OPPOSITIONAL LION OPTIMIZATION ALGORITHM AND DEEP NEURAL NETWORK BASED MULTI DOCUMENT SUMMARIZATION FROMLARGE-SCALE DOCUMENTS

K. Chandra Kumar

Research scholar in Acharya Nagarjuna University and Asst. Professor, Department of CSE, Kakinada Institute of Engineering and Technology, Korangi, E.G., Andhra Pradesh, India.

Dr. Sudhakar Nagalla

Professor, Department of CSE, Bapatla Engineering College, Bapatla, Guntur, Andhra Pradesh,India.

Abstract

Multi-document summarization (MDS) is an automatic process where the essential information is extracted from the multiple input documents.However, it has many limitations such as inaccurate extraction to essential sentences, low coverage, poor coherence among the sentences, and redundancy.So, in this paper, oppositional lion optimization algorithm (OLOA) and deep neural network (DNN) based multi-document summarization is presented. For optimal sentence selection from the pre-processed documents, OLOA algorithm is proposed. Based on the extracted features, sentence score is calculated using DNN. Finally, based on the rank of the sentences, the multi-documents are summarized. Simulation results show that the performance of the proposed summarization outperforms the state of the arts in terms of precision, recall and F-measure.

Keywords:Multi-document summarization, oppositional lion optimization algorithm, deep neural network and sentence score calculation.

1.Introduction

Multi-document summarization is the way toward producing a short form of offered materials to demonstrate their fundamental thoughts. As the quantity of archives depicting a similar occasion on the web increments exponentially, summaries of multi-document can assist individuals with getting a handle on the point inside a brief timeframe. The principle task of multi-document summarization is to separate the main sentences from various archives and structure them into a summary. The manner by which summary is produced either is an extraction or an abstraction strategy [1] [2][3]. Extraction based summaries are created by choosing the significant parts of the original text. While, abstraction based summaries requires semantic investigation to build new sentences from the original text [4-7].Contributions of this paper are described as follows:

- Initially, the documents from the web sources are pre-processed. Then, from the pre-processed documents, we choose optimal sentences using oppositional lion optimization algorithm (OLOA).
- After the selection of optimal sentences, appropriate features are extracted. Then, these extracted features are given as input to the deep neural network(DNN) which is used to calculate the score of each sentence in the documents.
- Finally, the multi-documents are summarized based on the rank of the sentences.

Rest of this paper is organized asfollows. Section 2 reviews some recent literature which focused research on document summarization. Section 3 proposes Oppositional Lion Optimization Algorithm and Deep Neural Network Based Multi Document Summarization from Large-Scale Documents. Results of the proposed approaches are discussed in section 4. Section 5 concludes this research work.

2. Related works

In this section, some recent literatures that focused research on multi-document summarization are reviewed. YE Feiyue and XU Xinchen [8] had presented Sentence-Word Graphs and Keyword Density based automatic multi-document summarization. Aim of the authors was to produce multi-document summarization. So, the authors had proposed a sentence-word two layer graph algorithmthat was combined with keyword density. This proposed approach was also called as Graph & Keyword. They had modified the sentence graph by developing multiple word graph and extracting suitable keywords from each document. Besides, the keyword density was utilized to attain rich content when utilizing less number of words. By presenting this proposed approach, the authors had achieved better accuracy.

ZeynabKhaleghi, Mohammad Fakhredanesh and Maryam Hourali [9] had presented a novel Sentences Overlapping Criterion based Extractive Multi-document Summarization. In this approach, the authors had calculated 12 features of every sentence. They had assigned a score for every sentence using the learned model. Besides, they had calculated pair wise overlappingamong the sentences. Then, they had chosen the sentences with less redundancy and maximum score. Final output summary was attained under a constraint of length. Because of this proposed document summarization, the authors had achieved betterROUGE score.

Seyed Hossein Mirshojaee, BehroozMasoumi and EsmaeilZeinali [10] had proposed aa multi-agent meta-heuristic optimization algorithm based extractive summarization. Aim of the authors was to solve the problems of document summarization. The proposed algorithm was the hybrid of multi-agent systems concept and biogeography-based optimization algorithm that was abbreviated as BBO. Because of this proposed algorithm, the authors attained the optimum summary. By presenting this approach, the authors achieved better recall and ROUGE accuracy.

Amreen Ahmad and Tanvir Ahmad [11] had presented multi-document summarization with the game theory scheme. The authors aimed to solve the problem of relevant information extraction for summarization. So, the authors had presented multi-document summarization system based on game theory method along with the Wikipedia ontology. Using this system, the authors had exploited the hidden sub modularity from the ontology and were optimized utilizing enhanced algorithm. Because of the proposed system, the authors had improved the performance of summarization in terms of ROUGE.

