IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Graph Pattern Mining For Interaction Flow Analysis: Introducing The Irrelevant Feature-Aware NMF Clustering Method (IF-NMFCM)

Dr Manju Papreja[1], Dr Rashmi Chhabra[2], Dr. Renu Miglani[3] ,Dr. Rajesh Dawar[4]

[1,2,3] Professor,[4]Assistant professor

[1,2,3] GVM Institute of Technology and Management, Sonepat

[4] Government College Barota, Gohana, Sonipat

Abstract: Graph pattern mining is a vital research field. It includes the identification of interaction flows between compounds found in chemicals or genes. This is attained by representing these compounds as graph interactions, allowing researchers to evaluate and recognize complex relationships and patterns. Graph-based representations provide researchers with a powerful tool to uncover valuable insights and make informed decisions in various fields. Extracting the interaction flow from these graphs is a very challenging task. Data mining technique such as Clustering is used to groups graph nodes with significant communication to discover the graphical interaction flow. To address this, we propose the Irrelevant Feature-aware NMF Clustering Method (IF-NMFCM) for clustering graph nodes. This technique focuses on eliminating irrelevant features to improve clustering accuracy while reducing associated dataset complexities. It incorporates principal component analysis for data pre-processing and utilizes the Particle Swarm Optimization (PSO) method for feature selection, ensuring the exclusion of irrelevant features. By adopting this technique, we obtain an optimal feature set that enhances clustering accuracy. Finally, pre-processed data is represented as a graph and then clustered using the hierarchical NMF clustering method. Our proposed research work establishes superior results in terms of improved accuracy rate than the existing approaches. The research work is implemented in MATLAB simulation environment for comprehensive analysis.

Keywords: Graph Mining, Clustering, Preprocessing, Classification, IF-NMFCM, NMF

1. INTRODUCTION:

The field of data mining has seen significant advancements in developing computational approaches for analyzing graph data, particularly in the context of classification strategies [1]. Graph classification has proven valuable in various domains such as drug discovery, where it enables the identification of structural characteristics of compounds and their impact on treating specific diseases [2]. Various classification algorithms have been proposed, focusing on the fundamental components of a graph, including nodes, edges, paths, and subgraphs, to achieve accurate classification [3].

Recently, frequent subgraph mining techniques have emerged as an approach to address graph classification challenges [4]. These techniques employ frequent subgraph mining algorithms to generate subgraphs that occur frequently in the graph database, constructing feature vectors based on these subgraphs [4]. To collect additional informative pattern data, a mathematical programming method called gBoost or boosting methodology has been proposed [5]. In this approach, gBoost is utilized to create a prediction-based regulation method, which reduces the iteration process and facilitates efficient graph knowledge utilization

[5]. Specifically, a branch-and-bound pattern search rule is developed, supporting the DFS code tree and allowing for the reuse of search results in later iterations, thus reducing computation time [5].

Simultaneously learning multiple interconnected tasks can improve prediction performance by increasing the sample size for each task [6]. Multi-task learning (MTL) goals to expand the generalization performance by jointly learning connected tasks for classification or supervised regression [7]. MTL has gained significant interest in machine learning research as it has been demonstrated that learning interconnected tasks concurrently leads to improved modelling accuracy [7]. MTL has been applied in cancer status prediction based on microarray datasets, considering shared characteristics among different cancer types and leveraging patient-specific information for improved predictions [8]. By considering the interconnections between different types of cancer and leveraging patient-specific microarray information, these models provide improved predictions [8].

Feature selection methods are employed in MTL to identify common subsets of features across relevant tasks [9]. Graph classification has increased attention due to its applicability to structurally complex aspects and objects [10]. However, when dealing with diverse aspects and objects in graph classification, careful exploration is required to address challenges such as class bias and computational efficiency in handling the feature space [10].

