Applied Computing and Informatics 14 (2018) 88–97

Contents lists available at ScienceDirect

Applied Computing and Informatics

journal homepage: www.sciencedirect.com

Original Article

A self-adaptive k-means classifier for business incentive in a fashion design environment

O.R. Vincent *, A.S. Makinde, O.S. Salako, O.D. Oluwafemi

Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria

ARTICLE INFO

Article history:
Received 18 January 2017
Revised 25 April 2017
Accepted 7 May 2017
Available online 9 May 2017

Keywords:
Data mining
K-means classification
Fashion design
Clustering and fashion business

ABSTRACT

An incentive mechanism to target market for fashion designers is proposed. Recent researches have been focused on the art, style or the design; while a few were based on traditional practice. In this study, economy is considered as a major liberation in the fashion world by analyzing six attributes, namely, style, color, fabric, brand, price and size that could bring about commercial success. Dataset of 1000 customers' records were used and categorized as original, combined and new designs using self-adaptive *k-means* algorithm, which extract common attributes that would foster better business from the dataset. The results would be useful to designers in knowing the type of designs usually ordered by customers with the design code, and which combinations of the attributes have high patronage. In addition, customers would have easy access to the best and current designs invoke from a combination of highest patronized designs.

1. Introduction

Design encourages communication by representing ideas usually through images. In today's business world, design cuts across many professions including architectural, interior decoration, footwear, graphics, landscape, furniture, fashion designs, etc. [39,29]. Over the years, the fashion industry has been engaged in different marketing strategies with traditional cataloguing being made referenced to on proxy without adequate segmental provisioning [38]. Thus, fashion industry relies on previous data and mental intuition on imaginations in order to predict and meet customers' demands. However, with the recent fast-dynamic fashion trends, the design market is highly competitive and therefore renders the present mechanism obsolete and irrelevant [33,13,17].

Traditional brainstorming on designs are time wasting and reduces productivity. Hence, it is difficult for fashion designers to place their designs on mannequins due to outdated styles [17]. Very recently, customers' needs and demands have been analyzed using different methods through comments, shares and re-tweets [6]. Mok et al. [29] proposed an Interactive Genetic Algorithm (IGA) to customised fashion design to enable customers create their preferred fashion designs in a user-friendly manner [29]. Unfortunately, busy customers would not have the time to start creating new designs themselves. In these days of e-payment, some customers even send money to designers and expect the fashion designers to deliver appropriately. In another development, set-theoretic qualitative comparative analysis (QCA) was used to discover patterns that hold reliably across some cases [27,26,15,17].

Other researchers used expert systems, Computer Aided Design (CAD) and neural networks to focus mainly on creating garments, targeting garment appearance and fittings without considering commercial success of the designers and satisfaction of the customers [21,36,7,28,37,39]. Consequently, there are no incentive mechanisms to entice customers to patronize fashion designers on frequent basis. Jacobs et al. [17] argued that the design art should be supported with business inclination. The business itself should be given the same attention as the style. This implies that the design mechanisms and other fabricates should be done with the intent of bringing more gains to the designers and greater satisfaction to the customers [16].

This paper proposes an incentive mechanism to target business success in fashion design using a novel self-adaptive *k-means* (SAk-means) by analyzing some attributes that would attract customers. *K-means* was used due to its ability to assume a balanced cluster

* Corresponding author.
E-mail addresses: vincent.rebecca@gmail.com, vincentor@funaab.edu.ng (O.R. Vincent).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2017.05.002
2210-8327/

size with the dataset. It also assumes that the joint distribution of features within each clusters is spherical; meaning that the features within a cluster have equal variance, and are independent of each other [14]. Each clusters have similar density. *K*-means is faster, because order of time complexity is linear with respect to the number of data and efficient in terms of computational cost. It uses unsupervised learning method to resolve known clustering issues and work well with large datasets [25,44].

Six attributes considered as market incentives include style, color, fabric type (texture), brand, price and size. We categorized fashion design in three ways: original, combine and new designs. A merging by extracting these attributes into significant insight of market strategies was done using the proposed algorithm [19,31]. Though, *k-means* has been criticized as a result of its high storage maintenance requirements and high computational cost, we argue that the proposed self-adaptive *k*-means is valuable and is dynamic by updating new classified contents and removing old unclassified contents from the dataset. By this, the problem of high storage are resolved.

Business challenges are met through meaningful analysis through possible combinations of these attributes. This was done by ranking design's databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. We used the proposed SAk-means to classify the original designs and assign new designs from different combinations to a set of predefined classes of the old designs on the basis of the available training designs. This helps in clustering of heterogeneous designs into a number of more homogeneous sub-designs. New designs either emanated from a combination or modification of the highest patronized designs. With this, the customers would maximally allocate their resources in the right channel with time [1]. Hence, the challenge of not getting the proper designs invoke is addressed by continuous dynamic updating of the dataset of new creative designs as it unfolds from the system. In this study, customers are not to design their own clothes or designs as opposed to the work of Mok et al. [29], rather, the proposed model ensures automatic creation of new designs from the combinations of frequently chosen designs, such that the new designs would automatically replace the old ones.

This paper also builds on the fact that fashion design could not only be based on the style but commercial success implications. Jacobs et al. [17] states that fashion companies need to maintain balance between artistic and economic considerations. In this perspective, this paper analyses the attributes that would account for more and better business gains for designers as it helps designers to know the type of designs usually ordered for, by customers with the design code, and which combinations have the highest patronage. Our findings show that in order to attract more customers to patronize designers, the style, fabric, color combinations are considered important as these revealed the level of projective creativity in fashion design. In addition, it also projects the designers as a creative entrepreneur, balancing business and the art.

The rest of this paper is organized as follows: Section 2 briefly presents the literature overview of data mining and fashion world. Section 3 considers market incentives for fashion designers using *k*-means. In Section 4, implementation and result are presented, while section 5 discusses the results. Section 6 concludes the paper and provides pointers for future work.

2. Overview of data mining

Data mining techniques involve decision tree, ensembles, linear regression, naïve Bayes, *k*-means, neural networks, logistic regression, fisher's linear discriminant, Support Vector Machines (SVM) and quadratic classifiers [18,22,35,30]. These are classification algorithms used for extracting models to predict categorical class labels; and are often much more accurate than human crafted rules since it is data driven [20,24,23].

