Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 4788 – 4804

doi:10.3906/elk-1611-16

Discovering the relationships between yarn and fabric properties using association rule mining

Pelin YILDIRIM[1], **Derya BİRANT**[2,*], **Tuba ALPYILDIZ**[3]
[1]Graduate School of Natural and Applied Sciences, Dokuz Eylül University, İzmir, Turkey
[2]Department of Computer Engineering, Faculty of Engineering, Dokuz Eylül University, İzmir, Turkey
[3]Department of Textile Engineering, Faculty of Engineering, Dokuz Eylül University, İzmir, Turkey

Received: 02.11.2016 • **Accepted/Published Online:** 15.09.2017 • **Final Version:** 03.12.2017

Abstract: Investigation of the effects of yarn parameters on fabric quality and finding important parameters to achieve desired fabric properties are important issues for the design process with the aim to meet the needs of the textile industry and the consumer for complex and specific requirements of functionality. Despite many statistical and mathematical studies that predict and reveal specific properties of utilized yarn and fabric materials, a number of challenges continue to exist when evaluated in many perspectives, such as discovering complex relationships among material properties in data. Data mining plays an important role in discovering hidden patterns from fabric data and transforming it into knowledge. Therefore, the aim of the study is to uncover relationships between yarn parameters and fabric properties using an extended FP-Growth algorithm in association rule mining. This study extracts different types of frequent itemsets (closed, maximal, top-k, top-k closed, top-k maximal) that have not been determined in textile sector before. This article also proposes two novel concepts, closed frequent item and maximal frequent item, to identify significant items in data. In the experimental studies, the algorithm was executed on a real-world textile dataset with different support threshold values to compare the different types of patterns. Experimental results show that proposed approach is very useful for discovering rules related to yarn and fabric properties.

Key words: Data mining, association rule mining, FP-Growth algorithm, closed and maximal frequent itemsets, yarn and fabric properties

1. Introduction

Recent technological innovations and developments in the textile industry bring out a need to process textile data to discover hidden and valuable knowledge from it. Textile studies using classical mathematical and statistical models for analyzing textile materials have been presented in previous literature. However, sometimes these methods are inadequate to identify complex and nonlinear relationships in textile datasets and cannot predict unknown values for a new instance. This challenge leads to a need for data mining techniques.

Data mining is the process of discovering previously unknown and potentially useful information from data. Data mining has a wide range of application areas, such as marketing, education, healthcare, and finance, and it has started to be used in the textile sector [1,2] in recent years, as an interdisciplinary approach.

Techniques used in data mining can be grouped under three tasks: classification, clustering, and association rule mining (ARM). ARM finds frequent patterns and interesting relationships among set of items

*Correspondence: derya@cs.deu.edu.tr

in dataset. The most commonly used ARM algorithms are Apriori, FP-Growth, and Eclat. FP-Growth is a scalable algorithm that discovers large volumes of frequent itemsets efficiently using an extended prefix-tree structure (FP-tree). While there are a number of classification and clustering studies in textile domain, ARM has only been used in several textile studies [3–7] so far. The present study is the first study where an ARM algorithm (FP-Growth) is implemented on a real-world textile dataset to discover relationships among yarn parameters and fabric properties.

Textile datasets can be of any end-product in the textile industry, such as a fiber, yarn, fabric, or garment. Knowing what is expected from a raw material is important to both the supplier of raw material and the purchaser. A cotton grower and fiber manufacturer would like to know what sort of yarn quality can be produced from their crop so that they can ask the right price for the fiber. The buyer, a yarn manufacturer, would be interested in knowing whether the desired yarn properties can be obtained from a particular variety of cotton it intends to buy. The user of the yarn, the fabric manufacturer, will be interested in knowing the performance of the yarn from its physical and mechanical properties. Thus, one of the major concerns in the fabric-manufacturing process is to determine settings of design parameters that result in a satisfactory combination of quality characteristics. The fabric structure and properties are primarily influenced by fiber properties (length, fineness, etc.), spinning methods (ring, rotor, air jet, etc.), yarn parameters (count, twist, single and doubled, etc.), fabric structural parameters (warp and weft density, weave, etc.), and finally finishing treatments. The relationship between fabric structure and properties is complex and inherently nonlinear.

The novelty and main contributions of this paper are as follows. First, it is the first study that proposes using ARM to identify the complex relationships between significant yarn parameters (i.e. hairiness, irregularity, diameter) and their effects upon fabric quality (i.e. pilling, abrasion). Second, it implements the FP-Growth algorithm in the textile domain for the first time. Third, it presents an extended FP-Growth algorithm that has the ability to find the different types of patterns such as closed, maximal, and top-k frequent itemsets. Fourth, this study also proposes two novel concepts, cf-item and mf-item, to identify significant items in data. In contrast to previous studies, the present study focuses on single-item based analysis because it can be used to solve different types of problems in different areas such as feature selection, the determination of important parameter values, and discretization.

An extended FP-Growth algorithm, which is proposed in this study, was executed on a real-world textile dataset with different support values to compare the different types of patterns and to demonstrate the applicability of ARM algorithms on yarn and fabric data. The results show that proposed approach is very useful for discovering novel and interesting rules related to fabric quality.

This article is structured as follows: Section 2 summarizes previous works on the subject. Section 3 gives brief background information on ARM and the types of frequent patterns. This section also proposes novel concepts with their definitions. Section 4 explains the extended (proposed) version of the FP-Growth algorithm. Section 5 presents the obtained experimental results. Finally, concluding remarks and future directions are presented in Section 6.

2. Related work

With significant innovations and improvements in the textile domain, it became necessary to analyze textile data, including material properties, machine settings, and process parameters. Thus, several studies on the prediction of yarn [8] and fabric [9] properties have been conducted by using traditional methods such as regression model [10], kernel estimation [11], and correlation analysis [12]. Although all of these traditional methods provide predictions about yarn and fabric properties and influences, they unfortunately remain incapable of analyzing complex relationships among data attributes, estimating unknown attribute values, or investigating hidden patterns among instances. At this point, data mining plays an important role and has been started to be used in the textile area. In the literature, there are several articles about implementation of classification [13] and clustering [14] methods on fabric studies to tackle different problems where traditional methods are not useful.

In the textile sector, the most commonly used data mining technique is classification, which assigns a label to an unknown target attribute using a constructed predictive model [15]. There are five different classification algorithms that are broadly implemented in the textile area to process fabric data: artificial neural networks (ANNs) [16], support vector machines (SVMs) [17] Bayesian classifiers [18], decision trees [19], and k-nearest neighbors [20]. For example, Bahadir et al. [21] determined the drape behavior of woolen fabrics treated with different dry-finishing processes using an ANN. The results of their study indicated that the ANN had a high prediction rate for woolen fabrics. Jing et al. [22] developed a predictive model for the wrinkling appearance of fabric using modified wavelet coefficients and optimized SVM classifications.