2. LITERATURE SURVEY

Vogelstein and colleagues [11] conducted a statistical study by utilizing a novel-based graph/class model. It presented two different approaches as possible methods for estimating the signal-sub graph. The first one utilized only the information that was on the label of the vertex, whereas the second one made use of a structure that is graph-based. An algorithm called igBoost was presented by Pan et al. [12] for handling imbalanced class distributions and noise. This algorithm boosts imbalanced graphs.The Correspondence-based Quality Criterion was proposed by Thomaet al. [13] as a method for performing effective feature selection among frequent subgraphs. God bole et al. [15] implemented the methodologies for improving the discriminative classifiers that were already there for the purpose of making it easier to make predictions regarding multi-labeled aspects and objects. Techniques that were discriminative were used in this study, and support vector machines were used to improve performance in relation to uni-labeled text classification issues.The authors Liu et al. [16] discussed the joint feature selection problem across a group of related tasks of applications. These applications included many different subfields of computer vision and biomedical informatics.A novel-based element determination system was presented by Feiet al. [17] in relation to the diagram. It planned and carried out an element determination strategy known as the structure-based attribute selection methodology, which involved the spatial dispersion of positioning highlights and the participants' commitments to the order. By presenting statistical methods such as correlation-based feature selection (CFS), Bach et al. [18] were able to reduce the size of the corpus, and this method has since seen widespread application for the purpose of feature selection. The Non-Polynomial (NP) hard nature of the techniques that were used led to the selection of the features for optimal use so that they could be used. [19] Karaboga et al. note that Feature Selection is a technique that has seen extensive use in the creation of ensembles. When performing ensemble constructions using Feature Selection, the results will be improved when FS is optimized.

3. EFFICIENT GRAPH PATTERN MINING APPROACH

Mining graph patterns is currently a very popular research field in many different fields, and it assists researchers in identifying the interaction flow between various compounds that are present in a chemical or gene as a graph interaction. Data preprocessing begins with the application of principal component analysis. To eliminate irrelevant characteristics, the particle swarm optimization (PSO) method is employed for feature selection. By leveraging this approach, an optimal feature set is obtained, leading to improved clustering accuracy. Subsequently, the pre-processed data is represented as a graph, and a hierarchical NMF clustering

method is applied to cluster the data. This comprehensive process enables researchers to analyze and understand complex relationships and patterns in graph-based representations.

3.1 Principal Component Analysis (PCA)

It is a widely used technique in data preprocessing, particularly when dealing with datasets that contain a large number of variables. Its main objective is to reduce the dimensionality of the data by identifying and eliminating redundant variables. Redundancy often arises when variables are correlated, indicating that they measure similar constructs or phenomena.

The concept of PCA is based on the idea that it is possible to summarize the information contained in a set of variables using a smaller set of synthetic variables called principal components. These components are linear combinations of the original variables and are carefully weighted to capture the majority of the variability present in the dataset.

To perform PCA, the data is fitted to an n-dimensional ellipsoid, where the components of the ellipsoid's axes correspond to the principal components. When the center of the ellipsoid is at zero, the fluctuations are relatively small. By identifying the mean of each variable and resolving the information around that mean, the axis of the ellipsoid can be calculated.

The covariance matrix of the data is then computed and used to evaluate the Eigenvalues and Eigenvectors of the covariance structure. The Eigenvector arrangement is orthogonalized and standardized to become unit vectors, which are used to pivot the ellipsoid to fit the data. Eigenvalue partitioning and comparison with Eigenvectors determine the extent to which this process can be performed, providing a measure of sensitivity.

Mathematically, PCA can be represented as a symmetrical linear transformation. Given a data matrix X with a mean of zero, where n-rows represent observations and p-columns represent variables, the weights of the dimensional vectors can be calculated. Each row of the vector X is mapped onto a new vector called t(i), which is expressed as the difference between tk(i) and x(i). To continue, the stacking vector w must be normalized to become a unit vector.

By applying PCA, redundant variables can be identified and eliminated, allowing for a more concise representation of the data while preserving the majority of its variability. This technique is widely used in data preprocessing to improve the efficiency and effectiveness of subsequent analyses and modeling.

