Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 23, Issues 4, pp. 275–289, DOI 10.12694/scpe.v23i4.2030

# A META HEURISTIC MULTI-VIEW DATA ANALYSIS OVER UNCONDITIONAL LABELED MATERIAL: AN INTELLIGENCE OCMHAMCV

SRINIVAS KOLLI [*], PRAVEEN KRISHNA A.V [†], AND M. SREEDEVI[‡]

**Abstract.** Artificial intelligence has been provided powerful research attributes like data mining and clustering for reducing bigdata functioning. Clustering in multi-labeled categorical analysis gives huge amount of relevant data that explains evaluation and portrayal of qualities as trending notion. A wide range of scenarios, data from many dimensions may be used to provide efficient clustering results. Multi-view clustering techniques had been outdated, however they all provide less accurate results when a single clustering of input data is applied. Numerous data groups are conceivable due to diversity of multi-dimensional data, each with its own unique set of viewpoints. When dealing multi-view labelled data, obtaining quantifiable and realistic cluster results may be challenge. This study provides unique strategy termed OCMHAMCV (Orthogonal Constrained Meta Heuristic Adaptive Multi-View Cluster). In beginning, OMF approach used to cluster similar labelled sample data into prototypes of dimensional clusters of low-dimensional data. Utilize adaptive heuristics integrate complementary data several dimensions complexity of computational analysis data representation data in appropriate orthonormality constrained viewpoint. Studies on massive data sets reveal that proposed method outperforms more traditional multi-view clustering techniques scalability and efficiency. The performance measures like accuracy 98.32%, sensitivity 93.42%, F1-score 98.53% and index score 96.02% has been attained, which was good improvement. Therefore it is proved that proposed methodology suitable for document summarization application for future scientific analysis.

**Key words:** Clustering, Document summarization, Data mining, Meta heuristic technique.

**AMS subject classifications.** 68T05

## 1. INTRODUCTION.

Large amounts data gathered from several study fields, such image processing, computer vision data fusion, natural language and processing in real time as result of fast computer-related technology deployment. A wide range of dimensions associated to a wide variety of properties are examined in these data, which include many high-dimensional features with complicated structures [1, 2]. High-dimensional data represents the abundance of data curse dimensionality, therefore managing high-dimensional data general concern big challenge for optimizing the dimension's dimensions. Using hidden data to represent low dimensionality and reducing dimensionality in relation to input data is an effective method for large amounts of data [3].

Theoretically optimized matrix factorization has emerged as the research hotspot with the easiest implementation for multi-labeled data reduction. It's possible extract low-dimensional attribute relations high-dimensional data relations using factorization matrix-related methodologies such ICA (Independent Component Analysis), PCA (Principal Component Analysis), & VQ (Vector Quantization). No components in matrices are decomposed; this implies that in order to maximize matrix representation, negative elements must be included in the low-dimensional representations of data [4]. Deep learning has recently proven exceptional performance in include representation projects [5]. Lattice factorization has been enriched by various analysts who have incorporated substantial learning into the process [6]. A multi-layer non-negative MF technique presented (MNMF) [7]. First, MNMF degraded the grid many times to produce the fundamental part-based representation that may remove profoundly different degrees of information from the original information. To propose

[*]Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India (kollisreenivas@gmail.com).,

[†]Associate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India (praveenkrishna@kluniversity.in).

[‡]Professor, Department of CSE, Amrita Sai Institute of Science and Technology, Andhra Pradesh, India. (srivijaysujay@gmail.com).

thorough semi-non-negative grid factorization strategy [8, 9] employed Semi- non-negative MF (SNMF) and coordinated deep Factorization [10, 11]. But the deterioration of the coefficient network for preparation information in both MNMF and deep semi-NMF can only be seen as a profound decay in this network [12, 13]. The premise grid was used to minimize dimensionality new data problems construct new data issues [14, 15]. When premise framework applied to the deep representation, it had a direct impact on the outcome. Using factorization, investigated the accurate reduction of dimensions with depth of initial information framework, and they presented a profound NMF technique that relies on premise picture learning [16, 17].

Invariant data acquired from several data domains/sources may create this issue. Represent data several dimensions depending their representation features and relationships with multiple features, this challenge tackled using multi labelled clustering [18, 19]. Different techniques to multi-view clustering have been explored before, but they have not been analyzed in terms of dimensionality reduction in the representation of multi labelled data in supervised learning. In order to identify the text, the supervised learning technique defines the labelled information based on its features [20, 21]. It is thus crucial to find data sources for unsupervised learning with multiple labelling [22, 23]. Because finding quantifiable and realistic cluster outcomes multi-view labelled data still challenge, new technique called OCMHAMVC being offered representing Data as cluster with many kinds. To begin, first suggested methodology examines low-dimensional data using OMF model, clusters comparable labeled sample data prototype clusters data related several dimensions [24, 25].

The following are the main goals of the suggested method:

1. Unsupervised multi-labeled clustering method uses orthonormality matrix factorization (mix normality constraints and orthogonal constraints) is first proposed.
2. Objective model we implement an objective model, which provides and expresses the minimums of the suggested implemented model
3. In order to demonstrate efficiency of the suggested technique compared to standard approaches accuracy, other metrics multi-labeled cluster data sets, we conduct tests on numerous real time datasets.

**2. Review of Related Work.** Using multi-labeled data, this section explores the relationship between standard clustering methods and their results.

Prior to this time, a wide variety of single-view data grouping solutions had been discussed and implemented. Three typical single-sight collection strategies piece bunching [22, 23, 24], paranormal groups [26, 27] & subspace groups [28, 29]. Most part, bit-based methods employed build primary commitments High-Dimensional piece space where proffered grouping successfully [30]. employ Gaussian piece design commitments split region and wire pair- savvy constraints into part sorting out some methods co-ordinate pattern collecting [31] pre-shown bits used design data sources and improvement piece game plan better encourage bunching execution [31, 32], optimum portion space picked from social event predetermined portions. In the run-up to the terrible grouping, they normally produce a partiality diagram to describe information similitude and analyses Eigen structure this affection diagram acquire clustering conclusion. To create proclivity chart, existing clustering methods [33, 34] show efficient strategy.