A multiclass Support Vector Machine (SVM) with classifier based on a type-2 fuzzy C-means algorithm was used for the recognition of control chart patterns due to its generalization capability [40]. In another perspective, naïve Bayes classifier was used to develop a prediction model of mutagenicity [34]. Analysis shows that the prediction performance of naïve Bayes classifier yielded average overall prediction accuracies [41]. However, there was a problem of space complexity and risk assessment emanated from its lack of practical evaluation [9].

Recently, data mining techniques have been applied to creative industries, namely, entertainment, advertising, architecture, film, fine art, software, data science, music, design and fashion, to mention but a few [32]. One palpable limitation of big data in fashion design is the difficulty encountered in accurate access and uncertainty during classification processes. To address the problem, ensemble methods were proposed [19]. Ensemble has been proven to be highly effective but require repeated resampling of the training data, making them inappropriate in a data mining context. Due to imbalanced in datasets, which is based on oversampling, under sampling or cost-sensitive classification, Bartosz et al. [5] introduced an effective ensemble of cost-sensitive decision trees for imbalanced classification.

Due to the failure encountered in the traditional Linear Regression Classification (LRC) method when the number of data in the training set is greater than their dimensions, Koç and Barkana [22] proposed a new form of LRC, which was proven to be balanced even with low-dimensional excessive number of data. In addition, another work proposed a global linear regression coefficient (GLRC) classifier that uses the test sample vector and whole class subspaces to calculate the GLRC rather than the use of test sample and the class subspace to calculate the distance, which is used for classification [12]. Very recently, *k*-means was introduced to fashion design with a dynamic reduction through homogeneous clusters (dRHC) [31]. Therefore, data mining techniques are evolving from modifications of the previous approaches with time.

In this study, constant observations are needed with automatic upgrade of new fashion design. Therefore, *k-means* was used to perform the clustering because it represents an unsupervised method, which permits the use of real numbers in the values of the vectors [10]. Though, *k-means* is a well-known clustering algorithm that is frequently used in clustering of patterns, it was employed in this study because more recent studies have shown its usefulness in exploratory stage where information on initial knowledge of what obtains in the fashion design environment is necessary [8].

2.1. State of the art: mining fashion design environment

Industrial design is the professional service of creating and developing concepts and specifications that optimize the function, value, appearance of products and systems for the mutual benefit of both users and manufactures. The roles of industrial designers are to create and execute design solutions for problem of form, usability, ergonomics, marketing, brand development and sales. Fashion design merges sketching and general visual skills with the power of imaginations. The interest is in knowledge of materials creative designs with lay outs, ordering supervising and budgeting [4].

Catalogue, a systematic lists of names, books/pictures/designs, is a means by which designers keeps records. It is a list of the contents of a design library arranged according to various styles serving as inventory to the designs' contents. It involves collecting, abstracting, and coding of printed written information for future purposes with instruction and references. Good decisions on what design to document, invoking season, styles preferred and process surrounding cataloguing are important. Therefore, targeting market interest involves segmental solutions. It encourages direct marketing and provides focus to all design marketing activities. Data mining can be used to improve targeting by selecting which people to contact, figuring out what design, when to offer and to whom it should be offered.

In today's competitive business environment, a typical fashion catalogue contains exploding dataset as shown in Fig. 1. This contains very large image data of designs that requires classifications. For proper guidance, categorization of the designs, fabrics, brand and color combination are needed for better returns in marketing, investments and in order to enhance business management in strategizing and delving into other new designs [16].

3. The proposed incentive mechanism to target market in a fashion design

A typical fashion business environment is considered. The design consideration requires that a fashion designer should normally have automated catalogue. Customers' classifications were based on three classes of objects: original, combine and new designs. There are also the ready or initial designs with the designers, called original in this case. In combined design, two or more designs from the catalogue can also be coupled to create/obtain new design. A customer may also come up with his own entirely new design.

3.1. Data collection

Different fashion dresses were collected from Jumia Nigeria. The dataset contains 1000 customers' records. The data comprises of customer ID, name, gender, preferred color, fabric, style, brand, and code. Due to some missing attributes from the 1000 data, 938 were used. The dataset was randomly categorized using six attributes: color, style, brand, fabric type, size and price. The first four are classified as major while size and price were considered as minor attributes for effective selection and combination due to their low impact on design selection and combination.

3.2. Fashion design incentives procedure

We classified the customers under three categories in Fig. 2 as users X, Y and Z using the following characteristics:

3.2.1. USER X

Customers in this category chooses dresses from the fashion designer initial catalogue, negotiate with the fashion designer and choose from his existing database. The dresses are represented with a name called 'original design' which symbolizes the fashion designers initial designs. In Fig. 2, a mutual relationship is established between this user and the designer.

3.2.2. USER Y

Customers in this category choose from two or more different original designs to make up entirely new design. They either choose original top with different original design of skirt and or combine part of the designs. Any adjustment or addition to the original design is considered as a combined design by the designer because a new design product will emerge from the existing designs.

3.2.3. USER Z

Customers in this category brings in entirely different designs that the fashion designer does not have in the catalogue.

Fig. 3 shows a fashion design business scenario where three set of customers are identified as customers X, Y, and Z. A Designer-Analyst-Customer relationship is presented where the analyst stands as the middleman between the designer and the customer. In some cases, the analyst could also be the designer. An analyst, in this case, is considered as the vendor who sells dresses and to whom payments are made. The analyst has an automated documentation of his stock by categories of designers, styles, color, brand, size; and as expected, prices are given by these different cadres. A Customer orders for dress, whether original, combined or new. He pays for his dress and receives the dress from the ana-