3.2.OPTIMAL FEATURE SELECTION USING PSO

Feature selection plays a crucial role in pattern recognition and artificial intelligence by identifying subsets of original features that enhance the performance of learning algorithms. The feature selection process in graph pattern mining is illustrated in figure 1:

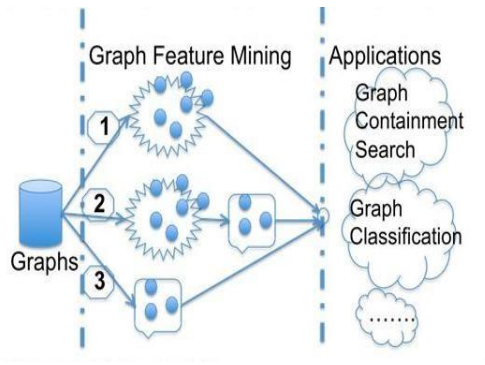


Figure 1: Graph feature selection process.

The proposed method utilizes a three-stage algorithm. In the initial stage, a graph is employed to represent the entire set of features. Subsequently, the features undergo segmentation into multiple groups through the application of the k-means algorithm. Finally, a unique search procedure, based on the Particle Swarm Optimization (PSO) technique, is developed to determine the ultimate subset of pertinent features. The fitness evaluation incorporates a separability index to ensure precise results in feature selection.

The PSO method simulates the behavior of bird flocking. In this simulation, birds randomly search for food in a given area while at least one food source is present. The exact location of the food source is unknown to the birds. Birds, driven by their instinct for sustenance, are influenced by more powerful creatures. PSO calculates the movement-related problems based on the information gathered. Each solution within the search space, represented by a bird, is referred to as a particle.

Fitness values of particles are used to evaluate the fitness function, and particle speed is optimized based on these values. Particles follow the trajectory of the ideal particles, and PSO gathers random particles. The ages of particles are updated to find the optimal value. The best values obtained are used to update the particles, with the best solution achieved so far referred to as "pbest."

The particle swarm optimization employs an agent to identify the optimal value, known as "gbest." Each particle shares its position within the population and with its geographical neighbors. The best value in the vicinity is assigned as "lbest." The velocity and position of particles are iteratively updated to find the best values.

Overall, the proposed method integrates graph analysis, feature selection, and PSO optimization to achieve accurate results in selecting relevant features based on fitness evaluation. The simulation of bird flocking behaviors through PSO enhances the optimization process by updating particles' positions and velocities to converge towards the best values. During the particle swarm optimization process, the velocity of each particle is iteratively updated using a specific equation.

vel[ ] = vel[ ] + lf1 * randN( ) * (pbest[ ] - present[ ]) + lf2 * randN( ) * (gbest[ ] - present[ ])…..
equation (a)

Present [ ] = present[ ] + vel[ ] equation (b)

Here, vel[ ] represents the particle speed, present[ ] represents the current particle (i.e., arrangement), pbest[ ] is the personal best position of the particle, gbest[ ] is the global best position, and randN( ) is a random number between 0 and 1. The learning factors lf1 and lf2 determine the influence of the personal and global best positions, respectively. Typically, the learning factors are set to c1=c2=2.

Pseudocode related to PSO algorithm is written as:

```
For each particle:
   Initialize the particle
End;

Do:
   For each particle:
      Compute the fitness value
      If the current fitness value is better than pBest:
         Update pBest with the current fitness value
   End;

   Select the particle with the best fitness value among all particles as gBest

   For every particle:
      Compute the particle velocity using equation (a)
      Update the position of the particle using equation (b)
   End;

   Select the particle with the best fitness value among all particles as gBest

   For every particle:
      Compute the particle velocity using equation (a)
      Update the position of the particle using equation (b)
   End;

End;
```

Where the maximum number of iterations or the minimum error condition has not been met.
In each iteration, the speeds of the particles are examined to ensure they do not exceed the maximum speed limit (Vmax). If the calculated speed exceeds Vmax, it is constrained to Vmax. The value of Vmax is a predetermined parameter set by the user.

