Bandwidth Allocation Mechanism based on Users' Web Usage Patterns for Campus Networks

Article *in* International Journal of Communication Networks and Information Security · August 2018

3 authors:

Rolysent Paredes
Misamis University
10 PUBLICATIONS **9** CITATIONS

SEE PROFILE

Ariel M. Sison
Emilio Aguinaldo College
91 PUBLICATIONS **114** CITATIONS

SEE PROFILE

Ruji Medina
Technological Institute of the Philippines
125 PUBLICATIONS **149** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

DIT Dissertation View project

Association Rule Mining Project View project

Bandwidth Allocation Mechanism based on Users' Web Usage Patterns for Campus Networks

Rolysent K. Paredes[1], Ariel M. Sison[2] and Ruji P. Medina[3]

[13]Graduate Programs, Technological Institute of the Philippines, Quezon City, Philippines
[2]School of Computer Studies, Emilio Aguinaldo College, Manila, Philippines

Abstract: Managing the bandwidth in campus networks becomes a challenge in recent years. The limited bandwidth resource and continuous growth of users make the IT managers think on the strategies concerning bandwidth allocation. This paper introduces a mechanism for allocating bandwidth based on the users' web usage patterns. The main purpose is to set a higher bandwidth to the users who are inclined to browsing educational websites compared to those who are not. In attaining this proposed technique, some stages need to be done. These are the preprocessing of the weblogs, class labeling of the dataset, computation of the feature subspaces, training for the development of the ANN for LDA/GSVD algorithm, visualization, and bandwidth allocation. The proposed method was applied to real weblogs from university's proxy servers. The results indicate that the proposed method is useful in classifying those users who used the internet in an educational way and those who are not. Thus, the developed ANN for LDA/GSVD algorithm outperformed the existing algorithm up to 50% which indicates that this approach is efficient. Further, based on the results, few users browsed educational contents. Through this mechanism, users will be encouraged to use the internet for educational purposes. Moreover, IT managers can make better plans to optimize the distribution of bandwidth.

Keywords: Bandwidth Allocation, Web Usage Patterns, Neural Network for LDA/GSVD, Linear Discriminant Analysis.

1. Introduction

One of the essential challenges in handling multi-service networks such as campus networks is to manage the bandwidth efficiently which has a significant part in various internet facilities [1]. Thus, assigning bandwidth in an equal and effective way is a main concern in networking due to its limitation [2] and because of that experts have thought of some ways to attain fairness goals in managing the bandwidth [1]. Also, bandwidth control is crucial in designing, deploying, and maintaining a network [3].

Due to the importance of bandwidth allocation, this study aims to provide a new mechanism for allocating bandwidth to network users by utilizing the weblogs that are stored in the proxy servers [4]. These weblogs comprise necessary data of the network users which can be utilized to discover user patterns through the use of data mining techniques such as classification which includes K-Nearest Neighbor (KNN) classifier, Artificial Neural Network (ANN), ID3, C4.5, Naive Bayes, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) [5-7].

To discover users' web usage patterns from the weblogs, LDA was used since it outperformed other classification models and algorithms and has been utilized widely in the previous years for dimensionality reduction, detection, and supervised learning [6, 8]. Further, LDA has the benefit of looking for projection vector that produces ideal discrimination among different collections of observations

[9]. However, it has a problem in dealing with unlabeled datasets [8, 10] and fails when there exists a singularity of matrix due to small sample size (SSS) problem [11-13]. Thus, this study aims as well to address the two issues of LDA before it is applied to the weblogs to identify patterns. Typically weblogs are unlabeled, and due to the dynamic data that composes it, SSS problem can occur.

There were several ways to manage the unlabeled dataset as mentioned in the literature; one is making LDA unsupervised by utilizing clustering algorithms. With that, Self-Organizing Map (SOM) was employed in creating the clusters or making of the class labels because it is an unsupervised learning method to determine dataset's patterns and excellently work with high dimensional data [14]. It is fast and robust since it is an example of an artificial neural network architecture [15, 16].

In dealing with SSS problem in LDA, there were several solutions proposed in the literature. One of them is Generalized Singular Value Decomposition (GSVD) [17] which is generally applied by various discriminant analysis approaches [18, 19]. It is also a typical method for computing the matrix singular problems in different mathematical solutions [18, 20]. Further, GSVD on LDA (LDA/GSVD) provides extraordinary recognition accuracy [21] that is why many researchers used and developed variance of it. However, GSVD suffers from computational cost [21-23] which can cause a longer time in classifying datasets when applied to LDA. Moreover, if there is a new instance for classification in the existing LDA/GSVD, the whole process of the algorithm will be repeated from the very start.

In solving the problems of the existing LDA/GSVD, the Artificial Neural Network (ANN) algorithm was developed. The ANN algorithm eliminates the mathematical computations and numerous iterations that are involved in the existing LDA/GSVD algorithm which compromise time complexity which makes it less efficient [21-23]. Further, with this method, learning can be done from the weblog and classification to each instance, whether new or previously part of the training or testing, will be faster because it will not go back to the start of the whole procedure. The use of ANN in developing the algorithm has the benefit of accuracy. Also, ANN was equipped with the uniqueness of concurrent processing, can learn and recall data relationships, and mapping of non-linear instances [24, 25].

Hence, the primary purpose of the study is to introduce a new mechanism for allocating bandwidth to the users which can be used mainly on the university's networks. This proposed bandwidth control approach can help the network administrator or those who manage the network in school in

identifying the users who educationally utilized the internet. This approach can also assist them in their decision and policy-making concerning the utilization of the internet. This new technique utilized the weblogs from the proxy server which was used for discovering the users' web usage patterns through the use of ANN for unsupervised LDA/GSVD. The idea is that users will be assigned with higher bandwidth if they are classified or inclined in accessing more on educational websites; otherwise, they will have a lower bandwidth. With this approach, it encourages the users in the campus to engage more in accessing educational contents, be focused in their studies especially the students, and the faculty and staff will be more active in research and development. For LDA to work efficiently in determining users who browsed more educational websites or not, SOM was used for class labeling, and ANN for LDA/GSVD was employed for faster classification and computation of the feature subspaces. These feature subspaces separate well the users according to their web usage patterns. The results of this study will be an excellent contribution to the continuous development in the area of computer networking.

A simulation of the proposed bandwidth allocation mechanism was presented in the study.