Asad Abdi*et al* [12] had proposed soft computing based automatic sentiment depend multi-document summarization. They had included to processes in the proposed approach. They were sentiment summarization and sentiment analysis. The sentiment analysis was presented to solve the issues of sentence types, polarity of contextual and word coverage restriction. For sentiment summarization, the authors had presented ranking model based on graph that integrated sentiment data, linguistic and statistical schemes for enhancing the result of sentence ranking. By presenting this proposed approach, they had improved the Average ROUGE Score of summarization.

Pradeepika Verma and Hari Om [13] had presented hybrid sentence scoring schemes based text summarization. In the research work, the authors had extracted relevant and salient sentences from the set of documents. Besides, they had generated hybrid of sentence scoring schemes. Also, they had extracted sentences using the optimization based on teaching–learning. The authors had compared their proposed approach with the various meta-heuristic schemes and they proved that the accuracy of the proposed summarization method was improved than the state of the arts.

V. Priya and K. Umamaheswari [14] had proposed the improved continuous and discrete multi objective particle swarm optimization for summarization. Aim of the authors was to solve the inefficient of summarize the huge content. To solve this issue, the authors had presented two multi objective optimization schemes known as Discrete and Continuous. For extractive summarization, the proposed technique performed with the particle swarm optimization algorithm. They had evaluated the performance of the proposed technique in terms of Inverted Generational Distance, ROUGE and Success Counting.

Minakshi Tomer and Manoj Kumar [15] had presented Fuzzy with LSTM based ensembled scheme for enhancing text summarization. In this approach, the authors had combined fuzzy logic rules with the Bi-LSTM for producing abstractive summary. Weights of the network were updated using Adam optimizer and attention mechanism in BI-LSTM. For extracting the textual information from the set of documents, measures of fuzzy and inference were utilized. Based on the proposed approach, most appropriate sentences were found. By presenting this approach, the authors had achieved better ROUGE.

3. OLOA and DNN Based Multi Document Summarization

3.1. Overview

For efficient multi-document summarization, Oppositional Lion Optimization Algorithm (OLOA) and Deep Neural Network(DNN) are presented in this approach. Initially, the web documents are pre-processed. Then, from the pre-processed documents, the optimal sentence is chosen by methods for oppositional lion optimization algorithm (OLOA).Special lifestyle of lions and their cooperation characteristics has been the basic motivation for development of the lion optimization algorithm (LOA). Here the traditional LOA algorithm is altered by means of opposition based learning algorithm.

After selecting the relevant sentence, feature values are extracted. Finally, the sentences are scored by applying deep learning algorithm for the automatic summary creation from Large-scale Documents. For sentence score calculation, deep neural network (DNN) is presented. Finally, the multi-documents are summarized based on the rank of the sentences.

3.2. Pre-processing

The HTML/XML Tag Images Removal, Sentence Tokenization, Stop word Removal and the Stemming processes take place in Document Pre-Processing step.

***HTML/XML Tag Images Removal*:**From any online origins, the process of removal of HTML/XML Tag Images is done when the input documents are gathered. Moreover, the extra white space, figures, equations and the special characters like ‘{[(<!@#$%^&*~`:+;>)]}?’ are also removed.

***Sentence segmentation*:**After the removal of unnecessary elements introduces on the documents, in summary, creation, each sentence is to be segmented from the documents for the ease of processing. Let us conceivethe document database, $N = \{n_1, n_2, \mathrm{K}\ n_T\}$, where, n_k represents the k^{th} document in the database, N. For each text document ‘n_k’, the number of sentences is segmented separately as, $A = [A_1, A_2, \mathrm{K}\ , A_r]$. Where, A_q depicts q^{th} the sentence in the document for simple extraction of the short sentence and n is the number of sentences in the document.

***Tokenization*:**Next, the Sentence Tokenization is executed to divide the set of sentences from each document. Hence, the words/terms of each sentence are tokenized as, $X = [X_1, X_2, \mathrm{K}\ , X_m]$, where $X_p \in X$ for $p = 1,2,\mathrm{K}\ m$ represents all the distinct terms founding in the document ‘n_k’ and ‘m’ is the number of unique terms.

***Stop word Removal*:**After tokenizing the Sentences, the stop words are dispatched; where, the most usual words applied in the English language like ‘a’, ‘an’, ‘the’ etc. which has less significant centrality as for the document are taken out.

***Stemming*:**Stemming is a procedure of slashing off the parts of the bargains a typical base structure. Here, the stemming is done set up on the Porter Stemming Algorithm. The algorithm is extremely straightforward in idea, with around 60 additions, two recoding rules and a solitary kind of setting the touchy standard to decide if a postfix ought to be taken out.

3.3. Optimal sentence selection using OLOA

The alignment of the proposed Multi-Document Summarizer can be modeled utilizing simply maximizing the input sentence set utilizinganoppositional lion optimization algorithm (OLOA) before calculating the sentence score.