In addition to classification studies, several studies implemented clustering techniques on fabric data, especially k-means [23], fuzzy c-means [24], and hierarchical [25] algorithms. In these studies, fabric images were first preprocessed and parameters were defined by a feature extractor, and then a clustering algorithm was applied.

The aforementioned textile studies only used classification and clustering tasks of data mining for fabric data to get valuable information required for fabric quality. Although ARM has a very long history in many areas, especially in marketing, only a few studies [4–7] have been done so far in the textile industry. However, these studies do not aim to determine textile parameters; instead, they focus on different purposes such as textile marketing [4], ARM method verification by using a textile dataset [5,6], and environmental strategies developed by fashion companies [7]. One of the studies in the literature [3] proposed a quality management system based on hybrid OLAP-association rule mining for extracting defect patterns in the garment industry. Similarly, ARM can be used to discover useful patterns and interesting relationships among set of yarn and fabric parameters in a textile data. Considering this motivation, the work presented in this paper focuses on the application of frequent pattern mining on yarn and fabric data for the first time in the literature. Differently from the previous studies related to quality management in textile sector, in this study, frequent pattern mining is used to make the right decisions to increase quality and productivity for textiles using an extended FP-Growth algorithm. It is the first study that investigates complex relationships by obtaining frequent, top-k, closed, and maximal frequent itemsets in the textile sector. The proposed algorithm was used for discovering interesting relationships among a set of yarn parameters and fabric properties in this study. The proposed approach can also be implemented for other textile domains, such as fault detection in embroidery, garment and fabric defects, prediction of utility properties and consumption of textile materials, and textile printing.

3. ARM

ARM discovers interesting correlations, frequent patterns, or associations among items in a dataset [26]. An association rule could be in the form X $\Rightarrow$ Y, meaning "if itemset X occurs in a transaction, then itemset Y will also likely occur in the same transaction." There are two main constraints for association rules: support and confidence. The support of an itemset is the percentage of transactions containing this itemset and indicates the frequency of appearance of it. For example, if the support of rule X $\Rightarrow$ Y is 50%, it means that fifty percent of the transactions contain X and Y itemsets together. The confidence of a rule is the fraction of transactions containing itemset X that also contain itemset Y. Association rules are extracted among items by satisfying minimum support, denoted by *minsup*, and minimum confidence, denoted by *minconf*, thresholds.

3.1. Frequent pattern mining (FPM)

FPM discovers patterns from data that are more frequent than the specific threshold [27]. An itemset I is called a frequent itemset (FI) if its support value, which is denoted by $\sigma(I)$, in the dataset D is greater than or equal to the user-defined minimum support, i.e. if $\sigma(I) \geq$ *minsup*. Frequent pattern analysis on a large volume of data is a challenging process, since there is usually a large number of distinct single items, and their combinations may form a huge number of itemsets; thus, it requires a significant amount of time. In addition, the necessary storage capacity plays an important role. Due to the large amount of frequent itemsets that can be generated from a dataset, some studies [28,29] revealed the need for concise representations of all frequent itemsets such as closed and maximal frequent itemsets.

3.1.1. Closed frequent itemset (CFI)

An itemset I is a closed itemset if there exists no itemset I' such that 1) $I \subset I'$ and 2) $\forall$ transaction T, $I \in T \longrightarrow I' \in T$.

CFI is a subset of frequent itemsets that has no superset with the same support, as shown in Eq. (1) [28].

$$\forall X \supset C : \sigma(X) < \sigma(C) \, \Lambda \, \sigma(X) \geq \, minsup \, \Lambda \, \sigma(C) \geq \, minsup, \tag{1}$$

where C is a closed frequent itemset whose supersets X have a strictly smaller support.

3.1.2. Maximal frequent itemset (MFI)

A closed frequent itemset is an MFI if it is not a subset of any other frequent itemset, as shown in Eq. (2) [29].

$$\forall X \supset M : \sigma(X) < \, minsup \, \Lambda \, \sigma(M) \geq minsup, \tag{2}$$

where M is a maximal frequent itemset that has no frequent superset like X.

3.1.3. Top-k frequent itemsets (TFI)

An itemset I is a TFI if I is the k most frequent itemset for a specified value k [30,31]. Users will need to give the desired count of frequent itemsets, which is an easily understood parameter.

Given an itemset I, let $f(I)$ be frequency of I in dataset D. Assume that the complete list of itemsets is denoted as δ, which is sorted in descending order according to their frequencies such that $\delta = \{I_1, I_2, I_3, ..., I_p\}$. For a given k, with $1 \leq k \leq p$ and $p = |\delta|$, $f(I_k)$ represents the frequency of kth itemset. Top-k frequent itemsets can be represented as shown in Eq. (3).

$$TFI(k) = \{(I, f(I)) : I \in \delta, f(I) \geq f(I_k)\}, \tag{3}$$

where $TFI(k)$ refers to the k most frequent itemsets in dataset D.

A closed itemset I is a top-k frequent closed itemset (TFCI) if there is no more than $(k-1)$ closed itemsets whose frequency is higher than that of I.

A maximal itemset is a top-k frequent maximal itemset (TFMI) if its frequency count is among the k highest frequencies of maximal itemsets, where k is the desired number of frequent maximal itemset.

3.2. Item mining: cf-item and mf-item

Item mining is a part of traditional frequent pattern mining with the goal of identifying items that are essential for the analysis. It focuses on discovering frequent items whose length is equal to 1. In the subject of item mining, this study proposes two novel concepts, cf-item and mf-item, to distinguish the types of items in the dataset. Zaki and Hsiao [28] indicated closed and maximal itemsets that are subsets of frequent itemsets and include more than one item. However, the present study introduces cf-item and mf-item concepts because single-item identification is particularly important where there is a need to investigate the significances of the attributes.

3.2.1. CF-item

Cf-item is a single item that is both closed and frequent. Cf-items can be determined by finding the subset of closed frequent patterns whose lengths are equal to 1.

Definition 1 *Let $\mathcal{I} = \{i_1, i_2, ..., i_m\}$ be a set of m items. Let C be closed and frequent, and $||C||$ denotes its length. The item C is said to be cf-item if it has no superset with the same support value, its support count is higher than minimum support denoted by $\sigma(C) \geq$ minsup, $C \in \mathcal{I}$ and its length is equal to 1, so $||C|| = 1$. Cf-item is defined in the Eq. (4).*

$$\forall X \supset C : \sigma(X) < \sigma(C) \ \Lambda \ \|C\| = 1 \ \Lambda \ \sigma(X) \geq minsup \ \Lambda \ \sigma(C) \geq minsup, \tag{4}$$

where C is a cf-item whose supersets X have a strictly smaller support and the maximum length is limited to 1.

