



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

Human Interaction Pattern Mining Using Enhanced Artificial Bee Colony Algorithm

S. Uma¹, Dr. J. Suguna²

Assistant Professor, School of Computer Science, C.M.S. College of Science & Commerce, Coimbatore, India¹

Associate Professor, Dept of Computer Science, Vellalar College for Women, Erode, India²

ABSTRACT: Interactions between human beings occur often in meeting discussions. Semantic knowledge of such meetings can be revealed by discovering interaction patterns from them. Artificial Bee Colony (ABC) algorithm is a relatively new swarm intelligence algorithm that has attracted great part of attention from researchers with the advantage of less control parameters and strong global optimization ability than other optimization algorithms. In this research work ABC algorithm is enhanced using Partial Least Square (PLS) mechanism in order to extract frequent interaction among patterns. The objectives of the research are three-fold such as detecting more number of interactions made, more number of frequent interactions made and reducing the execution time of the algorithm. The experimental results shows that by using Enhanced Artificial Bee Colony algorithm more interesting patterns that are useful for the interpretation of human behavior in meeting discussion can be extracted.

KEYWORDS: Tree Based Mining, Enhanced Artificial Bee Colony, Human Interaction Pattern Mining, Partial Least Square, Partial Ancestral Graph meet.

I. INTRODUCTION

Optimization techniques play a major role in many scientific and engineering fields. Among them, Evolutionary Algorithms (EAs) have attracted researchers during the past decades in solving optimization problems as these methods do not force assumption on continuity and differentiability like traditional methods. Also, they do not easily fall into local optima.

Artificial Bee Colony (ABC) algorithm is most prominent among the population-based algorithms. It is an effective algorithm proposed for global optimization. Numerical performance demonstrated that ABC algorithm is competitive to that of other population-based algorithms with the advantage of employing fewer control parameters and the need for fewer function evaluations to arrive at an optimal solution [1]. Due to its simplicity and ease of implementation, ABC has captured much attention and has been employed to solve many numerical as well as practical optimization problems since its inception [2, 3, 4]. An upgraded artificial bee colony (ABC) algorithm for constrained optimization problems is discussed in [5].

In this paper, Human interaction flow in a discussion session is represented as a tree. Human cooperation is identified by whether the gathering was decently composed or not. It has been considered as one of the primary issues in the gatherings. The interaction flow is coined as semantic knowledge that projects the pattern of the knowledge. The formation of well-defined dictionary for the relevant meetings is one among the challenging research dimensions that is based on the people's interaction in the meetings, forums, discussions and so on. Some performance metrics in order to resolve the interesting pattern in such meetings include session percentage, execution time and number of frequent subtrees discovered. This research work aims in detecting more number of interesting patterns with minimum execution time. So, in this work ABC algorithm is enhanced using Partial Least Square (PLS) mechanism in order to extract frequent interaction among patterns. In order to have powerful algorithm, we have used PLS mechanism in population initialization, so that solutions are generated uniformly within the search space. This helps to generate at least some points in the neighbourhood of global solution.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

The paper is organized as follows. In section II, the related works are given. In section III, methodologies used for generation of interaction flow are introduced. In section IV, the results of the comparison with EPCA-ABC-PAG Meet and EPCA-EABC-PAG Meet are presented and discussed. Finally, a conclusion is provided.

II. RELATED WORK

ABC algorithm for the first time was adapted by Karaboga and Bastruck [6] to solve unconstrained optimization problems. In recent years there have been many efforts to develop a constrained ABC algorithm possessing balanced exploration and exploitation behaviour. Constrained ABC algorithm is also applied to solve many real-world engineering problems in recent years. A chaotic search mechanism was applied in [7] to initialize population for constrained ABC. A modified constrained ABC algorithm was proposed in which chaotic mechanism as well as opposition based method was applied for population initialization to enhance the global convergence of algorithm [8]. Modified constrained ABC by applying multiple onlooker bees was developed in [9] to improve constrained ABC. A genetically inspired ABC algorithm was presented for Constrained Optimization Problem (COP). In this algorithm uniform crossover and mutation operators from Genetic Algorithm (GA) are applied to scout bee phase to improve the performance of ABC algorithm [10]. ABC was used to solve large scale optimization problems as well as engineering design problems [11] and it outperforms other optimization algorithms.