For a wild felid lion, inhabitant and migrant are the social organization. Pride is a group of residents. The combination of five females, their cubs with both genders and minimum one adult male is called as pride of a lion. Until from becoming adult, the young males are forbidden from their basis of the pride. The second valid behavior is called as migrant who moves infrequently either sets or complete. The males who have been rejected from their parental pride are observed by the sets. The lions may change their habitat behavior as either be a migrant or reside over a particular place. Normally lions hunt together with different characters for their pride but in contrast, the cats search over individually. The prey has been cornered and focused by few lions in different ways and attacks the victim quickly. The probability of the success rate is higher while hunting the prey by forming the group.

The phases of this algorithm for selecting optimal RN are described as follows:

Initialization: Each solution $S_b\ (b=1,2,...,S)$ is a D- dimensional vector wherever D is the number of optimization limitations (i.e. the number of words inaugurating in a sentence). The b^{th} solution at v^{th} dimension can be generated by applying the below equation,

$$S_{bv} = S_v^{\min} + rand\left[0,1\right]\left(S_v^{\max} - S_v^{\min}\right) \tag{1}$$

Where, $rand\left[0,1\right]$ is the regularly scattered accidental numeral, $S_v^{\min}$ and $S_v^{\max}$ are the bounds of S_b at v^{th} dimension.

Fitness calculation: The primary fitness of the population is estimated after initialization. Here, the objective function is found from the maximum value of the term frequency similarity. The fitness of a sentence (S_x) can be estimated utilizing the below equation.

$$fitness(S_x) = \max\left(\frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(S_x(j) - S_i(j)\right)^2\right) \tag{2}$$

Where, I represents the total number of sentences from multiple documents; J depicts the number of unique words. Also, $S_x(j)$ represents the j^{th} unique term in the sentence 'S_x' and $S_i(j)$ is the j^{th} unique term in the other remaining sentences.

Update the solution: After the fitness calculation, we update the lion using lion algorithm. The process of Lion algorithm is explained below;

***Hunting*:** To offer food for their pride, femalelions in each pride search for prey.For encircling prey and catch them,thehunters follows particular policies. There are three subgroups which classifies the hunters. The highest fitness in the group is conceived as the center and the other group is conceived as the right and left wings. In the centre of hunters a dummy prey (Py) is considered and is evaluated as follows:

$$Py = \frac{\sum Hunters or\ solutions}{Number of\ hunters} \tag{3}$$

In the event of hunting, this $P_y$is attacked by the hunters which are selected randomly. The P_y(prey) will escape from the hunter during hunting when the hunter develops its own fitness and the novel position of the prey can be shown below,

$$\mathrm{P}_y^{'} = (rand(0,1) \times \%PI \times (P_y - H)) + Py \tag{4}$$

Where, P_y represents the current position and H refers the hunter. In the left and right wings the new position of the hunters can be calculated as follows,

$$H^{'} = \begin{cases} rand((2\times P_y - H), P_y), & (2\times P_y - H) < P_y \\ rand(P_y, (2\times P_y - H)), & (2\times P_y - H) > P_y \end{cases} \tag{5}$$

The centre hunter new position was represented as follows,

$$H^{'} = \begin{cases} rand(H, P_y), & H < P_y \\ rand(P_y, H), & H > P_y \end{cases} \quad (6)$$

Where, rand(x, y) stands for random values between y and x. The x and y stands for upper and lower bound respectively.

Moving toward a safe place**:** The rest of the females apart from pride are move towards the safe territory.As each territory has best solutions, the solutions in LOA can be improved. In equation (7), the new position of the female lion is given,

$$\mathrm{PF}_{\mathrm{L}}^{'} = PF_L + 2d \times rand(0,1)\{S_1\} + U(-1,1) \times \tan(\theta) \times d \times \{S_2\} \quad (7)$$

$$\{\mathrm{S}_1\} \cdot \{S_2\} = 0, \ \ \|\{S_2\}\| = 1 \, (8)$$

Where, the female line position is denoted as PF_{L}, the point chosen during the tournament selection among the pride territory and d the distance among the F_L is denoted as d, the starting point of the previous location of a female lion and the vector with its direction towards the chosen point is denoted as $\{\mathrm{S}_1\}$ and $\{S_2\}$ is at right angles to $\{\mathrm{S}_1\}$.

When the position of the lion (male & female) is developed, the success of a lion is determined during the tournament selection in the lion optimization algorithm and it is given below,

$$\mathrm{SU(i, t, G)} = \begin{cases} 1 & \mathrm{Bst}_{\mathrm{i,G}}^{\mathrm{t}} < Bst_{i,G}^{t-1} \\ 0 & Bst_{i,G}^{t} = Bst_{i,G}^{t-1} \end{cases} \quad (9)$$

Where, the best position of the lion I at the iteration t can be denoted as $\mathrm{Bst}_{\mathrm{i,G}}^{\mathrm{t}}$.

