https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

A Comparative Study of Text Summarization Based on Synchronous and Asynchronous PSO

R. Pallavi Reddy[1], Kalyani Nara[2], S. Sravani Reddy[3]

[1]Assistant professor, Dept of CSE, GNITS, Hyderabad, Telangana, India.
[2]Professor, Dept of CSE, GNITS, Hyderabad, Telangana, India.
[3]PG Scholar, Dept of CSE, GNITS, Hyderabad, Telangana, India.

***Abstract*—*Text summarization is the process of extracting the most important sentences from the original document without its meaning change. The paper focus on Extractive summarization technique which chooses the important sentences from the document and integrates into summary. An extractive summarization technique, Particle swarm Optimization performs arithmetic operations that enhances a problem, by iteratively trying to improve possible solution with regard to input data. It determines a problem by having a population of possible solutions moving around the search space according to arithmetic formulae over the particles position and velocity. The sequence of modernized particles of PSO can be categorized into Synchronous PSO (S-PSO) and Asynchronous PSO(A-PSO). In synchronous PSO, after calculating the whole performance, velocities and positions of the particles are modernized, this increases the performance. In A-PSO after calculating its performance, velocities and positions of the particles are modernized using partial data which leads to extreme analysis. The comparative study on the synchronous PSO and asynchronous PSO with the precision and recall values for different datasets is considered. Asynchronous PSO has higher precision and recall values compared to synchronous PSO. Asynchronous PSO leads to extreme analysis of data.*

Keyword*—*Text Summarization, particle swarm optimization, Synchronous PSO (S-PSO), Asynchronous PSO (A-PSO).

I. INTRODUCTION

Text summarization is the process of distilling the most important information from the source document to produce a abridged version of text. Automatic text summarization is to present the input text into a summary. The main advantage of using a summary is abating the reading time. Text summarization techniques can be classified into extractive and abstractive summarization. An extractive summarization method elites important sentences, paragraphs etc. from the original document and concatenating them into short data. An Abstractive summarization is an adapting of the main concepts in a document and then expresses those concepts in clear natural language.

Generally Extraction methods use sentence extraction technique to create the summary. In 1995 Kennedy and Eberhart introduced Particle swarm optimization (PSO) [1]. PSO is stochastic optimization algorithm depends on the swarm that simulate the social behavior of organisms such as birds and fishes. These organisms' benefits in search for food sources through distinctive work with neighbors. In PSO, the distinctive agents depicted by a swarm are called particles. The particles move within the search space to find the optimum solution by modernizing their velocity and position. These values are affected by the participation of the particles. PSO has drawn a lot of attentions from the researchers all over the world. PSO has sustained many evolutionary processes. Many variations of PSO have been proposed to improve the performance of the algorithm. The particles update sequence effects on the efficiency of PSO. In PSO, after evaluating the whole performance the best found solution is chosen as PBest from the Particle information. This method of PSO algorithm is known as synchronous PSO (S-PSO). The update method leads to the exploitation of the data.

In Asynchronous PSO (A-PSO), the position and velocity are modernized as soon as a particle's performance is evaluated. Therefore, a particle's search is directed by the partial or flawed information from its neighbor. This method leads to distinctness in the swarm [3]. In the beginning of iteration, the particles are updated using previous iterations while particles are updated at the end of the iteration based on the existing iteration [4]. A-PSO has been asserted to perform better than S-PSO. Xue et al. [8] reported that asynchronous update leads to a shorter execution time. Asynchronous method attempt on the incomplete information of the current best found solution communicated to the particles more slowly, thus lead to more exploration.

A comparative study is performed on the two algorithms to determine which algorithm support for a better summary. The paper is further organized as follows: The text summarization technique correspond to

https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

Preprocessing and Feature Extraction as their initial stage. These steps are briefly explained in section II. The synchronous PSO (S-PSO) algorithm is explained in detail in section III. The asynchronous PSO (A-PSO) algorithm is highlighted in section IV. Various input documents relating to different domains are given as input data. The results obtained from algorithms are used to calculate precision and recall values. The analysis of results are given in section V. The conclusions stated in section VI based upon the experimental evaluations from section V.

