Indo American Journal of Multidisciplinary Research and Review
ISSN: 2581 - 6292 / Website: www.iajmrr.com

PARTICLE SWARM OPTIMIZATION (PSO) BASED FEATURE SELECTION METHOD FOR TEXT DOCUMENT CLUSTERING USING K-NEAREST NEIGHBOR (KNN)

Dr. P. Venkateshkumar
Assistant Professor, Department of Computer Science, Government Arts and Science College, Idappadi, Salem, Tamil Nadu

Cite This Article: Dr. P. Venkateshkumar, "Particle Swarm Optimization (PSO) Based Feature Selection Method for Text Document Clustering Using K-Nearest Neighbor (KNN)", Indo American Journal of Multidisciplinary Research and Review, Volume 3, Issue 1, Page Number 29-36, 2019.

Abstract:

The increased volume of text data on the internet web pages makes it challenging to deal with this amount of data. One way in which this humungous amount of text data can be dealt with is by using text clustering scheme that groups a set of documents into categories. However, the performance of text clustering diminishes due to the huge amount of redundant/irrelevant information present in the text documents. An unsupervised scheme that chooses relevant features by formulating a new sub-set of relevant features is feature selection. The throughput of the underlying algorithm can be improved using this scheme. Of late, meta heuristic algorithms have been implemented for successfully solving many of the complicated optimization issues. This work suggests the use of Particle Swarm optimization (PSO) algorithm for solving the feature selection issue and optimizing the K-Nearest Neighbour (KNN) text clustering is explored. Term Frequency-Inverse Document Frequency (TF-IDF) is used for extracting the features. Population based meta-heuristic algorithm, PSO is utilized for optimizing the KNN. Experimental results showed that the proposed method achieves better performance than other methods.

Key Words: Text Mining, Text Documents, Text Clustering, Feature Selection, Hierarchical Agglomerative Clustering (HAC), Term Frequency (TF), Inverse Document Frequency (IDF), Particle Swarm Optimization (PSO) and K-Nearest Neighbor (KNN).

1. Introduction:

Dealing with the huge volume of test data is crucial in text analysis. Also, the volume of data held by the web pages and other modern applications is required by the users. The division of the text document into subsets of predetermined cluster numbers forms the underlying basis for text clustering which is one of the best schemes. There are many fields in which this scheme is comprehensively applied for performing the comprehensive analysis of text data: data clustering, text mining, detection and disease clustering, open source clustering software, clustering the search engine outcomes, time series clustering and wireless sensor clustering. The crucial challenge that exists in text mining and text clustering is the huge amount of redundant data present in text documents. The text clustering process is severely affected by these challenges as these redundant features are noisy, uncorrelated and affect the performance of the text clustering scheme [1].

Thus, dealing with high dimensionality of the feature space is one of the basic issues in text clustering. The computational complexity increases and the throughput of the clustering scheme diminish with the huge number of irrelevant features. The incorporation of digital document processing has exacerbated this issue. Thus, decrease in dimensionality in the area of text clustering is now mandatory where earlier it had been optional. Selecting a differentiating subset of features from a feature space of high dimensionality without compromising on the throughput of the underlying scheme is the primary goal of dimension reduction scheme. Conventionally, there are two dimension reduction schemes- feature extraction and feature selection. The former, also referred to as feature construction scheme, converts high dimensional feature space into a distinct low dimensional feature space. This is achieved by fusing or converting the original feature space [2].

The three sub-categories of feature selection comprise filter, wrapper and embedded schemes. Statistical evaluation of the feature set is performed by the filter schemes for selecting a distinct feature sub-set. However, for evaluating the quality of a given feature set, the wrapper and the embedded schemes make use of machine learning technique. The benefits of these schemes is that compared to the filter schemes, their precision is higher. However, the downside is that, computationally, these schemes are costlier and the feature subsets that are procured lean towards the learning scheme employed. Filter schemes take into consideration only the inherent traits of documents for selecting the feature subset. Hence in comparison, they are quicker and also general meaning, the feature subset that is procured does not lean towards any particular learning scheme. This is why the feature schemes are used extensively for dimension reduction, especially where the feature space dimension is huge [3].