3.2.2. MF-item

Mf-item is a single item such that it is both maximal and frequent. Mf-items are the subset of maximal frequent patterns whose lengths are equal to 1. This study is the first study that proposes mf-item.

Definition 2 *Let $\mathcal{I} = \{i_1, i_2, ..., i_m\}$ be a set of m items. Let M be maximal and frequent, and $||M||$ denotes its length. The item M is said to be mf-item if it has no frequent superset, its support count is higher than minimum support denoted by $\sigma(M) \geq$ minsup, $M \in \mathcal{I}$, and its length is equal to 1 and so $||M|| = 1$. Mf-item is defined in Eq. (5).*

$$\forall X \supset M : \sigma(X) < minsup \ \Lambda \ \|M\| = 1 \ \Lambda \ \sigma(M) \geq minsup, \tag{5}$$

where M is an mf-item that has no frequent superset like X and the maximum length is limited to 1.

Figure 1 shows the relationship among the different types of frequent patterns. The general relationship is MFI $\subseteq$ CFI $\subseteq$ FI. Cf-items, mf-items, and top-k patterns are located in the related areas as a subset. In addition, all MFIs are closed because they have no frequent superset and so they cannot have the same support count as their supersets. Top-k patterns are located in the related areas as a subset.

Table 1 shows an example of obtaining closed, maximal, top-k frequent itemsets, cf-item, and mf-item from a sample dataset. The example dataset contains five items $\mathcal{I}$ = {A, B, C, D, E} and consists of six transactions, which are uniquely identified by an ID. Minimum support threshold was chosen as 50% and so items or itemsets that occur in the dataset three or more times will be selected as frequent.

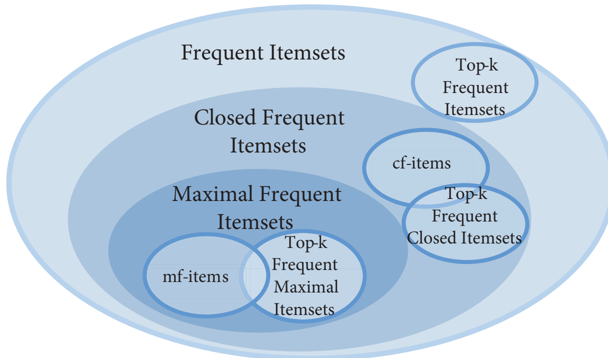

Figure 1. The relationships among the different types of frequent patterns.

Table 1. A sample dataset and different types of frequent patterns obtained from it.

ID	Transactions	FI	CFI	MFI	Top-4 FI	Items
1	A, C, D	1-itemset	1-itemset	1-itemset	1-itemset	cf-item
2	B, C, E	{B} : 4	{C} : 4	{D} : 3	{B} : 4	{C} : 4
3	A, B, C, E	{C} : 4	{D} : 3	3-itemset	{C} : 4	{D} : 3
4	D, B, E, C	{D} : 3	2-itemset	{B, C, E} : 3	{E} : 4	mf-item
5	D	{E} : 4	{B, E} : 4		2-itemset	{D} : 3
6	B, E	2-itemset	3-itemset		{B, E} : 4	
		{B, C} : 3	{B, C, E} : 3			
		{B, E} : 4				
		{C, E} : 3				
		3-itemset				
		{B, C, E} : 3				

Step 1: Frequent itemsets

To find frequent itemsets, all transactions are traversed and the support values of items are evaluated first. Items whose support values are greater than or equal to the minimum support are selected as frequent. In this example, all items, except item A, are frequent because they appear in more than three transactions. In the next iteration, candidate 2-itemsets are generated using only the frequent 1-itemsets and evaluated by counting their supports. For example, the candidate {B, D} is found to be infrequent after computing their support values. Next, the algorithm will iteratively generate new candidate k-itemsets using the frequent (k − 1)-itemsets found in the previous iteration.

Step 2: Closed frequent itemsets

Itemset {C} , with support 4, is a closed frequent itemset because its supersets ({B, C} : 3, {C, E} : 3, and {B, C, E} : 3) have a smaller support count (3). However, itemset {B, C} is not a closed frequent itemset because its superset {B, C, E} has the same support of 3. For the same reason, itemsets {B} , {E} , and {C, E} are also not closed frequent itemsets.

Step 3: Maximal frequent itemsets

Itemsets {D} and {B, C, E} are maximal frequent itemsets because none of their supersets are frequent. In contrast, itemset {B, E} is nonmaximal because one of its immediate supersets {B, C, E} is frequent. For the same reason, itemset {C} is also not a maximal frequent itemset.

Step 4: Top-k frequent itemsets

When k parameter is assigned to 4, the k most frequent itemsets in this example could be {B} , {C} , {E} , and {B, E} without any constraints.

Step 5: Top-k frequent closed itemsets

The frequency counts of the itemsets {C} and {B, E} are among the $k = 2$ highest frequencies of closed itemsets.

Step 6: Top-k frequent maximal itemsets

When $k = 2$ is the desired number of frequent maximal itemsets, {D} and {B, C, E} are selected in this example.

Step 7: Cf-item

The closed frequent itemsets whose lengths are equal to 1 are {C} and {D}. Cf-items in data can be found if there is a need to investigate the significances of the items.

Step 8: Mf-item

The item D is both closed and maximal, because none of the supersets of this item are frequent. Determining a mf-item is important when performing single item-based data analysis.

3.3. Advantages of proposed method

The novel concepts (cf-item, mf-item) and the extended FP-Growth algorithm can be used to solve different types of problems in different areas, such as feature selection, the determination of important parameter values, and discretization.

3.3.1. Feature selection

Feature selection, one of the important data-preprocessing stages, is performed to choose a subset of relevant items in the dataset. This process decreases the number of features and increases the accuracy of the categorization. Cf-item and mf-item can be useful when identifying the most frequent single features in the data set.

3.3.2. Determination of parameter values

Cf-item and mf-item specify the significant of the parameters and so they can be used to determine the important of items with their values. For example, a cf-item discovered from the dataset {YarnHairiness = (3–4]} and interpreted as *Yarn Hairness* is one of the significant parameters, with a range between 3 and 4.

3.3.3. Discretization

Discretization converts numeric values of attributes to nominal/ordinal values by using a categorization strategy. The key point of discretization is the determination of a set of optimal split points and intervals. The presented method in this study discovers the patterns containing same attributes with different range of values. Suppose cf-items {YarnHairinessH = (3–5]} and YarnHairinessH = (8–10]} were obtained when the algorithm was executed. According to these patterns, the optimal split points can be found to define the interval boundaries in the discretization process.