II. PREPROCESSING AND FEATURE EXTRACTION

A. Preprocessing:

Preprocessing is important as it provides summarization systems with a clean and adequate representation of source document. The pre-processing helps in interacting the most important information of a document. The text file is taken as the input document which is given for pre-processing. Pre-processing consists of four main steps: Segmentation, stop word removal, tokenization, stemming.

Sentence segmentation is the process of dividing the input file into number of sentences. The stop words such as I, a, the. .etc. are removed from the segmented lines. After stop word removal, each word is divided into tokens. base words are obtained by removing the prefixes and suffixes.

B. Feature Extraction

After Pre-processing, it is subjected to feature extraction by which the properties of the sentences are extracted to score the sentence. Eight features are considered. Values for each Feature are between 0 and 1.The eight features are:

Title Feature:

The sentences that contain title words are important as they are more relevant to theme. These sentences have a more chance of getting constituted in the summary. The title feature (T_F) can be calculated as below:

$$T_F = \frac{no\ of\ title\ words\ in\ sentence}{no\ of\ title\ words\ in\ title} \quad (1)$$

Sentence Length:

Sentence Length (S_L) is important in creating the summary. Short sentences such as names, date lines etc., are not added to the summary. This feature is used to isolate the short sentences.

$$S_L = \frac{no\ of\ words\ in\ sentence}{no\ of\ words\ in\ longest\ sentence} \quad (2)$$

Term Weight:

The term occurrences within a document have often been used for calculating the weight of each sentence. The sentence score can be calculated as the sum of the score of words in the sentence. Each word weight is given term frequency. The term weight is given by:

$$\mathbf{Wt = tfi * isfi = tfi * \log \frac{n}{N}} \quad (3)$$

Tfi :Term Frequency of word i

N: Number of sentences in the document

ni :Number of sentences in which the Word i occur

The Total Term Weight (T_W) is given by the formula

$$T_W = \frac{\sum_{i=1}^{k} Wi(s)}{Max(\sum_{i=1}^{k} wi(si))} \quad (4)$$

K: Number of Words in Sentences

Sentence Position:

The sentence position (S_P) also plays an important role in determining whether the sentence is appropriate or not. If there are 5 lines in document the sentence positions are given by

S_P= 5/5 for 1st, 4/5 for 2nd, 3/5 for 3rd, 2/5 for 4th, 1/5 for 5th (5)

Sentence to Sentence Similarity:

Similarity between the sentences is very important in generating the summary. The Similar sentences should not repeat in the summary that is to be generated.

$$SS_{sim} = \frac{\sum sim(si,sj)}{Max(\sum sim(si,sj))} \quad (6)$$

Si: sentence i

Sj: sentence j

Sim (si, sj): is the similarity of 1 to n terms in sentence si and sj

Proper Noun:

The sentences which have more proper nouns are mostly to be included in the summary. The Proper noun(N_P) feature is calculated as below:

$$N_P = \frac{no\ of\ proper\ nouns\ in\ sentence}{length\ of\ sentence} \quad (7)$$

Thematic word:

The terms that occur more frequently are more related to the topic. We consider top 10 most frequent words as thematic words. Thematic words(W_T) are calculated as below:

$$W_T = \frac{no\ of\ thematic\ word\ in\ sentence}{Max(\ no\ of\ thematic\ words)} \quad (8)$$

Numerical Data:

This Feature is used to identify the statistical data in every sentence. Numerical data(D_N) is calculated as follows:

$$D_N = \frac{no\ of\ numerical\ data\ in\ senence}{length\ of\ sentence} \quad (9)$$

III. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization performs arithmetic operations which enhances a problem by iteratively trying to improve possible solution with regard to given input data. A conventional approach called synchronous method is a more precise natural model which increases the possibility of parallelization of an algorithm [7], [8].