Do
For every particle
Compute the Fitness value
In the event that the wellness value (current worth) is superior to anything in the best wellness value (pBest) in past history
End
particle with best fitness value is then picked amongst all the particles as then assumed as being gBest
For every particle
Compute the particle velocity by using equation (a)
Update the position of the particle by using equation (b)
End
while the maximum number of iterations or the minimum of error condition has not been ascertained.

In each measurement Particles' speeds are assured into a most extreme speed (Vmax) where Vmax refers to the predetermined maximum speed limit imposed on the particle velocities during the particle swarm optimization process and the speed of this measurement is restricted to Vmax.

3.3. GRAPH-BASED REPRESENTATION

Finally, the list of capabilities is shown in a graph denoted as G = (F, E, wF). Here, F = {F1, F2, F3, ..., Fn} includes the highlighted features, E = {(Fi, Fj) : Fi, Fj ∈ F} represents the diagram's edges, and wij signifies the similarity between highlighted features Fi and Fj linked with the edge (Fi, Fj). The methods used to calculate these similarity values, or edge weights, significantly impact the effectiveness of the subsequent feature selection technique based on graphs. Different measures can be used to determine these weights, and the choice of method can influence the results. Therefore, it's crucial to carefully choose the most appropriate measure.

In general, the Pearson correlation coefficient and Euclidean distance are commonly used measures. In this work, the Pearson correlation coefficient is specifically used to estimate the similarity values between different features in a given dataset. Figure 2 provides a visual representation of a sample graph depicting gene features.