From the above positive measures, the lions have coverage even at the long distance from its ideal point. The negative measure is that the lions are roll through optimum solution without critical enhancement. So the size of the competition is assessed applying the success values

$$\mathrm{T}_{\mathrm{S_j}}(SU) = \sum_{i=1}^{N} SU(i, t, G) \quad j = 1, 2, \ldots N \quad (10)$$

Where, the number of lions in the pride j which has the fitness improvement during last iteration is denoted as $\mathrm{T}_{\mathrm{S_j}}(SU)$.

For each iteration, the size of the tournament may vary, i.e. $\mathrm{T}_{\mathrm{S_j}}(SU)$ is increased which leads to diversity when SU value decreases. So that the tournament size is computed applying below equation,

$$\mathrm{T}_{\mathrm{S}_j} = \max\left(2, ceil\left(\frac{T_{S_j}(SU)}{2}\right)\right) \quad j = 1,2,...G \tag{11}$$

***Roaming*:** In this section, each male lions are a move to territory due to some problems.To enhance the solution for lion optimization algorithm, a solid local pursuit called roaming is used.The lion pushes toward the chose zone of a territory by n units, where n stands for consistent distribution with irregular number.

$$n \sim U(0, 2 \times d) \tag{12}$$

Where,the distance between male lion position in chosen area of territory is denoted as d. Also, nomad lion moving randomly in search space. The migrant lion novel position is generated as follows,

$$L_{ij} = \begin{cases} L_{ij} \ \ if\ rand_j > prob_i \\ RAND_j \ \ otherwise \end{cases} \tag{13}$$

Where,

L_{ij} Stands for migrant lion current position

$rand_j$ Stands for the consistent random number between [0, 1]

$RAND_j$ Stands for the vector that are generated randomly

$prob_i$ Stands for probability value

Here the probability value of every migrant lion independently calculated as follows,

$$prob_i = 0.1 + \min\left(0.5, \frac{(N_{O_i} - Bst_{N_O}}{Bst_{N_O}}\right) \quad i = 1,2,...M \tag{14}$$

Where,

N_{O_i} *and* Bst_{N_O} Stands for ith migrated lion current position with minimum delay

M Stands for the no of nomad's lion

***Mating*:** The mating operator is the linear blend of parents who produces two new offspring. The offsprings are calculated using equation (15) and (16).

$$Offspring_j 1 = \alpha \times F_{L_j} + \sum \frac{1-\alpha}{\sum_{i=1}^{N_R} SU_i} \times M_{L_j} \times SU_i \tag{15}$$

$$Offspring_j 2 = (1-\alpha)\times F_{L_j} + \sum \frac{\alpha}{\sum_{i=1}^{N_R} SU_i} \times M_{L_j} \times SU_i \qquad (16)$$

Here, the dimension is j. When $SU_i = 1$ or $SU_i = 0$, the i^th^male is selected for mating.

The resident males during a pride are denoted as N_R.

The mean value and the standard deviation of the random number with consistent distribution are 0.5 and 0.1 respectively and where it is denoted as α .

The lion algorithm shares the information among the genders and the new cubs inherit the characters from both of the genders using mating.

***Defense*:** The male lions when they become adult, they battle against various others with force during their pride. The defeated fellow kept alone from their and turn into a migrant. Still, the lions which are defeated by the migrant male lions will be out of the pride and the migrant lions may take control of the pride. The defense operator was limited into two steps in lion optimization and they are;

a. Defense across a novel mature resident males
b. Defense across the migrant males

The powerful male lions which play an important role in lion optimization algorithm are retained by assisting the defense operator with the lion optimization algorithm.

***Migration*:** Some of the female lions chose arbitrarily and forwarded towards migrants during the migration process. According to fitness, the new migrant and old migrant lions are arranged. Then the unfilled places of migrated females are filled by selecting the best females to allocate in the pride arbitrarily. The variety of whole people and their data shares among the pride are kept by this technique.

***Lions Population Equilibrium*:** At the end of every cycle,the number of alive lions are been controlled by balancing the lions populace.The migrant lions with minimum fitness will be removed during their sexual orientation.

Termination Criteria: The solutions are updated until finding the optimal solution or RN in the network. The algorithm will be stopped when the optimal solution is obtained. The LOA algorithm for RN selection is shown in figure.

3.4. Sentence score calculation using DNN

The extracted features are given as input to the proposed ODNN which is used to calculate the score for each sentence from the documents. The calculation of sentence score using ODNN is explained in the following sections.

3.4.1. Deep Neural Network (DNN)

A deep neural network (DNN) is an artificial neural network (ANN) between the layers of the input and the output has plenty of hidden units.

Training in ODNN**:** To streamline production, use the Deep Neural Network (DNN), an in-depth program, and a regular feedforward system, in which information travels from the input layer to the output layer through several different hidden layers. The DBN model empowers the structure to make obvious origins subject to its hidden layers' states, which depicts the system conviction. Here, fixing the problem in the above Restricted Boltzmann machine (RBM) exercised. DNN of the system shown in Figure 1.