https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

In PSO, the search for the optimum solution is directed by a swarm of P particles. At time t, the ith particle has a position, $p(t)$, and a velocity, $V(t)$. A solution is represented by the particle position and velocity. Velocity represents the rate of change from the current particle position to the next particle position. The position and velocity values are initialized by random numbers at the beginning. In consecutive iterations, the search process is directed by updating the position and velocity using the following equations:

$$V(t) = V_i(t-1) + c_1 r_1(p\text{Best}_i - x_i(t-1)) + c_2 r_2(g\text{Best} - x_i(t-1)) \quad (1)$$

$$x_i(t) = V(t) + x_i(t-1)\,. \quad (2)$$

To prevent the particles from attempting too far from the feasible region, the $V(t)$ value is clasped to $\pm V\text{max}$. If the value of $V\text{max}$ is too large, then the exploration range is too wide. Conversely, if the value of $V\text{max}$ is too small, then the particles will favor the local search [10]. In (1), c_1 and c_2 are the learning factors that control the effect of the logical and social impact on a particle. Typically, both c_1 and c_2 are set to 2. Two independent random numbers r_1 and r_2 ranges from 0.0 to 1.0 are consolidated into the velocity equation. These random terms provide hypothetical behavior to the particles, thus strengthen them to explore a wider area.

A distinctive progress in PSO influenced not only by the particle's endeavor and experience but also by sharing the information to its neighbors. The particle's involvement is represented in equation (1) by $p\text{Best}i$, the best position which is found until, by the ith particle. The neighbors' influence is represented by $g\text{Best}$, the best position found by the swarm till the current iteration. The particle's position, $x_i(t)$, is updated using equation (2), in which a particle's next search is started from its previous position and the new search is involved by the past search[4]. Typically, $x_i(t)$ is limited to prevent the particles from searching in an infeasible region [5]. The quality of $x(t)$ is appraised by a problem-dependent fitness function. Each particles is evaluated to determine its current fitness. If a new position fitness is better than the current fitness then $g\text{Best}$ or $p\text{Best}i$ or both are found, then the new position value will accordingly be saved as $g\text{Best}$ or $p\text{Best}i$; otherwise the old best values will remain same. This process continues till the stopping benchmark is met, when the maximum iteration limit, T, is attained or the target solution is accomplished. Therefore, the maximum number of fitness evaluation for a swarm with number of particles P in a run is $(P\times T)$.

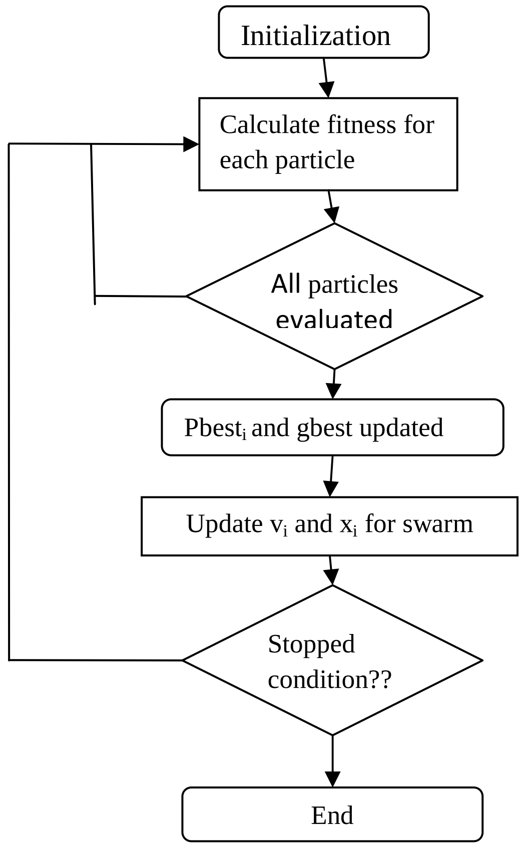

Fig.1: Synchronous PSO

The flowchart of figure 1 represents the original PSO algorithm. As shown in the algorithm, the updated values of the Best$i$ and gBest are evaluated after the fitness of all the particles has been evaluated. Therefore, this approach of PSO is known as Synchronous PSO (S-PSO). The pBest$i$ and gBest are modernized after all the particles fitness is evaluated, S-PSO assure that all the particles receive accurate and complete information about their neighbors , leads to a better choice of gBest and thus allowing the particles to exploit this information so that a better solution can be found. The summary is generated based the gBest values that are arranged in the descending order and the sentences are extracted from the source document and concatenated. However, this possibly leads the particles in S-PSO to converge faster, resulting in a untimely convergence.