The first and the foremost step in looking for a new accurate subset of relevant features for every document is the Unsupervised Feature Selection. When the class label of the document is unknown beforehand, this scheme improves the clustering scheme. Appropriate outcomes are obtained when the feature selection scheme is defined as an optimization problem. There are two goals for its basis: (1) To enhance the efficacy of the subsequent clustering algorithm. (2) The technique used for obtaining fewer irrelevant features. Text retrieval, predicting the load and associated cost of the electrical power systems, clustering feature selection-based text categorization, anomaly detection in earth dams and text clustering are some of the text mining areas that draw advantages from the feature selection scheme [4].

One of the most reliable schemes for dividing a digital text document text into a cluster of groups is Text clustering which improves the ease of access for the users. These text clustering algorithms strive to arrive at an optimal solution for disseminating a set of text documents. Evaluation conditions such as objective function and fitness function are used. Nowadays, the main source of data is internet. Many of the hospitals, digital libraries, universities, learning centres etc have umpteen text documents that have highly unorganized data. A user can conveniently retrieve what he requires by the aid of clustering schemes.

Data mining, image segmentation, text document clustering, computer vision, predicting, retrieving data and image clustering issues in the medical field are some of the tough optimization problems for which several meta-heuristic optimization algorithms have been suggested. Meta heuristic schemes are designed with the goal of finding optimal solutions based on certain roles which are used for resolving optimization issues. Several solutions are obtained by means of potential values and these are iteratively used to attain an optimal result. The terminal solution is the optimal result [5].

Of late, the issue of decrease in the throughput of global search has been noticed sue to the early convergence of the meta-heuristic algorithms. This is hugely attributed to the quality of the initial solutions. Global optimum can be easily attained if the clustering algorithm procures high quality solutions. Combined schemes have been suggested by many researchers. These fuse the local (exploitation) search with the global (exploration) search so that the diversity of the algorithm is enhanced during the search process. The effectiveness of the PSO based feature selection schemes has been improved by employing these combination methods on meta-heuristic algorithms. This is obtained by improving the original version of the KNN classifier for solving many of the optimization problems. The remainder of the literature is arranged as follows: The related works in literature are discussed in the second section. The schemes employed in the work are discussed in third section. The empirical outcomes are discussed in the fourth section and the work is concluded in the fifth section.

2. Related Works:

A new filter selection scheme was suggested by labani et al., [6]. This is referred to as Multivariate Relative Discrimination Criterion (MRDC) for text classification. Redundant features such as minimal redundancy and maximal relevancy concepts are used by the suggested scheme for decreasing the redundant features. This scheme considers the frequency of documents for every term while evaluating their usefulness. This scheme chooses features that are the most relevant. It also uses a correlation matrix for decreasing the redundancy of the features. As no learning algorithm is used by the MRDC for evaluating the utility of the chosen features, it can be called a filter scheme. There are many experiments that have been performed on the 3 real datasets for evaluating the efficacy of the suggested scheme. The outcomes that are attained have been contrasted with the state-of-the-art filter schemes. It has been shown by the outcomes that in majority cases, the classification performance in MRDC is better than the others.

Of late, a memetic feature selection scheme that combines a filter and an evolutionary filter wrapper has become very popular, displaying effective outcomes. There is one issue with the traditional memetic feature selection- it needs problem conversion which results in diminishing the search ability and leads to feature subsets that are irrelevant and are imprecise. This leads to the low performance of the scheme. A new scheme based on an innovative feature filter has been suggested by Lee et al., [7]. This is highly advanced to multi-label text classification. This scheme has been shown to perform much better than the traditional schemes, by means of experimental outcomes.

This scheme suggests a bi-level sentences selection prototype for text summarization. Proposed by Alguliyev et al., for clustering and optimization schemes known as COSUM form the basis for this [8]. Initially, the K-means scheme is used for clustering the sentences set for finding all the topics in a text. The second level uses optimization model for selecting the salient features from clusters. An objective function is optimized by this prototype. This function is expressed as a harmonic mean of the objective functions. These will strive for coverage and the distinctiveness of the chosen sentences in the summary. A summary's readability is enabled by this model controlling the length of the sentences that are chosen in the candidate summary. An adaptive differential evolution algorithm which comprises an innovative mutation scheme has been formulated for solving the optimization problem.