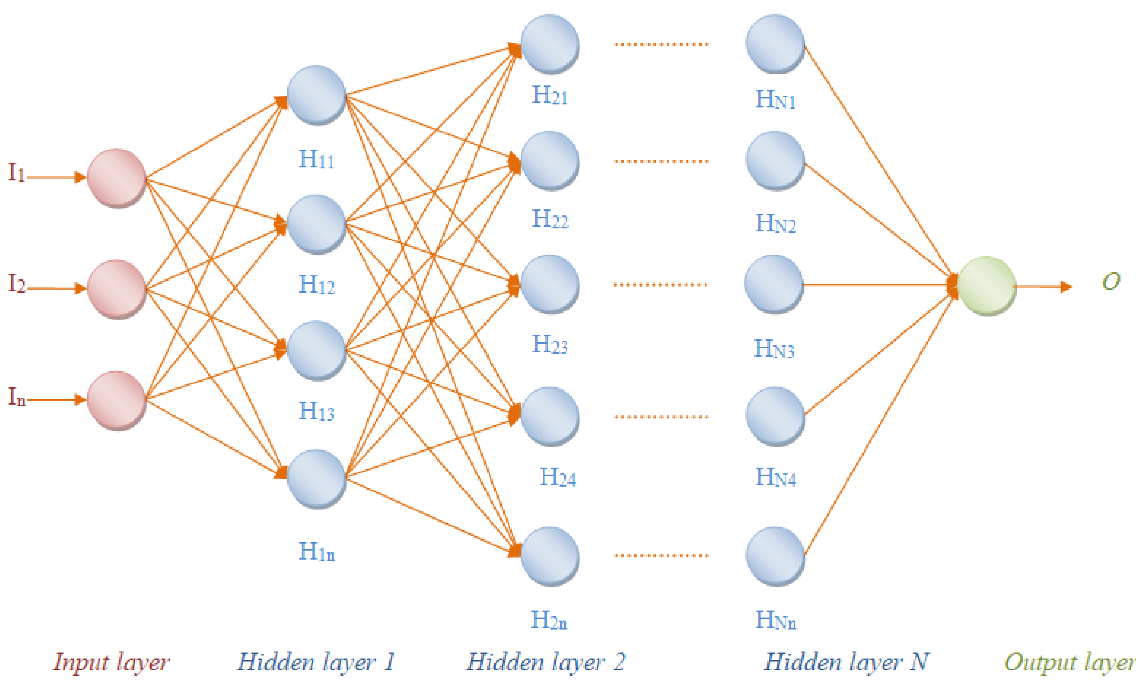

Figure 1: Deep Neural Network structure

Restricted Boltzmann Machine (RBM):

Assume $[R_m]$ is the input dataset where $1 \le \mathrm{m} \le \mathrm{N}$ and 'C' represents the output dataset. In DNN, there are more hidden layers; the first hidden layer increases the weights on the individual element inputs. Likewise, the individual output of the first hidden element in the second hidden layer multiplied by another set of weights.

In the first hidden layer, the input weighted values are summed with the bias value of the neuron and the output of the first hidden layer is dented as follows:

$$C_{H\_1}(x = 1,2..,K) = \left(\sum_{m=1}^{M} w_{xm} R_m \right) + b_x \quad (17)$$

Hence, b_x steady worth goes about as predisposition, w_{xm} is the interconnection weight between the information and concealed layer with M and K talking about the amount of information and the hidden nodes in the main hidden layer.

Efficiency of the main hidden layer is given as,

$$F\left(C_{H\_1}(x)\right) = \frac{1}{\left(1 + e^{-C_{H\_1}(x)}\right)} \quad (18)$$

Hence sigmoid activation function is $F(\cdot)$,

Consequently, the hidden layer y^{th} operation is generalized as,

$$C_{H\_y}(p) = \left(\sum_{z=1}^{K} w_{pz} F\left(C_{H\_(y-1)}(z)\right) \right) + b_p \quad (19)$$

Where, w_{pz} the intermediate is weight connecting the $(y-1)^{th}$ hidden layer and $(y)^{th}$ hidden layer with K hidden nodes, b_p is the bias of p^{th} hidden node.

The actuation work which is the efficiency y^{th} of the hidden layer is specified as,

$$F\left(C_{H\_y}(p)\right)=\frac{1}{\left(1+e^{-C_{H\_y}(p)}\right)} \quad (20)$$

In the output layer, the output of y^{th} wrapped hidden layer is doubled again with interconnecting loads (i.e. weight between the hidden layer and output layer is represented as y^{th}) and afterward summarized with the inclination (b_q) as

$$C(q)=F\left(\sum_{p=1}^{K} w_{qp}\ f\left(C_{H\_y}(p)\right)+b_q\right) \quad (21)$$

Where, w_{qp} talks about the interconnection weight in the y^{th} hidden layer and the output layer that has p^{th} and q^{th} nodes separately. Getting started work on the output layer develops as the output of the entire model.