IV. ASYNCHRONOUS PSO (A-PSO)

In S-PSO, a particle has to wait for the complete swarm to be evaluated before it can progress to a new position and continue its search. Thus, the particle is idle for the longest time after evaluating and waiting for the entire swarm to be modernized. A-PSO is an alternative approach to S-PSO, in which the particles are modernized based on the present state of the swarm. In A-PSO, A particle position, velocity, $p$best and $g$best are modernized as soon as its fitness is evaluated. The

https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

particle chooses gBest using a combination of information from the present and the prior iteration. In A-PSO, particle in the same iteration uses various values of gBest as it is preferred, based on the accessible information during a particle's updating process.

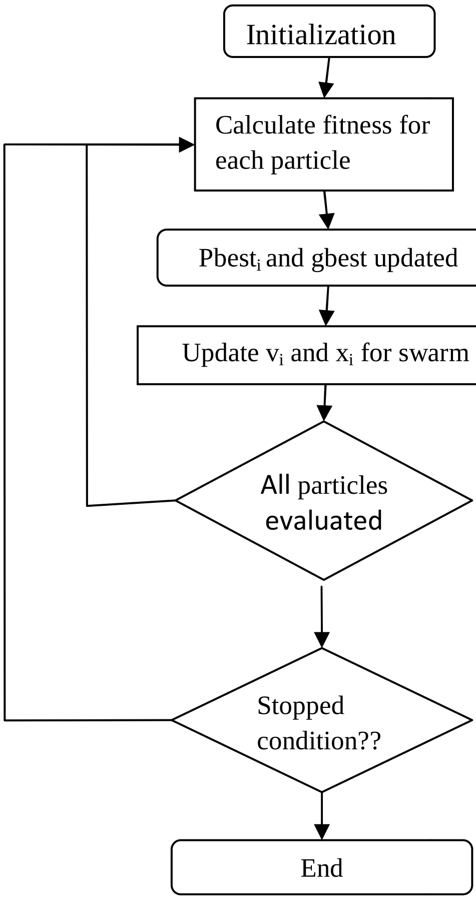

Fig.2: Asynchronous PSO

The flowchart in Figure 2 represents A-PSO algorithm. The flow of A-PSO is unlike S-PSO, however the fitness function is evaluated for P times per iteration, once for each particle. Therefore, the maximum number of iterations for fitness evaluation is ($P \times T$).This is alike to S-PSO. Using the same equations as S-PSO, The velocity and position are evaluated.

Other than the type of information, the lack of coexistence in A-PSO resolves the issue of ineffective particles faced in S-PSO. An asynchronous update also allows the modernize sequence of the particles to alter dynamically or a particle to be modernized more than once.

V. ANALYSIS OF RESULTS

The analysis of the result is done considering the domains relating to Economy, Secularism, Earth, Nature, Forest, and Metadata. For every document, a manually generated relevant summary is compared to obtain the precision and recall values. Summary is generated using synchronous PSO and asynchronous PSO. Precision, recall and F-measure values are calculated for each document as shown in the below table1. In the graphs, precision, recall and F-measure values are represented to determine the performance of the systems.. The graph are drawn for each dataset as shown in the below. Recall is also known as sensitivity. Recall is gradually increasing as shown in the figure. The increase in recall suggests that the system performs better compared to other systems. Compared to synchronous PSO and Asynchronous PSO the recall value of Asynchronous PSO is higher than the Synchronous PSO. This leads to more exploration of data.