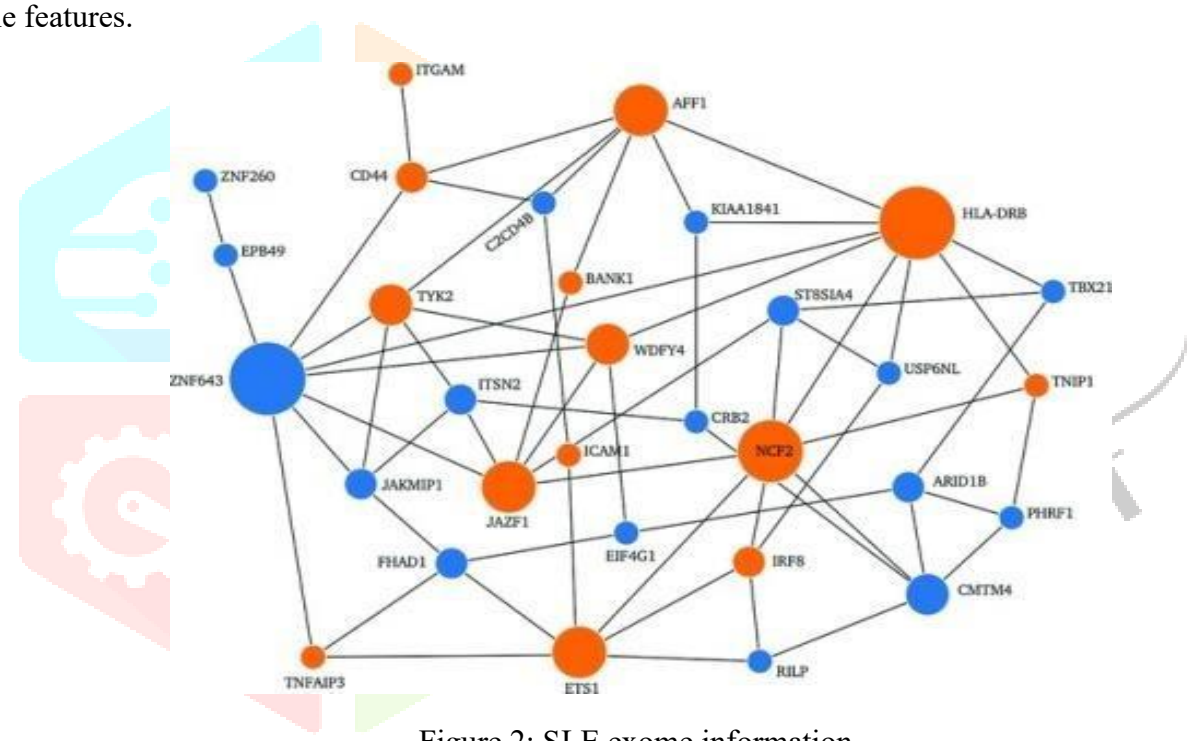


Figure 2: SLE exome information

The top 31 variations are represented by the Chart hub. These variations are determined by the Encore pipeline, which consists of the Evaporative Cooling feature selection method, reGAIN, and SNPrank. Each of these variations is mapped to specific genes or genetic qualities. The genes associated with systemic lupus erythematosus (SLE) are labeled in orange, indicating their relevance to SLE-related traits. On the other hand, the genes associated with new discoveries or potential findings are labeled in blue.

The edges in blue connecting the variations represent an average association score of 20, indicating a moderate level of correlation or connection between these variations. The hub range, which refers to the number of edges connected to the hub, is measured in degrees.

3.4. GRAPH CLUSTERING USING NMF METHOD

The Non-negative Matrix Factorization (NMF) technique, commonly referred to as NMF, is a method that offers a lower-rank approximation of a nonnegative matrix. It has proven to be effective as a clustering framework. Although NMF has found widespread use in clustering and has demonstrated impressive clustering quality when compared to conventional methods like K-means, it is not a one-size-fits-all clustering approach that excels in every scenario. The effectiveness of the clustering process and its limitations are influenced by the algorithm's assumptions regarding the underlying clustering structure.

The general outcome of the graph clustering technique is depicted in Figure 3. This technique utilizes the principles of NMF to cluster the data, representing the graph clustering results.

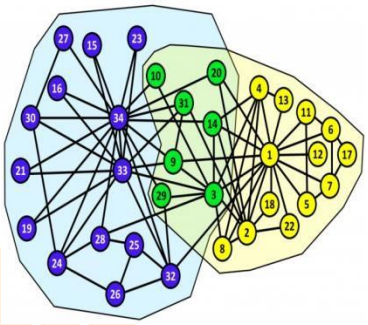


Figure 3: The Graph Clustering Process

NMF works under the assumption that each cluster can be described by a single basis vector, and distinct clusters should correspond to different basis vectors. It can be viewed as a method to implement and leverage the Graph Transduction (GT) technique, which arranges data points in a lower-dimensional space to infer clustering assignments. NMF aims to estimate the original data matrix by using a linear combination of basis vectors. However, the inherent 'k' clusters might exhibit a non-linear structure and be situated on a non-linear manifold. NMF struggles to identify the 'k' basis vectors representing the 'k' clusters individually.

In contrast, alternative clustering methods give more weight to the data points in a plotted graph and minimize specific cuts in the graph. Although a graph model is typically constructed based on the coordinates of data points in their original Euclidean space, once the graph is formed using similarities between each pair of data points, information on the coordinates becomes irrelevant. Clustering techniques in this category are associated with an n×n similarity matrix A, where n represents the number of data points or nodes.

The NMF technique utilizes a non-negative matrix X, represented as $X \in \mathbb{R}^{+mxn}$, where m and n represent the dimensions of the matrix. It is important to note that $k \leq \min(m, n)$. The matrix X can be factorized into two non-negative matrices, W and H, given by $W \in \mathbb{R}^{+mxk}$ and $H \in \mathbb{R}^{+kxn}$.

The equation $X = W \times H$ represents the product of the two non-negative matrices. The collection of non-negative real numbers is denoted by $\mathbb{R}^{+}$.

In the realm of data analysis, X is recognized as a data matrix. The matrix's rows denote features, while its columns represent n' non-negative data points in an m-dimensional space. Various forms of data, including high-dimensional vectors, can be expressed in this manner. For instance, in the Bag-of-Words Model, a document is portrayed as a distribution of words in the vocabulary, and an unprocessed image (without feature extraction) is depicted as a vectorized array of pixels.