Another modification on ABC algorithm was in [12]. What makes this algorithm different from the original ABC is the probability selection mechanism and parameter setting process. In this algorithm a new probability selection mechanism is presented to enhance diversity by allowing infeasible solutions in the population. The infeasible solutions were introduced inversely proportional to their constraint violations and feasible solution were defined based on their fitness values.

A modified ABC [13] introduced four modifications related with the selection mechanism, the equality and boundary constraints, and scout bee operators to improve the behaviour of ABC in constrained search space. A smart bee [14] was introduced to solve constrained problems which apply its historical memories for the solution. ABC-BA [15] is a hybrid algorithm that integrates ABC and Bee Algorithm (BA). It can perform as an ABC individual in ABC sub-swarm or a BA individual in the BA sub-swarm. In addition, the population size of the ABC and BA sub-swarms change stochastically based on current best fitness values achieved by the sub-swarms.

III. METHODOLOGY

3.1 Dimension Reduction

A. Enhanced Principal Component Analysis (EPCA)

In human interactions, patterns can be derived from multiple interactions. Such interaction dataset will have confusing and irrelevant words in sentences. EPCA will reduce such complex data to a lower dimension and eliminate the variations present in the data [16]. When EPCA completes the reduction of deviated discussion contents, ABC algorithm is applied.

B. Artificial Bee Colony (ABC) algorithm

ABC is a relatively new population-based algorithm [17] emulating the foraging behaviour and waggle dance of honey bee swarm. Artificial bee colonies are classified into three groups, employed bees, onlooker bees and scout bees. Half of the colony includes employed bee and the other half consist of onlooker bees. In ABC, the position of food source denotes a possible solution to the optimization problem and the nectar amount of food source represents fitness value of the associated solution. The number of employed bees or the onlooker bees is equal to the number of solutions (SN) in the population. Each solution x_i ($i = 1, 2, \dots, SN$) is a d -dimensional vector and $x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$ represents the i th solution in the population.

At initialization step, ABC generates a randomly distributed initial population of SN solutions using Equation (1).

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

$$x_{ij} = x_{\min,j} + rand(0,1)(x_{\max,j} - x_{\min,j}) \text{ -----} > (1)$$

where each solution x_i , $i = 1, 2, \dots, SN$ is d -dimensional vector for $j = 1, 2, \dots, d$. In addition, $x_{\min,j}$ and $x_{\max,j}$ are the lower and upper bounds for the dimension j respectively. These food sources are randomly assigned to SN number of employed bees and their fitness are evaluated. At the initialization phase the human interaction dataset is taken as the input.

After initialization, the population of the solutions is subjected to repeat the search processes for employed bee, the onlooker bees and the scout bee phases. The process continues until the algorithm reaches the Maximum Cycle Number ($M CN$). In employed bee phase each employed bees produces a modification on the solution x_i where only one dimension of this solution is changed using Equation (2) and the rest keep the same as x_i

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \text{ -----} > (2)$$

where $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, d\}$ are randomly chosen indexes and k has to be different from i . ϕ_{ij} is a random number in the range $[-1, 1]$. After v_i is obtained, its fitness value is evaluated and a greedy selection mechanism is applied comparing x_i with v_i . If the fitness value of the new solution v_i is less than the current solution then, the solution is replaced with the x_i , otherwise the current solution remains. Indexes are assigned for the data present in the human interaction dataset and the fitness value is given by using the frequently used discussions.

After the employed bee phase, the solution information is transferred to the onlooker bee phase. In this phase a solution is chosen depending on the probability value p_i associated with that solution calculated using the following formula,

$$p(x_i) = \frac{fit(x_i)}{\sum_{j=1}^{SN} fit(x_j)} \text{ -----} > (3)$$

The $fit(x_i)$ is defined in Equation (4).

$$fit(x_i) = \begin{cases} \left(\frac{1}{1 + f(x_i)}\right) & f(x_i) \geq 0 \\ 1 + |f(x_i)| & f(x_i) < 0 \end{cases} \text{ -----} > (4)$$

where $f(x_i)$ is the objective value of solution x_i . Once the onlooker has selected solution x_i , a modification is done on this solution similar with employed bee using Equation (2). Then fitness values of generated solutions are evaluated and greedy selection mechanism is employed. If new solution has better fitness value than current solution, the new solution remains in the population and the old solution is removed. The new solution consists of the repeatedly used contents of the meeting minutes.

