(2.81)	(6.28)	(8.38)	(0.77)
Cluster 3	11.54	0	23.85	46.15	69.88	4.36
(n=26)	(0.37)	(0)	(7.25)	(13.42)	(19.34)	(0.52)
Cluster 4	13.68	5.36	73.67	104.55	180.36	5.59
(n=42)	(0.93)	(5.83)	(37.33)	(67.76)	(103.41)	(0.94)

Values in parantheses indicate standard deviation.

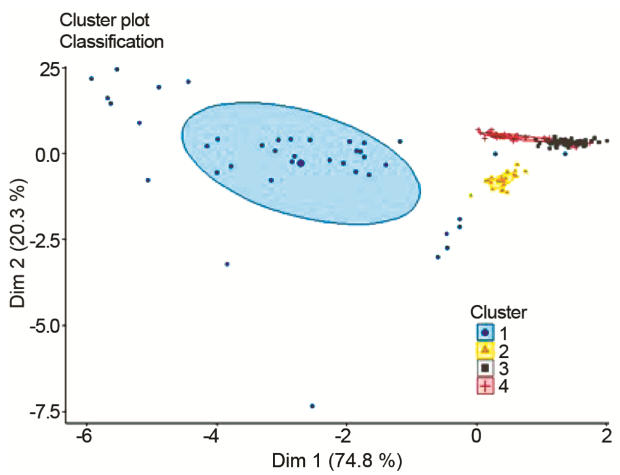

Fig. 2 — Clustering results using two first components

various parameters allow fitting model on different structure on the dataset , thus finding the best fitting for classification. In this study, Bayesian information criterion was applied to select the best fitting and to find final clustering. ClustMD package was applied for clustering the dataset using EM algorithm by free statistical software R studio. Then, based on this reality that the label of clustering is assigned to the samples of the dataset using model-based clustering, the agreement between the results of clustering and the real rank was measured using Cohen's kappa with a corresponding 95% confidence interval. The highest value for Kappa represents a high agreement to match between the clustering result and real rank.

3 Results and Discussion

3.1 Cluster Analysis

For model based clustering, all existed methods for mixture in ClustMD package are fitted to the dataset for different options of groups (1 to 9 groups). BIC is evaluated for selecting best number of clusters and how to classification patients (Fig. 1). BIC is maximum by model with identity covariance matrix (VEV) and for 4 clusters.

Based on model based clustering and BIC criteria, samples are divided into four clusters, viz 24 samples in cluster 1, 58 samples in cluster 2, 26 samples in cluster 3 and 42 samples in cluster 4. In Table 4, descriptive statistics for clustering results are mentioned. According to the results in Table 4, all variables are the highest for cluster 4.

In Fig. 2, clustering results are displayed as four distinct groups. First, principal component analysis is performed on five variables, and the new components are calculated as a linear combination of them. According to Fig. 2, the first component represents 74.8% of the variation of five original variables, and the second component illustrates 20.3% of it. As a result, both first components are capable of evaluating 95.1% of the total variation.

Despite the first two principal components, this reduces the number of variables and converted the original variables into two components with 95.1% variation. Figure 3 shows the classification of ring-spun yarns for all two combinations of variables. In addition, the uncertainty of clustering results for all two combinations of variables is also presented in Fig. 4. In the second cluster, there is more uncertainty. Cluster 2 is a combination of two major yarn groups. According to Figs 3 and 4, the yarns in this cluster are more heterogeneous and there is more uncertainty.

Table 5 shows that 24 samples of the first rank are in cluster 1 and others put in cluster 4, which is equal to 6. For the second rank, 28 samples are in cluster 2 and others are in cluster 4. From 30 samples of the third rank, 26 samples are in cluster 3, 3 samples are in cluster 2 and only one sample is in cluster 4. Twenty seven samples of the fourth rank are in cluster 2 and