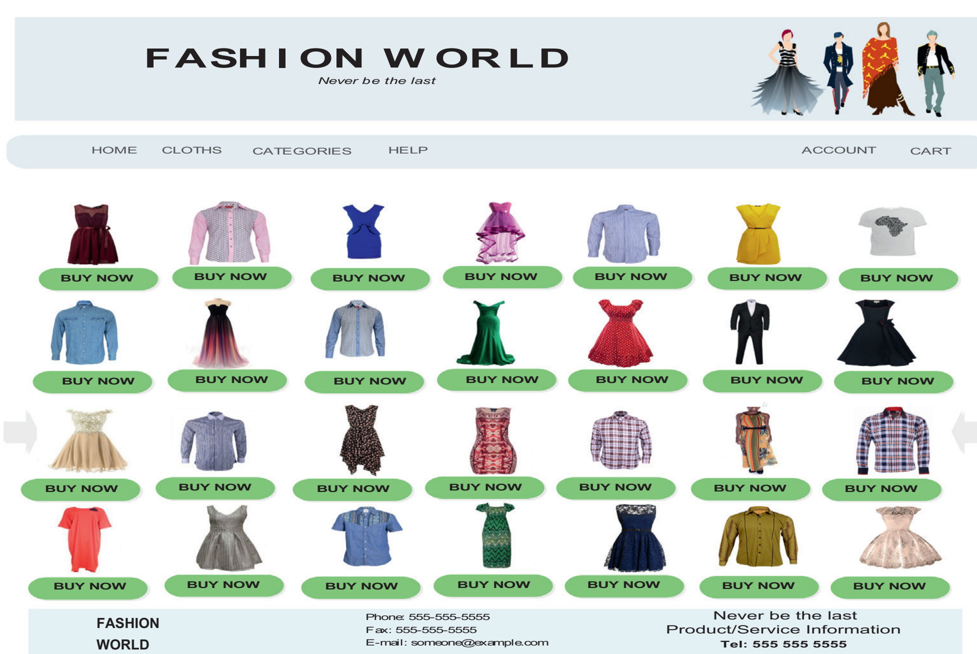


Fig. 1. A typical fashion catalogue.

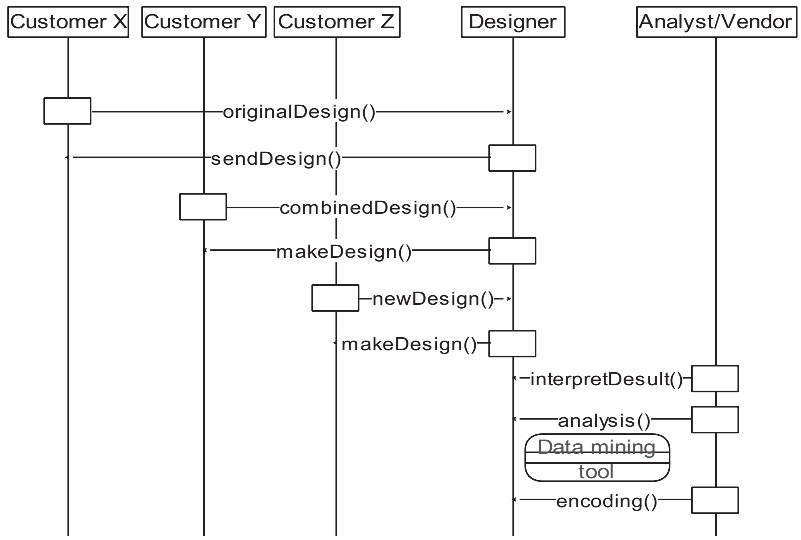


Fig. 2. A sequence of marketing strategy in fashion design.

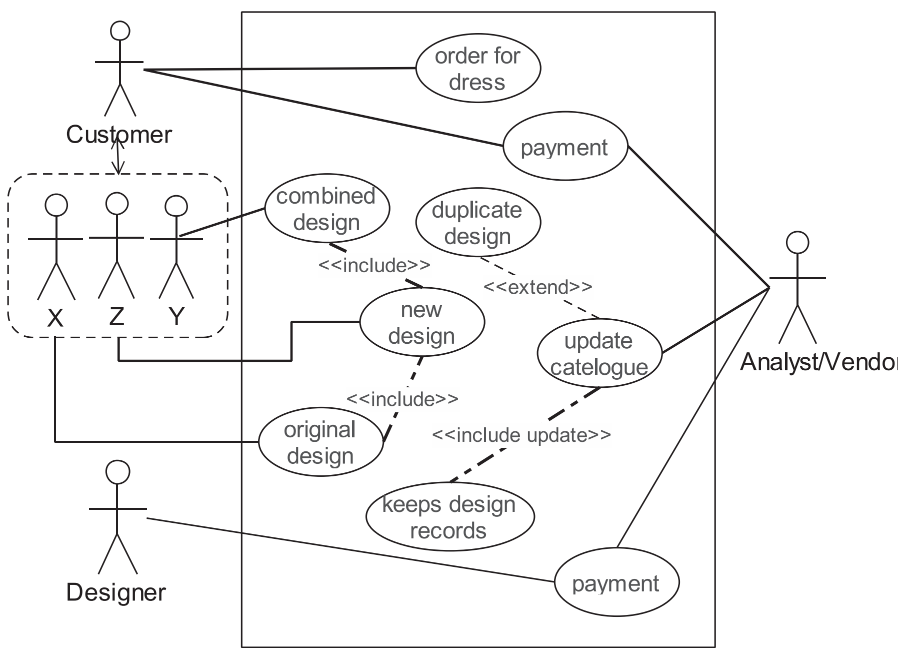


Fig. 3. A fashion design business use case.

lyst. The analyst on the other hand, receives the order, payment and keeps record of sales. The designers, having supplied the dresses, received payment from the analyst and update his record.

3.3. Cataloguing of homogeneous clusters using dynamic K-means

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem, which work incredibly well in practice. *K-means* clustering is a method of vector quantization, originally employed in signal processing, and very popular for cluster analysis in data mining. In this paper, *k-means* is modified by introducing a novel self-adaptive approach for classification of the identified six fashion attributes. If we let x_{ip} be the ith discrete point of a dress P, and y_{jp} be the ith discrete point of dress Q, and $i = 1, \ldots, k$, the classification follows that of Euclidian distance (D) given as

$$D(i,j) = \sqrt{|x_{i1} - y_{j1}|^2 + |x_{i2} - y_{j2}|^2 + \cdots + |x_{ip} - y_{jp}|^2}. \tag{1}$$

These points represent the same features as measured and recorded from both P and Q dresses. The set D of Euclidean distances between dress x and all successive dress $y \in Y$ is calculated using Eq. (1). From the distances in the set D, the smallest k distances are selected [11,43,42]. The verified designs are classified into three: original, combined and new designs by the classifier. The final classification are established on the basis of a voting score so as to reduce the noise, which depends on the number n of neighbors, given as

$$n = \left[\left(\binom{c}{2} + (c.d) \right)^{1/2} \right], \tag{2}$$

belonging to any of the three classes [32,3]. The number of data is a function of the cluster number and the mean vector. For example, if the number of data is less than the number of clusters then we assign each data as the mean vector of the cluster. Therefore, each mean vector will have a cluster number. However, if the number of data is bigger than the number of clusters, for each data, we calculate the distance to all mean vector and obtain the minimum distance. Thus, the data is said to belong to the cluster that has minimum distance from this data [19,2,31].