Two new text clustering scheme have been suggested by Abualigah et al., [9]. These use the Krill Herd (KH) algorithm as the basis for improvising the text document clustering. The first scheme makes use of the KH algorithm along with all of its operators while the second scheme makes use of the genetic operators that have been ignored in the basic KH algorithm. The throughput of the suggested KH schemes has been evaluated and contrasted with the k-mean algorithm. 4 standardized text datasets were used for experimentation. The outcomes have proven that the suggested KH schemes perform better than K-mean scheme in terms of the quality of clusters – entropy and purity were the two clustering measures used for analysis.

A new weighting method has been suggested by Abualigah et al., [10]. This is referred to as Length Feature Weight (LFW). This scheme aims to enhance the performance of text document clustering schemes by making use of new factors. Based on the information obtained from the document, the suggested method allocates a suitable term. weight. It identifies the terms that are characteristic of every cluster and increases their weight on the basis of the suggested factors at the document level and this scheme is evaluated using β-hill climbing technique. For validating the outcomes of the suggested scheme in that domain, the suggested method us contrasted with the existing weight scheme- TF-IDF. 8 benchmark standardized datasets from Laboratory of Computational Intelligence (LABIC) are used for experimentation. The outcomes have shown that the limitations of the existing scheme are overcome using the suggested weighting scheme LFW. Also the outcomes of the text document clustering scheme are enhanced in terms of F-measure, precision and recall.

3. Methodology:

In this section, the HAC method, TF-IDF feature extraction method, KNN classifier, and PSO based feature selection optimized with KNN are discussed.

3.1 Hierarchical Agglomerative Algorithms (HAC):

Better clustering outcomes are generally obtained using hierarchical techniques. However, their complexity makes it a deterrent process. O(n2) is the lower bound on time complexity in hierarchical clustering. There are different levels at which the outcomes procured by hierarchical clustering are visible. Every stage is regarded as an independent clustering outcome with variant clusters, despite the fact that these levels are connected closely due the outcome at every level being dependent on the outcome of the previous level. A graphical representation of this outcome hierarchy is depicted using a tree known as a dendrogram. This tree just has one node at the top and several nodes below. Each node below shows a cluster containing one document. Initially in this agglomerative scheme, the initial number of clusters is 'n' as every document is in its own cluster. Every iteration merges the two of the most identical clusters [11]. This process goes on until there is only one cluster present and this scheme is known as the bottoms-up scheme. The conventional agglomerative scheme can be thus summarized:

- Similarity matrix A is computed. Its entry [a_{ij}] is the similarity between ith and jth cluster.
- The two most identical clusters are fused.
- The similarity matrix is updated such that only the column representing the new cluster is updated.
- Steps 2 and 3 are iterated until one cluster comprises all documents.

3.2 Term Frequency-Inverse Document Frequency (TF-IDF):

The best and the most accurate term weighting method is the TF-IDF which is widely used for eliminating terms having lower weights from the documents so that the efficacy of document retrieval is enhanced. TF-IDF refers to a figure which conveys the importance of a word to a document in a body or a collection [12]. This is widely used as a weighing factor in many of the processes that are employed in retrieving information and text mining. The frequency of a word occurring in a document and the increase in the TF-IDF (1) value of the word are directly correlated; however, this is nullified by the frequency of the word in the body. This can balance the words that appear with greater frequency.

$$TF.IDF = (Term\ Frequency * Inverse\ Document\ Frequency) \quad (1)$$

The number of times a term appears in a document is referred to as Term Frequency (TF). Shorter documents may have fewer terms while longer documents may have many more, as every document has a different length. Thus, to normalize it, the term frequency (2) is mostly.

$$TF = \frac{Total\ number\ of\ items\ in\ a\ document}{Number\ of\ times\ a\ term\ appears\ in\ a\ document} \quad (2)$$

The importance of a term is denoted by Inverse Document Frequency (IDF). All the terms are given equal importance when TF is computed. There are some terms that appear often but are not significant, such as "the", "that" and "is". Thus, there is a need to decrease the weights of the terms that occur often and enhance the weights of those that are infrequent by computing in (3):

$$IDF = \log_2 \frac{(Number\ of\ document\ with\ term\ t\ in\ them)}{Total\ number\ of\ documents} \quad (3)$$