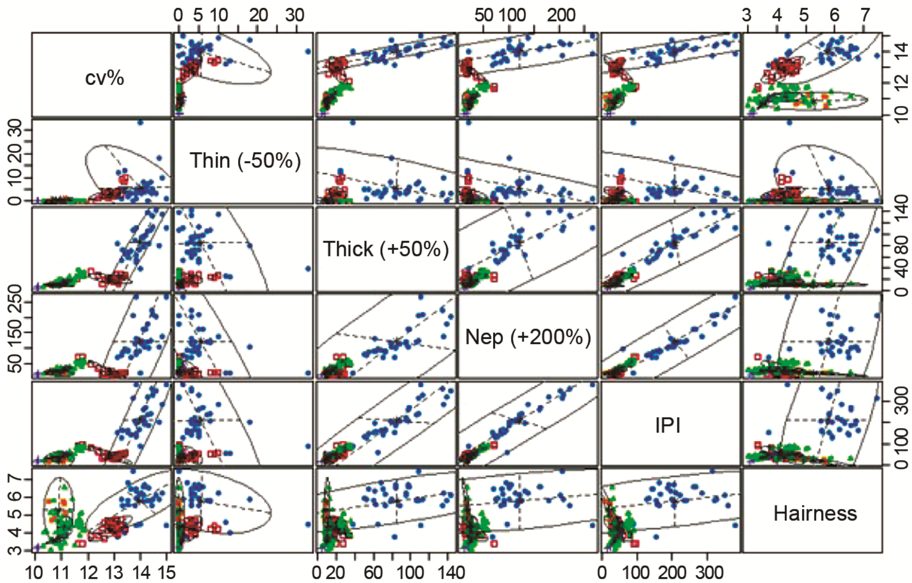

Fig. 3 — Results of clustering for classification in five parameters

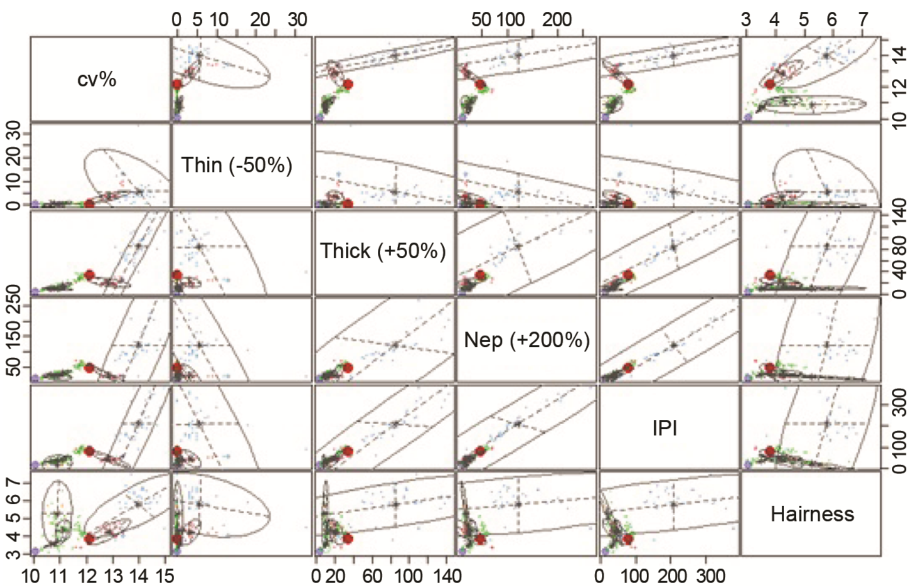

Fig. 4 — Results of clustering for uncertainty in five parameters

others are put in cluster 4. For the final rank, all samples are in cluster 4.

Therefore, we conclude that rank 1 which is related to the polyester/cotton (65/35%) yarn with 40 Ne and combed type; rank 3 which is related to the cotton (100%) yarn with 40 Ne and combed /compact type, and rank 5 which is related to the cotton (100%) yarn with 30 Ne and carded type have different characteristics, and hence can be classified from each other. However, for rank 2 and rank 4, there is a similar feature, and hence model-based clustering doesn't recognize the differences between these two ranks and find the same cluster for them. The Cohen's kappa for this result is 0.59 (95% CI: 0.38-0.80). As we mentioned, the highest value for Kappa represents a high agreement to match between the clustering result

Table 5 — Cross table for real rank and clustering results

Clusters	Real ranks				
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Cluster 1	24	0	0	0	0
Cluster 2	0	28	3	27	0
Cluster 3	0	0	26	0	0
Cluster 4	6	2	1	3	30

Table 6 — Clustering validation

Parameters	Indices	Value
Internal measures	Silhouette width	0.425
	Dunn index	0.256
Stability measures	Average proportion of non-overlap (APN)	0.017
	Average distance between means (ADM)	0.004

and real rank. Both rank 2 and rank 4 are related to the 30 Ne ring-spun yarns with approximately similar twist level according to the values in Table 2. The only difference with yarns in ranks 2 and 4 is the type of yarns which is compact for rank 2 and combed for rank 4. So, the result of uster tester is approximately similar and in modeling these two ranks of yarns classify in the same cluster.