Since the location of the mean vector cannot be predicted, the mean vector is adjusted based on the current updated data. Thus, we assign all the data to the new mean vector and update the mean vector using a vector update law given as

$$VP(x_1,x_2,x_3,\ldots,x_k) = \left(\frac{\sum_{i=1}^{k} x1st_i}{k}, \frac{\sum_{i=1}^{k} x2nd_i}{k}, \frac{\sum_{i=1}^{k} x3rd_i}{k}, \ldots, \frac{\sum_{i=1}^{k} xnth_i}{k}\right). \quad (3)$$

In Eq. (3), x represents the set of data for clustering. The clusters were then evaluated using the Sum of Squared Error

$$SSE = \sum_{i=1}^{K}\sum_{x=C_i} dist^2(m_i, x). \quad (4)$$

For each point, the error accounts for the distance to the nearest cluster [19,31]. In order to get the SSE, the error is squared and sum together in Eq. (4). Let the dimension of the matrix X be $l(c \times d)$ as given by Eq. (2), algorithm 1 describes the generation of new x_i.

Algorithm 1 (*Dynamic K-means for Fashion Design*).

Input: (X, Y)
Output: Y_i
1: **for** i = 1 **to** k do
2: compute distance $D(X_i, n)$
3: end for
4: Compute set I containing indices for the k smallest distances $D(X_i, n)$
5: **Return** majority label for $\{Y_i \text{ where } i \in I\}$
6. Newdesign ← K-means(X, Y)
7. For x_i = 1 to k do
8. Update (X)
8. End for

Algorithm 1 classifies a new design x by searching in the training set for the k nearest designs called neighbours to x according to the distance metric in Eq. (1). Then, x is assigned to the most patronised designs (common class) determined by calculating the percentage patronage on designs of the retrieved k nearest neighbours. Ties are settled using single nearest neighbour of the color, style, fabric type, and brand noting, the error rates. We place k points, that is, the initial group centroids, into the space represented by the objects that are being clustered and assign each object to the group that has the closest centroid. When all objects have been assigned, the position of k centroids is re-calculated. This procedure is repeated until the centroids no longer change or move. By this, clusters are constructed until all of them contain items that belong to only a particular class. This resulted in the separation of the objects into groups from which the metric to be minimized is calculated. The result is then combined severally and a dynamic upgrade forming a new dataset of designs is obtained.

4. Implementation and results

The dataset was categorized into three classes: original, new and combined designs. The proposed self-adaptive k-means was developed in Java with every chosen design coded with a unique identifier. The coding was done to differentiate between the three designs with x, y, and z for each design of original, combined and new, respectively using ZeroR as the classifier rule. Euclidean distance was used for the determination of centroid on which the clustering was based. Each object was based on minimum Euclidean distance. Out of the six attributes of the dataset, four (color, style, size and price) were found to be mainly constraints for categorizing the dataset, while brand and fabric could not be distinctively classified. Fig. 4 shows the identified fashion attributes where a compilation of the six attributes are presented in their classes showing how size and price changes.

Fig. 5 shows how multiple original styles were combined and new styles emerged. Dresses were identified by styles where two or three styles were combined to form other new styles. The newly created styles were made to update design catalogue automatically. We classify the original designs and assign new designs from different combinations to a set of predefined classes of the combined designs on the basis of the available training designs. From Fig. 5, new designs either originated from a combination or modification of predefined designs.

4.1. Experimental evaluation of attributes

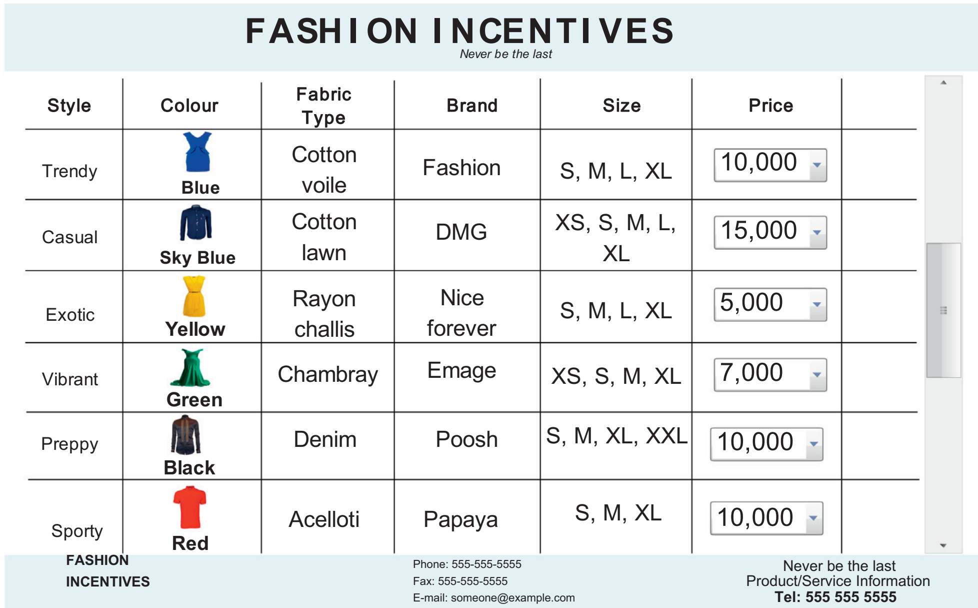
This section presents experimental results with four of the six attributes to demonstrate the effect on customers' choices. In this section, we illustrate the behavior of our SAk-*means* algorithm on some of the fashion design attributes, which is seen as market strategy to foster improved economy on the designers, vendor and the customers.

4.1.1. Color and style as marketing strategies

Color is always a baseline in fashion attraction, which could foster customer's interest and fashion design choice making. We therefore compared how color could be a main factor on price, size and style of designs. The goal of these experiments is to study the effects of color and color combinations on customers' choices of designs and the impact on sales. Fig. 6 shows a classifier result when color and prices effect were examined.

Fig. 6 shows what happens when color was compared to price. Clearly, the increase in price influences the decision of color by customers. In Fig. 6, it is shown that color, especially the scarce ones, is a factor that determines what price a particular dress is sold. To this end, customers choose dresses with lesser prices even though it may not be their color of choice. Hence, more customers chose red, multi-color and black because they are less costly and probably as a result of easy combination with other colors. This is depicted in the bottom left corner of Fig. 6 where the clusters form around lower cost colors.