In the scout bee phase, if solution x_i cannot be improved further through a predetermined number of cycles ($limit$), then that solution is abandoned and replaced with a new solution generated randomly by using Equation (1). The cycle is repeated until all the unwanted contents are removed that achieves maximum convergence of the minutes of the meeting. The ABC main procedure is summarized as below.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

Artificial Bee Colony Algorithm for Human Interaction Pattern Mining
Initialize the population of solution i.e. human interaction dataset
Evaluate the initial population
cycle=1
Repeat
Employee bee phase – repeatedly use contents of the minutes of the meeting
Apply greedy selection process
Calculate the probability values for the solution
Onlooker bee phase
Apply greedy selection process
Scout bee phase
Memorize the best solution achieved so far
cycle=cycle+1
until cycle=maximum cycle number

When ABC completes removal of unwanted contents, the PAG meet takes the pre-processed meeting contents and constructs the PAG graph.

C.Enhanced Artificial Bee Colony (EABC) algorithm

In most of the unconstrained ABC algorithms the role of population initialization is ignored. In order to have powerful algorithm preliminary solutions must be generated uniformly within the search space. The uniformly distributed initial solutions help to produce at least some points in the neighbourhood of global solution. To obtain that, this section introduces Partial Least Square (PLS) dimension reduction mechanism. PLS is based on correlation. PLS is very popular in areas like chemical engineering, where predictive variables often consist of many different measurements in an experiment and the relationships between these variables are ill-understood [18].

The basic idea of partial least squares is that, the characteristic variable matrix X is compressed, and given consideration to the correlation of the dependent variable matrix Y . Suppose there are n characteristic variables, x_1, x_2, \dots, x_n and p dependent variables y_1, y_2, \dots, y_p , then X is decomposed into

$$X = TP^T + E \text{ -----> (5)}$$

where, T is the score matrix, P is the load matrix, E is the residual error matrix. Matrix multiplication of TP^T can be expressed as the sum products of score vector t_i (thei-th column of matrix T) and load vector p_i (the i-th column of matrix P). Then the above formula can be written as

$$X = \sum_{i=1}^n t_i p_i^T + E \text{ } i=1,2,\dots,n \text{ -----> (6)}$$

Similarly, matrix Y is decomposed with Q as the load matrix and u_j as score vector.

PLS analysis separately extracts the score t and u from corresponding X and Y . They are the linear combination of characteristic variables and dependent variables. And both score satisfies the maximum load of variation information of characteristic variables and dependent variables. The regression equation is established as

$$u_j = b_k t_i \text{ -----> (7)}$$

where, b_k is regression coefficient.

The formula can be expressed in matrix form as

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

$$Y = BX \text{ -----} > (8)$$

where $B = W(P^T W)^{-1} Q^T \text{ -----} > (9)$

here B is the coefficient matrix and W is the weight matrix.

PLS follows an iterative process to extract the factors or scores for X and Y until the residual matrix element of absolute value approximates to zero. After initialization, the main loop consisting of employed bees, onlooker bees and scout bees phase are subjected to be repeated until the stopping criterion is met.

In main loop after producing a new solution in employee bee phase, Deb's constrained handling method is adopted in order to adapt ABC algorithm for solving constrained optimization problems instead of using greedy selection in unconstrained ABC [19]. Applying Deb's rules, the bee either memorizes the new solution by forgetting the current solution or keeps the current solution. Deb's method uses a tournament selection mechanism where two solutions are compared at a time by applying following rules.

- Any feasible solution is preferred to any infeasible solution,
- Among two feasible solutions, the one having better objective function value is preferred,
- Among two infeasible solutions, the one having smaller constraint violation is preferred.

Algorithm 2: Enhanced ABC algorithm

Initialize the population using PLS
Evaluate the initial population
Cycle=1
Repeat
 Employed bee phase
 Apply Deb's mechanism to select between V_i and X_i
 Calculate the probability values for the solution
 Onlooker bee phase
 Apply Deb's mechanism
 Scout bee phase
 Memorize the best solution achieved so far
Cycle = cycle + 1
Until cycle = maximum cycle number

After completion of the search by all employed bees, the information of the solutions are shared with the onlooker bees. Then the probability value for onlooker bee phase is calculated. If the new solution is better, it will be kept in the population and the current solution is removed, otherwise the current solution is retained.