4. Algorithm

Pattern-mining algorithms have a wide range of applications, such as cross-marketing, website click stream analysis, and biomedical applications. The most commonly used algorithms in these applications are Apriori, FP-growth, and Eclat. FP-Growth stands for "frequent pattern growth" and was proposed for discovering sets of frequent patterns using an extended prefix-tree structure named FP-tree. FP-Growth was developed as an alternative for the Apriori algorithm to handle large volumes of frequent itemsets with high performance utilizing a divide-and-conquer strategy. The algorithm consists of two steps: building an FP-tree and obtaining frequent itemsets from this tree. An FP-tree has a compact prefix tree structure that stores and represents the transaction database horizontally and vertically. While horizontal representation of the tree indicates a prefix tree of transactions, vertical representation shows links between the prefix tree branches. The root of the tree is labelled as "null" and each node holds an item's name, an item's transaction count, and node links.

A simple FP-tree construction example with minimum support 50% is illustrated in Figure 2 by considering a sample dataset given in Table 2. To construct the tree, the algorithm passes over the dataset two times. First, the support count of each item is calculated and frequent items are sorted in decreasing order. At the next pass, each transaction is read and mapped to a path in the tree. Paths in the FP-tree overlap when different transactions share common items. Cf-items and mf-items are highlighted in gray on the tree.

Table 2. Transactions in a sample dataset and their frequent items.

ID	Items	1-itemset	Ordered frequent items
1	A, C, D	{B} : 4	C, D
2	B, C, E	{C} : 4	B, C, E
3	A, B, C, E	{E} : 4	B, C, E
4	D, B, E, C	{D} : 3	B, C, E, D
5	D		D
6	B, E		B, E

This study presents an extended version of the FP-Growth algorithm that has the ability to find the different types of patterns, such as frequent, closed, maximal, top-k frequent, top-k closed, top-k maximal, cf-item, and mf-item. The pseudocode of the extended FP-Growth algorithm is presented in Figure 3. The algorithm first computes a list of frequent items sorted by frequency in descending order ($F[]$). After that, the FP-Tree is constructed by scanning each transaction in the dataset. Then the FP-Growth method [32] starts to mine the FP-tree for each frequent item by recursively building conditional trees. The algorithm also mines closed and maximal patterns on frequent itemsets. The lengths of each itemset in CFI and MFI are also controlled to determine cf-items and mf-items, respectively. In addition, it also determines the k most frequent, closed, and maximal itemsets. The time complexity of computing the list $F[]$ is $O(n)$, where n is the number of the transactions in the dataset. However, the computational cost of procedure *Growth()* is at least polynomial.

5. Experimental study

In this study, the extended FP-growth algorithm was applied on a real-world fabric dataset [33] to discover the relationships between selected yarn parameters with selected fabric properties.

The algorithm was executed on the dataset with varying support threshold values to compare the different types of patterns. Most relevant yarn parameters with fabric properties that were obtained as the outcomes of this algorithm are explained and the number of frequent items/itemsets are shown with the help of charts. In

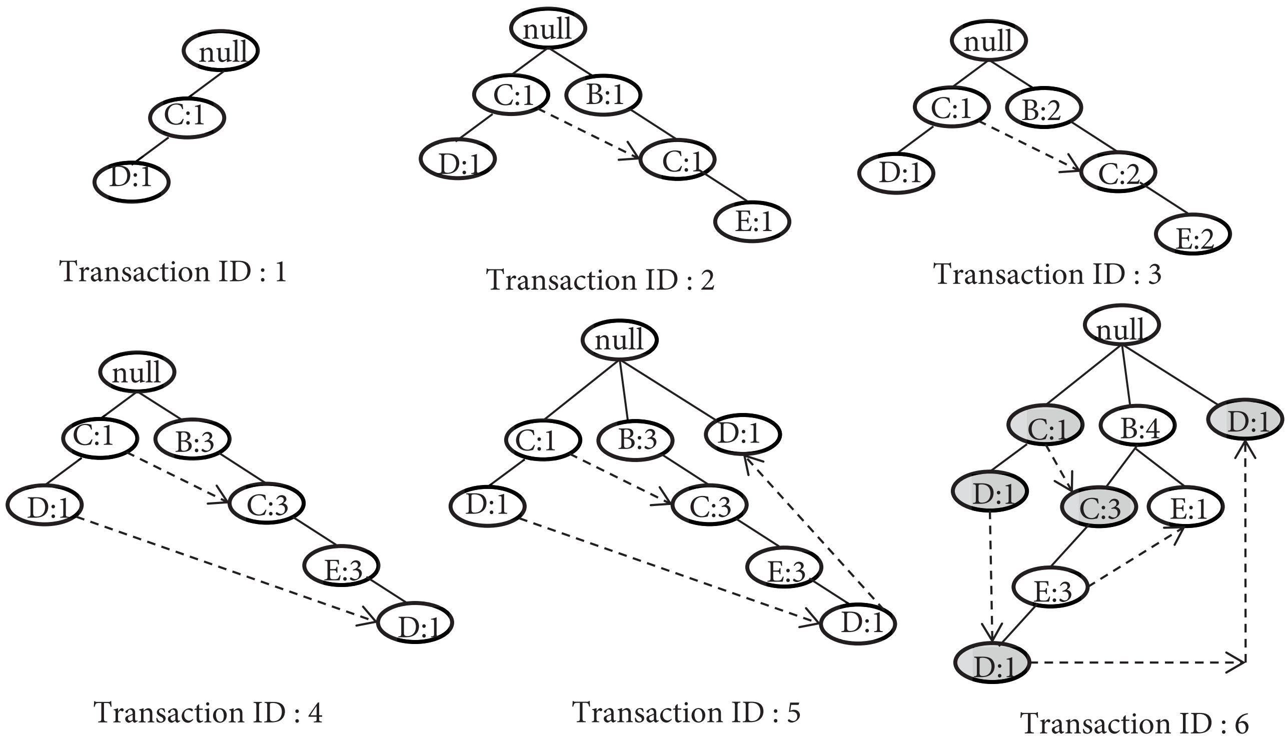

Figure 2. Illustrated FP-tree, cf-items, and mf-items in the tree.

this experimental study, cf-item and mf-item were utilized for the first time to perform single item-based data analysis.