Fig. 7 shows the relationship between the style, color and size. In Fig. 7, a distinguished grouping of original, combine and new designs are shown when style and color were considered. Fig. 7a shows that customers' decisions of style clearly influence their color of choice. The clumsiness of the clusters at the left hand side of the graph is an indication that customers prefers to choose style over color. With respect to the impact of size on color and the decision of design chosen by the customer, Fig. 7b illustrates customers' choice with respect to size and color. The Figure shows that choice of color does not depend on size. This implies that no matter how attractive some colors are, customers would not select bigger or smaller sizes than the actual choice of size. This is reflected in the disparity of customers' choices as the clusters are gathered around some particular sizes regardless of the colors. We can therefore conclude that color and style are functions that determined customer choices of dress selection and thereby determines patronage.



Style	Colour	Fabric Type	Brand	Size	Price
Trendy	Blue	Cotton voile	Fashion	S, M, L, XL	10,000
Casual	Sky Blue	Cotton lawn	DMG	XS, S, M, L, XL	15,000
Exotic	Yellow	Rayon challis	Nice forever	S, M, L, XL	5,000
Vibrant	Green	Chambray	Emage	XS, S, M, XL	7,000
Preppy	Black	Denim	Poosh	S, M, XL, XXL	10,000
Sporty	Red	Acelloti	Papaya	S, M, XL	10,000

Fig. 4. Mining fashion attributes.

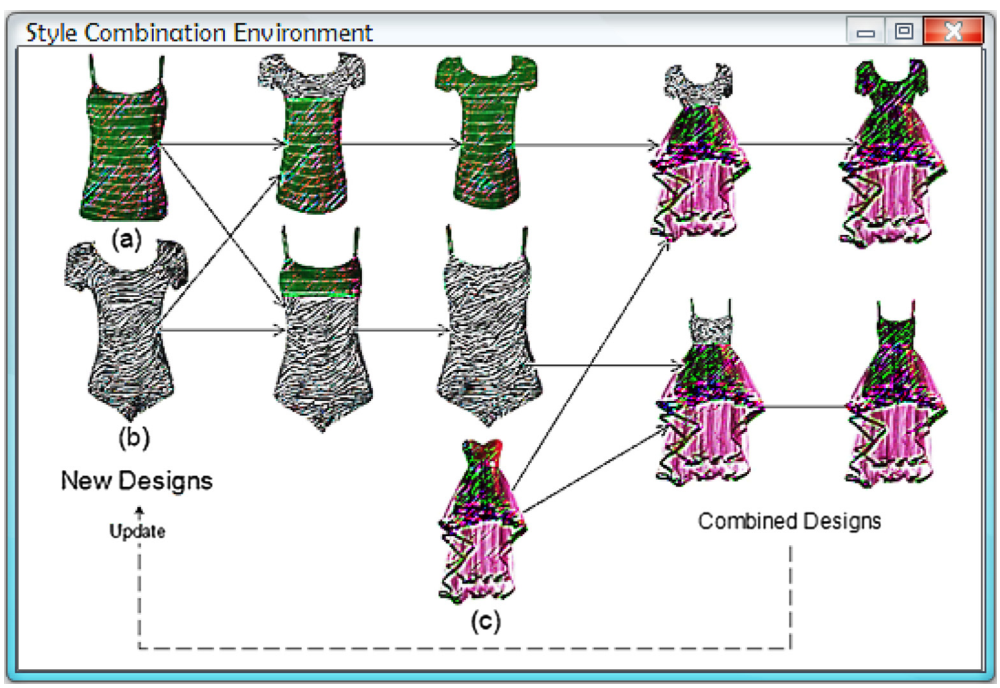


Fig. 5. Styles' combination.

Fig. 8a shows the impact of size on price for new, combined and original designs. It can be observed that increase in prices of dresses increases customers' interest towards a choice of combined designs. Observations from Fig. 8a and b also show that combined designs has the highest number of customers. Fig. 9 shows the effect of color combinations on fashion. For the color strength, 649 instances were used, out of which 212 were black dress choices. This implies that black color would always invite highest patronage in any fashion market. Multi-colors, white and blue are other colors that show closely related market attractions. However, when it comes to making dresses to target customers' choice of colors such as pale-pink, blue-grey, black-orange, etc. can be done away with as the market strength shows negligible sales.

4.1.2. Validation of classifier

To validate our classifier, we studied the root mean square error, the Ignore Class Unknown Instance (ICUI) and the Receiver Operator Characteristics (ROC) area curve measures. In Table 1, we illustrate the correlation of colors with price and size; and the results clearly show that the correlations between the variables are not significant. Due to diverse input variations, correlation for style, fabric type and brand could not be examined. But when color and size, color and price were examined, negative relationships were observed between the variables. While we could not establish a correlation for brand, style and fabric type, the correlation coefficient of color and price is −0.0937 and that of color and size is −0.0467, which show that the degree of linear dependence

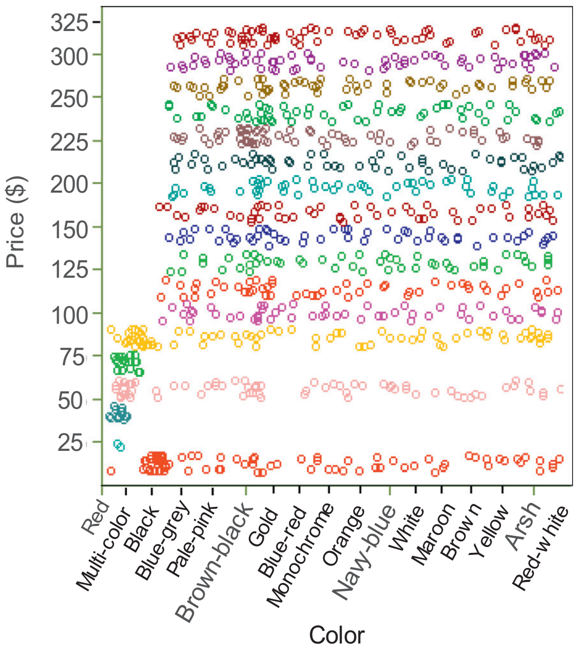


Fig. 6. Effect of customers' price decision on choice of color.

between the variables are in opposite directions. This reveals that as color and size change, price increases or decreases. Therefore, the relationship between color and price, size and price lie on opposite directions as shown in Fig. 6 and this is clearly seen in Fig. 7b.