As an example, [35] construct the inborn diagram and the punishment chart independently using pair-wise requirement data. In [36], an incredible affection graph is fostered by using discriminative component subspaces. Mishandling multi-layer layers in the form of a pyramid-style structure by [37, 38, 39] creates a reformist bipartite diagram. Besides the foregoing, the collecting structure makes good use of a help vector machine (SVM). In [40] introduced Twin Help Vector Machine (TWSVC) framework, discover bundle plane close contrasting pack places avoiding characteristics other groups. Matrix factorization (NMF applied range MVC assessment methodologies non-threat impediments consider improved inter-predibility (Guan et al 2020; Trigeorgis et al 2018). Using non-negative structural factorization of multiple datasets, an average inert factor is discovered (Liu et al 2018; Zhang et al 2019, 2020). Semi-NMF has been projected to broaden NMF by relax the factorized premise structure to be actual performances. Semi-NMF is one of the most prevalent variants of Non-negative Matrix Factorization. Because of this preparation, semi-NMF may be employed in a far wider variety of applications than NMF [41, 42]. Additionally, our method offers many advantages over NMF-based MVC techniques, including the ability to analyze Semi-NMF in more depth [43, 44]. Data tests from a related class may be brought closer together using the Semi-NMF structure, as mentioned. This dish has a particular flavor since it is based on extensive education [45]. Even though our approach isn't exactly the same as the current MVC auto-encoder-based solutions, we still have a major strategy (Andrew et al 2019; Wang et al 2020). When compared to previous research, we don't employ Canonical Correlation Analysis (CCA), restricted two cases. An Andrew et al 2020, as well as Wang et al, 2018. In this metaheuristic nature-based algorithm, an improved meta-heuristic methodology motivated by different researcher's studies are provided. One important model called Oppositional based Harris Hawk Optimizer-OHHO technique; it is an unsupervised algorithm. The following approach is constructed on the speculative Harris Hawk Optimize-HHO algorithm, which has no inherent dependent variables. The search space is later changed by integrating HHO with the Oppositional Based Learning-OBL method in order to provide better estimation for the dominant approach. A clustering strategy is also been discussed on unsupervised learning, known as OCMHAMVC [46]. The deep learning related scientific paper [47] has been taken as input files and applied various clustering techniques on it getting various clustered index values. The all literature survey section giving limitations of earlier studies related to document indexing. The survey which was analyzed has been taken as reference and proposed an advanced technology.

**3. Preliminaries.** This section explains fundamental preliminaries employed suggested strategy, along with the relevant procedures that are required.

**3.1. Optimized Matrix Factorization (OMF):.** This data is provided in $A = \{a_1, a_1, .......a_n\} \in M_+^{n\times d}$, where $n$ indicates no. of various examples, and d indicates its dimension feature vector $a_j$ $(1 \le q \le n)$ represented number of distinct samples in **(** $H = \{h_1, h_2, ....., h_d\} \in Q_+^d \& W = \{w_1, w_2, ....., w_d\} Q_+^d$ $(k \prec\prec n \& d \prec D)$**).**

To find reduced rank matrices inside non-negative relationships, OMF investigates, i.e., which may be characterized $W = \{w_1, w_2, ....., w_d\}$ Input data examined as $a_j = \sum_{i-1}^{d} h_i w_{ji}$ stated using combination of linear matrix construction and impact factor wi after exploring matrix relations (i.e. H & W). Here is a breakdown of the purpose of non-negative matrix formation:

$$\min_{H,W} ||A - HW||_F^2 \;\; w.r.t \quad H \ge 0, W \ge 0 \tag{3.1}$$

$||||_F$ Fresenius standard procedure with variety functions
$W$ and $H$ are described as variables in the Karush-Kuhn-Tucker (KKT) ruling condition,

$$H_{il} = H_{il} \frac{\left(AW^T\right)_{il}}{\left(HWW^T\right)_{il}} \tag{3.2}$$

With associative parameters,

$$W_{ij} = W_{ij} \frac{\left(H^T A\right)_{ij}}{\left(W^T HW\right)_{ij}} \tag{3.3}$$

Factorization DL described as

$$\begin{gathered} W^m \simeq H_1^m W_1^m \\ H_1^m \simeq H_2^m W_2^m \\ ......... \\ H_{1-2}^m \simeq H_{l-1}^m W_{l-1}^m, \\ H_{l-1}^m \simeq H_l^m W_l^m \end{gathered} \tag{3.4}$$

where $H_1^m, H_2^m, ...., H_{1-1}^m, H_1^m \& W_1^m, W_2^m, ...., W_{1-1}^m, W_1^m$ Coefficient and basis matrices m-dimensionality referred matrices. As result of combining the two equations above,

$$\min_{W_l^m, H_l^m} ||A^m - H_l^m W_l^m W_{l-1}^m, ...., W_2^m W_l^m||_F^2 \;\; w.r.t \quad H_l^m \ge 0, W_l^m \ge 0 \tag{3.5}$$

It is defined as: based on numerous viewpoints with objective functionality

$$\min_{W_l^m, H_l^m} \sum_{m-1}^{M} ||A^m - H_l^m W_l^m W_{l-1}^m, ...., W_2^m W_l^m||_F^2 \;\; w.r.t \quad H_l^m \ge 0, W_l^m \ge 0 \tag{3.6}$$

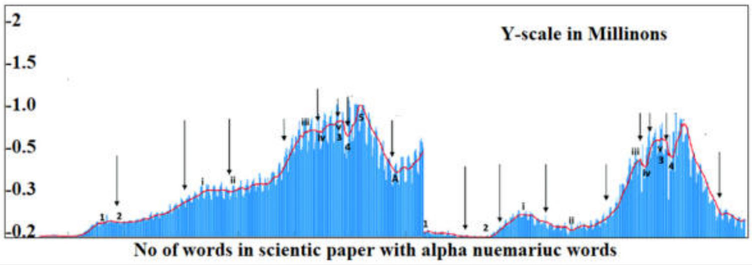


Fig. 4.1: Number of words scientific paper with unstructured scale

Objective functionality multiple attribute describes

$$\min_{H_l^m, W_l^m} ||A^m - H_l^m W_l^m W_{l-1}^m, ...., W_2^m W_l^m||_F^2 \ \ w.r.t \quad H_l^m \geq 0, W_l^m \geq 0 \tag{3.7}$$

**3.2. In-depth Matrix Factorization Indices.** Associative modularity's associative modularity may be used to design or investigate structure in complicated procedures. Deep learning is used to describe the process of discovering optimal matrix functions.

$$\begin{gathered} A \simeq Z_1 W_1 \\ A \simeq Z_1 Z_2 W_2 \\ \dots\dots\dots \\ A \simeq Z_1 Z_2 .... Z_m W_m \end{gathered} \tag{3.8}$$

$$Z_l \in Q^{K_{t-1} \times K_t} \ \ Be \quad l-th(l \leq m) \tag{3.9}$$

$$W_l \in \Re^{k_i \times n} \, (> 0) \tag{3.10}$$

Matrixes are clustered into clusters or factored out of a matrix with$Z_l, ...., (Z_1, Z_2, Z_t) \in \Re^{d \times k_i}$ m-dimensional layers by taking this connection into account, which may be done in a variety of ways. In this way, distinct data sets may be represented using the same group procedures, but from different viewpoints. For multi-labeled clustering numerous attribute relations, deep matrix factorization approach is appropriate. Our first findings lead propose new heuristic approach investigating multi-labeled data clustering enhanced matrix construction.