3.2 Validation Test of Model

An important part of this clustering method is evaluating the framework used to classify clusters. We first evaluated the performance of clustering results by the model-based method using BIC. This result shows a BIC of -3055.98, which is the lowest value for fitting the model. In the next step, the clustering result is evaluated by two indices for internal clustering validation including silhouette width and dunn index as well as two indices for clustering stability validation including average proportion of non-overlap (APN) and average distance between means (ADM). Clustering validation is shown in Table 6.

Silhouette width indicates the average level of confidence in the clustering. Silhouette width is in the range of [-1, 1] and values near 1 show good clustering. Dunn Index is another index that divides the smallest distance between two observations in the same cluster by the largest distance within the cluster. According to Dunn, the index is between [0, ∞] and is preferable to maximum values. Average proportion of non-overlap (APN) partitions the observations into different clusters following clustering in full data. The APN is in the interval [0, 1] with minimum values preferred. Average distance between means (ADM) measures the distance between observations in the same cluster and clustering in full data. The ADM is in the interval [0, ∞] and smaller values show well clustering[17].

4 Conclusion

The categorization of ring-spun yarn is very essential in quality control of yarns in spinning process and marketing values. In this paper, a method using statistically clustering analysis is presented for the purpose of classifying ring-spun yarns. The unevenness (CVm), imperfections and hairiness parameters of yarns are measured using uster instrument. Statistical methods are used to classify yarns based on measurements. Therefore, in this study, model-based clustering has been used as a strong method based on distributions of each variable that is the number of clusters as a mixture of distribution. According to the results of classification, five ranks of different types of ring-spun yarns are classified in four clusters. Because of the similar features of two ranks, model-based clustering does not recognize the differences between them and find the same cluster. According to the Cohen's kappa value, there is a high agreement to match between the clustering result and real rank.

References

1 Rong GH, Slater K & Fei RC, *J Text Inst*, 85 (1992) 389.
2 Rwawiire S, Kasedde A, Nibikora I & Wandera G, *JTATM,* 8 (2014) 1.
3 Nikolić M, Stjepanovič Z, Lesjak F & Štritof A, *Fibres Text East Eur,* 11 (2003) 30.
4 Cyniak D, Czekalski J, Jackowski T & Popin L, *Fibres Text East Eur,* 14 (2006) 33.
5 Canoglu S & Yukseloglu SM, *Fibres Text East Eur,* 16 (2008) 34.
6 Teklehaymanot FK, Muma M & Zoubir AM, *IEEE Trans. Signal Processing,* 66 (2018) 5392.
7 Dahiya S, Nanda H, Artwani J & Varshney J, *Int J Adv Trends Comput Sci Eng Inf technol,* 9 (2020) 2138.
8 Akogul S & Erisoglu M, *Math Comput Appl,* 21 (2016) 34.
9 Banerjee A & Shan H, *Encyclopedia of Machine Learning*, (Springer, Boston, MA), 2011.
10 Wu H, Cheung S-F & Leung S-O, *Multivariate Behav Res*, 55 (2020) 1.
11 Kiruthika C & Chandrasekaran R, *J Appl Statistics,* 39 (2012) 1129.
12 Hamdi AA, Sayed MS, Fouad MM & Hadhoud MM, *International Conference on Innovative Trends in Computer Engineering* (ITCE), Aswan University, Egypt, 19-21 February 2018.
13 Pan R, Kumah C & Zhang N, *Curr Trends Fashion Technol Textile Eng*, 1 (2017) 79.
14 Huang ML & Fu CC, *Fibers,* 6 (2018) 73.
15 Hossain A & Samanta AK, *J Text Eng Fashion Technol,* 5 (2019) 213.
16 Fraley C & Raftery AE, *J Am Stat Assoc,* 97 (2002) 611.
17 Brock G, Pihur V, Datta S & Datta S, *J Stat Softw*, 25 (2008) 1.