Table 1 shows the error values on color, size, and price. The relative absolute error (RAE) and root relative squared error (RRSE) is very high for the three attributes. The root mean squared ratio (RMSR) measures the average magnitude of the errors. The RMSR gives a relatively high weight to expensive errors, which makes the RMSE most useful because of vast errors that are particularly undesirable. The Mean absolute error (MAE) delineates the direct score and this implies that individual differences are weighted similarly on the average. Out of 1000 instances, 62 were ignored, which implies 938 instances were classified in all.

Table 2 shows the range, under a Receiver Operator Characteristics (ROC) curve measures, of the general capacity of the test to distinguish the effect of style, size, and color on designs' sales. The result in Table 2 describes the effect of the three attributes on sales. The ROC region likewise test for the level of precision among the three attributes, which shows high performance in relation to sales.

Next, we evaluate the accuracy of the clustering decision on fashion design based on new, combine and original design by comparing the results achieved by the SA*k*-means, SVM, and Bayes methods. Using the optimal parameter values obtained on the training datasets, the accuracy was examined using independent test datasets. The experimental results on datasets is given in Table 3. From the results in Table 3, it can be observed that SA*k*-means applied on the dataset improves the accuracy of the clustering decision-making. Except for the combined design where SAk-means and Bayes method have no significance performance measure. Regarding the SVM method, results on combined and original designs showed no significance; whereas, better results on clustering decision are obtained from Bayes and particularly, the proposed SA*k*-means methods.

Finally, the impact of classification time was examined by comparing the time used by each method based on the fashion design dataset. Table 3 presents the total running time based on learning and testing time using the original, combined and new designs. It was observed that the running time of the classifiers were significantly reduced in the experiments. In general, SA*k*-means is faster than SVM and Bayes, due to space complexity of SVM method which is proportional to $n * Q^2$ and Bayes method, which is $O(pqr)$, where p is the number of features, q is values for each feature, and r is alternative values for the class.

5. Discussion

Analysis of the attributes that would account for more and better business gains for designers was carried out in this paper. Fig. 7 (a) shows the relationship between style and color. It can be clearly seen that 54% of customers preferred combined designs because they have full customization control over the design, while 31% of the customers shows preference for new designs by the fashion designer, because new designs are produced based on the acquired data from the original designs. However, 15% of the customers preferred the original called old designs. This could be attributed to the obsoleteness of the designs. From Fig. 7, it was found that more customers' preferred combined designs than other two designs.

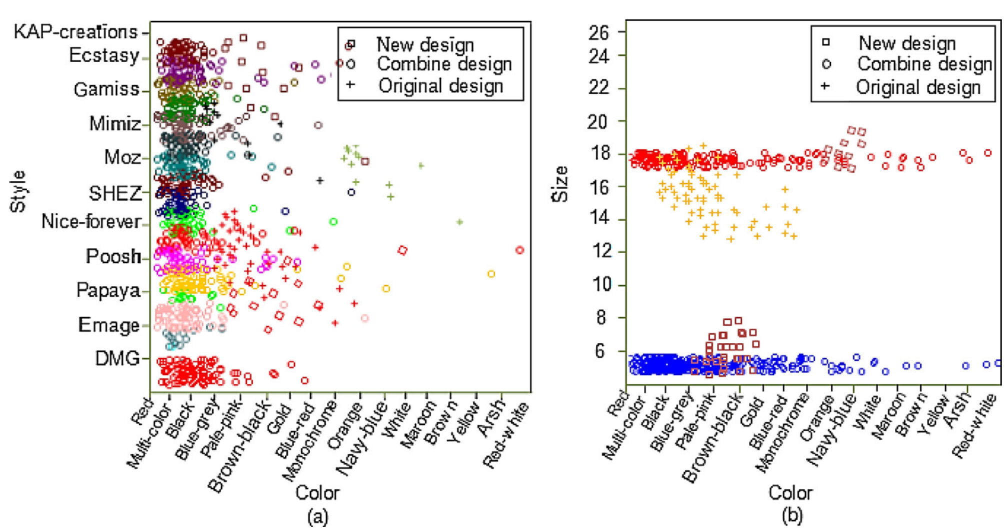


Fig. 7. Effect of (a) customers' style decision on color and (b) customers designs (new, combine and original) on size and color.

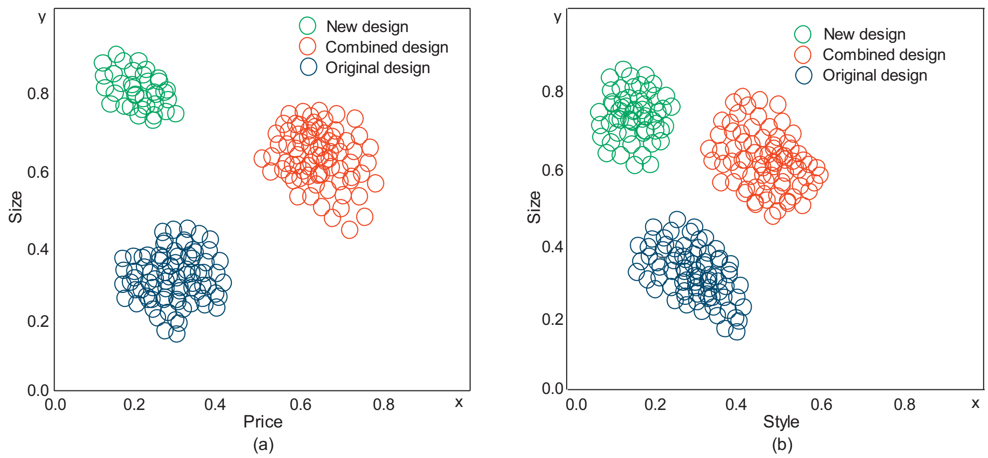


Fig. 8. Self-adaptive clustering of designs. (a) Size against price and (b) size against Style.

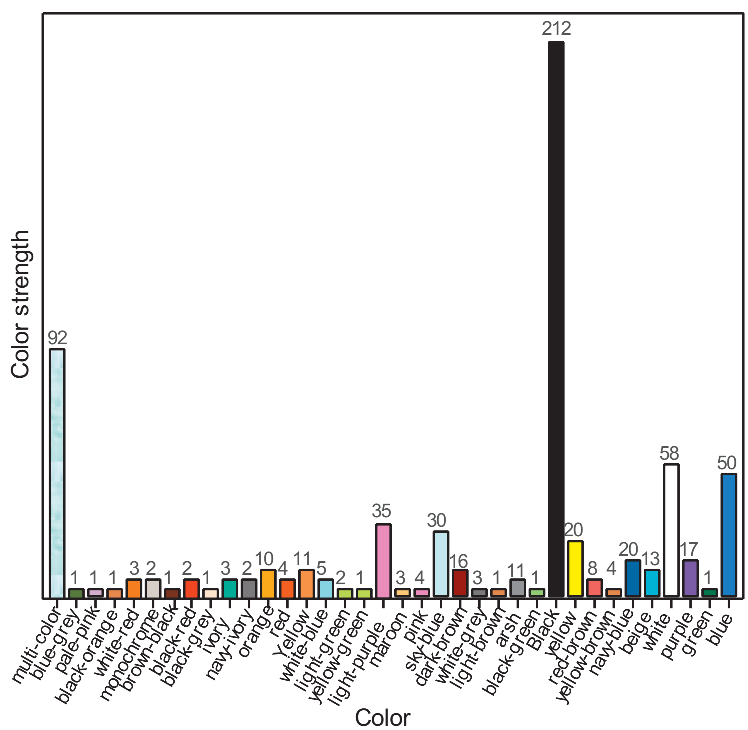


Fig. 9. Customers' decision on choice of color.

Table 1
Cross validation of classifier output for three attributes.

Attribute	MAE	RMSR	RAE (%)	RRSE (%)	ICUI
Color	0.0213	0.1029	96.87	100	8
Size	0.0528	0.1619	96.56	100	0
Price	3345.8867	5953.124	89.07	100	0

Table 2
Detailed accuracy by attribute under the ROC curve.

Attribute	TP Rate	FP Rate	Precision	Recall	F-measure	ROC area
Style	0.488	0.218	0.691	0.488	0.129	0.854
Size	0.647	0.375	0.633	0.647	0.188	0.872
Color	0.684	0.428	0.615	0.684	0.085	0.736

Table 3
Accuracy and running time.

Design	Accuracy (%)			Running time (s)		
	SA*k*-means	SVM	Bayes	SA*k*-means (s)	SVM (s)	Bayes (s)
New	0.87	0.76	0.81	1.2	8.7	5.77
Combined	0.90	0.84	0.90	1.37	10.9	6.0
Original	0.97	0.84	0.92	0.3	5.6	3.14

This result is valuable for the fashion designer, enabling him to target on producing more of the combined design rather than the other designs.

Fig. 7a also shows that style is a major factor on any color choices. Consequently, design style should be given a major priority during fabric consideration. This would address the challenge of not getting the proper designs invoke by continuous dynamic updating of the dataset of new creative designs as it unfolds from the system. From Fig. 8, a fashion designer could deduce what color and color combinations would attract customers' interest. Black, multi colors, blue and white are seen in this case to be of highest interest. Fig. 8 shows that black color dominates the fashion world market, hence design should not only be based on the style but commercial success implications.

6. Conclusions

A mechanism for commercial success in fashion design has been proposed in this paper. Here, economy was considered as a major liberation in the fashion world by analyzing six fashion attributes, which includes style, color, fabric, brand, price and size. The dataset of 1000 customers' records used were categorized as original, combined and new designs using *k-means* algorithm to extract common attributes that would foster better business bargains. The dataset of the designs was first ranked for hidden patterns. Then, SA*k-means* was used to classify the original designs and assign new designs from different combinations to a set of predefined classes of the original designs on the basis of the available training designs. This helps in clustering of heterogeneous designs into a number of more homogeneous sub-designs categorized with the six attributes. New designs emanated either from a combination or modification of the highest chosen designs.

The SA*k-means* is a system of automatic creation of new designs as a result of the combinations of frequently chosen designs. The new designs automatically replaced the old original designs. In addition, this study also builds on the fact that fashion companies need to maintain balance between artistic and economic considerations. From the study, it was found that in order to get more customers patronizing designers, the style and color combinations are considered important as these revealed the level of projective creativity in fashion design. Hence, the study revealed that more customers would likely patronize fashion designers primarily because of the style and color combinations as creativity is projected through these combinations.

During the study, the computations were a little limited with the number of data analyzed. When the data is increased, the result is expected to be more precise. However, immediate next research steps should be to explore other criteria that would foster fashion business development.

Acknowledgment

The authors acknowledge the anonymous reviewers for their constructive comments, which has tremendously improved the quality of the paper. We also acknowledge Dr. U. E. Vincent, Lancaster University, United Kingdom for reading through and editing the manuscript and providing valuable suggestions.