In the scout bee phase, if solution x_i cannot be improved further through a predetermined number of cycles (*limit*), then that solution is abandoned and replaced with a new solution generated randomly by using Equation (1). The cycle is repeated until all the unwanted contents are removed that achieves maximum convergence of the minutes of the meeting.

When EABC completes removal of unwanted contents, the PAG meet takes the pre-processed meeting contents and constructs the PAG graph.

3.2 Partial Ancestral Graph (PAG) meet

Human interaction flow in a discussion session is represented as a tree. The PAG meet will find frequent interactions among patterns [20]. It will generate a set of all frequent nodes and then expand these nodes with new root, new level, new node and new edge. So that no duplicate Partial Ancestral Graph (PAG) or sub PAG is generated. Sub-graphs containing siblings may not be connected without the presence of their common ancestor in a tree. So, if a common ancestor is not frequent, tree based mining method [21] will fail to mine them as frequent pattern. Also tree

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

based mining will not distinguish multiple interactions but PAG meet will identify such multiple interactions and will extract both temporal and triggering relations.

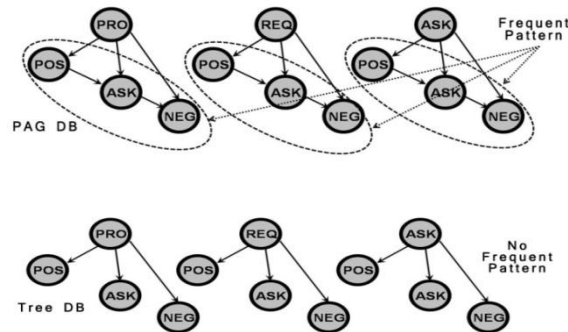


Figure 1. PAG representation vs Tree representation of meetings

The previous three methodologies EPCA, ABC and EABC remove irrelevant data from meeting contents and give the preprocessed content as input for the PAG meet.

IV. EXPERIMENTAL RESULTS

Dataset is gathered from meetings that are conveyed through on-line forums and documented discussions from various data sources. Table 1 shows the categorization of datasets.

Table 1. Dataset Information

Document Type	Number of documents	Maximum number of conversations per document
Mini dataset	2000	8
Large dataset	2000	250
Very large dataset	2000	5000

The datasets are focused on the topics namely corruption, democracy, defense, expenditure, social expenditure, backwardness, women quota bill, politics, communal and religious reservations, retirement age, elections etc. The datasets have certain amount of irrelevant data and hence EPCA mechanism followed by ABC mechanism is used to remove such irrelevant data and then PAG meet is applied to generate interaction flow in meetings. This work is named as EPCA-ABC-PAG meet. Another observation is done by applying EPCA mechanism followed by EABC mechanism and then applying PAG meet. This work is named as EPCA-EABC-PAG meet.

The Table 2 shows the interaction percentage made during the sessions and Figure 2 clearly depicts that EPCA-EABC-PAG meet detects more number of interactions in the conversation made during the sessions than EPCA-ABC-PAG Meet.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

Interactions	EPCA-ABC-PAG Meet	EPCA-EABC-PAG Meet
1	19	22
2	24	27
3	32	34
4	26	30
5	20	23
6	13	16
7	12	15
8	14	17
9	5	7
10	11	14
11	5	7
12	10	12
13	4	6
14	8	10

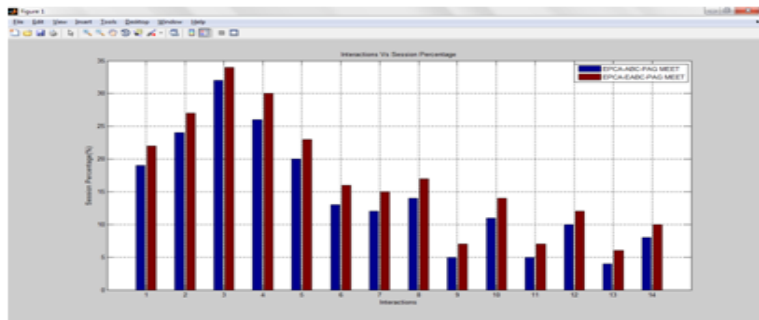


Figure 2. Interactions Vs Session Percentage

Table 2. Interactions Vs Session Percentage

The Table 3 illustrates execution time of EPCA-ABC-PAG Meet and EPCA-EABC-PAG Meet . It can be clearly understood that the EPCA-EABC-PAG meet consumes less execution time than EPCA-ABC-PAG Meet and it is noteworthy fact that when support threshold value increases the execution time is decreased and the same is flashed in Figure 3.