**4. Proposed Method.** A clustering strategy based on unsupervised learning, known as OCMHAMVC, is discussed in this section. It is associated restrictions related to orthogonal and combined frameworks with co-regularization, and this approach is referred to here. First define multi objective functions suggested and then examine optimum maximization strategy cluster multi labelled data. Lastly, efficient complexity analysis, computational analysis suggested approach. The scientific data is applied to proposed methodology interims of alpha numeric unstructured data.

Figure 4.1 clearly explains about scientific data analysis, millions of words are applied to proposed OCMHAMVC and get easy indexing. This analysis giving multi-view clustering with more accuracy and throughput. The multi objective function of an optimization issue with numerous objective functions is referred to as a multi-objective optimization problems. A multi-objective optimization problem might be described mathematically as

$$min_{x \epsilon X} (f_1(x), f_2(x), ..., f_k(x)) \tag{4.1}$$

**4.1. Multi objective functions of OCMHAMVC.**

$$\{A^u \in \Re^{m_u n}\}_{u=1}^{n_u} \tag{4.2}$$

For the sake of argument, let's say we have an efficient multi-labeled data collection, which is made up of n classes of samples each with a separate multi-label. Labeled constraints are represented in multiple perspectives as l samples, whereas unlabeled constraints are represented as l-1 samples.

$$X = \begin{pmatrix} C_{c*l} & 0 \\ 0 & I_{n-1} \end{pmatrix} \tag{4.3}$$

$I^{th}$ and $j^{th}$-class attribute data $C_{ij}$=0 as a result, $C_{ij}$ the data were given new labels, and $(n-1)(n-l)$ unlabeled sample $l_n$-1 was assigned to process of building identity matrix. If number samples unlabeled data exceeds predetermined threshold, then an identity matrix may be used to look for previously labelled constraint data. When a cluster is not identified, labelled matrix data generation is referred to as a "labelled matrix."

$$X = \begin{pmatrix} C_{c*l} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_{n-l-1} \end{pmatrix} = \begin{pmatrix} C_{c*l} & 0 \\ 0 & I_{n-1} \end{pmatrix} \tag{4.4}$$

Define auxiliary matrix $Z$, which expresses dimensions using sample data and numerous goal functions for orthogonal matrix creation as follows.

$$MO_F = \sum_{u=1}^{n_u} \theta_u ||A^u - V^u (Z^u)^T X^T||_F^2 + \lambda \sum_{u=1}^{n_u} ||Fo(Z^u(Z^u)^T) - I||_F^2 + \sum_{u=1}^{n_u}\sum_{s=1}^{n_u} \frac{1}{2}\theta_{us}||Z^u - Z^s||_F^2 \tag{4.5}$$

Explore low-dimensional feature representation executed data using dimension multi-objective function. Investigate representations of desired characteristics for each and every one of the dimensions. It is essential that the qualities across classes have an effective distinguishing factor, and that the scalability of all chosen features is the same. The introduction of orthogonal constraints in multi-labeled clustering ensures the desired feature representation is met. Associative clustering prototypes with various parametric notations are used in conjunction with low-dimensional feature representations under orthogonal constraint to effectively discriminate between classes and attributes. A joint constraint matrix is offered as a means to migrate the performance of orthogonal constraints. The padding of clustered information has been providing less congestion as well as getting fast dimensionality grouping. The following analysis helping to remove congestion and providing document analysis effectively.

$$F_{ij} \begin{cases} 1 & j = i \quad\quad 1 \leq j, i \leq c \\ 0 & otherwise \quad\quad 0 \end{cases} \tag{4.6}$$

As a result, the orthogonal constraint framework $\lambda \sum_{u=1}^{n_u} ||Fo(Z^u(Z^u)^T) - I||_F^2$

We can next determine whether or not an orthogonal constraint relation has controllability by looking at how o is represented in various notational systems. Under the clustering specification structure, representation many forms various dimensions include unity data.

**4.2. Convex Feature Optimization.** As a result objective non-convex functions based global minimum relations, variables examined applying them in conjunction. Using the most recent constraints, recalculate the optimization, such that one attribute relationship is linked to other attributes that are existent but are not changed in any way. To include non-negative matrix relations into Lagrange matrices, non-negative attribute relations with convex optimization are stated as

$$\begin{aligned} La\_r = {} & \textstyle\sum_{u=1}^{n_u} \theta_u ||A^u - V^u(Z^u)^T X^T||_F^2 + \lambda \sum_{u=1}^{n_u} ||Fo(Z^u(Z^u)^T) - I||_F^2 + \sum_{u=1}^{n_u}\sum_{s=1}^{n_u} \frac{1}{2}\theta_{us}||Z^u - Z^s||_F^2 \\ & + \textstyle\sum_{u=1}^{n_u} tr(\beta^u (V^u)^T) + \sum_{u=1}^{n_u} tr(\alpha^u (Z^u)^T) \end{aligned} \tag{4.7}$$

Implemented multi-labeled characteristics are given the KKT treatment by way of non-negative matrix relations are very important in documents analysis as well as Lagrange matrices. The convex optimization methodologies have been briefly concentrating on attributes and providing accurate document analysis. The non-negative Lagrange matrix is providing computations very smoothly compared earlier matrixes

$$\left(A^u X Z^u - V^u (Z^u)^T X^T X Z^u\right)_{ij} v^u_{ij} = 0 \tag{4.8}$$

$$\begin{aligned}&\theta_u X^T (A^u)^T V^u - \theta_u X^T X Z^u (V^u)^T V^u - 2\lambda(Fo(Z^u(Z^u)^T)Z^u + 2\lambda Z^u - 4\lambda F(Z^u.Z^u.Z^u)\\ &+4\lambda F(Z^u.Z^u.Z^u) - \textstyle\sum_{s=1}^{n_u} \theta_{us} Z^s_{ji} Z^u_{ji} = 0\end{aligned} \tag{4.9}$$

Finally, the multi-objective function is said to as being achieved, Descriptor extraction & utilization has been involved in document clustering. The Word groups known as descriptors which are used to characterize, elements of a cluster. In general, documents clustering is viewed as a centralized procedure also web document clustering for customers of search engines is an example of Document clustering. Online & offline applications of document clustering could be distinguished, especially comparing to offline apps, performance issues typically limit online applications. Text clustering could be utilized for a variety of purposes, including gathering related documents (news, tweets, etc.), analyzing customer & employee feedback, as well as identifying significant implicit subjects in all scientific datasets.