2. Literature Review

Allocating Bandwidth is concerned with a vital question on how much resource should be allotted to the network's user and flow, in such a way that there is a fair and efficient use of bandwidth [3]. [26] proposed an approach to raising the cellular network users' throughput by exploiting network and users' data gathered from their devices. Also, an algorithm which established on past data from telehealth service for forecasting and assigning the Virtual Machines' future bandwidth was proposed by [27]. Further, [1] developed a hybrid data mining scheme which utilized clustering and classification for the allocation of bandwidth in a priority-based manner. The main motivation is to distinguish, study and forecast students' behavioral patterns in a campus network and determine the primary aspects that influence the students in browsing the internet. Besides, [28] recommended that schools must a policy concerning the utilization of the internet. However, policies have no value if not properly communicated and employed. Moreover, [29] improved the bandwidth of the campus by implementing squidGuard in a proxy server which normally employed by most network managers.

In computer networks, web server normally saves the essential data every time a user requests a website resource and this data is called weblog. These weblogs can also be found in proxy servers and web browsers [4]. Thus, preprocessing is needed to remove unnecessary data or noise from the weblogs [4, 30]. However, identifying the individual user of each instance from the weblog is difficult [30]. Because of that, in this study, captive portal logs were used. Table 1 presents the essential attributes of the weblog.

The captive portal logs to identify the users associated with the weblogs. This captive portal log can be retrieved from a local database or some backend service. Its records are made when a network applies captive portal which traps internet's traffic so that only validated clients can access [31]. These captive portals are generally employed in public Wi-Fi

hotspots. It informs typically users about the internet access sponsor, probable limitations as well as the possible mode of payment, and the terms and conditions should be accepted by the users [32]. Further, it prompts the users to input their credentials, and traffic is dropped until a network user is authenticated [31].

Every time a user enters his/her username and password in the captive portal, the action is being recorded in the captive portal database. Usually, this authentication process is done through RADIUS. RADIUS accounting logs are stored in detailed which describe the users and the devices used. Further, those are necessary for recording the users who used the network, when; and for ascertaining security and accountability compliance [33]. Figure 1 shows the entity-relationship diagram for a RADIUS database.

Table 1. Essential attributes of the weblog

| Attribute	Description
Timestamp	Date and time when the user made a request to access certain URL or website
Client IP	IP address of the user's device
Client server methods	Method or model of request, can be GET, POST, or HEAD
Client server URL stream	Targeted default web page or domain that is accessed by the user.
Server IP	IP address of the web server
Server client status	Status code returned by server link
MIME	Standard way of classifying file types such as images, html, javascript, etc.
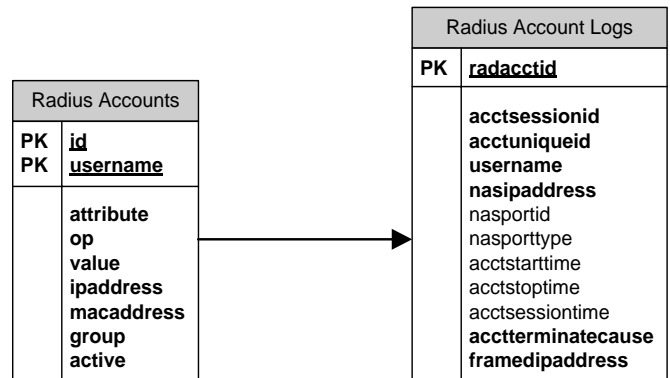
Figure 1. RADIUS database entity-relationship diagram

In identifying each website's category accessed by each user, a ShallaList file can be used. A ShallaList file is freely available and downloadable on the internet. It is frequently updated for the list of blacklisted websites [34, 35]. It comprises a pool of Uniform Resource Locators (URLs) and website domains utilized by web content filters to support web servers and browsers filter contents which are harmful or dangerous. Website domains are classified based on education, social, movies, music, dating, finance, etc. [34].

To discover patterns from the weblogs, the Linear Discriminant Analysis is one of the methods that can be employed. LDA is a classification model used in data mining which has been extensively employed in numerous applications [12, 23, 36-41]. It is a supervise learning [10, 42] which intends for ideal conversion directions by reducing

usual within-class scatter and make most of the average between-class scatter, i.e., the converted data samples are adjacent to their group centroids, and each group centroid is distant from the entire centroid. Explicitly, suppose matrix $X = [X_1, X_2, \cdots, X_k] \in R^{n \times m}$ which characterized the dataset, where frequency of samples m, dimension n, the number of groups or classes (k), a matrix $X_i \in R^{n \times m}{}_i$ which is the ith class with the index as N_i $(1 \leq i \leq k)$ [13]. LDA searches such a path $w \in R^n$, equation 1 describes it.

$$\max_{w} = \frac{\sum_{i=1}^{k} m_i \| c_i^T w - c^T w \|^2}{\sum_{i=1}^{k} \sum_{j \in N_i} \| x_j^T w - c_i^T w \|^2}, I \qquad (1)$$

LDA has other advantages such as (a) inexpensive application; (b) adapts in discriminating non-linear datasets; and (c) coherence to Bayesian classification [43, 44]. It surpasses PCA concerning classification performance [45]. However, LDA has a problem with unknown class labels [8, 10] and has an issue with SSS problem [11-13] which happens if the quantity of training vectors is less than the dimensions. Because of that, the calculation of eigenvalues and eigenvectors becomes unbearable [8, 12, 13, 18].

It is easy to collect unlabeled data, however, labeling it is time-consuming [10]. With that, experts suggested or established means to automatically create the class labels for LDA such as PCA-LDA which used Principal Component Analysis (PCA) to separate the classes [46], and ULDA which used target generation process (TGP) for the targets and classes [47]. LDA-Km [48], TRACK [49], Semi-LDC [10], and modified-2DLDA [50] utilized k-means clustering to make the class labels. Further, LDA-basis sequence calculates the projected class numbers and produces a sequence of bases that meets to the practical LDA solutions [8].

This study introduces a new technique in class labeling which is utilizing Self-Organizing Map (SOM). SOM is much similar with K-means in generating clusters [51], but the primary benefit of SOM is the capability to deal both subproblems together and compute them concurrently through the used of unsupervised learning. SOM can also work with multidimensional inputs [14].