References

[1] M.A. Abebe, A. Angriawan, Organizational and competitive influences of exploration and exploitation activities in small firms, J. Bus. Res. 67 (2014) 339–345.
[2] D.A. Adeniyi, Z. Wei, Y. Yongquan, Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, Appl. Comput. Inform. 12 (1) (2016) 90–108.
[3] V.M. Aradhya, M.S. Pavithra, A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video, Appl. Comput. Inform. 12 (2) (2016) 109–116.
[4] B. Borukhovich, Fashion design: the work of art that is still unrecognized in the United States, Wake Forest Intell. Prop. LJ 9 (2009) 155–238.
[5] K. Bartosz, W. Michał, S. Gerald, Cost-sensitive decision tree ensembles for effective imbalanced classification, Appl. Soft Comput. 14 (2014) 554–562.
[6] B. Bury, M. Tiggemann, A. Slater, Disclaimer labels on fashion magazine advertisements: impact on visual attention and relationship with body dissatisfaction, Body Image 16 (2016) 1–9.
[7] Y. Cho, T. Komatsu, S. Inui, M. Takatera, Y. Shimizu, H. Park, Individual pattern making using computerized draping method for clothing, Text. Res. J. 76 (2006) 646–654.
[8] M. Capó, A. Pérez, J.A. Lozano, An efficient approximation to the K-means clustering for massive data, Knowl.-Based Syst. 117 (2017) 56–69.
[9] E.O. Cannon, A. Amini, A. Bender, M.J. Sternberg, S.H. Muggleton, R.C. Glen, J.B. Mitchell, Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds, J. Comput. Aided Mol. Des. 21 (5) (2007) 269–280.
[10] Chévez, P., Barbero, D., Martini, I., Discoli, C., 2017. Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina. Sustainable Cities and Society.
[11] I. Davidson, S.S. Ravi, M. Ester, Efficient incremental constrained clustering, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007, pp. 240–249.
[12] Q. Feng, X. Zhu, J.-S. Pan, Global linear regression coefficient classifier for recognition, Optik-Int. J. Light Electron Opt. 126 (2015) 3234–3239.
[13] S. Guercini, A. Runfola, How western marketers respond to the new middle class in emerging market cities: the case of Italian fashion marketers, Int. Bus. Rev. 25 (2016) 691–702.
[14] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (Mar) (2003) 1157–1182.
[15] D. Heuer, M. Brettel, J. Kemper, Brand competition in fashion e-commerce, Electron. Commer. Res. Appl. 14 (2015) 514–522.
[16] K. Hjort, B. Lantz, The impact of returns policies on profitability: a fashion e-commerce case, J. Bus. Res. 69 (11) (2016) 4980–4985.
[17] S. Jacobs, B. Cambré, M. Huysentruyt, A. Schramme, Unraveling Belgian fashion designers' high perceived success: a set-theoretic approach, J. Bus. Res. 69 (2016) 1407–1411.
[18] K. Kameshwaran, K. Malarvizhi, Survey on clustering techniques in data mining, Int. J. Comput. Sci. Inform. Technol. 5 (2014) 2272–2276.
[19] S. Kanj, F. Abdallah, T. Denœux, K. Tout, Editing training data for multi-label classification with the k-nearest neighbor rule, Pattern Anal. Appl. 19 (2016) 145–161.
[20] H. Kim, A. Stumpf, W. Kim, Analysis of an energy efficient building design through data mining approach, Automat. Constr. 20 (2011) 37–43.
[21] S. Kim, C.K. Park, Parametric body model generation for garment drape simulation, Fibers Polym. 5 (2004) 12–18.
[22] M. Koç, A. Barkana, Application of Linear Regression Classification to low-dimensional datasets, Neurocomputing 131 (2014) 331–335.
[23] B. Krawczyk, M. Woźniak, G. Schaefer, Cost-sensitive decision tree ensembles for effective imbalanced classification, Appl. Soft Comput. 14 (2014) 554–562.
[24] S.H. Liao, P.H. Chu, P.Y. Hsiao, Data mining techniques and applications – a decade review from 2000 to 2011, Expert Syst. Appl. 39 (2012) 11303–11311.
[25] J. Magidson, J. Vermunt, Latent class models for clustering: a comparison with K-means, Can. J. Market. Res. 20 (1) (2002) 36–43.
[26] J.G. Martínez, Selling avant-garde: how Antwerp became a fashion capital (1990–2002), Urban Stud. 44 (2007) 2449–2464.
[27] P. Matthyssens, K. Vandenbempt, L. Berghman, Value innovation in business markets: breaking the industry recipe, Ind. Mark. Manage. 35 (2006) 751–761.

[28] Y. Meng, P. Mok, X. Jin, Computer aided clothing pattern design with 3D editing and pattern alteration, Comput. Aided Des. 44 (2012) 721–734.
[29] P. Mok, J. Xu, X. Wang, J. Fan, Y. Kwok, J.H. Xin, An IGA-based design support system for realistic and practical fashion designs, Comput. Aided Des. 45 (2013) 1442–1458.
[30] S.S. Nikam, A comparative study of classification techniques in data mining algorithms, Oriental J. Comput. Sci. Technol. 8 (2015) 13–19.
[31] S. Ougiaroglou, G. Evangelidis, RHC: a non-parametric cluster-based data reduction for efficient K-means classification, Pattern Anal. Appl. 19 (2016) 93–109.
[32] P. Porwik, R. Doroz, T. Orczyk, The *K*-means classifier and self-adaptive Hotelling data reduction technique in handwritten signatures recognition, Pattern Anal. Appl. 18 (2015) 983–1001.
[33] F. Roubelat, J. Brassett, M. Mcallum, J. Hoffmann, D. Kera, Probing ephemeral futures: scenarios as fashion design, Futures 74 (2015) 27–36.
[34] I. Sushko, S. Novotarskyi, R. Körner, A.K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K.R. Müller, L. Xi, Applicability domains for classification problems: benchmarking of distance to models for Ames mutagenicity set, J. Chem. Inf. Model. 50 (12) (2010) 2094–2111.
[35] V. Vagizova, K. Lurie, I.B. Ivasiv, Clustering of Russian banks: business models of interaction of the banking sector and the real economy, Probl. Perspect. Manage. 12 (2014) 83–93.
[36] C.C. Wang, Parameterization and parametric design of mannequins, Comput. Aided Des. 37 (2005) 83–98.
[37] Y. Wu, Development of an Intelligent Patternmaking System for Skirt Design Doctoral dissertation, The Hong Kong Polytechnic University, 2012.
[38] C.L. Yang, R.H. Huang, Key success factors for online auctions: analysis of auctions of fashion clothing, Expert Syst. Appl. 38 (2011) 7774–7783.
[39] Y. Yu, C.L. Hui, T.M. Choi, An empirical study of intelligent expert systems on forecasting of fashion color trend, Expert Syst. Appl. 39 (2012) 4383–4389.
[40] A. Khormali, J. Addeh, A novel approach for recognition of control chart patterns: Type-2 fuzzy clustering optimized support vector machine, ISA Trans. 63 (2016) 256–264.
[41] H. Zhang, Y.L. Kang, Y.Y. Zhu, K.X. Zhao, J.Y. Liang, L. Ding, T.G. Zhang, J. Zhang, Novel naïve Bayes classification models for predicting the chemical Ames mutagenicity, Toxicol. In Vitro 41 (2017) 56–63.
[42] N. Zhou, W.M. Gifford, J. Yan, H. Li, End-to-end solution with clustering method for attrition analysis, in: Services Computing (SCC), 2016 IEEE International Conference on, IEEE, 2016, pp. 363–370.
[43] Yan, J., Ren, Z., Zha, H., Chu, S., 2016. A constrained clustering based approach for matching a collection of feature sets. arXiv:1606.03731.
[44] O. Younis, S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, IEEE Trans. Mob. Comput. 3 (4) (2004) 366–379.