3.3 K-Nearest Neighbor (KNN):

One well known classification scheme is the KNN which can be practically implemented as well. An input parameter k and a class labelled data set are the prerequisites for KNN. The number of nearest neighbours that are considered for tuple classification is resolved by the value of the input parameter k. The process of tuple classification is simple. The data set is divided into two sets in the beginning. These subsets are referred to as training set and test set. These subsets have identical classification. The next stage is to determine K nearest data tuples to unseen tuple from the training set. The class comprising the maximum of these extracted k data tuples is allocated to the unknown tuple that is yet to undergo classification. The precision of the KNN algorithm is computed by the test set. the pseudo code for the standard KNN algorithm[13]:

Input Parameters : Data set, k

Output : Classified test tuples

Step 1: Store all the training tuples.

Step 2 : for each test tuple

A. Compute distance of it with all the training tuples using (1).

B. Find the k nearest training tuples to the test tuple.

C. The class which is most common in the k nearest training tuples to the test tuple is assigned to the test tuple.

End for

A data point in n-dimensional space can be represented by a tuple in the dataset. Here, n is the number of attributes that are used for defining the dataset. Euclidean distance is used for calculating the distance between the data point Euclidean distance between two data tuples x and y is given in (4):

$$\text{Euclidean distance} = \sqrt{\sum_{1\le i\le n}(x_i - y_i)^2} \quad (4)$$

Here, x_i *and* y_i refer to the attribute values of i present in the data tuples x and y correspondingly. 'n' refers to the number of attributes. Some of the other distance metrics that can be employed are Manhattan distance and Minkowski distance

'k' regarded as 1 is the most straightforward case of a KNN algorithm which is referred to as the nearest neighbour rule. In this case, the class that is allocated to the unknown tuple is that of the tuple which is at its closest proximity. Another characteristic of KNN is that it can forecast a continuous valued attribute as well as a categorical attribute. The continuous valued attribute is referred to as regression in which the value of the class attribute of an unknown tuple is the mean of the class attribute values of the k closest tuples to the unknown tuple.

3.4 Particle Swarm Optimization (PSO) Based Feature Selection:

One of the evolutionary optimization schemes based on swarm intelligence is the PSO that was proposed in 1995 by Kennedy and Eberhart. The social conduct of animals such as flock of birds and schools of fish form the basis for PSO. This scheme comprises a swarm of particles wherein every particle refers to a feasible solution for an optimization issue [14]. Each particle adjusts its path so as to search the problem space to progress towards the final most optimal solution. Thus, it advances towards its best personal experience as well as the best swarm experience. Let us say that N is the size of the population. For particle i (1 ≤ i ≤ N) in D-dimension space, current position is $x_i = (x_{i1}, x_{i2}, ..., x_{iD})$ and velocity is $v_i = (v_{i1}, v_{i2}, ..., v_{iD})$ .During optimization process, velocity and position of each particle at each step is updated by (5) and (6):

$$v_{i,j}(t+1) = wv_{i,j}(t) + c_1R^1_{i,j}(Pbest_{i,j}(t) - x_{i,j}(t)) + c_2R^2_{i,j}(Gbest_j(t) - x_{i,j}(t)) \quad (5)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (6)$$

Here, w refers to the inertia weight. This may either be a constant number or a positive function. The component j of particle i is referred to as $x_{i,j}$. The acceleration coefficients are c_1 and c_2. A random number that has uniform distribution in interval [0,1] is the parameter R. The optimal location discovered by particle i till time t which is the best personal experience of the particle i is $Pbest_i(t)$. The optimal location discovered by the members of the entire group is referred to as Gbest(t) which happens to be the best personal experience of the whole group. The best individual experience of a particle i, at every iteration is described by (7):

$$Pbest_i(t+1) = \begin{Bmatrix} Pbest_i(t) & if\ f(x_i(t+1)) \ge f(Pbest_i(t)) \\ x_i(t+1) & if\ f(x_i(t+1)) < f(Pbest_i(t)) \end{Bmatrix} \quad (7)$$

Here, the fitness value of vector x is given by f(x). (8) represents the best swarm experience.

$$Gbest(t+1)=\arg\min_{p_t} f(Pbest_i(t+1)),\ 1\le i\le N \quad (8)$$

The initial PSO has been formulated for real value continuous search spaces and for discrete spaces there are many different versions that are formulated. In case of a Binary PSO (BPSO) [15], a binary position represents the location of a particle. Due to the differing velocities that correspond to a particle, this concept initially seems infeasible. The equations that reconcile the PSO velocities are retained in the technique by Kennedy and Eberhart's. The difference here, however, is that, in case of BPSO, a set of possibilities, one for each component is represented by a velocity vector or a real valued vector where the value of every component is between 0 and 1. Sampling from this vector gives particle positions. However, as binary coding is inherent of a feature selection task, BPSO is both a feasible and an attractive option to use here.

3.5 Proposed PSO Optimized KNN:

This section describes the feature selection scheme that is a combination of the BPSO and KNN. The stepwise process is depicted below. Here, from a collection of documents, a total of N terms or features are determined beforehand [16].

Step (1): A population of particles on n dimension in feature space is created. There are 3 vectors used for demonstrating every particle. The velocity, current position and best previous position of the particle are respectively represented by (V_i), (X_i) and (P_i). Random binary values are used for initializing X_i.0 means a feature is unselected and 1 means the opposite. A duplicate of X_i is used for initializing P_i. After every particle is analyzed, the index of the particle with the best fitness value is used for initializing global gbest.

Step (2): The K-NN is used for evaluating the fitness of every particle followed by updating the personal best of the particle.

Step (3): The global best of the Gbest is evaluated for every particle.

Step (4): In accordance to the standard scheme in BPSO, the velocity and the position of every particle in the population is updated.

Step (5): Upon satisfying the termination criteria, the process if truncated. The selected subset of features results which is denoted by the present global best particle. If not revert to step (2).

The fitness of a particle is calculated using (9):

$$Fitness=(\alpha\times ACC)+(\beta\times((N-T)/N)) \quad (9)$$

Where

- Acc is the classification accuracy of the particle found using K-NN.
- α and β are two parameters used to balance between classification accuracy and feature subset size, where α is in the range [0, 1] and $\beta=1-\alpha$.
- N is the total number of features.
- T is length of the selected subset of features.

The classification accuracy of a particle (P) is calculatedusing the following procedure:

- Filter the subset of features selected by P.
- Set C=0.
- For each instance in the training set
- Calculate the Euclidean distance from the current instance to all instances in the training set.
- Classify the current instance according to its K-NN in the training set.
- If the predicted classification matches the known classification of the instance, increase C by 1.

Finally the classification accuracy of P is recorded as C divided by the total number of instances in the training set.

4. Results and Discussion:

In this section, the Reuter's and 20 newsgroups datasets are used. The HAC, KNN, PSO optimized KNN, TF-IDF, TF-IDF with PSO feature selection and TF-IDF with concept expansion and PSO feature selection methods are used. The entropy and F measure (for both Reuters and 20 news group) as shown in tables 1 to 4 and figures 1 to 4.

Table 1: F Measure for Reuters

	TF-IDF	TF-IDF with PSO Feature Selection	TF-IDF with Concept Expansion and PSO Feature Selection
HAC	0.714	0.748	0.8
KNN Clustering	0.698	0.722	0.774
PSO Optimized KNN Clustering	0.728	0.764	0.852

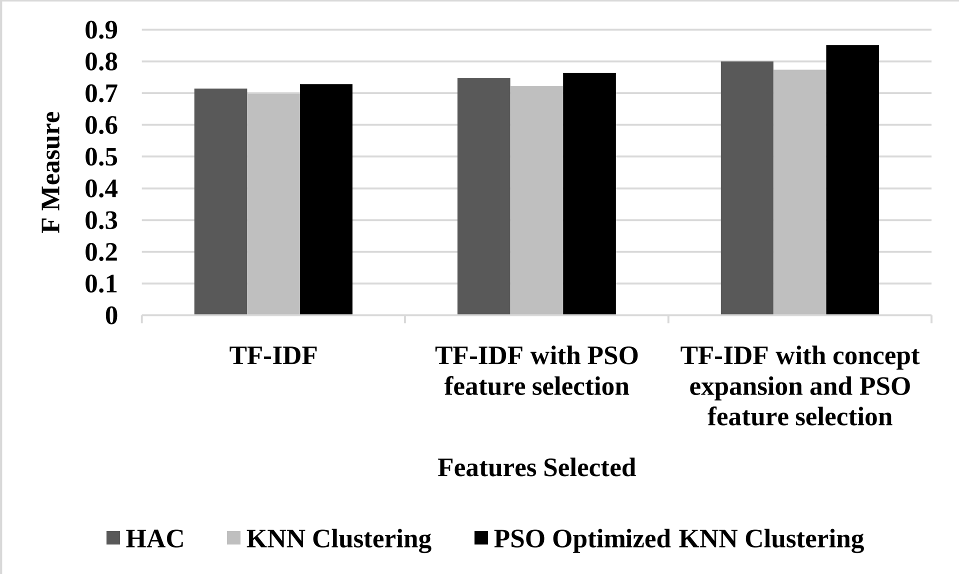

Figure 1: F Measure for Reuters

From the figure 1, it can be observed that the TF-IDF with concept expansion and PSO feature selection has higher f measure for Reuters by 11.36%, 10.32%& 15.69% for TF-IDF and by 6.72%, 6.95%& 10.89% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively.

Table 2: Entropy for Reuters

	TF-IDF	TF-IDF with PSO Feature Selection	TF-IDF with Concept Expansion and PSO Feature Selection
HAC	0.251	0.261	0.204
KNN Clustering	0.262	0.27	0.212
PSO Optimized KNN Clustering	0.214	0.222	0.14

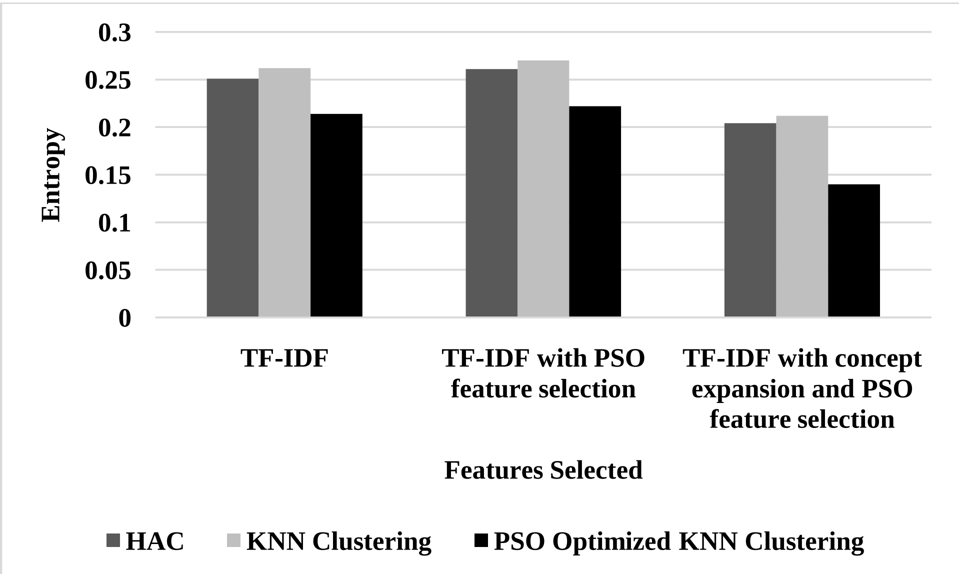

Figure 2: Entropy for Reuters

From the figure 2, it can be observed that the TF-IDF with concept expansion and PSO feature selection has lower entropy for Reuters by 20.66%, 21.09%& 41.81% for TF-IDF and by 24.52%, 24.06%& 45.3% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively.

Table 3: F Measure for 20 News Groups

	TF-IDF	TF-IDF with PSO Feature Selection	TF-IDF with Concept Expansion and PSO Feature Selection
HAC	0.562	0.589	0.693
KNN Clustering	0.532	0.551	0.677
PSO Optimized KNN Clustering	0.614	0.638	0.754

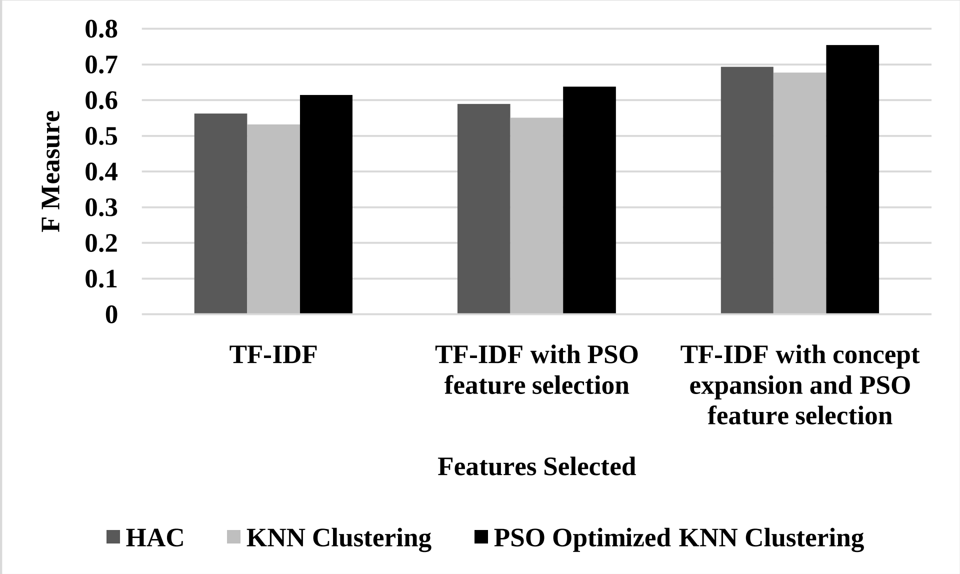

Figure 3: F Measure for 20 News Groups

From the figure 3, it can be observed that the TF-IDF with concept expansion and PSO feature selection has higher f measure for 20 news groups by 20.87%, 23.98%& 20.47% for TF-IDF and by 16.22%, 20.52%& 16.67% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively.

Table 4: Entropy for 20 News Groups

	TF-IDF	TF-IDF with PSO Feature Selection	TF-IDF with Concept Expansion and PSO Feature Selection
HAC	0.326	0.337	0.148
KNN Clustering	0.348	0.361	0.17
PSO Optimized KNN Clustering	0.287	0.296	0.143

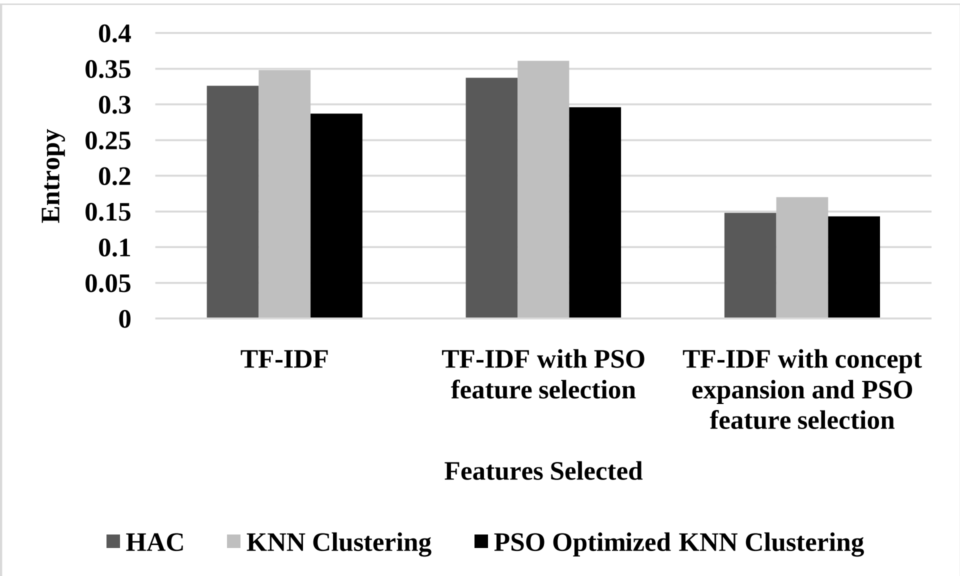

Figure 4: Entropy for 20 News Groups

From the figure 4, it can be observed that the TF-IDF with concept expansion and PSO feature selection has lower entropy for 20 news groups by 75.1%, 68.72%& 66.97% for TF-IDF and by 77.93%, 71.93%& 69.7% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively.