There were many techniques proposed to overcome the SSS issue on LDA such as Regularized LDA (RLDA) which offers a computation concerning the relationship among the dilemmas of multi-class discriminant analysis and multivariate regression [52]. An exponential discriminant analysis (EDA) technique was suggested to solve the undersampled problem [53]. While Spectral Regression Discriminant Analysis (SRDA) casts discriminant analysis into a regression framework. Also, direct LDA (D-LDA) [54], multiple between-class linear discriminant analysis (MBLDA) [13], and LDA/QR [55] are for classifying multidimensional data with fast learning capability. A two-stage method utilizing bidirectional LDA and RLDA was designed for two-dimensional data only [42]. A split and combined approaches for LDA (SC-LDA) was developed to replace the full eigenvector decomposition [56]. But the widely used method is the application of GSVD on LDA (LDA/GSVD) since it can overcome the mathematical problems integral in establishing the scatter matrices and more accurate [17, 21]. Further, GSVD has been proven

useful in most mathematical calculations regarding singularity of the matrix on discriminant analysis [18-20]. The matrix pair's generalized singular values (GSVs) (A, L) are $(A^T A L^T L)$'s square roots, where $L \in R^{p \times n}$ suffices $m \geq n \geq p$. It formulates GSVD as an expansion or usual simplification of singular value decomposition (SVD) [20, 57, 58]. It can be written as:

$$A = U \begin{pmatrix} \sum & 0 \\ 0 & I_{n-p} \end{pmatrix} X^{-1}, L = V(M, 0) X^{-1} \qquad (2)$$

where $U \in R^{mxn}$, $U^T U = I$ and $V \in R^{pxp}$, $= V^T V = I$ are orthonormal columned matrices, $X \in R^{nxn}$ is nonsingular, $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots \sigma_p)$ and $M = \mathrm{diag}(\mu_1, \mu_2, \cdots \mu_p)$ are $p \ x \ p$ non-negative diagonal elements as $1 \geq \sigma_p \geq \cdots \geq \sigma_2 \geq \sigma_1 \geq 0$, $1 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_p \geq 0$, $\sigma_i^2 + \mu_i^2 = 1$ $(i = 1, 2, \cdots, p)$. The GSVs of (A, L) are marked as the scales $\gamma_i = \sigma_i / \mu_i$ $(i=1, 2, \cdots, p)$ and the least GSVs are set to zero [20].

Table 2. Existing LDA/GSVD Algorithm

Algorithm: **Existing LDA/GSVD**
For the matrix $A \in R^{m \times n}$ with k groups, it calculates the matrix's columns $G \in R^{m \times (k-1)}$, which maintains the configured cluster dimensionally narrowed space, and determines $(k - 1)$-dimensional depiction Y of A.   Step 1: Calculate $H_w \in R^{m \times n}$ and $H_b \in R^{m \times k}$ from A Step 2: Solve the $K = (Hb, Hw)T \in R(k+n) \times m$ for its orthogonal decomposition. $$P^T KQ = \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix}$$ Step 3: Let $t = \mathrm{rank}(K)$.   Step 4: Calculate W from the SVD of P(1 : k, 1 : t), which is UTP(1 : k, 1 : t) W = ΣA.   Step 5: Solve the first $k - 1$ columns of $$X = Q \begin{pmatrix} R^{-1}W & 0 \\ 0 & I \end{pmatrix}$$ and allocate those to G.   Step 6: $Y = G^T A$.

To enhance the existing LDA/GSVD algorithm, the Artificial Neural Network (ANN) can be utilized. The used of ANNs in several real-world purposes is because of their capability concerning resiliency and stableness even in noisy data and fault tolerance [59]. Thus, the widely employed method is Back Propagation Neural Network (BPNN). It is composed of input layer, hidden layers, and output layer [60]. An example of BPNN is Bayesian Regularization Back Propagation (BRBP). BRBP offers robust approximation for difficult and noisy inputs. Thus, it works excellently by removing network weights which have no impact on the problem solving and presents improvements on evading the problems of local minima [59]. Furthermore, it delivers weights into a training function while advancing the simplification performance of the old BPNN automatically [60]. Moreover, ANN is one of the techniques that can be used to have an energy efficient network [61].

3. Research Methodology

In attaining this new bandwidth allocation approach, the ANN for LDA/GSVD algorithm must be developed first because it will be used in discovering users' web usage

patterns. Thus, this algorithm will be used to visualize appropriately the groups of users who are inclined to accessing educational sites and those who are not. The process of developing algorithm is presented in figure 3 which shows four (4) major stages these are preprocessing, class labeling, computation of feature subspaces, and ANN training.

3.1 Preprocessing

In this step, the unnecessary instances from the weblogs will be removed. This also includes the identification of the category of the website and who requests for it. There are four (4) objects involved in preprocessing. These are the weblog, captive portal log, shallalist files, and the predefined educational sites. The weblog and captive portal log were taken from the proxy servers in Misamis University, Ozamiz City, Philippines. It is composed of the logs staring 1st semester up to the end of the 2nd semester of 2017-2018. It mainly contains the websites accessed by the wired and wireless users while the captive portal log includes the date and time when the users login or use the internet. This captive portal log will be used to identify the specific user who requests the particular website. The shallalist file contains the category of websites which will be utilized in determining where the requested website belongs. Lastly, the predefined educational sites is a file that contains lists of educational websites which will prevail even though they are differently categorized in the shallalist. Table 3 presents the algorithm for the preprocessing.

Table 3. Preprocessing algorithm

Algorithm: **Preprocessing Algorithm**
1. Read log entry from weblog file
2. if(weblog.mime="text/html"   AND   weblog.status=200   AND   weblog.method="GET") {
3.   site=GetDomain(weblog.URL);
4.   user=GetUserAccount();
5.   if(IsNewLogin(user)=true) {
6.   if(GetSiteCategory(site)="Education" OR IsPredefinedSite(site)=true) {
7.   num_educ=GetNumEduc(user)+1;
8.   } else {
9.   num_not_educ= GetNumNotEduc(user)+1;
10.   }
11.   UpdateSitesAccessed(user);
12.   }
13.   }
14.   Repeat 1 and 2 until EOF encountered
15.   Generalize all the users with respect to their number of educational and noneducational sites.
16.   End the process

3.2 Class Labelling

After the preprocessing, the generalized data is unlabeled and because of that class labeling is imperative for the LDA/GSVD to work. SOM is employed to create these clusters which become the class labels.

3.3 Computation of Feature Subspaces

At this point, the dataset has been labeled and ready for actual computation of the feature subspaces. For the ANN architecture to learn, predict, and classify the users' web usage patterns, the feature subspaces must be derived from that existing LDA/GSVD algorithm. These feature subspaces are actual points where the users belong, and these are used for further visualization of the classification of the users.