$$v^u_{ji} \leftarrow v^u_{ji} \frac{(A^u X Z^u)_{ji}}{(V^u (Z^u)^T X^T X Z^u)_{ji}} \tag{4.10}$$

$$z^u_{ji} \leftarrow z^u_{ji} \frac{\left(\theta_u X^T (A^u)^T V^u + 2\lambda Z^u + 4\lambda F(Z^u.Z^u.Z^u) + \sum_{s=1}^{n_u} \theta_{us} Z^s\right)_{ji}}{(\theta_u X^T X Z (V^u)^u V^u + 2\lambda(Fo(Z^u(Z^u)^T))Z^u + 4\lambda F(Z^u.Z^u.Z^u) + \sum_{s=1}^{n_s} \theta_{us} A^u)_{ji}} \tag{4.11}$$

Above, following description how objective algorithm created:

**Algorithm for Multi Labeled Clustering**

**Input:** multi labeled data set $\{A^1, A^2, ......, A^{n_u}\}$**,** No. of Clusters, No. of Samples, dissimilar variables$\theta_v, \theta_{vs}$

**1** Form constraint labeled matrix $X$

**2** Form optimized constraint matrix $F$

**3** For $u = n - 1$ then

- **a.** Normalization factors i.e. $A^u \left(||A^u(:, j)||^2\right)$ ;
- **b.** Update primary parameters V$^u$ & Z$^u$ plotted region [1, 0]

**4** E-For

**5** For u=1 n$_u$ then

- **a.** execute matrix building based on no. of iterations < T
- **b.** Constraint V$^u$ then update Z$^u$
- **c.** Constraint Z$^u$ then update V$^u$

**6** E-For

**7** estimate indication low-dimensional data i.e. U$^v$=XZ$^u$

**8** calculate final indication low-dimensional data $U^* = \frac{\sum_{u=1}^{n_u} U^u}{n_v}$

**Output:** final multi labeled cluster result

Convergence multi-dimensional data clustering using this approach.

**4.3. Multi labeled Dimensional clustering.** Our first cluster function is based on the similarity measure, and our average similarity weight measure is derived from documents in the same cluster. We then use this measure to create a multi-label clustering. Figure 4.2 depicts an architecture for exploring Multi labeled Dimensional clustering.

Multi-dimensional cluster results collection is shown in step-by-step detail in Figure 4.2. Weighted similarity was obtained using this method.

$$CS = \sum_{r=1}^{n} m_r \left[\frac{1}{m_t^2} \sum_{t_i, t_j S_r} S(t_i, t_j)\right] \tag{4.12}$$

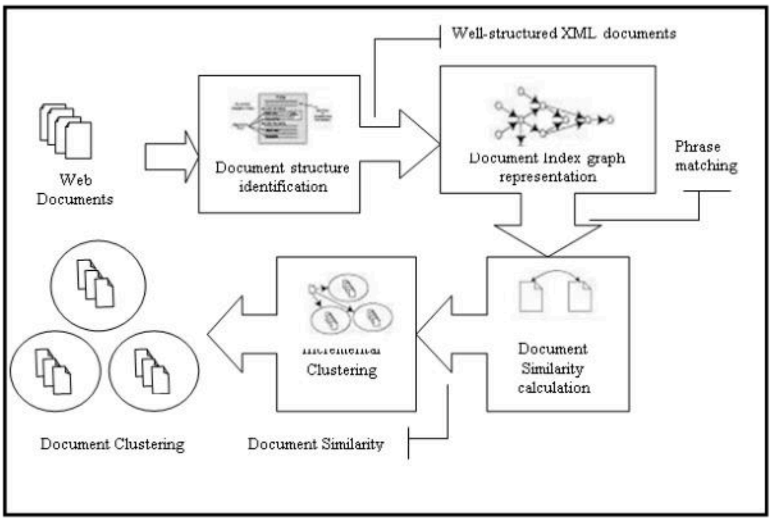


Fig. 4.2: Architecture proposed method

Class labels are n, and m is number of documents. D want to improve functionality so that we can locate functions that are comparable to

$$\begin{aligned}&\textstyle\sum_{t_i,t_j\in S_r} S(t_i,t_j)=\sum_{t_i,t_j\in S_r} t_i^d, t_j-\frac{2m_r}{m-m_r}\sum_{t_i\in S_r} t_i^t \sum_{t_m\in S\setminus S_r} t_h+n_r^2\\&=T_r^dT_r-\tfrac{2m_r}{m-m_r}T_r^d(T-T_r)+m_r^2\\&\tfrac{m+m_r}{m-m_r}||T_r||^2-\tfrac{2m_r}{m-m_r}T_r^dT+m_r^2\end{aligned} \tag{4.13}$$

Using weighted cluster functions to maximize the similarity of the document relations.

$$CF=\sum_{r=1}\frac{1}{m_r}\left[\frac{m+m_r}{m-m_r}||T_r||^2-\left(\frac{m+m_r}{m-m_r}-1\right)T_r^dT\right] \tag{4.14}$$

Comparing min-max functionality between words input documents to maximized weighted cluster functions, it assesses the depending on the maximum weight of the attribute. Incorporating cluster-related documents' functionality is represented as

$$\overline{CF}=\sum_{r=1}^{k}\frac{\lambda_r}{m_r}\left[\frac{m+m_r}{m-m_r}||T_r||^2-\left(\frac{m+m_r}{m-m_r}-1\right)T_r^dT\right] \tag{4.15}$$

Weighted functionality (Y) related document clustering rated most efficient.

$$Y=\sum_{r=1}^{k}\sum_{t_i\in S_r}\frac{1}{m-m_r}\sum_{t_h\in S\setminus S_r} S\left(t_i-t_h,\frac{C_r}{||C_r||}-t_h\right) \tag{4.16}$$

Discuss optimization multi-labeled clustering weighted cluster functions based similarity underlying cluster creation numerous multi-labeled capabilities.

$$MLC_{opt}=\sum_{r=1}^{k} I_r(m_rT_r) \tag{4.17}$$

Create optimum clusters outcome convergence clusters matrix relation measure depending number of iterations applied various input label data during multi-labeled clustering distinct sources. Because of the client base's exponential development has been imported, it is important to identify the users who contribute the most useful data. Although the relevance of individual participants in the recommender systems could increase the recommender system's robustness & suggestion efficiency, there has n't been much study done along this line to find a more efficient approach. They suggest an approach with multi-label clustering to identify the core user as well as establish the idea of correlation among user & label cluster in order to address this issue. The following issue has been solved with this proposed methodology and results are giving proofs.

**5. Experimental Evaluation of OCMHAMVC.** A comparison of OCMHAMVC's performance with that of standard techniques is shown in this section. Experimental data, OCMHAMVC used measure different weighted cluster functions clustering multi-labeled document clusters, measure clustering functions work based on Euclidean similar distance, similar cosine and relative jacquard co-efficient similar measurement.