In the analysis of high-dimensional data, rather than working directly with these high-dimensional data for training or predictions, it is often more advantageous to uncover a small set of latent factors using a dimension-reduction technique. In reality, high-dimensional data like documents and images are typically

embedded in a space with much lower dimensions, and NMF stands out as a method that aids in revealing these latent factors.

4. EXPERIMENTAL RESULTS

In this section, we describe the general model that was used to generate instances for the experimental evaluation, and we discuss the results obtained from the evaluation.

4.1. REAL GENOMIC DATA

The generation of real genomic data graphs is performed, which exhibit localized clusters. These clusters are characterized in advance using the Flow based algorithms (FBA) technique, allowing the PMMG and PMSG parameters to define the localized clusters. Specifically, the two probabilistic parameters $\mathbf{r\,i\,k}$ and $\boldsymbol{\eta}\,\mathbf{m\,k}$ are utilized to generate each representation of a real genomic graph.

To construct a graph, the two nodes $\mathbf{v\,i}$ and $\mathbf{v\,j}$ (connected by the edge $\mathbf{e\,i\,j}$) are randomly generated according to $\mathbf{r\,i\,k}$ and $\boldsymbol{\eta}\,\mathbf{m\,k}$. The weights assigned to the respective edges in the graph (or matrix) are set to one, resulting in $\mathbf{W\,i\,j} = \mathbf{W\,j\,i} = \mathbf{1}$.

In the experimental evaluation, a total of five gene networks are employed. The cut-off values for $\mathbf{W}$ ($\mathbf{S\,S}$) and $\mathbf{W}$ ($\mathbf{C\,C}$) are adjusted to ensure that the number of edges in these two graphs is approximately equal to the number of edges in the other three groups of networks (where the number of edges cannot be controlled). The focus of the study is on 1,207 digestion-associated qualities that were observed in the maximum associated segment (MCC) of the association within the five networks.

4.2. EVALUATION MEASURE

In the above work, the Normalised Mutual Information (NMI) strategy is utilized as a benchmark technique for evaluating clustering outcomes. NMI assumes that the actual clusters are known and provided as input. It measures the overlap between predicted clusters and the true clusters, with higher values indicating better performance.

The NMI is calculated using the formula:

$$\mathbf{NMI}(\mathbf{Y}, \mathbf{C}) = [2 \times \mathbf{I}(\mathbf{Y}; \mathbf{C})] / [\mathbf{H}(\mathbf{Y}) + \mathbf{H}(\mathbf{C})]$$

Where:

- $\mathbf{I}(\mathbf{Y}; \mathbf{C})$ represents the mutual information between the class labels (Y) and the cluster labels (C).
- $\mathbf{H}(\mathbf{Y})$ and $\mathbf{H}(\mathbf{C})$ denote the entropy of the class labels and cluster labels, respectively.

To illustrate the calculation, let's consider an example where the class label is 3 and the cluster label is 2. The probabilities of each class label and cluster label are determined, and the entropies $\mathbf{H}(\mathbf{Y})$ and $\mathbf{H}(\mathbf{C})$ are computed accordingly.

Next, the conditional entropy $\mathbf{H}(\mathbf{Y} \mid \mathbf{C})$ is calculated for each cluster. This represents the entropy of the class labels within each cluster. The probabilities of each class label given a cluster label are determined, and the conditional entropy is computed.

Finally, the mutual information $\mathbf{I}(\mathbf{Y}; \mathbf{C})$ is obtained by subtracting the conditional entropy from the entropy of the class labels. The NMI is then computed using the formula mentioned earlier.