5.1. Dataset description

The dataset considered in this study contains selected yarn parameters (yarn manufacturing method, elongation at break, irregularity, hairiness, bending rigidity, and capillary properties) and selected fabric properties (pilling, abrasion resistance, and bending rigidity) that were experimentally obtained in a previous study [33]. The raw dataset contains 1800 records and consists of fifteen attributes, including nominal and numerical values: nine of them are yarn parameters and the rest of them are fabric features. The statistical details of the dataset are presented in Table 3.

ARM algorithms require categorical data; they cannot directly deal with numeric attributes. For this reason, in this study, numeric attributes were discretized into intervals by finding a set of significant split points of distribution changes that define the interval boundaries of the discretization. The discretization process is divided into different categories from different perspectives, such as supervised or unsupervised, top-down or bottom-up, static or dynamic, local or global, nominal or ordinal, univariate or multivariate, direct or iterative [34] The discretization technique applied in this study was unsupervised, top-down, static, global, nominal, univariate, and direct. The split points were obtained by using both expert techniques approved by the textile community and an equal width binning method that divides numerical values into equal n intervals. Due to irregular distribution, three attributes (hairinessS3, yarn diameter, and abrasion resistance) were categorized by evaluating boundaries and binning widths by using a frequency table, while the rest of the numeric attributes in the dataset were discretized by equal width binning method. For example, attribute tenacity was discretized

Table 3. Statistical details of the dataset and the categories in the attributes.

Attributes	# of records	Min value	Max value	Mean	Std. Dev.	Categories
Yarn tenacity	100	16.98	21.87	18.97	1.6	[15.5–17.5], (17.5–19.5], (19.5–21.5], (21.5–23.5]
Yarn elongation at break (%)	100	14.54	18.47	16.66	0.99	[13.5–15.5], (15.5–17.5], (17.5–19.5]
Yarn irregularity	100	7.96	9.6	8.76	0.54	[7–9], (9–11]
Yarn hairiness (H)	100	3.63	5.5	4.16	0.69	[3–4], (4–5], (5–6]
Yarn hairiness (S3)	100	8	1947	594.08	784.27	[3–9], (9–27], (27–39], (39–462], (462–1419], (1419–2376]
Yarn capillary	200	2	4.3	3.18	0.56	[2–3], (3–4], (4–5]
Yarn bending rigidity	600	3.34	4.89	4.03	0.42	[3–3.5], (3.5–4], (4–4.5], (4.5–5]
Yarn diameter	200	0.51	0.7	0.58	0.06	[0.5–0.56], (0.56–0.63], (0.63–0.7]
Abrasion resistance	48	16.3	25.1	21.46	2.28	[16–18.75], (18.75–20.75], (20.75–23.125], (23.125–25.5]
Pilling resistance	36	3	5	4.17	0.49	[3–3.5], (3.5–4], (4–4.5], (4.5–5]
Wrinkle resistance	72	100.5	143.25	121.86	13.3	[100–111], (111–122], (122–133], (133–144]
Fabric bending rigidity	64	0.91	2.59	1.55	0.62	[0.5–1.3], (1.3–2.2], (2.2–3]
Capillary warp direction	40	1.6	3.6	2.44	0.55	[1–1.75], (1.75–2.5], (2.5–3.25], (3.25–4]
Capillary weft direction	40	1.2	3.4	2.1	0.61	[1–1.75], (1.75–2.5], (2.5–3.25], (3.25–4]
Yarn manufacturing method		MVS, RAJ, SIRO, RING				

```
Algorithm Extended_FP-Growth (D, minsup)
  Inputs:  D: dataset, minsup: minimum support, 𝒥: set of items in D,
           k: the desired number of patterns (top-k)
  Outputs: FI: frequent itemsets, CFI: closed frequent itemsets,
           MFI: maximal frequent itemsets, TFI: top-k frequent itemsets,
           TFCI: top-k frequent closed itemsets,
           TFMI: top-k frequent maximal itemsets, cf-item, mf-item
begin
  Define frequency list: F[ ] = {}
  foreach transaction T_i in D
    foreach item a_j in T_i
      F[a_j]++
  Sort F[ ]
  Define and clear the root of FP-tree: r
  foreach Transaction T_i in D
    Make T_i ordered according to F[ ]
    Call ConstructTree(T_i, r)
  foreach item a_i in 𝒥
    if F[a_i] ≥ minsup
      FI = FI ∪ Growth(r, a_i, minsup)
  foreach frequent itemset f_i in FI
    if f_i has no superset in FI with same support
      CFI = CFI ∪ f_i
      if |f_i| = 1
        cf-item = cf-item ∪ f_i
      if f_i has no superset in FI
        MFI = MFI ∪ f_i
        if |f_i| = 1
          mf-item = mf-item ∪ f_i
  for i = 1 to k
    TFI = TFI ∪ FI.sort[i]
    TFCI = TFCI ∪ CFI.sort[i]
    TFMI = TFMI ∪ MFI.sort[i]
end
```

Figure 3. Pseudocode of the extended FP-Growth algorithm.

into four categories as follows: [15.5–17.5], (17.5–19.5], (19.5 –21.5], (21.5–23.5] and each numeric value in this attribute was mapped in a category according to the range of value. The last column in Table 3 shows the categories of attributes with their interval values that were selected during the discretization process.

5.2. Comparison of different types of patterns

An extended FP-Growth algorithm was executed on the dataset with varying support thresholds from 10 to 60 in increments of 5. Only those items with support values greater than or equal to the threshold level were selected as frequent patterns and others were discarded. Table 4 shows the numbers of different types of patterns (i.e. frequent, closed, maximal, cf-item, and mf-item) separately varying from 1-itemset to 5-itemset. Results show that the number of frequent itemsets produced from dataset is large when the minimum support level is set to low, i.e. the number of 5-itemset patterns is 7166 when *minsup* = 10%. However, the algorithm produces a reasonable number of closed and maximal frequent itemsets, i.e. the number of 5-itemset closed patterns is 57 when *minsup* = 10%. Thus, it is possible to compress the collection of frequent itemsets in a more manageable size. In addition, it is also possible to determine the significance of the attributes with cf-item and mf-item concepts.

Table 4. The numbers of different types of patterns with different support thresholds.