The clustering data assortment is an advanced version data analysis, many earlier models has been unable provided. The proposed method has been optimizing the infromation as wells providing deep extraction of data on scientific document. The understandability and transferability of optimized document analysis was providing deep information about scientific document and giving good performance measures.

The cluster prototypes are not that much efficient but proposed model has been getting many rules from algorithm and giving solution to padding problems. The cooperative solution and regularization have been called through clustering steps.

**5.1. Input Clustering data.** We employed real-time benchmark data before we used Reuter's 08-10 versions of k1b for clustering of documentation, as well as additional standard datasets from efficient & exhaustive data sources, in this experiment for multi-labeled document clustering. Clustering applications may be conducted in real time using downloaded data sets and cloud-related data sources that have measurable similarity. For the most part we utilize the BBC Series data set, Reuter's datasets, Series 3 sources, and MSRC dataset (http://mlg.ie/datasets/3sources.html, http://lig-membres.imag.fr/grimal/data.html, http://www-vision-cal tech-edu/Image Datasets /Caltech101.html, https://pgram-com/dataset/msrc-v1/) (entertainment, politics, and sports, medical and business related applications). It was found that the suggested technique outperformed other multi-dimensional clustering methods when evaluated on all of the datasets mentioned above. We have GMNMF (multi view non-negative matrix factorization) , CoNMF-P [3], and MVCC [4], all of which use non-negative comments as a basis for their non-negative matrix factorization results.

The accuracy, NF, lacquard coefficient, precision, recall, F1 measure, presentation computing cost and memory consumption are performance measures which are deciding the application stability and comparing earlier models. The proposed methodology attains more improvement and suitable for traditional document analysis algorithm.

**5.2. Setting of Experiments.** Samples from numerous data sources are randomly gathered, according to the authors' original publications, and the tagged data is eliminated. Search parameter weights with various notations are then applied to each sample in order to compare traditional techniques with the novel approach. With the use of multi-labeled dimensional clustering and Euclidean distance metrics, we conduct our research. Using these metrics, evaluate accuracy, NF, lacquard coefficient, precision, recall, F-score, presentation computing cost & memory consumption each cluster dataset proposed to approach. Document retrieval accuracy multiple labels proposed method shown in Figure 5.1 in comparison other well-established strategies.

The confusion matrix is used to define measures like accuracy, sensitivity, recall and precision. The measures can be decided by collection of true positive, true negative rate, false positive rate and false negative rate via statistical information. The clustering process is performed through many ways but in our research class based analysis were performed. The selected scientific document is applied to our designed application it can differentiated the information into multi class variance, by using following analysis document has been clustered.

On a sample of 100-500 html text documents, five alternative clustering algorithms are shown in Figure 5.1, multi-labeled data collection showing best accuracy. Clustering results displayed variety ways depending on the accuracy values provided in table 5.1, every dataset contains row, best value shown boldly & remaining values second best results various algorithms. Table 5.2 shows the recall values of many suggested techniques with consistent multi-labeled clustering results. This is a good sign. Figure 5.2 depicts recall performance various multi-labeled html text content variations (left to right).

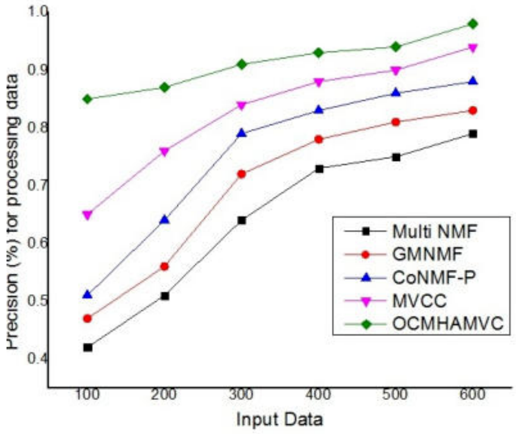


Fig. 5.1: Multi class label parameters performance

Table 5.1: MDC Result

| **Data input** | **Multi-NMF** | **GMNMF** | **Co-NMFP** | **MVCC** | **Proposed OCMHAMVC** |
|---|---|---|---|---|---|
| 100 | 00.43 | 00.480 | 052 | 0.66 | 0.86 |
| 200 | 0.52 | 0.57 | 0.65 | 0.77 | 0.88 |
| 300 | 0.65 | 0.73 | 0.80 | 0.86 | 0.92 |
| 400 | 0.74 | 0.79 | 0.84 | 0.90 | 0.94 |
| 500 | 0.76 | 0.82 | 0.87 | 0.91 | 0.95 |
| 600 | 0.80 | 0.84 | 0.89 | 0.95 | 0.99 |

Table 5.2: Recall MDC values

| **Input data** | **Multi-NMF** | **GMNMF** | **Co-NMFP** | **MVCC** | **Proposed OCMHAMVC** |
|---|---|---|---|---|---|
| 100 | 00.590 | 00.68 | 00.76 | 00.85 | 00.89 |
| 200 | 00.65 | 00.72 | 00.79 | 00.88 | 00.92 |
| 300 | 00.705 | 00.79 | 00.85 | 00.90 | 00.93 |
| 400 | 00.78 | 00.84 | 00.88 | 00.91 | 00.95 |
| 500 | 00.84 | 00.86 | 0.91 | 00.94 | 00.96 |
| 600 | 00.86 | 00.91 | 00.93 | 00.95 | 00.98 |

Multi-NMF, GMNMF, Co-NMFP, MVCC and OCMHAMVC methods are implemented on python 3.7.0 and it is identified that proposed OCMHAMVC was attains good improvement and suitable for future document summery applications.

It is demonstrated in figure 5.3 that OCMHAMVC delivers the best results when compared to conventional ways when the number of documents is increased; when number of attributes is increased, OCMHAMVC exhibits competent clustering results with multi attributes. The figure 5.3 clearly explains about MDC analysis using F1 score via data input. Here generating measures with multi labelled data gathering. The html text document is collected from normal text which is performed through proposed model. The generated statistical values are proven that proposed model is good at document summarisation applications.

With a strong preference for multi-labeled data gathering, five alternative clustering algorithms were performed to 100-500 html text documents in Figure 5.4.