Thus, it is the current algorithm that is used so that the accuracy of the developed ANN algorithm is the same.

3.4 ANN Training

The tansigmoid transmission function was utilized for the hidden layers' activation function. The ANN architecture is formed from 2 input variables which are the dimensions or the frequency of the educational websites and noneducational websites per user, and the corresponding three output variables which are the expected feature subspaces and class label. These dimensions, feature subspaces, and class labels will be used in training and testing. For the sampling, 70% of the instances of the dataset were allocated for training, and 30% for the testing. Moreover, in training of the network, Bayesian Regularization Back Propagation (BRBP) was employed.

After saving the trained network, it will become a module or subroutine that will be used to solve the expected new feature subspaces of the inputs. Thus, the algorithm shown in table 4 is for the proposed bandwidth allocation mechanism. Further, the architecture of the trained network which is employed to separate the groups of users according to their web usage patterns is presented in figure 2.

Table 4. Proposed Bandwidth Allocation Mechanism Algorithm

| Algorithm: **Bandwidth Allocation Mechanism Algorithm** |
| --- |
| 1. Open Dataset |
| 2. Get each user's total number of educational and noneducational websites. |
| 3. Compute the corresponding feature subspaces and determine the class label using the ANN for LDA/GSVD. |
| 4. Classify and visualize the user's web usage pattern. |
| 5. Allocating bandwidth to the user. |
| 6. Repeat 1 and 2 until EOF encountered |
| 7. End Process |
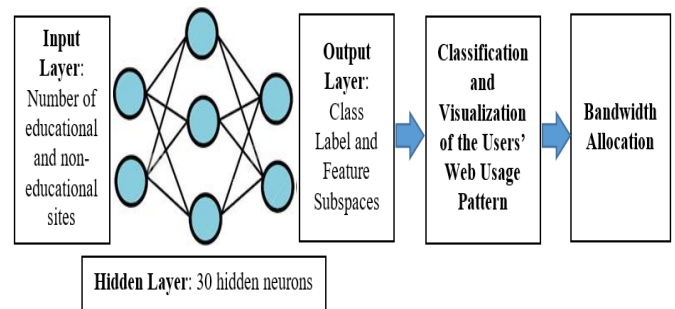
Figure 2. Architecture of the Trained ANN for LDA/GSVD used in determining Users' Web Usage Pattern

As shown in figure 2, the inputs to the trained ANN for GSVD are the dimensions or the total number of educational and noneducational websites accessed by each user. The trained network will compute the feature subspaces and classify the users' web patterns. In viewing the actual separation of the users according to their patterns, visualization will be presented as well to find out the density of the users who manage to browse educational websites and those who are not. This visualization can be a great help for the network administrators or IT managers in allocating the specific amount of bandwidth.
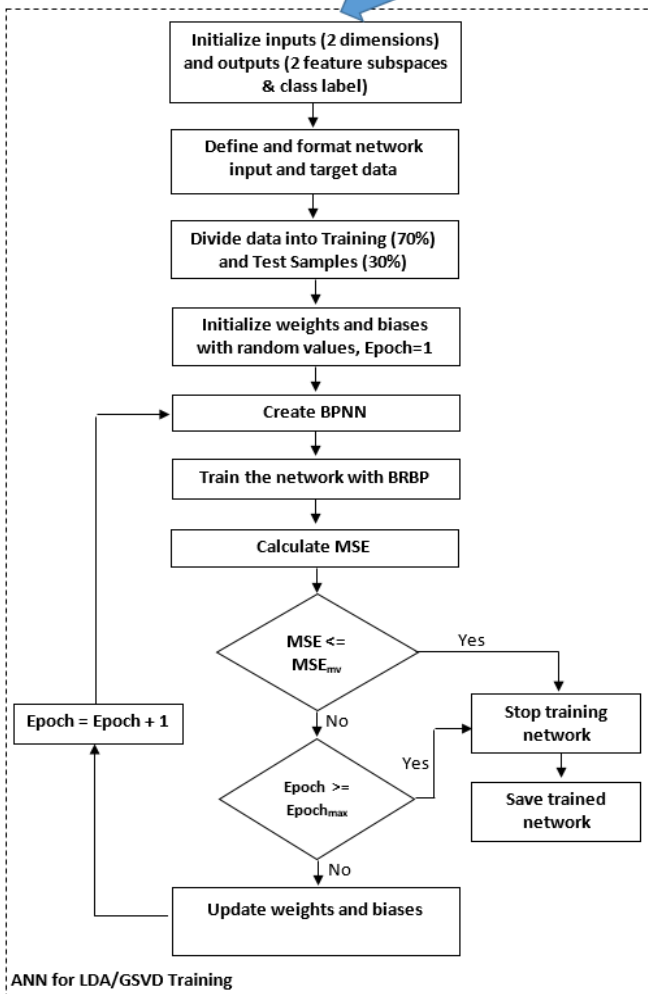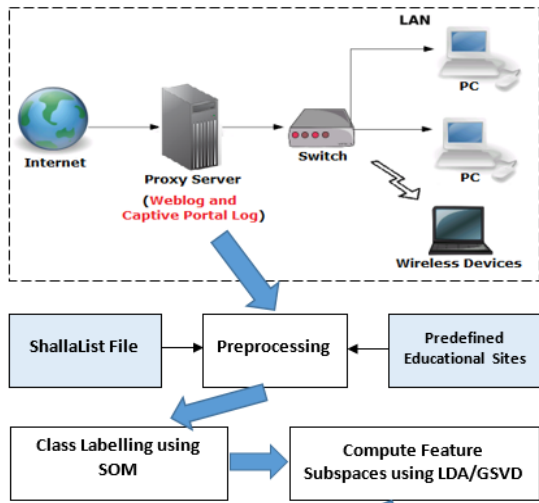
Figure 3. Training process for the development of the ANN for LDA/GSVD algorithm

4. Results and Discussions

Using MATLAB R2014a, both algorithms, existing and enhanced LDA/GSVD, were coded and ran on a PC with the processor of Intel® Core i5, 4GB RAM, and 2.7GHz speed.