	Frequent itemsets					Cf-item	Closed frequent itemsets				Mf-item	Maximal frequent itemsets			
Support	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
10	48	502	2084	4772	7166	14	41	58	67	57	0	0	1	7	12
15	41	306	805	1084	891	14	41	57	60	45	0	0	10	12	35
20	37	183	302	240	113	14	41	47	48	10	1	6	12	36	9
25	27	100	105	48	13	13	36	35	14	1	0	11	20	13	1
30	21	48	26	3	0	13	27	18	3	0	2	12	15	3	0
35	13	25	7	0	0	11	17	7	0	0	2	9	7	0	0
40	13	13	3	0	0	11	9	3	0	0	4	5	3	0	0
45	9	9	1	0	0	7	6	1	0	0	1	6	1	0	0
50	9	5	1	0	0	7	2	1	0	0	4	2	1	0	0
55	6	1	0	0	0	6	1	0	0	0	4	1	0	0	0
60	3	1	0	0	0	3	1	0	0	0	1	1	0	0	0

Figure 4 shows the comparative results of the numbers of frequent, closed and maximal itemsets for support threshold levels ranging from 20 to 50 in increments of 5. The figure presents the results for different lengths of patterns, i.e. 1-itemset, 2-itemset, 3-itemset, and 4-itemset. Results show that when the minimum support value decreases, the number of FI patterns increases almost exponentially. This means that a large amount of frequent itemset patterns are generated when the algorithm is executed with small support values. However, the algorithm produces a reasonable number of CFI and MFI patterns. When the minimum support value and size of the itemset increase (i.e. *minsup* = 25% and 4-itemset), the differences between the number of FIs, CFIs, and MFIs decrease. For this reason, the type of the pattern is not critical in the case of large parameter values.

The graph in Figure 5 shows the number of closed and maximal frequent items generated by the extended FP-Growth algorithm for varying support thresholds from 30 to 60 in increments of 5. From this graph, it is possible to see that the number of cf-items is always greater than or equal to the number of mf-item patterns because of the relationship MFI $\subseteq$ CFI. The obtained results also show that while minimum support value increases, the number of cf-items decreases about linearly. However, the numbers of obtained mf-item patterns are irregular because the supersets of the patterns changes according to support threshold levels. When the support value increases, the differences between the number of CFI and MFI decrease. For this reason, the type of the pattern is not as critical as in the case of large support values.

Figure 6 shows the lowest and highest support values of top-k patterns with k ranging from 1 to 10. It compares TFI, TFCI, and TFMI patterns with 2-itemset when the minimum support threshold is 35%. According to the results, the lowest and highest support values of TFI and TFCI patterns are generally close to each other, but the support values of TFMI patterns in top-k lists are lower than them. To select interesting relationships among data and determining frequent itemsets, the widely used parameter is minimum support value. However, specifying optimal minimum support threshold is a difficult and time-consuming task for users because selecting the threshold is somewhat unstable. For this reason, it is also possible to discover top-k frequent patterns without the minimum support specification. In this case, a specified itemset-length can also

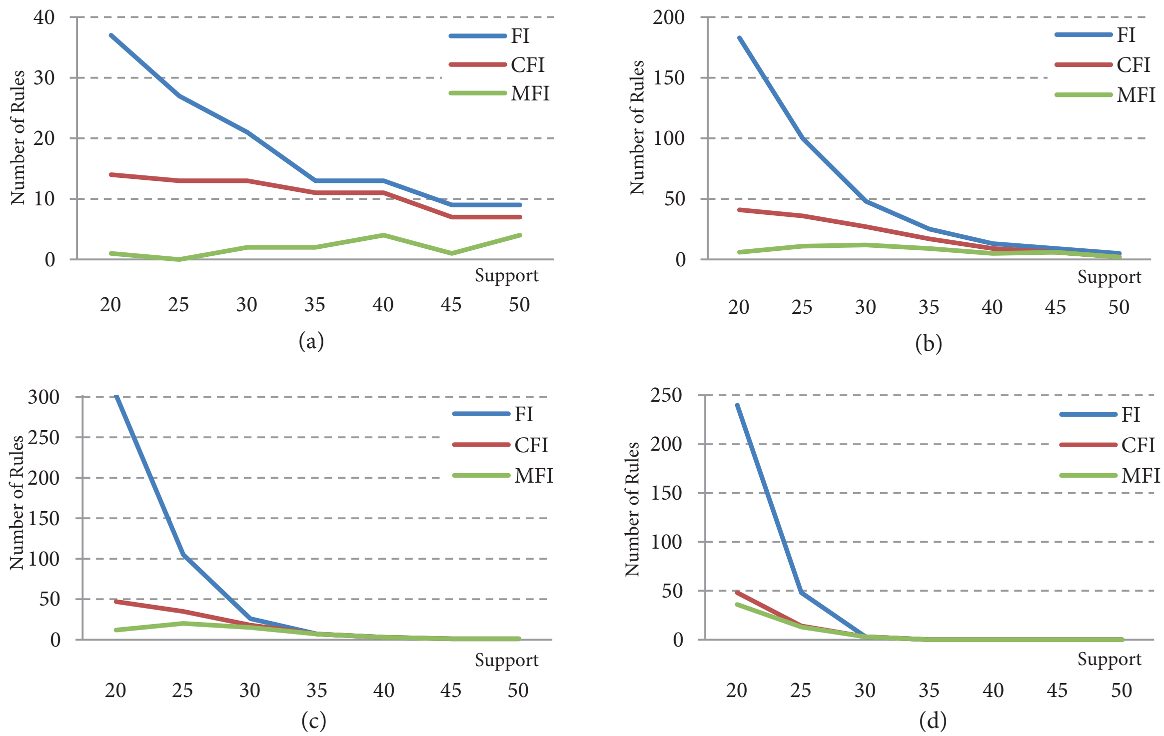

Figure 4. Comparison of the numbers of FI, CFI, and MFI patterns. (a) 1-itemset, (b) 2-itemset, (c) 3-itemset, (d) 4-itemset.

be used as a threshold to focus on the desired pattern size.

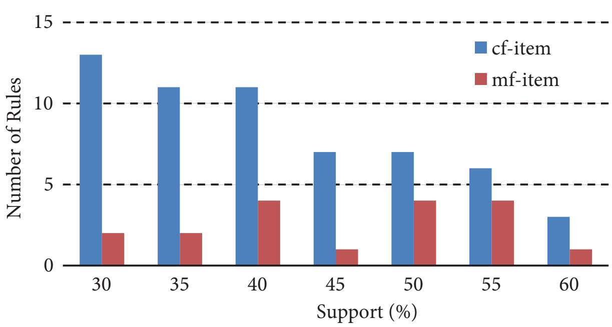

Figure 5. Comparison of the numbers of cf-items and mf-items.

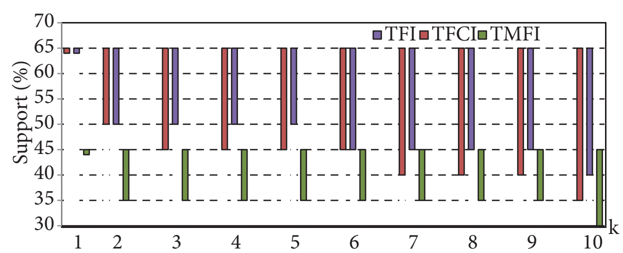

Figure 6. The lowest and highest support values of top-k patterns (2-itemset).