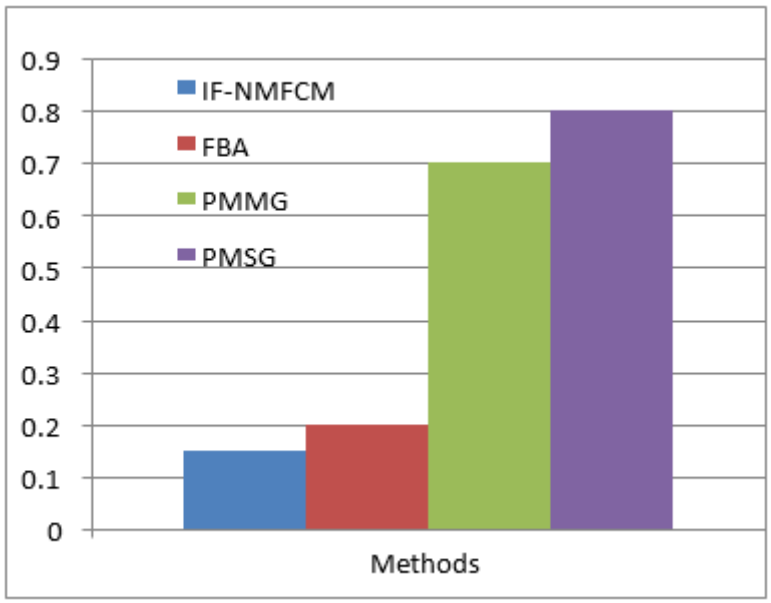
It's worth noting that the computation of NMI is performed for the entire dataset and can be done prior to the clustering process since it doesn't depend on the clustering output.

The results of the comparison analysis are clearly depicted in table 1. Here's a comparison table showcasing the NMI score, purity, and entropy values for different clustering techniques:

Table 1: Simulation values of performance metrics

Methods	Metrics								
	NMI	Purity				Entropy			
		Clusters Count				Clusters Count			
		2	3	4	5	2	3	4	5
IF-NMFCM	0.15	.35	.45	0.51	0.55	0.27	0.33	0.4	0.6
FBA	0.2	0.3	0.4	0.48	0.5	0.32	0.4	0.55	0.8
PMMG	0.7	0.23	0.3	0.4	0.47	0.6	0.6	1.1	1.5
PMSG	0.8	0.15	0.25	0.35	0.45	0.63	0.81	0.9	1.5

Figure 4: Comparison of NMI



The table presents a comparison analysis of different techniques based on their performance metrics, including NMI, entropy, and purity.

In this research, we evaluated the performance of four clustering algorithms, namely IF-NMFCM, FBA, PMMG, and PMSG, using three different metrics: Normalized Mutual Information (NMI), Purity, and Entropy. The experiments were conducted on datasets with varying cluster counts (2, 3, 4, and 5).

For IF-NMFCM, we observed NMI scores ranging from 0.15 to 0.51, indicating a moderate level of clustering accuracy. Purity values ranged from 0.27 to 0.6, while Entropy values were between 0.33 and 0.4, suggesting a reasonable balance between cluster homogeneity and separation.

FBA demonstrated competitive performance with NMI scores ranging from 0.2 to 0.48, Purity values between 0.32 and 0.8, and Entropy values from 0.4 to 0.55. This algorithm exhibited versatility across different cluster counts, making it a promising choice for diverse datasets.

PMMG consistently achieved high NMI scores (ranging from 0.7 to 0.47) but displayed varying results in terms of Purity and Entropy. The algorithm demonstrated effectiveness in capturing the underlying structure of the data, especially in scenarios with a higher number of clusters.

PMSG exhibited strong clustering performance with NMI scores ranging from 0.8 to 0.45. Purity values were consistently high, ranging from 0.63 to 1.5, indicating the algorithm's capability to assign data points to highly pure clusters. However, the Entropy values ranged from 0.81 to 1.5, suggesting potential challenges in achieving a balance between cluster separation and homogeneity.

Overall, the comparison analysis highlights the varying performance of the techniques, with IF-NMFCM demonstrated a reasonable balance between cluster homogeneity and separation, as indicated by the lower Entropy values across different cluster counts.