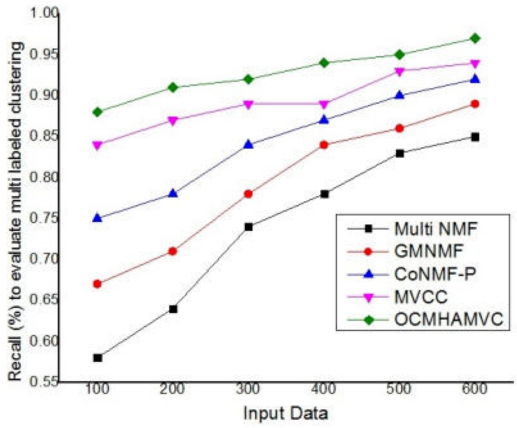


Fig. 5.2: MDC performance in recall

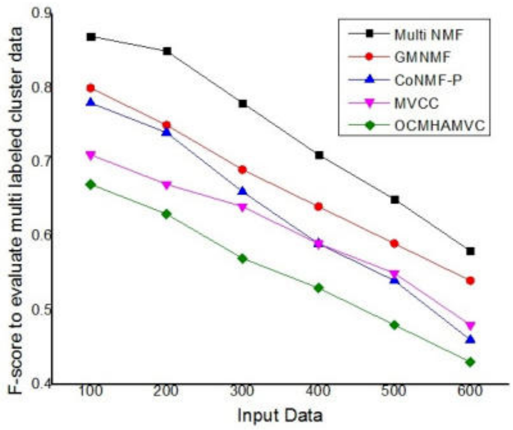


Fig. 5.3: MDC in F1- Score performance

Table 5.3: MDC's Values in F1- Score

| Data input | Multi-NMF | GMNMF | Co-NMFP | MVCC | Proposed OCMHAMVC |
|---|---|---|---|---|---|
| 100 | 00.88 | 00.81 | 00.75 | 00.72 | 00.68 |
| 200 | 00.86 | 00.76 | 00.76 | 00.68 | 00.64 |
| 300 | 00.79 | 00.70 | 00.67 | 00.64 | 00.58 |
| 400 | 00.72 | 00.65 | 00.60 | 00.60 | 00.54 |
| 500 | 00.66 | 00.60 | 00.55 | 00.56 | 00.49 |
| 600 | 00.60 | 00.55 | 00.47 | 00.50 | 00.44 |

For every dataset, the finest value is indicated in bold, while all of the other values are second-best findings from other methodologies.

OCMHAMVC shown figure 5.5 best time results when compared traditional approaches retrieving matched multi-dimensional documents html text documents. It takes less time to get multi-attribute relational documents from multiple domains using OCMHAMVC as the number of documents increases.

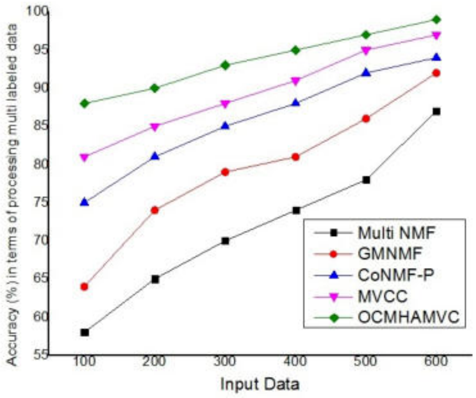


Fig. 5.4: Multi Dimension clustering Accuracy

Table 5.4: MDC's Accuracy

| Input data | Multi-NMF | GMNMF | Co-NMFP | MVCC | Proposed OCMHAMVC |
|---|---|---|---|---|---|
| 100 | 59 | 65 | 76 | 82 | 89 |
| 200 | 66 | 75 | 82 | 86 | 91 |
| 300 | 71 | 80 | 86 | 89 | 94 |
| 400 | 75 | 82 | 89 | 92 | 96 |
| 500 | 79 | 87 | 93 | 96 | 98 |
| 600 | 88 | 93 | 95 | 98 | 100 |

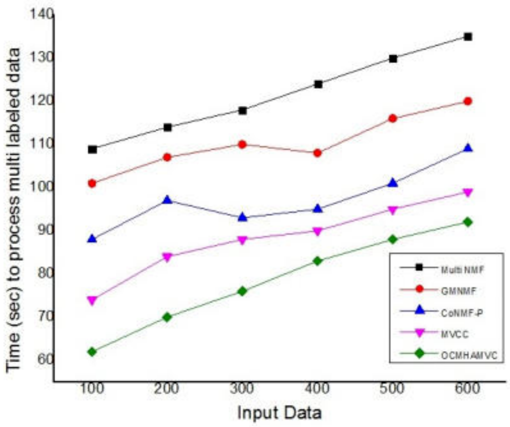


Fig. 5.5: MDC performance in time

Data set, best value indicated bold, other values second-best findings methodologies.

Data from 100-500 html text articles were used to test five alternative clustering algorithms, and the results shown in Figure 5.6 demonstrate that multi-labeled data collection is the most efficient in terms of computing cost when combined with the most desirable parameters.

For each data set in table 5.6, best value indicated bold and others second best findings from various methodologies.

Table 5.5: Multi Dimension Clustering in Time Values

| Input data | Multi-NMF | GMNMF | Co-NMFP | MVCC | Proposed OCMHAMVC |
|---|---|---|---|---|---|
| 100 | 110 | 102 | 89 | 75 | 63 |
| 200 | 115 | 108 | 98 | 85 | 71 |
| 300 | 119 | 112 | 94 | 89 | 77 |
| 400 | 125 | 109 | 96 | 92 | 84 |
| 500 | 131 | 115 | 102 | 96 | 90 |
| 600 | 136 | 122 | 110 | 100 | 93 |

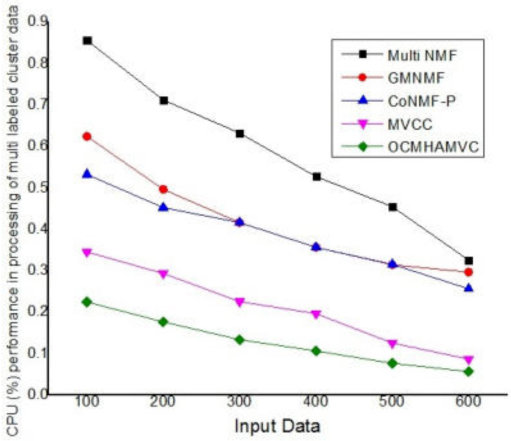


Fig. 5.6: Computational Cost Performance in CPU Processing

Table 5.6: Multi Dimension Clustering in Computational Cost Performance Values

| Data input | Multi-NMF | GMNMF | Co-NMFP | MVCC | Proposed OCMHAMVC |
|---|---|---|---|---|---|
| 100 | 00.857 | 00.625 | 00.533 | 00.346 | 00.226 |
| 200 | 00.713 | 00.497 | 00.454 | 00.294 | 00.177 |
| 300 | 00.633 | 00.417 | 00.417 | 00.226 | 00.134 |
| 400 | 00.528 | 00.357 | 00.357 | 00.197 | 00.107 |
| 500 | 00.454 | 00.315 | 00.316 | 00.127 | 00.077 |
| 600 | 00.326 | 00.297 | 00.258 | 00.088 | 00.058 |

Table 5.6 clearly explained about OCMHAMVC score and comparison has been performed with earlier models. In this context for various input like 100, 200.....600 at any instant proposed model got good improvement since 00.226 to 00.058.