Support Threshold	EPCA-ABC-PAG Meet	EPCA-EABC-PAG Meet
1	130	123
2	52	41
3	34	25
4	25	17
5	22	14
6	19	11
7	18	10
8	16	8
9	15	6
10	13	5

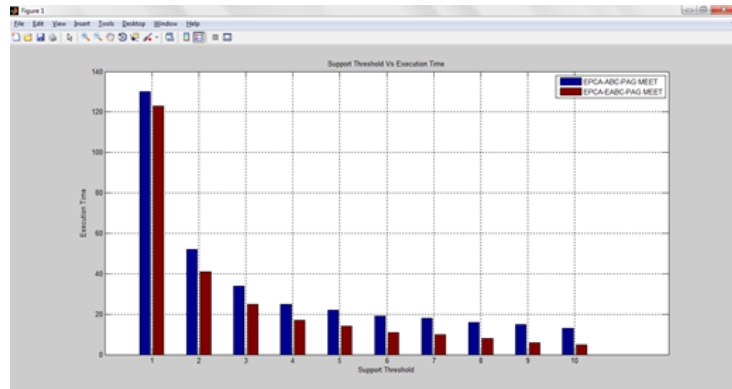


Figure 3. Support Threshold Vs Execution Time

Table 3. Support Threshold Vs Execution Time

Detecting more number of interactions gains advantage towards discovering frequent pattern mining. From Table 4 it is shown that the number of discovered frequent subtrees is increased in EPCA-EABC-PAG meet than EPCA-ABC-PAG Meet and it is exposed in Figure 4.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

Support Threshold	EPCA-ABC-PAG Meet	EPCA-EABC-PAG Meet
1	12667	14172
2	5100	5525
3	5185	6085
4	4601	5551
5	4409	5361
6	4143	5145
7	3640	4491
8	3718	4662
9	3522	4448
10	2941	3793

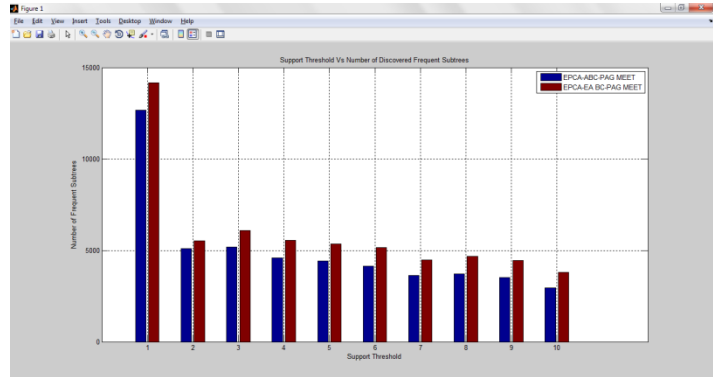


Figure 4. Support Threshold Vs Number of Discovered Frequent Subtrees

Table 4. Support Threshold Vs Number of Discovered Frequent Subtrees

V. CONCLUSION

This paper considers the problem of detecting more number of interaction patterns made in the meeting to gain semantic knowledge of meetings. The idea is to employ EPCA algorithm and EABC algorithm to reduce irrelevant data in meeting content and then construct the PAG graph. The performance is analysed by applying EPCA-ABC-PAG meet and proposed EPCA-EABC-PAG meet. The results indicate that the proposed method outperforms earlier mechanism in terms of session percentage, execution time and number of discovered frequent sub-trees. Future research work may apply any soft computing techniques like machine learning or fuzzy logic for identifying more frequent subtrees.

REFERENCES

- Karaboga D., Basturk B., "On the performance of artificial bee colony (ABC) algorithm", Applied soft computing, vol.8, no. 1, pp. 687-697, 2008.
- Gao W. F., Liu S. Y., Huang L. L., "Enhancing artificial bee colony algorithm using more information-based search equations", Information Sciences, 270, pp.112-133, 2014.
- Aydina D., