Reducing Dimensionality in Text Mining using Conjugate Gradients and Hybrid Cholesky Decomposition

Jasem M. Alostad

The Public Authority for Applied Education and Training (PAAET), College of Basic Education
P.O.Box 23167, Safat 13092, Kuwait

Abstract—Generally, data mining in larger datasets consists of certain limitations in identifying the relevant datasets for the given queries. The limitations include: lack of interaction in the required objective space, inability to handle the data sets or discrete variables in datasets, especially in the presence of missing variables and inability to classify the records as per the given query, and finally poor generation of explicit knowledge for a query increases the dimensionality of the data. Hence, this paper aims at resolving the problems with increasing data dimensionality in datasets using modified non-integer matrix factorization (NMF). Further, the increased dimensionality arising due to non-orthogonally of NMF is resolved with Cholesky decomposition (cdNMF). Initially, the structuring of datasets is carried out to form a well-defined geometric structure. Further, the complex conjugate values are extracted and conjugate gradient algorithm is applied to reduce the sparse matrix from the data vector. The cdNMF is used to extract the feature vector from the dataset and the data vector is linearly mapped from upper triangular matrix obtained from the Cholesky decomposition. The experiment is validated against accuracy and normalized mutual information (NMI) metrics over three text databases of varied patterns. Further, the results prove that the proposed technique fits well with larger instances in finding the documents as per the query, than NMF, neighborhood preserving: nonnegative matrix factorization (NPNMF), multiple manifolds non-negative matrix factorization (MMNMF), robust non-negative matrix factorization (RNMF), graph regularized non-negative matrix factorization (GNMF), hierarchical non-negative matrix factorization (HNMF) and cdNMF.

Keywords—*Data mining; non-integer matrix factorization; Cholesky decomposition; conjugate gradient algorithm*

I. INTRODUCTION

Computing application in several fields generates numerous data over several instances. In order to extract knowledge from such instances, solutions are used conventionally with data mining tools. However, the large datasets with numerous instances poses severe challenges and that leads to improper processing of such huge data volume. The reduction of datasets or improved mining algorithm can overcome such challenges [1]. The reduction of improper values from the datasets provides a greater impact and this increases the performance of processing the large data [2]; hence, the improved mining approach is not useful in some cases [3].

The data reduction is the process of reducing the size or dimensionality of the data, however, the representation of the data should be retained. Selection of instance is one better way to reduce the data by reducing the total number of instances. In spite of many efforts to deal with such instances, data mining algorithm, however undergoes severe challenges due to non-applicability of datasets with large instances. Hence, the computational complexity of the system increases with larger instances [3], [4] and leads to problems in scaling, increased storage requirements and clustering accuracy. The other problems associated with larger data instances include, improper association or interaction in the feature space, lack of ability to handle the large datasets with discrete variables, inability to classify the data and poor knowledge generation for a given query, and finally poor computation due to missing variables.

Recently, there are significant developments in NMF for various clustering problems in data mining, defined above. The NMF process is used to factorize the input matrix into two matrices of non-negative variables in a lower rank order [5]-[8]. Several applications of NMF include: chemometrics, environmetrics, pattern recognition, text mining and summarization [9], multimedia data analysis [10], analysis of DNA gene expression [11], analyzing the financial data [12], and social network analysis [13]. Several algorithms are designed to overcome the problems associated with objective functions [14], classification [15], collaborative filtering [16] and computational methodologies.

Thought, NMF is used for data analysis, the recent trends has improved the discoverability and learning ability of NMF in the data mining to solve the problems associated with larger datasets. In order to avoid limitations associated with larger dataset, the following consideration are made in the present study: This proposed method uses NMF to study the feature vector of a text document and Cholesky decomposition is used to avoid the non-orthogonality problem in the NMF. Further, to avoid poor decomposition using Cholesky decomposition, conjugate gradient is used, which avoids the rapid multiplication by the gradients in the feature space.

Since, the NMF algorithm learns both the data and feature vector in the dataset feature space, the proposed method implies the following contributions:

- First, the metric using data matrix is estimated in the feature space using trained feature vector.

- Second, the Cholesky decomposition process is applied over the metric and the upper triangular matrix is identified.

- Third, upper triangular matrix is used a linear mapper for the associated data vector.

- Finally, conjugate gradient is applied to reduce sparse matrix through reduced multiplication and that avoids the NP-hard problem which significantly reduces the computational complexity [17], [18].

The outline of the paper is presented as follows: Section 2 discusses the related works. Section 3 provides the NMF model for clustering the larger datasets. Section 4 provides the modifications in NMF using Cholesky decomposition. Section 5 provides experimental verification of the proposed system over the associated datasets and section 6 concludes the paper with future work.

II. RELATED WORKS

Several methods of NMF are discussed here, which include: Semi supervised constrained NMF [19], semi-supervised graph based discriminative NMF [20], Bayesian learning approach to reduce the generalization error in upper bound using NMF [21] and update rules [22], sparseness NMF, which provides better characterization of the features [23], sparse unmixing NMF [24], locally weighted sparse graph regularized NMF [25], graph-regularized NMF [26], graph dual regularization [27], multiple graph regularized NMF [28], graph regularized multilayer NMF [29], adaptive graph regularized NMF [30], hyper-graph regularized [31], graph regularization with sparse NMF [32], multi-view NMF [33], extended incremental NMF [34], incremental orthogonal projective NMF [35], correntropy induced metric NMF [36], multi-view NMF [37], patch based NMF [38], MMNMF [39], regularized NMF [40], FR conjugate gradient NMF [41]. However, these methods failed to address the problems associated with non-orthogonality due to the presence of non-negative elements in NMF.

III. IMPROVED NMF METHOD

NMF is a non-negative low-rank approximation method associated with certain constraints that relates to the non-negative elements in the data and feature vectors. Here, non-orthogonality problem exist due to the presence of non-negative elements between the vectors and addition of linear combination, results in part representations. This interpretable and intuitive method for representing the text data elements is divided into two parts:

1) Data vector representation using Cholesky Decomposition (CD) with Conjugate Gradients (CG.
2) Feature vector representation using NMF.

The detail of these is shown in following sections:

A. Fitness Function for NMF

In NMF, it is assumed that matrices contain non-negative elements, hence, factorization is approximated. Let input data matrix is $X = (x_1, x_2,..., x_n)$, which carries n elements of input data vectors and the data matrix is decomposed into two matrices,

$$X \approx FG^T \tag{1}$$

Where, $\mathbf{X} \in \mathbb{R}^{p \times n}$, $\mathbf{G} \in \mathbb{R}^{n \times k}$ and $\mathbf{F} \in \mathbb{R}^{p \times k}$ and $\mathbb{R}$ is the set of non-zero real numbers, $\mathbf{G} = (g_1, g_2,..., g_n)$ and $\mathbf{F} = (f_1, f_2,..., f_n)$. In general, the value of $p < n$ and the rank of $\mathbf{F}$ and $\mathbf{G}$ matrices is less than X i.e. $k \ll \min(p,n)$. The rank $\mathbf{F}$ and $\mathbf{G}$ is generated using minimization fitness function and the sum of squared errors is used to evaluate the fitness function, which is represented as:

$$\min_{F,G \geq 0} J_{sse} = \left\| X - FG^T \right\|^2 \tag{2}$$

The matrix normalization is obtained using Frobenius norm and the values of $\mathbf{F}$ and $\mathbf{G}$ are non-negative with non-orthogonal column vectors in its Euclidean space. The non-deficiency cases for rank $\mathbf{R}$ and $\mathbf{G}$ is generated using **I-divergence** fitness function:

$$\min_{F,G \geq 0} J_{ID} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[X_{ij} \log \frac{X_{ij}}{\left(FG^T\right)_{ij}} - X_{ij} + \left(FG^T\right)_{ij} \right] \tag{3}$$

Here, when $I(x) = x \log x - x + 1 \geq 0$, inequality holds i.e. $x \geq 0$ and when $x = 1$, the equality holds. Hence, **I-divergence** using inequality condition is expressed as: $I(u, v) = (u/v)\log(u/v) - u/v + 1$.

B. NMF Clustering

The initialization in NMF is an important process with clustering, similar to k-means clustering. However, the fitness function as a minimization function often undergoes local minimum problem [42], [43]. Due to such constraint, even if the minimization function is convex, the intrinsic alternating function is non-convex. If a random initialization is considered, then the factor matrices are initialized as random matrices and hence, it become ineffective due to slow convergence to attain the local minima. If clustering process is used in NMF, the initialization is obtained from fuzzy [44], divergence-k-means [45] and spherical k-means [46], [47]. However, the proposed method considers a simple strategy for document clustering, which is discussed below:

NMF is applied to cluster the documents and number of features vectors in the document of each dataset is set as total clusters in a document [48], [49]. Each cluster is assigned with individual instances and the representation **g** is considered maximum, which is represented as:

$$c_g = \arg\max_c g_c \tag{4}$$

Where, $\mathbf{g}_c$ is considered as the c th element in **g**.

C. NMF Representation Learning

The representative learning, **G** is carried out by many supervised or unsupervised method using NMF, since it

reduces the dimensionality in an effective manner. Certain other techniques uses Euclidean space to conduct learning on **G** [50]. However, the non-orthogonality problem during the representative learning process is not dealt and hence, the proposed study uses such problem to reduce the dimensionality in large datasets.

IV. Cholesky Decomposition

The main reasons for the non-orthogonality problem during the representative learning (**G**) is the formulation of distance (squared) between the paired instances (g_i, g_j) as $(g_i - g_j)^T (g_i - g_j)$. The squared distance is implicitly assumes that g_i lies in Euclidean space. In general, the learning of $(f_1,...,f_q)$ using NMF are considered non-orthogonal to each other and the use of squared Euclidean distance is not appropriate during the representative learning by **G**. To solve this, generalized squared distance metric using Mahalanobis distance (**M**) measurement is used to solve the non-orthogonality of feature vector, which is represented as:

$$(g_i - g_j)^T \mathbf{M}(g_i - g_j) \tag{5}$$

The NMF property is exploited to decompose the data matrix, **X** into,

a) **F** with column vectors $(f_1, f_2,...,f_n)$ spans the feature space of the matrices, and

b) **G** provides the feature space representation.

With such decomposition property, the cdNMF,

1) Initially, the estimation of the NMF metric is carried out in feature space using the feature vectors (trained).

2) Then, the Cholesky decomposition is applied over the NMF metric, which finds the upper triangular matrix.

3) Finally, upper triangular matrix is used to map linearly the data vectors.

A. NMF Metric Estimation

In NMF, the data matrix (**X**) is approximated and it is represented in the feature space as **G** and the feature representation in the data space is **F**. The normalization [8] of f results in $f^T f = 1$ as the metric **M** is estimated as gram matrix **FGF** of the feature vector.

$$\mathbf{M} = \mathbf{F}^T\mathbf{F}, \ s.t. \ u_l^T u_l = 1, \ \forall l = 1,...,q \tag{6}$$

The metric estimation do not use label information for estimating **M** and the data vector is approximated over the feature space through $u_1,...,u_q$ and it is seen that $\mathbf{M} = \mathbf{F}^T\mathbf{F}$ can be used to estimate the feature space metric.

B. Cholesky Decomposition over NMF Metric

The estimation of metric using (5) guarantees **M** as symmetric positive semi-definite matrix. Linear algebra guarantees **M**, which decomposes the upper triangular matrix **T** using Cholesky decomposition:

$$\mathbf{M} = \mathbf{T}^T\mathbf{T} \tag{7}$$

By substituting (7) into (5), the Cholesky function to represent the upper triangular matrix **T** is given as:

$$\mathbf{G} \rightarrow \mathbf{TG} \tag{8}$$

C. Conjugate Gradients (CG)

Assuming, upper triangular elements to be sparse, hence linear representation of the data vectors is not considered valid. The use of CG for removing the sparse value in the matrices is found with the set of linear equations. The CG is applied on upper triangular matrix to remove the sparse value. The proposed method is used to utilize the trained representation using cdNMF, without any modifications in the algorithms over the learning representation. The elimination of sparse matrix is avoided by eliminating the rapid multiplication and clustering such data leads to increased convergence rate with faster association of elements in the dataset. Here, **M** = $(\mathbf{TG}^T\mathbf{TG})^{-1}$ is the pre-conditioner to enhance the multiplication process, in case of incomplete Cholesky decomposition, where **M** = $\mathbf{TG}^T\mathbf{TG}$ defines the incomplete Cholesky decomposition.

Algorithm 1 cdNMF

cdNMF (X, NMF, *q*, *parameters*)

1: **Find X** $\in \mathbb{R}^{p \times n}$, NMF, *q* and *parameter*(NMF)
2: **F, G** := run NMF on X with *parameter* and *q* // metric estimation
3: **M** := $\mathbf{F}^T\mathbf{F}$
4: **T**:=CD(**M**) *s.t.* **M** = $\mathbf{T}^T \mathbf{T}$
5: **Apply** CF once the linear coordinates changes, $x = \mathbf{TG}y$ and det **TG**$\neq 0$
6: **Use** CG for solving $\mathbf{TG}^T A\mathbf{TG}y = \mathbf{TG}^T b$
7: **Set** $x' = \mathbf{TG}^{-1}y'$
8: **Set** the preconditioner **M** = $(\mathbf{TG}^T \mathbf{TG})^{-1}$
9: **Multiply TG** by **TG**$^{-1}$
10: **Compute** $x' = \mathbf{TG}^{-1}y'$
11: **Return M**, x', **TG**

This algorithm helps in reducing increased multiplication process and increases the convergence rate. The computation of x' = $\mathbf{TG}^{-1}$y' is carried out only at the end of multiplying **TG** by **TG**$^{-1}$ and the computation process is multiplied with **M**.

V. Experimental Results

In the proposed system, the cdNMF is used to cluster the documents and compared with other algorithms to prove it effectiveness. The cdNMF system for evaluating the datasets is compared with conventional algorithms and that include: NMF [51], GNMF [5], NPNMF [6], MMNMF [7] and RNMF [8].

A. Text Mining Datasets

The proposed cdMNF with conjugate gradient is evaluated on text datasets: 20 Newsgroups data (Table 1), Reuters 21578 data (Table 2) and R52 data (Table 3).

TABLE I.   Attributes of Newsgroups Data

Class	Number of train documents	Number of test documents	Total Number of documents
soc.religion.christian	598	398	996
talk.politics.guns	545	364	909
talk.politics.mideast	564	376	940
talk.politics.misc	465	310	775
talk.religion.misc	377	251	628

TABLE II.   ATTRIBUTES OF REUTERS 21578 DATA

Reuters 21578				
Topics	Number of training documents	Number of test documents	Number of other documents	Total documents
0	1828	280	8103	10211
1	6552	2581	361	9494
2	890	309	135	1334
3	191	64	55	310
4	62	32	10	104

TABLE III.   ATTRIBUTES OF R52 DATA

R52			
Class	Number of training documents	Number of test documents	Total documents
Crude	253	121	374
Earn	2840	1083	3923
Interest	190	81	271
money-supply	123	28	151
Trade	251	75	326

TABLE IV.   SELECTION OF 20 SAMPLES FROM THE DATASETS

Samples	20 News Group	Reuters 21578	R52
#1	7	6	7
#2	6	7	7
#3	7	7	6
#4	8	7	5
#5	8	5	7
#6	7	5	8
#7	7	8	5
#8	5	8	7
#9	5	5	10
#10	5	10	5
#11	10	5	5
#12	10	0	10
#13	0	10	10
#14	10	10	0
#15	15	0	5
#16	0	15	5
#17	15	5	0
#18	0	0	20
#19	0	20	0
#20	20	0	0

Each document is represented as standard vector model [1] that contains occurrence of classes and terms in a document. Each document is represented as single line in the file and represented using a word or document class with TAB character, delimiting spaces and the terms. A total of 5 classes are used from each dataset with a set of training documents, test documents and other documents. A cluster is created with 5 classes of 20 Newsgroups data, Reuters 21578 data and R52 data.

Hence, three clusters are used in this study that includes a set of 4248, 21453 and 5045 documents, respectively for 20 Newsgroups, Reuters 21578 data and R52 data. The clusters with sub-clusters are classes are used to create the samples and a total of 100 documents from each sub-clusters of all the classes form the sample. Likewise, 20 such samples are created from the text datasets.

Here, the each text sample is conducted with pre-processing operations that include: trunc5 stemmer [52] and POS Tagger [53] and removal of stop words and finally it selects a total of 30000 words with mutual information in a larger perspective. The selection of sub-clusters for the sample formation is shown in Table 4.

B. Clustering Metrics

The performance of the clustering metrics is evaluated using two metrics Accuracy (*acc*) and Normalized Mutual Information (*NMI*). The parameter, *acc* is used to estimate the overall performance of the cluster, which is defined in the form of a fraction metric, $acc = \eta_t / \eta_{ov}$, where, η_t is the correctly clustered documents sample and η_{ov} is the overall amount of samples. The Mutual Information (MI) finds the interdependency between the variables and if the text variables are equal, then MI is zero and it is defined as:

$$MI(x,y) = \sum_{\hat{y} \in y} \sum_{\hat{x} \in x} p(x,y) \log\left(\frac{p(\hat{x},\hat{y})}{p(\hat{x})\,p(\hat{y})} \right)$$

Where, $p(\hat{x},\hat{y})$ is the joint Probability Distribution Function (PDF) of x and y, $p(\hat{x})$ and $p(\hat{y})$ are the marginal PDF of x and y. The MI provides information related to the amount of uncertainty measured between documents x and y and one documents reduces the uncertainty of the other documents. Entire information is shared between the documents if the value of MI is zero and vice versa. The Normalized MI or NMI is denoted as:

$$NMI(x,y) = \frac{MI(x,y)}{\max\left(E(x), E(y)\right)}$$

Where, $E(x)$ and $E(y)$ are the entropy of the document x and y.