Currently, the output system stands out from the target and the contrast is obtained (i.e., error) to optimize the output of the system. The calculation of the error is a condition (22)

$$E=\frac{1}{M}\sum_{m=1}^{M}\left(Actual(C_m)-\Pr edicted(C_m)\right)^2 \quad (22)$$

Where, $\Pr edicted(C_m)$ is the assessed system yield and $Actual(C_m)$ is the true output. The error must be limited to getting ideal system organization. Consequently, the weight esteems must be balanced until the error gets reduced at each cycle.

***Testing phase*:** After the completion of training process, we attain the trained structure of DNN. In this trained model, 20% data from the documents are taken as input for testing it. The processes that have been done for training process are also followed for testing phase. Namely, the pre-processing, data conversion, feature extraction and sentence score calculation are also done for the testing data. By processing the testing data, the proposed DNN outputs the classification of sentences within the score [0, 1].

3.5. Summarization

Here, the positioning of sentence is done utilizing the sentence score acquired from the past phase. At first, sentences in the document are arranged in descending order depend on the score of final sentence. At that point, the top-sentences are chosen for the summary dependent on the rate of compression given by the input user. Finally, the chosen top-sentences are requested in a consecutive manner d to acquire the final summary.

$$N = \frac{C \times N_S}{100} \tag{23}$$

Where, N_S represents the total count of sentences in the document and C denotes the rate of compression.

4. Results and Discussions

The proposed Multi-Document Summarization is implemented in the platform of JAVA, the proposed algorithm is accomplished and the experimentation is done in a framework that contains 4 GB RAM and 2.10 GHz Intel i-3 processor.

The documents were gathered from different databases for analysis. An established on word count, this gained information was pre-handled and traded into numerical information. For selecting optimal sentence sets, then the proposed OLOA algorithm is utilized. For sentence score calculation, DNN is presented.

4.1. Performance analysis

This segment is instanced about the performance evaluated of the proposed OLOA-DNN optimization algorithm. In the below tables, the Precision, Recall and F-Measure values for the suggested and existing methods are tabulated.

Iterations	**Precision**	**Recall**	**F-measure**
10	72.2	70.2	71.17
20	72.1	71.02	71.6

30	73.21	72.5	72.8
40	74.4	72.9	73.7

Table 1: Precision, Recall and F-Measure values for the proposed OLOA-DNN technique

Iterations	Precision	Recall	F-measure
10	71.6	69.4	70.7
20	71.8	70.5	71.4
30	72.3	72.4	72.7
40	74.2	72.3	73.2

Table 2: Precision, Recall and F-Measure values for existing PSO-DNN technique

Iterations	Precision	Recall	F-measure
10	71.3	70.04	70.7
20	71.6	70.4	70.9
30	72.3	71.6	71.8
40	73.7	72.08	72.7

Table 3: Precision, Recall and F-Measure values for existing DNN technique

From the analysis, it is clear that the precision value is reliable for the proposed OLOA-DNN for all the iterations on comparing with the values of the existing PSO-DNN and DNN techniques. Table 1 & Figure 2 show the performance analysis of the proposed OLOA-DNN algorithm. As shown in the table and figure, precision, recall and F-measure of OLOA-DNN are analyzed by varying number of iterations. The precision of OLOA-DNN is increased when the number of iterations increases. Similarly, recall of OLOA-DNN is 61% at 10 number of iterations while it is 64% at 40 number of iterations. Besides, F-measure of OLOA-DNN is 65% at 40 number of iterations. From these analyses, we inferred that the performance of multi-document summarization is enhanced due to the hybridization of PSO-DNN and DNN algorithm.

The performance analysis of existing PSO-DNN is shown in table 2 & figure 3. From the table and figure, precision of PSO-DNN is reduced to 0.26% than the proposed OLOA-DNN at 40 number of iterations. Also, recall and F-measure of PSO-DNNare decreased to 0.96% and 0.62% respectively that of OLOA-DNNat 40 number of iterations. Table 3 and figure 4 show the performance analysis of existing DNN technique. Compared to the proposed OLOA-DNN, precision, recall and F-measure of the cuckoo search algorithm are reduced to 1.1.%, 1.2% and 1.14% respectively.