5.3. Association rules

Table 5 shows some rules discovered by the algorithm. According to the results, the patterns {YarnElongationAt Break = (15.5–17.5]} and {YarnIrregularity = [7–9]} are the most frequent 1-itemsets, which indicates that they are the most influential parameters among the ones considered. Following these two parameters, the important attributes and their range values were determined as cf-items and mf-items and given as, for example, {YarnHairinessH = (3–4]} and {YarnCapillary = (3–4]} , respectively. The results also express the relationships among yarn manufacturing methods (msv, raj, siro, ring), yarn parameters (i.e. hairiness, capillary, diameter), and fabric properties (i.e. pilling, wrinkle, abrasion), as some are indicated in Table 5 by example patterns

Table 5. Examples of rules discovered by ARM.

Length	Pattern	Support (%)	Minsup (%)	Type
1-Itemsets	{YarnElongationAtBreak = (15.5–17.5]}	75	55	TFCI
	{YarnIrregularity = [7–9]}	75		TFCI
	{YarnHairinessH = (3–4]}	70		cf-item
	{YarnBendingRigidity = (3.5–4]}	55		TFMI
	{YarnCapillary = (3–4]}	55		mf-item
2-Itemsets	{YarnHairinessH = (3–4], YarnIrregularity = [7–9]}	45	30	CFI
	{YarnHairinessH = (3–4], FabricPilling = (3.5–4]}	35		FI
	{YarnElongationAtBreak = (15.5–17.5], FabriWrinkle = [100–111]}	30		FI
	{YarnTenacity = (19.5–21.5], FabricBendingRigidity = [0.5–1.3]}	30		FI
	{YarnHairinessH = (3–4], FabricAbrasion = (20.75–23.125]}	30		MFI
3-Itemsets	{YarnIrregularity = [7–9], YarnDiameter = [0.5–0.56], FabricBendingRigidity = [0.5–1.3]}	50	30	MFI
	{YarnTenacity = (17.5–19.5], YarnBendingRigidity = (3.5–4], YarnIrregularity = [7–9]}	30		CFI
	{YarnBendingRigidity = (3.5–4], FabricCapillaryWarp = (1.75–2.5], YarnElongationAtBreak = (15.5–17.5]}	35		FI
4-Itemsets	{YarnManufacturing = SIRO, YarnIrregularity = [7–9], YarnDiameter = [0.5–0.56], FabricBendingRigidity = [0.5–1.3]}	25	20	CFI
	{YarnManufacturing = RAJ, YarnElongationAtBreak = (15.5–17.5], YarnHairinessH = (3–4], FabricPilling = (3.5–4]}	25		FI
	{YarnManufacturing = MVS, FabricCapillaryWarp = (1.75–2.5], YarnTenacity = (17.5–19.5], FabricBendingRigidity = (1.3–2.2]}	20		FI
	{YarnManufacturing = RING, YarnHairinessH = (5–6], YarnElongationAtBreak = (17.5–19.5], FabricWrinkle = (133–144]}	20		FI
	{YarnDiameter = [0.5–0.56], FabricAbrasion = (20.75–23.125] YarnIrregularity = [7–9], FabricBendingRigidity = [0.5–1.3]}	20		MFI

of two or more itemsets. For 2-itemsets, the patterns {YarnHairinessH = (3–4], FabricPilling = (3.5–4]} and {FabricAbrasion = (20.75–23.125], YarnHairinessH = (3– 4]} indicate that when the yarn hairiness index H values are between 3 and 4, the fabric pilling performance can be expected to be from 3.5 to 4 and the fiber loss values due to abrasion of the fabric will lie between 20.75 and 23.12 mg. When the yarn irregularity values lie between 7 and 9 and yarn diameter is between 0.5 and 0.56 mm the fabric bending rigidity is expected to be between 0.5 and 1.3 with the rule {YarnIrregularity = [7–9], YarnDiameter = [0.5–0.56], FabricBendingRigidity = [0.5–1.3]}. If the yarns are manufactured by the Rieter Air Jet (RAJ) method and have elongation at break values between 15.75% and 17.75% and yarn hairiness H index between 3 and 4, then the pilling performance of the fabric from such yarn will be between 3.5 and 4, as indicated by the pattern {YarnManufacturing = RAJ, YarnElongationAtBreak = (15.5–17.5], YarnHairinessH = (3–4], FabricPilling = (3.5–4]}. For the textile industry, where there are many parameters affecting the end-product performance, the relationships between the parameters of the yarn and the fabric is important to understand the product behavior so that the end-product can be shaped according to the customer. With these example patterns, the algorithm proves to be able to derive the relationships between the yarn and fabric parameters and also the significant values for the important parameters could be stated.

6. Conclusion and future work

A rapidly expanding and highly competitive textile industry creates a great demand for exploring interesting and potentially useful information from data. Because of this requirement, implementing data mining techniques in the textile sector has been proposed as the solution for the challenge of discovering hidden patterns among textile data. This study presents the application of an extended FP-Growth algorithm on a yarn and fabric dataset to discover interesting relationships among a set of yarn parameters (i.e. hairiness, capillary, and diameter) and fabric properties (i.e. pilling, wrinkle, and bending rigidity).

To the best of our knowledge, this is the first study that investigates relationships in textile data by obtaining frequent, top-k, closed, and maximal frequent itemsets. In addition, this article also proposes two novel approaches, cf-item and mf-item, to perform single item-based data analysis. In the experimental studies, the algorithm was executed on a real-world textile dataset with different support threshold values to compare the different types of patterns. Experimental results and discovered association rules show that proposed approach is very useful for discovering the rules on textile data.

As future research, negative association rules can be addressed to be able to describe the occurrences of some textile properties characterized by the absence of others. For example, it would be interesting to find out which factors are relevant and irrelevant. Recently, several ontologies have been developed for the textile, fashion, and clothing domains. In the future, ARM studies can be supported by the ontologies to extract semantic relationships. Mining sequential patterns in textile data could also be one of the future research areas.

Acknowledgments

The authors would like to express their gratitude to Tuğba Yaşar and her thesis supervisor Prof Dr Hüseyin Gazi Türksoy for making certain textile data available for use in this study.

References

[1] Chattopadhyay R. Artificial neural networks in yarn property modeling. In: Majumdar A, editor. Soft Computing in Textile Engineering. Delhi, India: Woodhead Publishing, 2011. pp. 105-125.

[2] Shen J, Zhou X, Ma H, Chen W. Spectrophotometric prediction of pre-colored fiber blends with a hybrid model based on artificial neural network and Stearns-Noechel model. Text Res J 2017; 87: 296-304.