4.1   Actual Dataset

After the preprocessing step, it was identified that there are a total of 193 active users in specific network nodes for the school year 2017-2018. Therefore, this study used a dataset

that has 193 instances and two dimensions. Figure 4 shows the graph that corresponds to the total number of educational websites and noneducational websites that were accessed by the users. Blue bars are for the educational while the orange is for the noneducational.
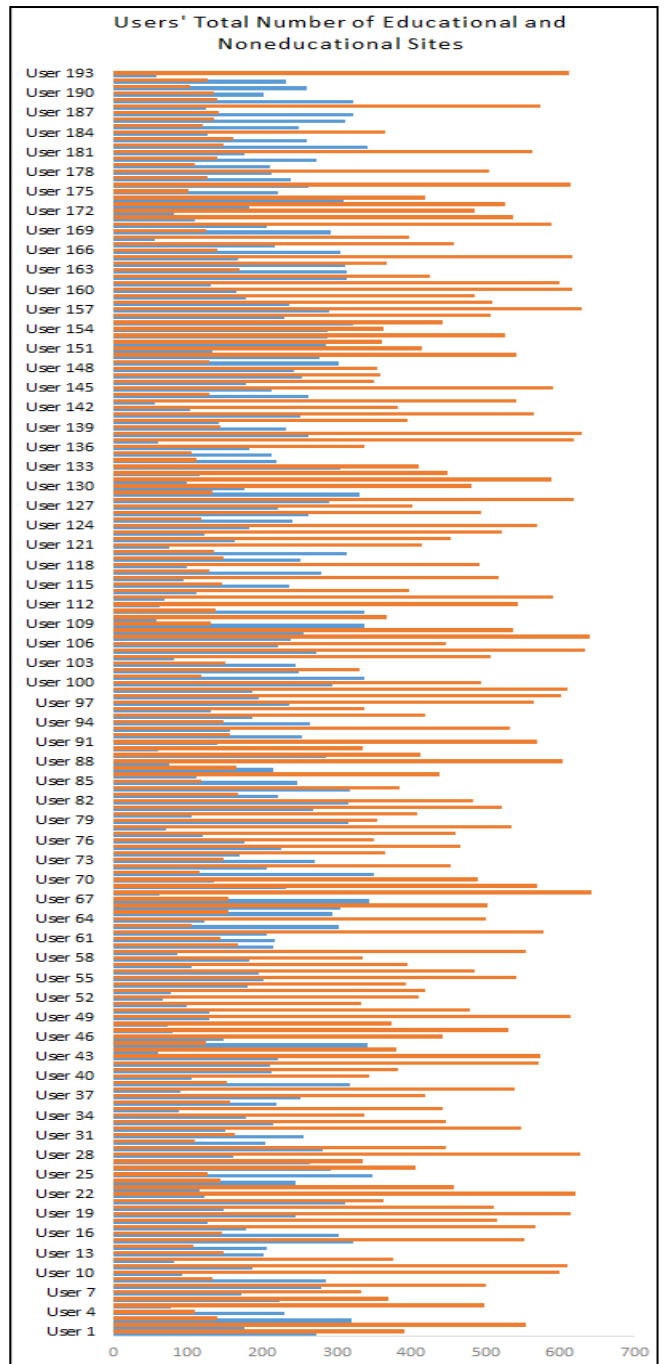

Figure 4. Total number of educational and noneducational websites accessed by the users

After preprocessing, class labeling comes in. Table 5 presents the number of users belongs to the classes.

Table 5. Frequency of the classes

Class Label	Number of Users
1	56
2	137

4.2 Computation and Visualization of the Feature Subspaces using the Existing LDA/GSVD Algorithm

Once all the instances of the dataset are labeled, the computation of the feature subspaces was done. It took 8 seconds for the existing LDA/GSVD to finish the whole process. Figure 5 shows the graph for the feature subspaces separating the classes. Each point in the graph represents a user. Thus, class 1 composed of fewer users compared to class 2. It is noticeable that all data points are appropriately separated for better visualization.
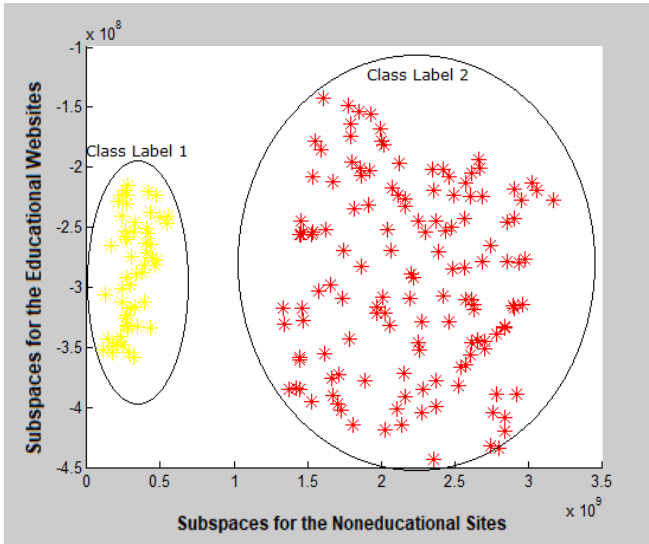

Figure 5. Graph of the subspaces after applying the existing LDA/GSVD

4.3 Training for the development of ANN for LDA/GSVD algorithm

After getting all the necessary values such as class labels and feature subspaces, training will be next. This study used the performance functions were used in the study which includes the Mean Squared Error (MSE) and Regression (R) to evaluate the performance of the ANN for LDA/GSVD algorithm. MSE is the average squared difference between experimental output values and the fed targets in training.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (t_i - a_i)^2 \qquad (3)$$

Where n is the sample set's size, a_i is the ANN experimental or observed output and t_i is the matching targets. Regression (R) computes the outputs and targets' correlation. When the value of R is 1, it signifies a good or close relationship, otherwise a random relationship [60].

Figure 6 depicts the performance of training and test samples using BRBP algorithm. The graph shows that the test and training samples almost overlap with each other. Further, training and test curves stabilized at epoch 403,525 which MSE error value is approximately 1.691.
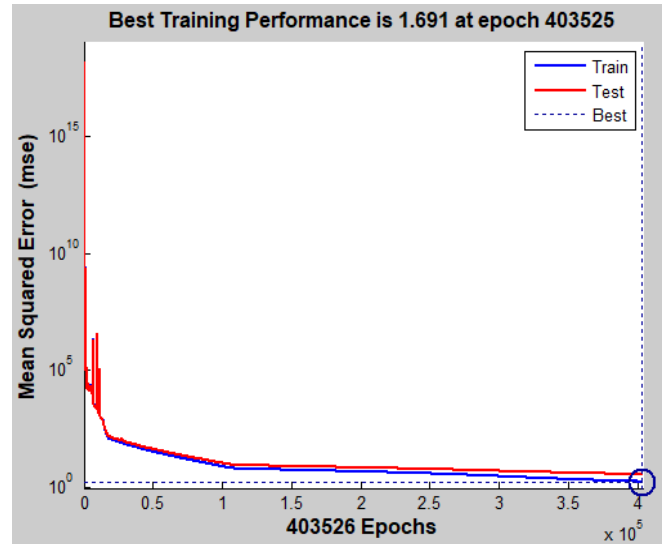

Figure 6. BRBP's Prediction Result

The histogram in figure 7 presents the frequency of the instances per error. The measurement of the error is by subtracting the targets and the resultant outputs. The most significant error in the training was at around 9.589.
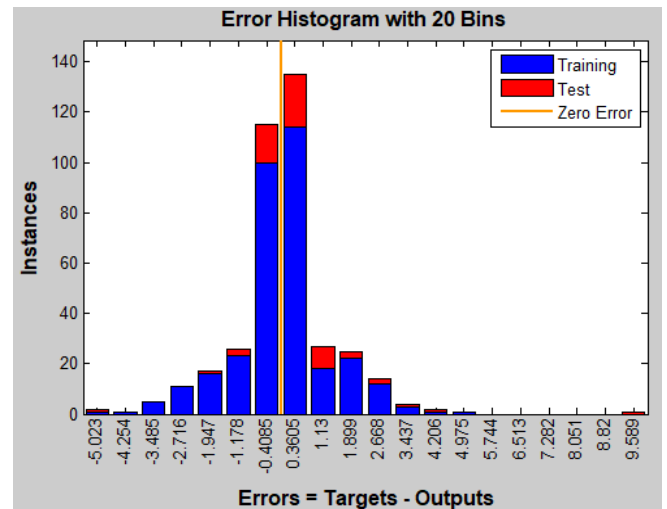

Figure 7. BRBP's Histogram of error sequences

Figure 8 shows BRBP algorithm's correlation. Thus, the graphs present that the algorithm is accurate and better because the MSEs are in minimal values, and the value of R for the training, test, and overall analysis is 1. Further, table 6 shows the performance of the enhanced LDA/GSVD.

Table 6. Performance of ANN for LDA/GSVD using BRBP

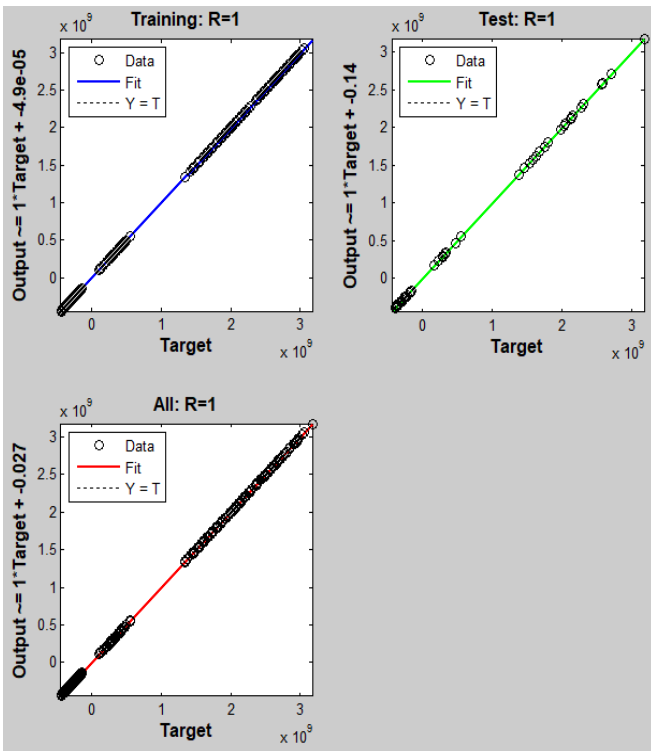| Dataset Sample	Mean Squared Error	Regression
Training	1.6910	1
Testing	3.4177	1

Figure 8. BRBP's Regression Analysis

4.4 Computation and Visualization of the Feature Subspaces using the ANN for LDA/GSVD Algorithm

It is noticeable that figure 9 which presents the classification of the data using the enhanced algorithm is very much similar to figure 5 which utilized the existing LDA/GSVD. It is a manifestation that the accuracy of the improved LDA/GSVD maintains the accuracy of current LDA/GSVD algorithm. Thus, if a new user joins the network, this developed algorithm can be used directly in identifying his/her web usage patterns whereas the existing LDA/GSVD, it is the whole algorithm that will be executed and all the instances of the dataset will be considered during the process. With that, the existing algorithm can be exhaustive especially when there is a new instance added to the dataset. The whole process took 4 seconds.
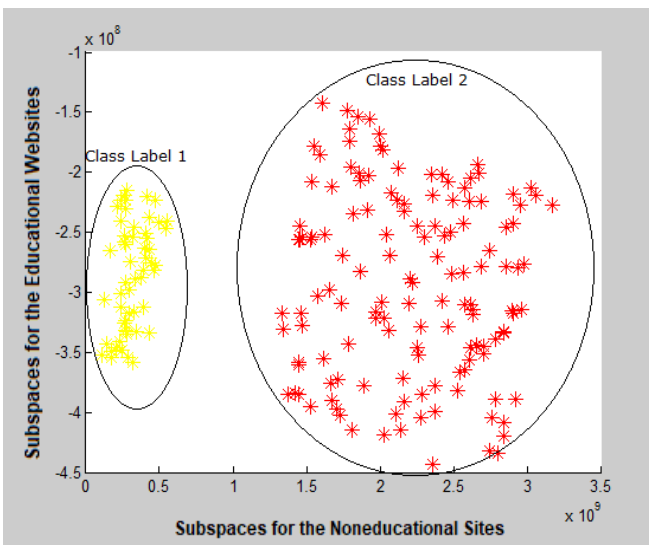

Figure 9. Graph of the subspaces after applying the ANN for LDA/GSVD Algorithm

Table 7 presents the computational costs of two algorithms. It is evident that the enhanced LDA/GSVD improved the computational cost by 50%. The values for the computational costs may be too small because there are only 193 instances that composed the dataset.

Table 7. Computational Costs of the Existing and Enhanced Algorithms

Algorithm	Computational Cost or Total Execution Time
Existing LDA/GSVD	8 seconds
ANN for LDA/GSVD	4 seconds
Improvement of the Enhanced LDA/GSVD	**50%**

4.5 Bandwidth Allocation

Based on the visualization of the feature subspaces, it can be seen that users who belong to class 1 are those individuals who are inclined to doing research or browsing educational websites. There are 56 users belong to class 1 who should be given a higher bandwidth compared to the users in class 2. For this case, the algorithm for assigning bandwidth to each user is presented in table 8; these are the particular lines of instructions to the step 5 of the algorithm shown in table 4. The IT manager must define the values for the higher and lower bandwidths.

Table 8. Allocating bandwidth to each user

Algorithm: **Allocating bandwidth to each user**
1. if(user belongs to class 1)
2.   setHigherBandwidth(user, higher_bandwidth_value);
3. else
4.   setLowerBandwidth(user, lower_bandwidth_value);

5. Conclusions

Allocation of bandwidth becomes an essential aspect of campus networks which various experts have focused on. In this study, focusing on the web usage patterns in allocating bandwidth using a hybrid data mining technique which composed of clustering, ANN, and discriminant analysis were presented and applied on a real weblog of a university network. The result of this method is useful in assigning bandwidth to the users. Simulation results also showed that discovering of the users' web usage patterns can be done faster using the developed ANN for LDA/GSVD algorithm. Besides, the enhanced algorithm is efficient while maintaining the accuracy of the existing algorithm. Through this proposed approach, network administrators, IT managers, or those who manage the school's network could identify those users that will be allocated with higher bandwidth.

In the future, this proposed technique must be implemented in campus' captive portal network where actual users are connected to the proxy servers for further testing, evaluation, and enhancement.

6. Acknowledgement

References

[1] E. A. Z. Noughabi, B. H. Far, and B. Raahemi, "Predicting Students Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method (Application Paper)," *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 2016.

[2] A. Sarigiannidis and P. Nicopolitidis, "Addressing the interdependence in providing fair and efficient bandwidth distribution in hybrid optical-wireless networks," *International Journal of Communication Systems*, vol. 29, no. 10, pp. 1658–1682, 2015.

[3] M. Hemmati, B. Mccormick, and S. Shirmohammadi, "Fair and Efficient Bandwidth Allocation for Video Flows Using Sigmoidal Programming," *2016 IEEE International Symposium on Multimedia (ISM)*, 2016.

[4] P. Sukumar, L. Robert, and S. Yuvaraj, "Review on modern Data Preprocessing techniques in Web usage mining (WUM)," *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 2016.

[5] S. G. Langhnoja, M. P. Barot, and D. B. Mehta, "Web usage mining using association rule mining on clustered data for pattern discovery," *International Journal of Data Mining Techniques and Applications*, vol. 2, no. 1, 2013.

[6] N. B. M. Zainee and K. Chellappan, "A preliminary dengue fever prediction model based on vital signs and blood profile," *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2016.

[7] S. S. Nikam, "A comparative study of classification techniques in data mining algorithms," *Oriental Journal of Computer Science and Technology*, vol. 8, no. 1, pp. 13–19, 2015.

[8] P. P. Markopoulos, "Linear Discriminant Analysis with few training data," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[9] S. Yu, Z. Cao, and X. Jiang, "Robust linear discriminant analysis with a Laplacian assumption on projection distribution," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[10] C. L. Liu, W. H. Hsaio, C. H. Lee, and F. S. Gou, "Semi-Supervised Linear Discriminant Clustering," *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 989–1000, 2014.

[11] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012.

[12] J. Shao, Y. Wang, X. Deng, and S. Wang, "Sparse linear discriminant analysis by thresholding for high dimensional data," *The Annals of Statistics*, vol. 39, no. 2, pp. 1241–1265, 2011.

[13] Z. Wang, Y.-H. Shao, L. Bai, C.-N. Li, L.-M. Liu, and N.-Y. Deng, "MBLDA: A novel multiple between-class linear discriminant analysis," *Information Sciences*, vol. 369, pp. 199–220, 2016.

[14] J. Faigl and G. A. Hollinger, "Autonomous Data Collection Using a Self-Organizing Map," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1703–1715, 2018.

[15] R. Mclean, A. J. Walker, and G. Bright, "An artificial neural network driven decision-making system for manufacturing disturbance mitigation in reconfigurable systems," *2017 13th IEEE International Conference on Control & Automation (ICCA)*, 2017.

[16] S. M. Sharif, F. M. Kusin, Z. H. Asha'Ari, and A. Z. Aris, "Characterization of Water Quality Conditions in the Klang River Basin, Malaysia Using Self Organizing Map and K-means Algorithm," *Procedia Environmental Sciences*, vol. 30, pp. 73–78, 2015.

[17] P. Howland and H. Park, "Generalizing discriminant analysis using the generalized singular value decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 995–1006, 2004.

[18] X. Jing, Y. Dong, and Y. Yao, "Uncorrelated optimal discriminant vectors based on generalized singular value decomposition," *International Conference on Automatic Control and Artificial Intelligence (ACAI 2012)*, 2012.

[19] C. H. Park and H. Park, "A Relationship between Linear Discriminant Analysis and the Generalized Minimum Squared Error Solution," *SIAM Journal on Matrix Analysis and Applications*, vol. 27, no. 2, pp. 474–492, 2005.

[20] Z. Chen and T. H. Chan, "A truncated generalized singular value decomposition algorithm for moving force identification with ill-posed problems," *Journal of Sound and Vibration*, vol. 401, pp. 297–310, 2017.

[21] W. Wu and M. O. Ahmad, "Orthogonalized linear discriminant analysis based on modified generalized singular value decomposition," *2009 IEEE International Symposium on Circuits and Systems*, 2009.

[22] S. Bahrami, "Three-dimensional inverse scattering approach using analytical singular value decomposition method," *2017 18th International Radar Symposium (IRS)*, 2017.

[23] H. Zhao and P. C. Yuen, "Incremental Linear Discriminant Analysis for Face Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 1, pp. 210–221, 2008.

[24] Y. Dash and S. K. Dubey, "Quality prediction in object oriented system by using ANN: a brief survey," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 2, 2012.

[25] S. S. Ranhotra, A. Kumar, M. Magarini, and A. Mishra, "Performance comparison of blind and non-blind channel equalizers using artificial neural networks," *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2017.

[26] B. Fan, S. Leng, and K. Yang, "A dynamic bandwidth allocation algorithm in mobile networks with big data of users and networks," *IEEE Network*, vol. 30, no. 1, pp. 6–10, 2016.

[27] J. Wang, M. Qiu, and B. Guo, "High reliable real-time bandwidth scheduling for virtual machines with hidden Markov predicting in telehealth platform," *Future Generation Computer Systems*, vol. 49, pp. 68–76, 2015.

[28] L. Chitanana and D. W. Govender, "Bandwidth Management in the Era of Bring Your Own Device," *The Electronic Journal of Information Systems in Developing Countries*, vol. 68, no. 1, pp. 1–14, 2015.

[29] S. Akpah, "Improving the Performance of a Network by Managing the Bandwidth Utilisation with squidGuard: A Case Study," *Ghana Journal of Technology*, vol. 1, no. 2, pp. 9-17, 2017.

[30] K. S. Reddy, M. K. Reddy, and V. Sitaramulu, "An effective data preprocessing method for Web Usage Mining," *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 2013.

[31] T. Kärkkäinen, M. Pitkänen, and J. Ott, "Enabling ad-hoc-style communication in public WLAN hot-spots," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 1, p. 4, Dec. 2013.

[32] A. Dabrowski, G. Merzdovnik, N. Kommenda, and E. Weippl, "Browser History Stealing with Captive Wi-Fi Portals," *2016 IEEE Security and Privacy Workshops (SPW)*, 2016.

[33] F. L. Aryeh, M. Asante, and A. E. Y. Danso, "Securing Wireless Network Using pfSense Captive Portal with Radius Authentication–A Case Study at UMaT," *Ghana Journal of Technology*, vol. 1, no. 1, pp. 40–45, 2016.

[34] Shalla Secure Services KG, " Shalla's Blacklists," *Shalla Secure Services*, 2018. [Online]. Available: http://www.shallalist.de. [Accessed: 17-Apr-2018].

[35] A. C. Squicciarini, G. Petracca, W. G. Horne, and A. Nath, "Situational awareness through reasoning on network incidents," *Proceedings of the 4th ACM conference on Data and application security and privacy - CODASPY 14*, 2014.

[36] X. Gao, X. Wang, X. Li, and D. Tao, "Transfer latent variable model based on divergence analysis," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2358–2366, 2011.

[37] X. Gao, X. Wang, D. Tao, and X. Li, "Supervised Gaussian Process Latent Variable Model for Dimensionality Reduction," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 2, pp. 425–434, 2011.

[38] J. J. D. M. S. Junior and A. R. Backes, "Shape classification using line segment statistics," *Information Sciences*, vol. 305, pp. 349–356, 2015.

[39] D. Tao, J. Cheng, X. Lin, and J. Yu, "Local structure preserving discriminative projections for RGB-D sensor-based scene classification," *Information Sciences*, vol. 320, pp. 383–394, 2015.

[40] D. Wang, X. Gao, and X. Wang, "Semi-Supervised Nonnegative Matrix Factorization via Constraint Propagation," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 233–244, 2016.

[41] L. Zhang, L. Wang, and W. Lin, "Generalized Biased Discriminant Analysis for Content-Based Image Retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 1, pp. 282–290, 2012.

[42] J. Zhao, L. Shi, and J. Zhu, "Two-Stage Regularized Linear Discriminant Analysis for 2-D Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1669–1681, 2015.

[43] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[44] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*.

[45] A. Sharma and K. K. Paliwal, "Linear discriminant analysis for the small sample size problem: an overview," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 3, pp. 443–454, Jul. 2015.

[46] S.-K. Oh, S.-H. Yoo, and W. Pedrycz, "Design of face recognition algorithm using PCA -LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks : Design and its application," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1451–1466, 2013.

[47] G. C. Lin, C. M. Wang, and W. J. Wang, "An Unsupervised Linear Discriminant Analysis Approach to Multispectral MRI Images Classification," *2007 International Conference on Machine Learning and Cybernetics*, 2007.

[48] C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and K-means clustering," *Proceedings of the 24th international conference on Machine learning - ICML 07*, 2007.

[49] D. Wang, F. Nie, and H. Huang, "Unsupervised Feature Selection via Unified Trace Ratio Formulation and K-means Clustering (TRACK)," *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, pp. 306–321, 2014.

[50] D. Conka, P. Viszlay, and J. Juhar, "Improved class definition in two dimensional linear discriminant analysis of speech," *2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA)*, 2015.

[51] S. Saraswathi and L. Sivakumar, "Comparative analysis of k-means and self organizing map clustering on boiler process data," *2014 International Conference on Computer Communication and Informatics*, 2014.

[52] Z. Zhang, G. Dai, C. Xu, and M. I. Jordan, "Regularized discriminant analysis, ridge regression and beyond," *Journal of Machine Learning Research*, pp. 2199–2228, Aug. 2010.

[53] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang, and B. Xu, "Generalized Discriminant Analysis: A Matrix Exponential Approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 1, pp. 186–197, 2010.

[54] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data — with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.

[55] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 929–941, 2005.

[56] J. K. P. Seng and K. L.-M. Ang, "Big Feature Data Analytics: Split and Combine Linear Discriminant Analysis (SC-LDA) for Integration Towards Decision Making Analytics," *IEEE Access*, vol. 5, pp. 14056–14065, 2017.

[57] H. Abdi, "Singular value decomposition (SVD) and generalized singular value decomposition," *Encyclopedia of measurement and statistics*, pp. 907–912, 2007.

[58] D. Senaratne and C. Tellambura, "Generalized Singular Value Decomposition for Coordinated Beamforming in MIMO Systems," *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 2010.

[59] K. Jazayeri, M. Jazayeri, and S. Uysal, "Comparative Analysis of Levenberg-Marquardt and Bayesian Regularization Backpropagation Algorithms in Photovoltaic Power Estimation Using Artificial Neural Network," *Advances in Data Mining. Applications and Theoretical Aspects Lecture Notes in Computer Science*, pp. 80–95, 2016.

[60] F. Dalipi and S. Y. Yayilgan, "The impact of environmental factors to skiing injuries: Bayesian regularization neural network model for predicting skiing injuries," *2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2015.

[61] P. Singh, V. Pareek, and A. K. Ahlawat, "Designing an Energy Efficient Network Using Integration of KSOM, ANN and Data Fusion Techniques," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 9, no. 3, 2017.