Table.1: values of Synchronous PSO(S-PSO) and Asynchronous PSO (A-PSO)

Data sets	Synchronous PSO			Asynchronous PSO		
	precision	Recall	F-measure	precision	Recall	F-measure
Nature	17.5	38.8	27	22.5	66.6	33
metadata	20	50	27.2	25.7	66.6	37
Forest	41	58	51	53	88	66
reservation	36	38	43	47	84	60
computer	26	46	34	34	61	44
economy	63.3	81	71	72	83	77

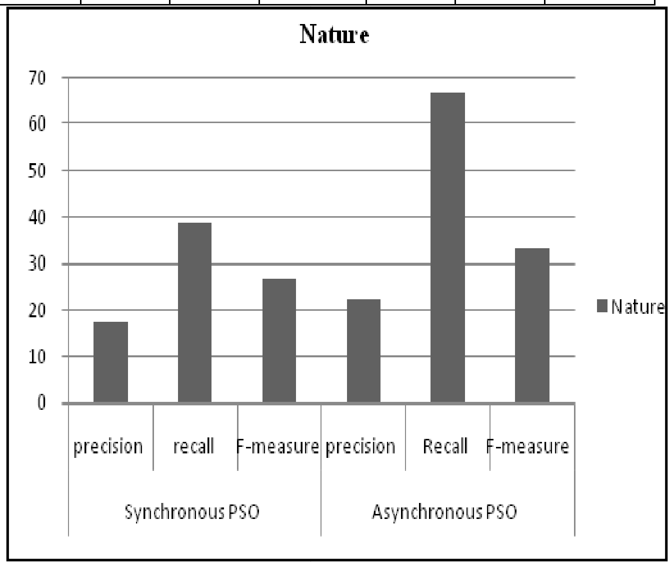

Fig.3: Comparison of S-PSO and A-PSO for Nature dataset

https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

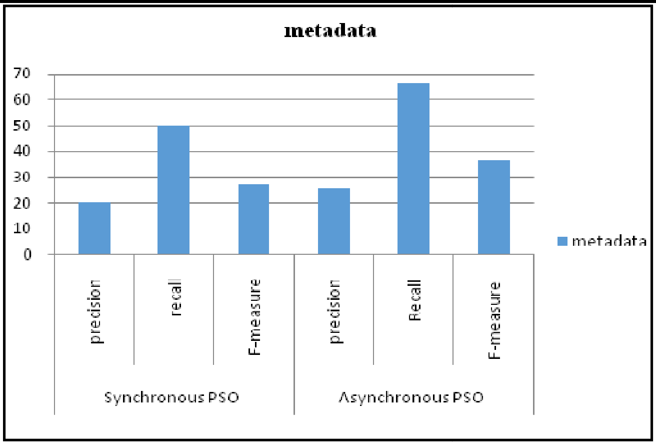

Fig.4: Comparison of S-PSO and A-PSO for Metadata dataset