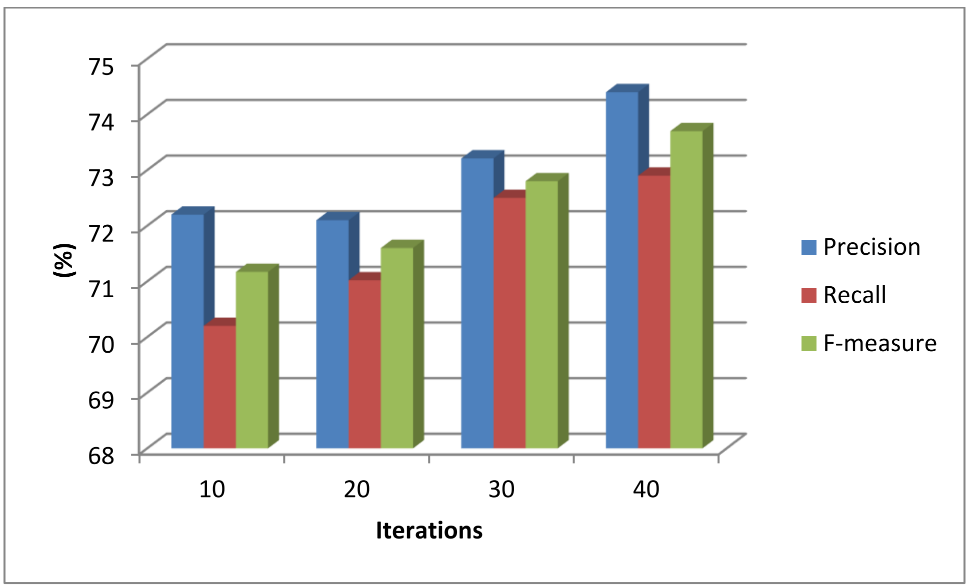

Figure 2: Precision, Recall and F-Measure values for the proposed OLOA-DNN technique

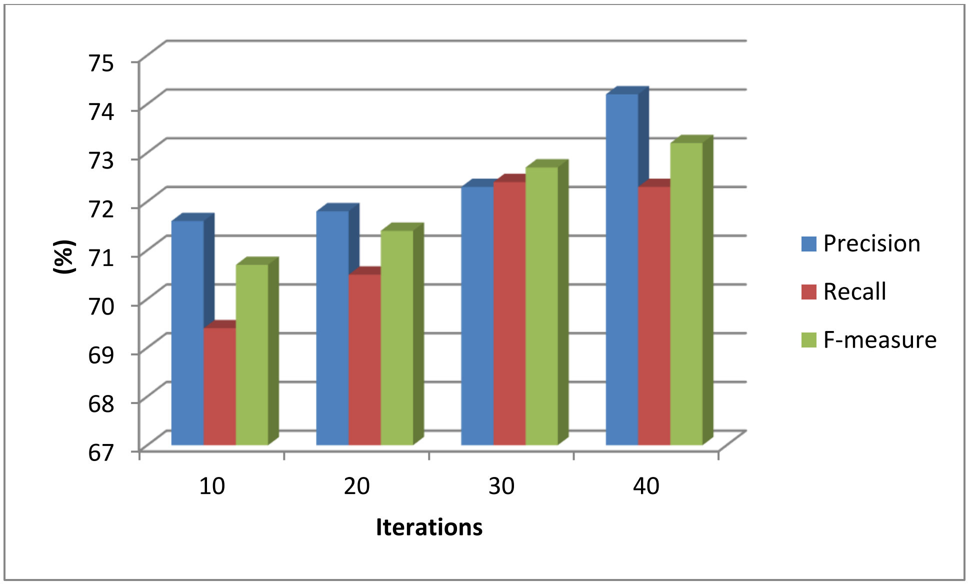

Figure 3: Precision, Recall and F-Measure values for the existing PSO-DNN technique