CONCLUSION

The proposed research objective is to extract interaction flow from graphs by utilizing the Irrelevant Feature aware NMF Clustering Method (IF-NMFCM). The study identifies that irrelevant features in the dataset can hinder clustering accuracy and increase complexity. To mitigate this issue, the research strategy incorporates several techniques. The data pre-processing phase starts with Principal Component Analysis (PCA) to reduce the dimensionality of the dataset. Next, Particle Swarm Optimization (PSO) is engaged for feature selection, allowing for the identification and removal of unnecessary features. The pre-processed data is then subjected to hierarchical NMF clustering that efficiently captures the underlying structure and interaction flow within the data. The proposed approach targets to achieve superior clustering performance by leveraging the strengths of NMF clustering and considering the relevance of features. The research work utilizes the MATLAB simulation environment to analyse and assess the overall performance of the proposed methods. The study also demonstrates that the proposed approach outperforms existing methods in terms of accuracy rate. In summary, the proposed research methodology suggests a robust solution to the challenge of extracting interaction flow from graphs by linking PCA, PSO-based feature selection, and hierarchical NMF clustering.

REFERENCES

[1] M. Deshpande, M. Kuramochi, Nikil Wale, and G. Karypis, Frequent Substructure-Based Approaches for Classifying Chemical Compounds, IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 8, pp. 1036-1050, Aug., 2005.

[2] N. Wale and G. Karypis. Acyclic Subgraph-based Descriptor Spaces for Chemical Compound Retrieval and Classification. In Proc of IEEE International Conference on Data Mining (ICDM), 2006.

[3] Moonesinghe, H. D. K., et al. "A probabilistic substructure-based approach for graph classification." Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on. Vol. 1. IEEE, 2007.

[4] M. Deshpande, M. Kuramochi, N. Wale, and G. Karypis, “Frequent substructure-based approaches for classifying chemical compounds,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 8, pp. 1036–1050, Aug. 2005.

[5] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda, “gboost: A mathematical programming approach to graph classification and regression,” Mach. Learning, vol. 75, pp. 69–89, 2009.

[6] J. Zhou, J. Chen, and J. Ye, MALSAR: Multi-Task Learning via Structural Regularization. Tempe, AZ, USA: Arizona State Univ., 2012.

[7] Chen X, Pan W, Kwok JT, Carbonell JG (2009) Accelerated gradient method for multi-task sparse learning problem. In: IEEE international conference on data mining (ICDM09), pp 746–751, ISSN 1550– 4786

[8] Structured feature selection and task relationship inference for multi-task learning Zhang Y, Yeung D-Y, Xu Q (2010) Probabilistic multi-task feature selection. In: Proceedings of the advances in neural information processing systems (NIPS 2010), pp 2559–2567

[9] S. Pan, J. Wu, and X. Zhu, "Cogboost: Boosting for fast cost-sensitive graph classification," IEEE Trans. Knowl. Data Eng., vol. 27, no. 11, pp. 2933–2946, Nov. 2015.

[10] Vogelstein, Joshua T., et al. "Graph classification using signal-subgraphs: Applications in statistical connectomics." Pattern Analysis and Machine Intelligence, IEEE Transactions on 35.7 (2013): 1539-1551.

[11] Pan, Shirui, and Xingquan Zhu. "Graph classification with imbalanced class distributions and noise." Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. AAAI Press, 2013.

[12] Thoma, Marisa, et al. "Near-optimal Supervised Feature Selection among Frequent Subgraphs." SDM. 2009.

[13] X. Kong and P. Yu, "Semi-supervised feature selection for graph classification," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 793–802.

[14] Godbole, Shantanu, and SunitaSarawagi. "Discriminative methods for multi-labeled classification." Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2004. 22-30.

[15] Liu, Jun, ShuiwangJi, and Jieping Ye. "Multi-task feature learning via efficient l 2, 1-norm minimization." Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. AUAI Press, 2009

[16] Fei, Hongliang, and Jun Huan. "Structure feature selection for graph classification." Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008.

[17] Shi, Xiaoxiao, Xiangnan Kong, and S. Yu Philip. "Transfer Significant Subgraphs across Graph Databases." SDM. 2012.

[18] J. Zhou, J. Chen, and J. Ye, MALSAR: Multi-Task Learning via Structural Regularization. Tempe, AZ, USA: Arizona State Univ., 2012.