OCMHAMVC has the best time results compared to traditional approaches when it comes to retrieving multi-attribute relational documents from html text documents, as shown in figure 5.7. When the number of documents is increased, OCMHAMVC provides efficient cluster results, which means less memory utilization. Shown table 5.7 depicts MDC values based on F-score enhance number of labelled text documents

Results from the figures and tables above are based on the experimental setup for the suggested technique and alternative approaches. Because of OCMHAMVC's fundamental convergence, it may be used on text-oriented documents several domains and yet satisfy text data connected to curves. OCMHAMVC additionally finds effective multi-dimensional clustering results decreasing iteration used on distinct text-oriented documents

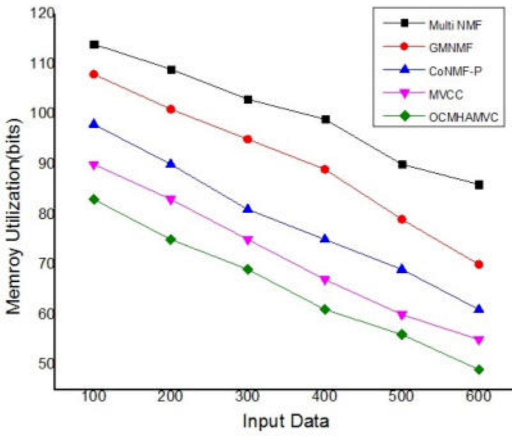


Fig. 5.7: Memory usage in procedure MLD performance

Table 5.7: Memory usage in MDC

| Data input | Multi-NMF | GMNMF | Co-NMFP | MVCC | Proposed OCMHAMVC |
|---|---|---|---|---|---|
| 100 | 115 | 109 | 99 | 91 | 84 |
| 200 | 110 | 102 | 92 | 84 | 76 |
| 300 | 104 | 97 | 82 | 76 | 70 |
| 400 | 100 | 90 | 76 | 68 | 62 |
| 500 | 92 | 80 | 70 | 62 | 57 |
| 600 | 8s7 | 72 | 62 | 56 | 50 |

while investigating documents using iterative functions. OCMHAMVC measures computational efficiency multi-dimensional document processing time, precision, recall, accuracy, memory use and CPU computational cost.

The OCMHAMVC model has getting computational efficiency, recall, accuracy, memory usage and CPU computational cost from confusion matrix and measures estimation on statistical data from scientific documents. The cluster analysis gives documents summarisation with easy analysis.

**6. Conclusion.** A new strategy representing data cluster of multiple categories using multi-labeled dimensional data is proposed in this study, namely OCMHAMVC. The suggested technique initially clusters the comparable label data that are important to similar cluster prototypes, given the fast evolution of labelled data. This cluster prototype has the same labels and attributes associated with the same classes as the original cluster. In order to gather comparable cluster prototypes, the suggested technique utilizes cooperative regularization with representative qualities to investigate associative aspects that are desirable from several perspectives. The OMF technique is used to assess low-dimensional data, and a prototype cluster of labelled sample data is formed by grouping related data into the OMF method's OMF evaluations. The proposed strategy is tested on a variety of real-time data sets, and it demonstrates how it compares well to currently available methods. Sampled class labelled cluster findings boundary relations compared multi-labeled dimensions to see how well they perform. Advanced machine learning multi-dimensional clustering would be a great addition to further enhance our suggested strategy.

REFERENCES

[1] Nguyen, D. T., Chen, L., & Chan, C. K. , *Clustering with multiviewpoint-based similarity measure*, IEEE transactions on knowledge and data engineering, 2011, 24 (6), pp. 988-1001.

[2] Liu, J., Wang, C., Gao, J., & Han, J. , *Multi-view clustering via joint nonnegative matrix factorization*, In Proceedings of the 2013 SIAM international conference on data mining Society for Industrial and Applied Mathematics, 2013, pp. 252-260.

[3] Wang, Z., Kong, X., Fu, H., Li, M., & Zhang, Y. , *Feature extraction via multi-view non-negative matrix factorization with local graph regularization*, In 2015 IEEE International conference on image processing (ICIP), 2015, pp. 3500-3504.

[4] He, X., Kan, M. Y., Xie, P., & Chen, X. , *Comment-based multi-view clustering of web 2.0 items*, In Proceedings of the 23rd international conference on World wide web, 2014, pp. 771-782.

[5] Wang, H., Yang, Y., & Li, T. , *Multi-view clustering via concept factorization with local manifold regularization*, In 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 1245-1250.

[6] Wei, S., Wang, J., Yu, G., Domeniconi, C., & Zhang, X. , *Multi-view multiple clusterings using deep matrix factorization*, In Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, No. 04, pp. 6348-6355.

[7] Li, Z., Tang, J., & Mei, T. , *Deep collaborative embedding for social image understanding*, IEEE transactions on pattern analysis and machine intelligence, 2018, 41(9), pp. 2070-2083.

[8] Liu, X., Zhu, X., Li, M., Wang, L., Zhu, E., Liu, T., ... & Gao, W. , *Multiple kernel k k-means with incomplete kernels*, IEEE transactions on pattern analysis and machine intelligence, 42 (5), pp.1191-1204.

[9] Luo, S., Zhang, C., Zhang, W., & Cao, X. , *Consistent and specific multi-view subspace clustering*, In Proceedings of the AAAI Conference on Artificial Intelligence, 2018, Vol. 32, No. 1.

[10] Mautz, D., Ye, W., Plant, C., & Böhm, C. , *Discovering non-redundant k-means clusterings in optimal subspaces*, In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 1973-1982.

[11] Nie, F., Cai, G., & Li, X. , *Multi-view clustering and semi-supervised classification with adaptive neighbours*, In Thirty-first AAAI conference on artificial intelligence, 2017.

[12] Tan, Q., Yu, G., Domeniconi, C., Wang, J., & Zhang, Z. , *Incomplete multi-view weak-label learning*, In Ijcai, 2018, pp. 2703-2709.

[13] Wang, X., Yu, G., Domeniconi, C., Wang, J., Yu, Z., & Zhang, Z. , *Multiple co-clusterings*, In 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 1308-1313.

[14] Wang, X., Wang, J., Domeniconi, C., Yu, G., Xiao, G., & Guo, M. , *Multiple independent subspace clusterings*, In Proceedings of the AAAI conference on artificial intelligence, 2019, Vol. 33, No. 01, pp. 5353-5360.