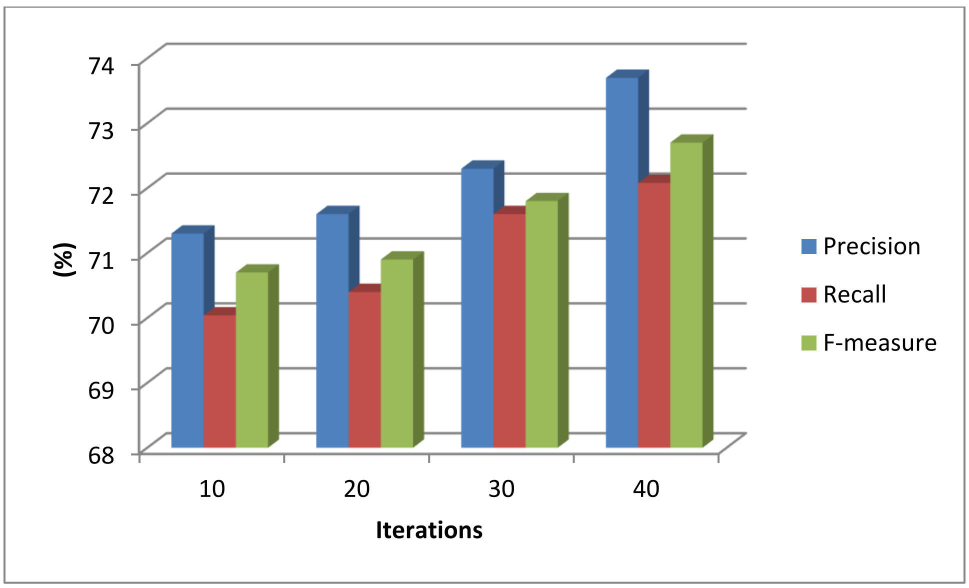

Figure 4: Precision, Recall and F-Measure values for the existing DNN

From the above comparison plot, the proposed OLOA-DNN technique is additional dependable than several extra existing techniques. While equating Precision, Recall with F-Measure, the planned technique demonstrates 100% than the existing PSO-DNN and DNN approaches.

5. Conclusion

For efficient multi-document summarization, oppositional lion optimization algorithm (OLOA) and deep neural network (DNN) have been presented in this paper. After pre-processing, optimal sentences have been chosen using OLOA algorithm. After the selection of optimal sentences, score is calculated for each sentence using DNN. Finally, based on the rank of the sentences, documents have been summarized.The performance of the proposed summarization has been evaluated in terms of precision, recall and F-measure. Simulation results showed that the F-score of the proposed multi-document summarization has been improved to 0.62% and 1.14% than existing PSO-DNN and DNN algorithm.

References

1. Liao, Tao, and Zongtian Liu. "Research of Summarization Extraction in Multiple Topics Document." In 2009 8thIEEE Int. Conf. on Dependable, Autonomic and Secure Comp., pp. 859-860. IEEE, 2009.
2. Kim, Heechan, and Soowon Lee. "A Context based Coverage Model for Abstractive Document Summarization." (2019).
3. Modi, Shivangi, and Rachana Oza. "Review on Abstractive Text Summarization Techniques (ATST) for single and multi documents." In 2018 Int. Conf. on Computing, Power and Communication Tech. (GUCON), pp. 1173-1176. IEEE, 2018.
4. Agarwal, Sanchit, Nikhil Kumar Singh, and Priyanka Meel. "Single-Document Summarization Using Sentence Embeddings and K-Means Clustering." In 2018 Int. Conf. onAdv. in Computing, Comm. Control and Netw. (ICACCCN), pp. 162-165. IEEE, 2018.
5. Hariharan, Shanmugasundaram, and Rengaramanujam Srinivasan. "Investigations in single document summarization by extraction method." In 2008 Int. Conf. on Computing, Communication and Netw., pp. 1-5. IEEE, 2008.
6. Mokhale, Shweta V., and Gauri M. Dhopawkar. "A Study on Different Multi-Document Summarization Techniques." In 2019 3rdInt. Conf. onInventive Systems and Cont. (ICISC), pp. 710-713. IEEE, 2019.
7. Meena, Yogesh Kumar, Ashish Jain, and Dinesh Gopalani. "Survey on graph and cluster based approaches in multi-document text summarization." In Int. Conf. onRecent Adv. and Innovations in Engg. (ICRAIE-2014), pp. 1-5. IEEE, 2014.

8. F. Ye and X. Xu, "Automatic Multi-Document Summarization Based on Keyword Density and Sentence-Word Graphs", Journal of Shanghai Jiaotong Univ. (Science), vol. 23, no. 4, pp. 584-592, 2018.
9. Khaleghi, Zeynab, Mohammad Fakhredanesh, and Maryam Hourali. "MSCSO: Extractive Multi-document Summarization Based on a New Criterion of Sentences Overlapping." Iranian J. of Sci. and Tech., Trans. of Electrical Engg. (2020): 1-11.
10. S. Mirshojaee, B. Masoumi and E. Zeinali, "MAMHOA: a multi-agent meta-heuristic optimization algorithm with an approach for document summarization issues", J. of Amb. Int. and Humanized Comp., vol. 11, no. 11, pp. 4967-4982, 2020.
11. Ahmad, Amreen, and Tanvir Ahmad. "A Game Theory Approach for Multi-document Summarization." Arabian J. for Sci. and Engg.44, no. 4 (2019): 3655-3667.
12. Abdi, Asad, SitiMariyamShamsuddin, Shafaatunnur Hasan, and JalilPiran. "Automatic sentiment-oriented summarization of multi-documents using soft computing." Soft Comp. 23, no. 20 (2019): 10551-10568.
13. P. Verma and H. Om, "A novel approach for text summarization using optimal combination of sentence scoring methods", Sādhanā, vol. 44, no. 5, 2019.
14. Priya, V., and K. Umamaheswari. "Enhanced continuous and discrete multi objective particle swarm optimization for text summarization." Cluster Comp., 22, no. 1 (2019): 229-240.
15. M. Tomer and M. Kumar, "Improving Text Summarization using Ensembled Approach based on Fuzzy with LSTM", Arabian J. for Sci. and Engg.,2020.