5. Conclusion:

Web search engines extensively make use of document clustering schemes so that the outcomes that correspond to the query are given. A scheme for unsupervised document arrangement is document clustering. A popular distance-measure based classifier is the KNN which categorizes a novel instance on the basis of the smallest class in its closest training instances. This work suggests using a PSO algorithm based feature selection scheme for determining a novel and best subset of relevant features in a text. Also, it is used to optimize the KNN classifier. In order to produce a feature subset which is the final gbest particle, BPSO-KNN was run on the training set. The test set was filtered using this feature subset. Only for the given set of terms, the test set was processed into TF-IDF vectors. Results show that the TF-IDF with concept expansion and PSO feature selection has higher f measure for Reuters by 11.36%, 10.32%&

15.69% for TF-IDF and by 6.72%, 6.95%& 10.89% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively. The TF-IDF with concept expansion and PSO feature selection has higher f measure for 20 news groups by 20.87%, 23.98%& 20.47% for TF-IDF and by 16.22%, 20.52%& 16.67% for TF-IDF with PSO feature selection when compared with HAC, KNN clustering and PSO optimized KNN clustering respectively.

References:

1. Abualigah, L. M., Khader, A. T., Al-Betar, M. A., Alyasseri, Z. A. A., Alomari, O. A., & Hanandeh, E. S. (2017, May). Feature selection with β-hill climbing search for text clustering application. In 2017 Palestinian International Conference on Information and Communication Technology (PICICT) (pp. 22-27). IEEE.
2. Bharti, K. K., & Singh, P. K. (2015). Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. Expert Systems with Applications, 42(6), 3105-3114.
3. Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.
4. Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A new feature selection method to improve the document clustering using particle swarm optimization algorithm. Journal of Computational Science, 25, 456-466.
5. Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis. Engineering Applications of Artificial Intelligence, 73, 111-125.
6. Labani, M., Moradi, P., Ahmadizar, F., & Jalili, M. (2018). A novel multivariate filter method for feature selection in text classification problems. Engineering Applications of Artificial Intelligence, 70, 25-37.
7. Lee, J., Yu, I., Park, J., & Kim, D. W. (2019). Memetic feature selection for multilabel text categorization using label frequency difference. Information Sciences, 485, 263-280.
8. Alguliyev, R. M., Aliguliyev, R. M., Isazade, N. R., Abdi, A., & Idris, N. (2019). COSUM: Text summarization based on clustering and optimization. Expert Systems, 36(1), e12340.
9. Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Awadallah, M. A. (2016, May). A krill herd algorithm for efficient text documents clustering. In 2016 IEEE symposium on computer applications & industrial electronics (ISCAIE) (pp. 67-72). IEEE.
10. Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A novel weighting scheme applied to improve the text document clustering techniques. In Innovative Computing, Optimization and Its Applications (pp. 305-320). Springer, Cham.
11. Tarczynski, T. (2011). Document clustering-concepts, metrics and algorithms. International journal of Electronics and Telecommunications, 57(3), 271-277.
12. Kumari, M., Jain, A., & Bhatia, A. (2016). Synonyms based term weighting scheme: an extension to TF. IDF. Procedia Computer Science, 89, 555-561.
13. Lamba, A., & Kumar, D. (2016). Optimization of KNN with Firefly algorithm. BVICA
14. Liu, J., Mei, Y., & Li, X. (2015). An analysis of the inertia weight parameter for binary particle swarm optimization. IEEE Transactions on Evolutionary Computation, 20(5), 666-681.
15. Chantar, H. K., & Corne, D. W. (2011, October). Feature subset selection for Arabic document categorization using BPSO-KNN. In 2011 Third World Congress on Nature and Biologically Inspired Computing (pp. 546-551). IEEE.M's International Journal of Information Technology, 8(2), 997.
16. Adeli, A., Ghorbani-Rad, A., Zomorodian, M. J., Neshat, M., & Mozaffari, S. (2012, November). Improving nearest neighbor classification using particle swarm optimization with novel fitness function. In International Conference on Computational Collective Intelligence (pp. 365-372). Springer, Berlin, Heidelberg.