C. Evaluation and Comparisons

NMF is the baseline algorithm, GNMF uses KNN graph with regularization term for preserving the structure of geometry, NPNMF uses local linear embedding and graph approach in NMF uses trained regularization term, MMNMF uses an eleven graph for exploring the multiple manifold data structure, RNMF adds noise in NMF and HNMF encodes the geometry into matrix factorization using hyper graph. These systems are tested against accuracy and Normalized Mutual Information (NMI) over the sample datasets.

The results of *acc* from Table 5 show that the cdNMF performs well than conventional schemes. Here, the performance of cdNMF increases gradually from samples 1 to 20. It is inferred that if the documents of similar dataset are more, the accuracy is more and it reduces when the 20 sample documents are equally distributed from similar clusters. The overall accuracy of cdNMF is slightly higher than HNMF and

GNMF but it has higher *acc* rate than MM-NMF, NPNMF and NMF. This is the same for NMI, shown in Table 6.

TABLE V.   ACCURACY OF DIFFERENT METHOD OF 20 SAMPLE DATA SETS

Samples	NMF	NPNMF	MM-NMF	RNMF	GNMF	HNMF	cdNMF
#1	0.4582	0.4801	0.4681	0.5142	0.6925	0.6841	0.7315
#2	0.4648	0.4872	0.5015	0.5345	0.7124	0.7241	0.7452
#3	0.4677	0.4877	0.5248	0.5913	0.7182	0.7518	0.7824
#4	0.4678	0.4908	0.5312	0.6011	0.7225	0.7622	0.7895
#5	0.4755	0.4912	0.5324	0.6025	0.7315	0.7693	0.7917
#6	0.4757	0.4979	0.5399	0.6104	0.7318	0.7721	0.7924
#7	0.5084	0.5012	0.5401	0.6201	0.7324	0.7741	0.7954
#8	0.5134	0.5032	0.5428	0.6215	0.7328	0.7779	0.8001
#9	0.5156	0.5089	0.5542	0.6225	0.7355	0.7815	0.8017
#10	0.5228	0.5206	0.5581	0.6241	0.7419	0.7826	0.8156
#11	0.5352	0.5273	0.5595	0.6255	0.7421	0.7892	0.8249
#12	0.5384	0.5478	0.5716	0.6258	0.7552	0.7912	0.8456
#13	0.5432	0.5501	0.5792	0.6293	0.7621	0.7924	0.8592
#14	0.5488	0.5591	0.5932	0.6309	0.7624	0.7927	0.8604
#15	0.5583	0.5858	0.6083	0.6385	0.7815	0.7935	0.8621
#16	0.5881	0.6078	0.6145	0.6515	0.7858	0.8019	0.8665
#17	0.5892	0.6378	0.6489	0.6922	0.8009	0.8245	0.8912
#18	0.5995	0.6584	0.6692	0.6945	0.8012	0.834	0.8942
#19	0.6087	0.6697	0.6745	0.7152	0.8241	0.8402	0.9018
#20	0.6724	0.6795	0.6845	0.7564	0.8512	0.8576	0.9156
Average	0.5326	0.5496	0.5748	0.6301	0.7559	0.7848	0.8284

TABLE VI.   NMI OF DIFFERENT METHOD OF 20 SAMPLE DATA SETS

Samples	NMF	NPNMF	MM-NMF	RNMF	GNMF	HNMF	cdNMF
#1	0.4789	0.5203	0.6113	0.753	0.7433	0.7854	0.7754
#2	0.4825	0.5504	0.6384	0.7586	0.7542	0.7884	0.7951
#3	0.5231	0.5648	0.6569	0.7769	0.7823	0.7885	0.7952
#4	0.5321	0.6212	0.6805	0.7872	0.7911	0.7946	0.8025
#5	0.5549	0.633	0.69	0.7875	0.8009	0.7952	0.8107
#6	0.6122	0.6749	0.7203	0.8124	0.8014	0.8095	0.812
#7	0.6199	0.6825	0.7254	0.8147	0.8306	0.8149	0.833
#8	0.6317	0.6842	0.7274	0.8164	0.8325	0.8412	0.8342
#9	0.6823	0.6924	0.7373	0.822	0.8336	0.8423	0.8466
#10	0.6842	0.7043	0.7504	0.8245	0.8365	0.8435	0.8587
#11	0.6902	0.7047	0.7582	0.8342	0.8424	0.8502	0.8629
#12	0.6948	0.7063	0.7592	0.84	0.8489	0.8567	0.8743
#13	0.7001	0.7137	0.7687	0.8412	0.8598	0.8676	0.8822
#14	0.7042	0.7158	0.7691	0.8489	0.8674	0.8781	0.8918
#15	0.7094	0.724	0.77	0.8695	0.8858	0.9001	0.8939
#16	0.7142	0.7288	0.7742	0.8731	0.8904	0.9079	0.9008
#17	0.7355	0.743	0.7745	0.9088	0.8998	0.9139	0.9153
#18	0.7412	0.7436	0.7769	0.9138	0.9335	0.9418	0.9441
#19	0.7419	0.7477	0.7787	0.9205	0.9356	0.9504	0.9514
#20	0.7496	0.765	0.8057	0.9309	0.9392	0.9505	0.9515
Average	0.6491	0.6810	0.7337	0.8367	0.8454	0.8560	0.8616

The average values of NMI results claim that the interdependence of documents belonging to similar cluster during testing is also high. It is seen further the documents are equal in number in sample dataset and the interdependency is less for other conventional algorithms, however, cdNMF performs well.

TABLE VII.   AVERAGE ACCURACY OF PROPOSED METHOD VS. EXISTING METHOD USING THREE DATA SETS

Cluster	NMF	NPNMF	MM-NMF	RNMF	GNMF	HNMF	cdNMF
20 News Group	0.8945	0.9342	0.8731	0.9261	0.9779	0.9822	0.9932
Reuters 21578	0.8591	0.8998	0.8779	0.8595	0.9286	0.9486	0.9524
R52	0.9023	0.9201	0.8769	0.8315	0.9599	0.9738	0.9869

TABLE VIII.   AVERAGE NMI OF PROPOSED METHOD VS. EXISTING METHOD USING THREE DATA SETS

Cluster	NMF	NPNMF	MM-NMF	RNMF	GNMF	HNMF	cdNMF
20 News Group	0.7602	0.6216	0.749	0.9382	0.9412	0.9551	0.9592
Reuters 21578	0.7605	0.7281	0.709	0.9186	0.9386	0.9491	0.9585
R52	0.7721	0.7054	0.6421	0.9282	0.939	0.9575	0.9621

The average values of the *acc* and the NMI test results for individual dataset is shown in Tables 7 and 8. It is seen that proposed cdNMF performs well with better accuracy to cluster the documents than conventional ones. Finally, the comparison with baseline NMF proves that the proposed cdNMF has better acc and NMI rate for the individual datasets.

VI.   CONCLUSIONS

In this paper, we present a new matrix factorization method called Cholesky Decomposition based non-negative matrix factorization. The Cholesky decomposition collects the data vector, specifically it avoids the non-orthogonality of the non-negative matrix factorization due to its local representation. Also, the presence of non-negative constraints is avoided finally with upper triangular matrix representation for mapping the data vectors. Further, the sparse matrix is eliminated using conjugate gradients, which takes hold of the complex conjugate values from the data vectors. Finally, better accuracy and normalized mutual information is obtained during the experimental validation and it enables better learning of the text data elements with reduced redundancy.

In future, we would like to improve the proposed approach on a graph based NMF framework that could generate better patterns to improve the learning representations of NMI for text mining.

REFERENCES

[1] http://www.cs.umb.edu/~smimarog/textmining/datasets/.

[2] Sang, Y. and Yi, Z., "Instance selection by using polar grids," ICACTE. IEEE 3rd International Conference on Advanced Computer Theory and Engineering, vol. 3, pp. V3-344, August 2010.

[3] García-Osorio, C., de Haro-García, A. and García-Pedrajas, N., "Democratic instance selection: a linear complexity instance selection algorithm based on classifier ensemble concepts," Artificial Intelligence, vol.174, no.5, pp.410-441, 2010.

[4]   Cano, J.R., Herrera, F. and Lozano, M., "Evolutionary stratified training set selection for extracting classification rules with trade off precision-interpretability," Data & Knowledge Engineering, vol.60, no.1, pp.90-108, 2007.

[5]   Liao, Q. and Zhang, Q., "Efficient Rank-one Residue Approximation Method for Graph Regularized Non-negative Matrix Factorization," In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, pp. 242-255, September 2013.

[6]   Gu, Q. and Zhou, J., "Neighborhood Preserving Nonnegative Matrix Factorization," In BMVC, pp.1-10, 2009.

[7]   Shen, B. and Si, L., "Non-Negative Matrix Factorization Clustering on Multiple Manifolds," In AAAI, pp. 575-580, July 2010.

[8]   Zhang, L., Chen, Z., Zheng, M. and He, X., "Robust non-negative matrix factorization," Frontiers of Electrical and Electronic Engineering in China, vol.6, no.2, pp.192-200, 2011.

[9]   Wang, D., Li, T., Zhu, S. and Ding, C., "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, pp. 307-314, July 2008.

[10]   Cooper, M. and Foote, J., "Summarizing video using non-negative similarity matrix factorization," In proceedings of 2002 IEEE Workshop on Multimedia Signal Processing, pp. 25-28, December, 2002.

[11]   Brunet, J.P., Tamayo, P., Golub, T.R. and Mesirov, J.P., "Metagenes and molecular pattern discovery using matrix factorization," Proceedings of the national academy of sciences, vol.101, no.12, pp.4164-4169, 2004.

[12]   de Fréin, R., Drakakis, K., Rickard, S. and Cichocki, A., "Analysis of financial data using non-negative matrix factorization," In International Mathematical Forum, Journals of Hikari Ltd, vol.3, no.38, pp.1853-1870, 2008.

[13]   Chi, Y., Zhu, S., Gong, Y. and Zhang, Y., "Probabilistic polyadic factorization and its application to personalized recommendation," In Proceedings of the 17th ACM conference on Information and knowledge management, pp. 941-950, October 2008.

[14]   Dhillon, I.S. and Sra, S., "Generalized nonnegative matrix approximations with Bregman divergences". In NIPS, vol.18, December 2005.

[15]   Sha, F., Saul, L.K. and Lee, D.D., "Multiplicative updates for nonnegative quadratic programming in support vector machines," In NIPS, pp.1041-1048, December 2002.

[16]   Wang, J., Zhou, J., Wonka, P. and Ye, J., "Advances in neural information processing systems," In Neural information processing systems foundation, 2013.

[17]   Ding, C.H., Li, T. and Jordan, M.I., "Convex and semi-nonnegative matrix factorizations," IEEE transactions on pattern analysis and machine intelligence, vol.32, no.1, pp.45-55, 2010.

[18]   Ding, C., Li, T. and Jordan, M.I., "Nonnegative matrix factorization for combinatorial optimization: Spectral clustering, graph matching, and clique finding," ICDM'08. Eighth IEEE International Conference on Data Mining, pp.183-192, December 2008.

[19]   Yu, D., Chen, N., Jiang, F., Fu, B. and Qin, A., "Constrained NMF-based semi-supervised learning for social media spammer detection," Knowledge-Based Systems, vol.125, pp.64-73, 2017.

[20]   Li, H., Zhang, J., Shi, G. and Liu, J., "Graph-based Discriminative Nonnegative Matrix Factorization with label information," Neurocomputing, 2017.

[21]   Hayashi, N. and Watanabe, S., "Upper Bound of Bayesian Generalization Error in Non-Negative Matrix Factorization," Neurocomputing. 2017.

[22]   Schachtner, R., Poeppel, G., Tomé, A.M. and Lang, E.W., "A bayesian approach to the lee–seung update rules for nmf," Pattern Recognition Letters, vol.45, pp.251-256, 2014.

[23]   Du, J.X., Zhai, C.M. and Ye, Y.Q., "Face aging simulation and recognition based on NMF algorithm with sparseness constraints," Neurocomputing, vol.116, pp.250-259, 2013.

[24]   Xu, Y., Deng, S., Li, X. and He, Y., "A sparse unmixing model based on NMF and its application in Raman image," Neurocomputing, vol.207, pp.120-130, 2016..

[25]   Feng, Y., Xiao, J., Zhou, K. and Zhuang, Y., "A locally weighted sparse graph regularized Non-Negative Matrix Factorization method," Neurocomputing, vol.169, pp.68-76, 2015.

[26]   Liao, Q. and Zhang, Q., "Local coordinate based graph-regularized NMF for image representation," Signal Processing, vol.124, pp.103-114, 2016.

[27]   Shang, F., Jiao, L.C. and Wang, F., "Graph dual regularization non-negative matrix factorization for co-clustering," Pattern Recognition, vol.45, no.6, pp.2237-2250, 2012.

[28]   Wang, J.J.Y., Bensmail, H. and Gao, X., "Multiple graph regularized nonnegative matrix factorization," Pattern Recognition, vol.46, no.10, pp.2840-2847, 2013.

[29]   Li, X., Shen, X., Shu, Z., Ye, Q. and Zhao, C., "Graph regularized multilayer concept factorization for data representation," Neurocomputing, vol.238, pp.139-151, 2017.

[30]   Wang, J.J.Y., Huang, J.Z., Sun, Y. and Gao, X., "Feature selection and multi-kernel learning for adaptive graph regularized nonnegative matrix factorization," Expert Systems with Applications, vol.42, no.3, pp.1278-1286, 2015.

[31]   Zeng, K., Yu, J., Li, C., You, J. and Jin, T., "Image clustering by hyper-graph regularized non-negative matrix factorization," Neurocomputing, vol.138, pp.209-217, 2014.

[32]   Sun, F., Xu, M., Hu, X. and Jiang, X., "Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation," Neurocomputing, vol.173, pp.233-244, 2016.

[33]   Ou, W., Yu, S., Li, G., Lu, J., Zhang, K. and Xie, G., "Multi-view non-negative matrix factorization by patch alignment framework with view consistency," Neurocomputing, vol.204, pp.116-124, 2016.

[34]   Qian, C., Zhuang, Y. and Xu, Z., "Visual tracking with structural appearance model based on extended incremental non-negative matrix factorization," Neurocomputing, vol.136, pp.327-336, 2014.

[35]   Wang, D. and Lu, H., "On-line learning parts-based representation via incremental orthogonal projective non-negative matrix factorization," Signal Processing, vol.93, no.6, pp.1608-1623, 2013.

[36]   Wang, Y., Wu, S., Mao, B., Zhang, X. and Luo, Z., "Correntropy induced metric based graph regularized non-negative matrix factorization," Neurocomputing, vol.204, pp.172-182, 2016.

[37]   Rad, R. and Jamzad, M., "Image annotation using multi-view non-negative matrix factorization with different number of basis vectors," Journal of Visual Communication and Image Representation, vol.46, pp.1-12, 2017.

[38]   Arabnejad, E., Moghaddam, R.F. and Cheriet, M., "PSI: Patch-based script identification using non-negative matrix factorization," Pattern Recognition, vol.67, pp.328-339, 2017.

[39]   Zong, L., Zhang, X., Zhao, L., Yu, H. and Zhao, Q., "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," Neural Networks, vol.88, pp.74-89, 2017.

[40]   Chung, H., Plourde, E. and Champagne, B., "Regularized non-negative matrix factorization with Gaussian mixtures and masking model for speech enhancement," Speech Communication, vol.87, pp.18-30, 2017.

[41]   Li, X., Zhang, W. and Dong, X., A class of modified FR conjugate gradient method and applications to non-negative matrix factorization. Computers & Mathematics with Applications, vol.73, no.2, pp.270-276, 2017.

[42]   Han, J., Pei, J. and Kamber, M., "*Data mining: concepts and techniques*", Elsevier, 2011.

[43]   Tan, P.N., "*Introduction to data mining*," Pearson Education India, 2006.

[44]   Zheng, Z., Yang, J. and Zhu, Y., "Initialization enhancer for non-negative matrix factorization," Engineering Applications of Artificial Intelligence, vol.20, no.1, pp.101-110, 2007.

[45]   Xue, Y., Tong, C.S., Chen, Y. and Chen, W.S., "Clustering-based initialization for non-negative matrix factorization," Applied Mathematics and Computation, vol.205, no.2, pp.525-536, 2008.

[46] Ding, C., Li, T., Peng, W. and Park, H., "Orthogonal nonnegative matrix t-factorizations for clustering," In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.126-135, August 2006.

[47] Wild, S., Curry, J. and Dougherty, A., "Improving non-negative matrix factorizations through structured initialization," Pattern recognition, vol.37, no.11, pp.2217-2232, .

[48] Xu, W., Liu, X. and Gong, Y., "Document clustering based on non-negative matrix factorization," In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp.267-273, July 2003.

[49] Ding, C., Li, T., Peng, W. and Park, H., "Orthogonal nonnegative matrix t-factorizations for clustering," In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 126-135, August 2006.

[50] Kamvar, K., Sepandar, S., Klein, K., Dan, D., Manning, M. and Christopher, C., "*Spectral learning*," In International Joint Conference of Artificial Intelligence, Stanford InfoLab, April 2003.

[51] Cai, D., He, X., Wu, X. and Han, J., "Non-negative matrix factorization on manifold," ICDM'08. Eighth IEEE International Conference on Data Mining 2008, pp.63-72, December 2008.

[52] Flores, F.N. and Moreira, V.P., "Assessing the impact of Stemming Accuracy on Information Retrieval – A multilingual perspective," Information Processing & Management, vol.52, no.5, pp.840-854, 2016.

[53] Ferro, M.V., Bilbao, V.M.D and Pena, F.J.R., "Modeling of learning curves with applications to POS tagging," Computer Speech & Language, vol.41, pp.1-28, 2017.