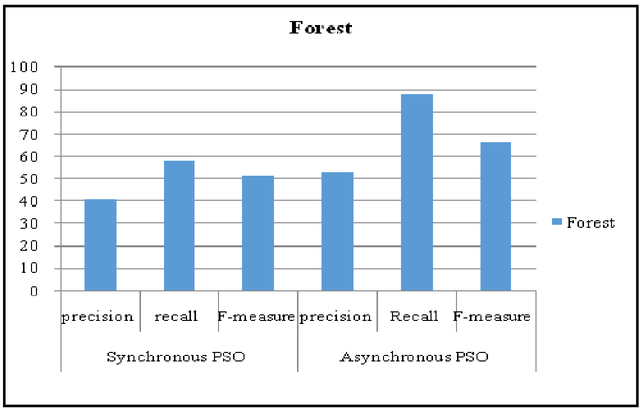

Fig.5: Comparison of S-PSO and A-PSO for Forest dataset

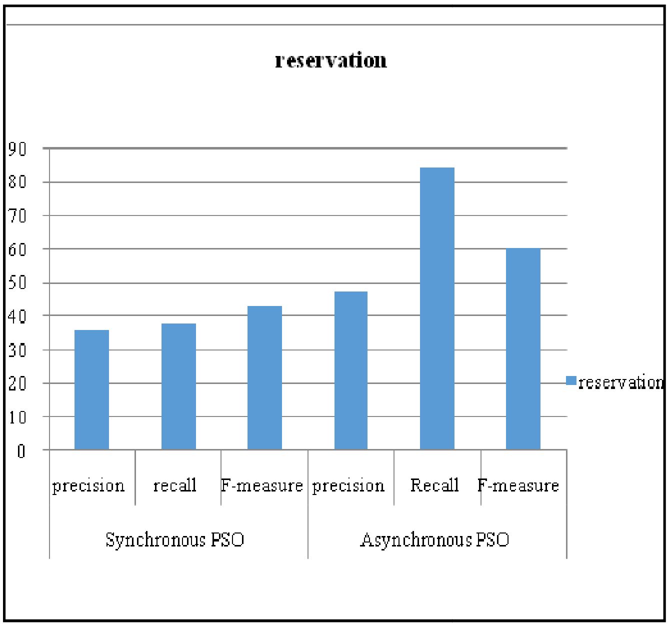

Fig.6: Comparison of S-PSO and A-PSO for Reservation dataset

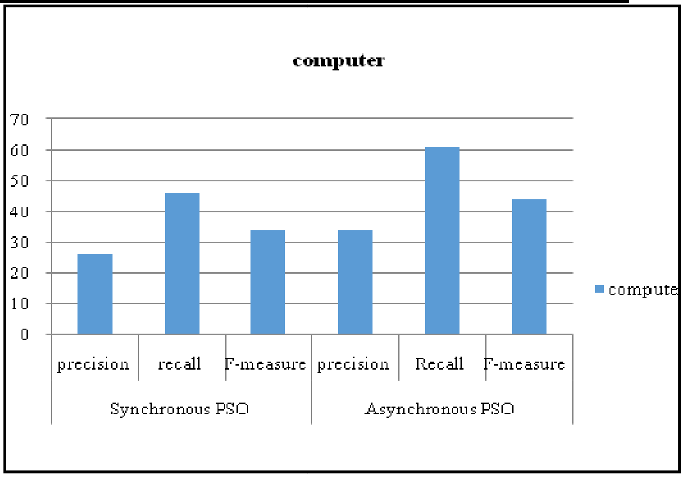

Fig.7: Comparison of S-PSO and A-PSO for Computer dataset

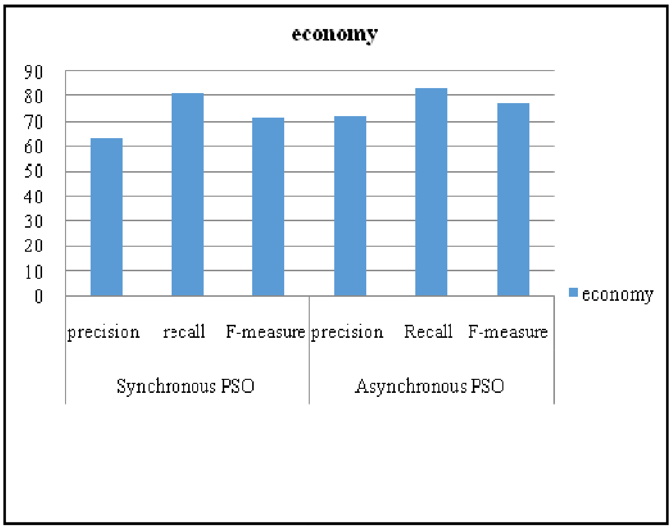

Fig.8: Comparison of S-PSO and A-PSO for Economy dataset

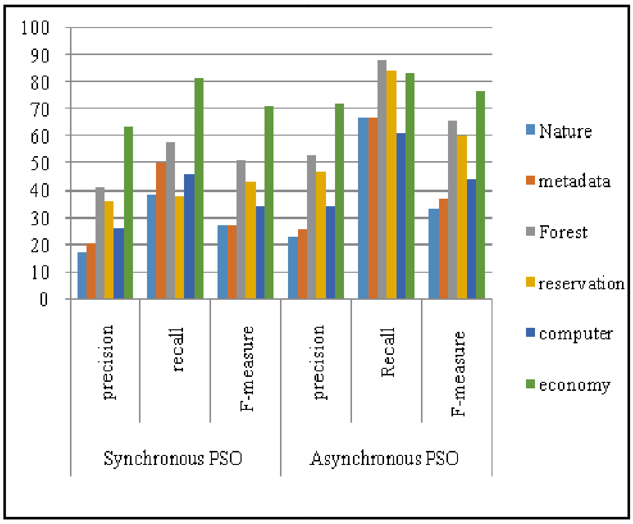

Fig.9: values of synchronous PSO(S-PSO) and asynchronous PSO(A-PSO)

https://dx.doi.org/10.22161/ijaers/3.11.22 ISSN: 2349-6495(P) | 2456-1908(O)