[3] Lee CKH, Choy KL, Ho GTS, Chin KS, Law KMY, Tse YK. A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry. Expert Syst Appl 2013; 40: 2435-2446.

[4] Huang Y, Qiu X, Yang J. The association rule mining technology in marketing decision-making and its application in logistics enterprises. In: Proceedings of the 4th International ISKE Conference on Intelligent Systems and Knowledge Engineering; 27–28 November 2009; Hasselt, Belgium: ISKE. pp. 252-257.

[5] Ingle MG, Suryavanshi NY. Association rule mining using improved Apriori algorithm. Int J Comput Appl 2015; 112: 37-42.

[6] Logeswari T, Valarmathi N, Sangeetha A, Masilamani M. Analysis of traditional and enhanced Apriori algorithms in association rule mining. Int J Comput Appl 2014; 87: 4-8.

[7] Ciarapcia FE, Sanctis ID, Resta B, Dotti S, Gaiardelli P, Bandinelli R, Fani V, Rinaldi R. Integrating sustainability in the fashion system using association rules. In: Rinaldi R, Bandinelli R, editors. Business Models and ICT Technologies for the Fashion Supply Chain, Lect Notes Electr En 2017; 413: 239-250.

[8] Cai Y, Cui X, Rodgers J, Thibodeaux D, Martin V, Watson M, Pang SS. A comparative study of the effects of cotton fiber length parameters on modeling yarn properties. Text Res J 2013; 83: 961-970.

[9] Xue Z, Zeng X, Koehl L, Chen Y. Extracting fabric hand information from visual representations of flared skirts. Text Res J 2013; 84: 246-266.

[10] Yang S, Gordon S. Accurate prediction of cotton ring-spun yarn quality from high-volume instrument and mill processing data. Text Res J 2016; 1: 1-15.

[11] Huang XX, Tao XM, Xu BG. A theoretical model of maximum hairiness of staple ring-spun yarns. Text Res J 2014; 84: 1121-1137.

[12] Park S, Yun C, Kim J, Park CH. The effects of the fabric properties on fabric movement and the prediction of the fabric movements in a front-loading washer. Text Res J 2013; 83: 1201-1212.

[13] Kumar TS, Sampath V. Prediction of dimensional properties of weft knitted cardigan fabric by artificial neural network system. J Ind Text 2012; 42: 446-458.

[14] Saharkhiz S, Abdorazaghi M. The performance of different clustering methods in the objective assessment of fabric pilling. J Eng Fiber Fabr 2012; 7: 35-41.

[15] Chenganmal M, Vijayakumar HL, Subramaniam V. Evaluation of bending rigidity of dressing materials using artificial neural networks. Journal of Textile and Apparel Technology and Management 2013; 8: 1-10.

[16] Rahnama M, Semnani D, Zarrebini M. Measurement of the moisture and heat transfer rate in light-weight nonwoven fabrics using an intelligent model. Fibres Text East Eur 2013; 21: 89-94.

[17] Sun J, Yao M, Xu B, Bel P. Fabric wrinkle characterization and classification using modified wavelet coefficients and support-vector-machine classifiers. Text Res J 2011; 81: 902-913.

[18] Zhang J, Xungai W. Objective pilling evaluation of wool fabrics. Text Res J 2007; 77: 929-936.

[19] Agarwal G, Koehl L, Perwuelz A, Lee KS. Interaction of textile parameters, wash-ageing and fabric softener with mechanical properties of knitted fabrics and correlation with textile-hand - Interaction of textile parameters with laundry process. Fiber Polym 2011; 12: 670-678.

[20] Yildiz K, Buldu A, Demetgul M. A thermal-based defect classification method in textile fabrics with k-nearest neighbor algorithm. J Ind Text 2016; 45: 1-16.

[21] Bahadir SK, Kalaoglu F, Jevsnik S, Eryuruk SH, Saricam C. Use of artificial neural networks for modelling the drape behaviour of woollen fabrics treated with dry finishing processes. Fibres Text East Eur 2015; 23: 90-99.

[22] Jing J, Zhang Z, Kang X, Jia J. Objective evaluation of fabric pilling based on wavelet transform and the local binary pattern. Text Res J 2012; 82: 1880-1887.

[23] Zhang J, Xin B, Fang H, Cao Y. Novel colour clustering method for interlaced multi-colored dyed yarn woven fabrics. Fibres Text East Eur 2015; 23: 107-114.

[24] Xiao Z, Nie X, Zhang F, Geng L, Wu J, Li Y. Automatic recognition for striped woven fabric pattern. J Text I 2014; 106: 409-416.

[25] Nourani G, Jeddi AAA, Moghadam MB. Determining the structural parameters and yarn type affecting tensile strength and abrasion of weft knitted fabrics using cluster analysis. Middle East J 2011; 8: 1008-1017.

[26] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of data; 25–28 May 1993; Washington, DC, USA: ACM. pp. 207-216.

[27] Aggarwal CC, Bhuiyan MA, Hasan MA. Frequent pattern mining algorithms: A survey. In: Aggarwal CC, Han J, editors. Frequent Pattern Mining. Basel, Switzerland: Springer International Publishing, 2014. pp. 19-64.

[28] Zaki MJ, Hsiao CJ. CHARM: An efficient algorithm for closed itemset mining. In: Proceedings of the Second SIAM International Conference on Data Mining; 11–13 April 2002; Virginia, USA: SIAM. pp. 457-473.

[29] Burdick D, Calimlim M, Gehrke J. MAFIA: a maximal frequent itemset algorithm for transactional databases. In: Proceedings of the 17th International Conference on Data Engineering; 2–6 April 2011; Heidelberg, Germany: IEEE. pp. 443-452.

[30] Pietracaprina A, Riondato M, Upfal E, Vandin F. Mining top-K frequent itemsets through progressive sampling. Data Min Knowl Disc 2010; 21: 310-326.

[31] Cheung YL, Fu AWC. Mining frequent itemsets without support threshold: with and without item constraints. IEEE Trans Knowl Data Eng 2004; 16: 1052-1069.

[32] Li H, Wang Y, Zhang D, Zhang M, Chang E. PFP: Parallel FP-Growth for query recommendation. In: Proceedings of the ACM Conference on Recommender Systems; 23–25 October 2008; Lausanne, Switzerland: ACM. pp. 107-114.

[33] Yaşar T. Hava jetli iplik eğirme sistemleri ile üretilmiş ipliklerin performanslarının değerlendirilmesi. MSc, Erciyes University, Kayseri, Turkey, 2015.

[34] Liu H, Hussain F, Tan CL, Dash M. Discretization: An enabling technique. Data Min Knowl Disc 2002; 6: 393-423.