[15] Yang, S., & Zhang, L. , *Non-redundant multiple clustering by nonnegative matrix factorization*, Machine Learning, 2017, 106(5), pp. 695-712.

[16] Yao, S., Yu, G., Wang, X., Wang, J., Domeniconi, C., & Guo, M. , *Discovering multiple co-clusterings in subspaces*, In Proceedings of the 2019 SIAM International Conference on Data Mining Society for Industrial and Applied Mathematics, 2019, pp. 423-431.

[17] Yao, S., Yu, G., Wang, J., Domeniconi, C., & Zhang, X., *Multi-view multiple clustering*, arXiv preprint arXiv:1905.05053. 2019.

[18] Zhao, H., Ding, Z., & Fu, Y., *Multi-view clustering via deep matrix factorization*, In Thirty-first AAAI conference on artificial intelligence, 2017.

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

[20] Li, S., Liu, Q., Dai, J., Wang, W., Gui, X., & Yi, Y., *Adaptive-weighted multiview deep basis matrix factorization for multimedia data analysis*, Wireless Communications and Mobile Computing, 2021.

[21] Ahn, J. H., Kim, S., Oh, J. H., & Choi, S., *Multiple nonnegative-matrix factorization of dynamic PET images*, In Proceedings of Asian Conference on Computer Vision, 2004, pp. 1009-1013.

[22] Trigeorgis, G., Bousmalis, K., Zafeiriou, S., & Schuller, B. W. , *A deep matrix factorization method for learning attribute representations*, IEEE transactions on pattern analysis and machine intelligence, 2016, 39(3), pp. 417-429.

[23] Zhao, Y., Wang, H., & Pei, J., *Deep non-negative matrix factorization architecture based on underlying basis images learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(6), pp. 1897-1913.

[24] Meng, Y., Shang, R., Shang, F., Jiao, L., Yang, S., & Stolkin, R. , *Semi-supervised graph regularized deep NMF with bi-orthogonal constraints for data representation*, IEEE transactions on neural networks and learning systems, 2019, 31(9), pp. 3245-3258.

[25] Tong, M., Chen, Y., Ma, L., Bai, H., & Yue, X. (2020) , *NMF with local constraint and Deep NMF with temporal dependencies constraint for action recognition*, Neural Computing and Applications, 2020, 32(9), pp. 4481-4505.

[26] J Li, J., Zhou, G., Qiu, Y., Wang, Y., Zhang, Y., & Xie, S. , *Deep graph regularized non-negative matrix factorization for multi-view clustering*, Neurocomputing, 2020, 390, pp. 108-116.

[27] Shu, Z. Q., Wu, X. J., Hu, C., You, C. Z., & Fan, H. H. , *Deep semi-nonnegative matrix factorization with elastic preserving for data representation*, Multimedia Tools and Applications, 2021, 80(2), pp. 1707-1724.

[28] Zhao, J., Xie, X., Xu, X., & Sun, S., *Multi-view learning overview: Recent progress and new challenges*, Information Fusion, 2017, 38, pp. 43-54.

[29] Li, Y., Yang, M., & Zhang, Z. , *A survey of multi-view representation learning*, IEEE transactions on knowledge and data engineering, 2018, 31(10), pp. 1863-1883.

[30] Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & de Albuquerque, V. H. C. (2021) , *A comprehensive survey of multi-view video summarization*, Pattern Recognition, 2021, 109, 107567.

[31] Xu, X., Yang, Y., Deng, C., & Nie, F. , *Adaptive graph weighting for multi-view dimensionality reduction*, Signal Processing, 2019, 165, pp. 186-196.

[32] Zhang, R., Nie, F., Li, X., & Wei, X. (2019) , *Feature selection with multi-view data: A survey*, Information Fusion, 50,

pp. 158-167.

[33] Luo, P., Peng, J., Guan, Z., & Fan, J. (2018) , *Dual regularized multi-view non-negative matrix factorization for clustering*, Neurocomputing, 2018, 294, pp. 1-11.

[34] Bai, L., Shao, Y. H., Wang, Z.,& Li, C. N., *Clustering by twin support vector machine and least square twin support vector classifier with uniform output coding*, Knowledge-Based Systems, 2019, 163, pp. 227-240.

[35] H Seung, H. S., & Lee, D. D., *The manifold ways of perception*, science, 2020, 290 (5500), pp. 2268-2269.

[36] Spielman, D. A. (2007, October) , *Spectral graph theory and its applications*, In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 2017, pp. 29-38.

[37] Hidru, D., & Goldenberg, A., *EquiNMF: Graph regularized multiview nonnegative matrix factorization*, arXiv preprint arXiv:1409.4018. 2014.

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

[39] Guan, Z., Zhang, L., Peng, J., & Fan, J. , *Multi-view concept learning for data representation*, IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11), pp. 3016-3028.

[40] Xu, C., Guan, Z., Zhao, W., Niu, Y., Wang, Q., & Wang, Z. , *Deep multi-view concept learning*, In IJCAI, 2018, pp. 2898-2904.

[41] Wang, J., Wang, X., Tian, F., Liu, C. H., Yu, H., & Liu, Y. , *Adaptive multi-view semi-supervised nonnegative matrix factorization*, In International conference on neural information processing. Springer, Cham, 2016, pp. 435-444.

[42] Kolli, S., & Sreedevi, M., *A novel index based procedure to explore similar attribute similarity in uncertain categorical data*, ARPN Journal of Engineering and Applied Sciences, 2019, 14(12), pp. 2266-2272.

[43] Kolli, S., & Sreedevi, M., *Adaptive clustering approach to handle multi similarity index for uncertain categorical data streams*, Jour of Adv Research in Dynamical & Control Systems, 2018, 10(04).

[44] Kolli, S. , *A novel granularity optimal feature selection based on multi-variant clustering for high dimensional data*, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(3), pp.5051-5062.

[45] Kolli, S., Krishna, P. R., & Reddy, P. B. , *A NOVEL NLP AND MACHINE LEARNING BASED TEXT EXTRACTION APPROACH FROM ONLINE NEWS FEED*, 2006.

[46] Gudadappanavar, S. S., & Mahapatra, S. , *Metaheuristic nature-based algorithm for optimal reactive power planning*,International Journal of System Assurance Engineering and Management, 2022, 13(3), pp.1453-1466.

[47] Saikumar, K., Rajesh, V., & Babu, B.S., *Heart disease detection based on feature fusion technique with augmented classification using deep learning technology*, Traitement du Signal, 2022, 39(1), pp.31-42.

*Edited by:* Vinoth Kumar
*Received:* Jun 30, 2022
*Accepted:* Dec 5, 2022