VI. CONCLUSION

Automatic summarization is a aggregate task that affects the performance to produce high quality summaries. A comparative study of synchronous PSO and asynchronous PSO summarization techniques are evaluated using different text documents related to different domains as inputs. In synchronous PSO, after calculating the entire performance the particles velocities and positions are modernized. This modernizing method improves the performances. In A-PSO after calculating the own performance, velocities and positions of the particles are modernized. Therefore, particles are modernized using partial data, leads to extreme exploration. The analysis of results show that the Asynchronous approach produces efficient results compared to Synchronous approach. The work can be further enhanced by using a hybrid approach which combines S-PSO and A-PSO.

REFERENCES

[1] J. Kennedy and R. Eberhart, “Particle swarm optimization,”in *Proceedings of the IEEE International Conference on Neural Networks (ICNN ’95)*, vol. 4, pp. 1942–1948, Perth, Western Australia, November-December 1995.

[2] G. J. Rath, A. Resnick, and T. R. Savage, “The formation of abstracts by the selection of sentences” American Documentation, vol. 12, pp.139-143.1961.

[3] R. Juan, M. Zhang, and W. Seah, “A performance study on synchronous and asynchronous updates in Particle Swarm Optimization,” in *Proceedings of the 13th Annual Genetic and Evolutionary Computation Conference (GECCO ’11)*, pp. 21–28,July 2011.

[4] J. Rada-Vilela, M. Zhang, and W. Seah, Random asynchronous PSO,” in *Proceedings of the 5th International Conference on Automation, Robotics and Applications (ICARA '11)*, pp. 220–225,Wellington, New Zealand, December 2011.

[5] B. Jiang, N. Wang, and X. He, “Asynchronous particle swarm optimizer with relearning strategy,” in Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society (IECON ’11), pp. 2341–2346, November 2011

[6] Voglis C.A., Parsopoulos K.E., and Lagaris I.E. 2012. Particle swarm optimization with deliberate loss of information. Soft Computing. Vol. 16, No. 8, pp. 1373–1392.

[7] M. N. Ayob, Z. M. Yusuf, A. Adam et al., “A particle swarm optimization approach for routing in VLSI,” in *Proceedings of the 2nd International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN '10)*, pp. 49–53,Liverpool, UK, July 2010.

[8] Y. Shi and R. Eberhart, “A modified particle swarm optimizer,” in *Proceedings of the IEEE International Conference on Evolutionary Computation and IEEE World Congress on Computational Intelligence*, (Cat.No. 98TH8360) , pp. 69–73, Anchorage, Alaska, USA, May 1998.

[9] M.A. Fattah and Fuji Ren, “Automatic Text Summarization” In proceedings of World Academy of Science, Engineering and Technology Volume 27. pp 192-195. February 2008.

[10] Shi Y. and Eberhart R. 1998. Parameter selection in particle swarm optimization. In Proceedings of the 7th International Conference on Evolutionary Programming VII (EP '98), V. William Porto, N. Saravanan, Donald E. Waagen, and A. E. Eiben (Eds.). pp. 591-600

[11] C. Tu, L. Chuang, J. Chang, and C. Yang, “Feature Selection using PSO-SVM”. IAENG International Journal of Computer Science. 33(1), 2006, pp.138-143

[12] Y. Liu, Z. Qin, Z. Xu and X. He, “Feature Selection with Particle Swarms”. In *Computational and Information Science, LNCS 3314*, J. Zhang, J.-H. He, and Y. Fu, Ed. Heidelberg: Springer-Verlag, 2004, pp. 425–430.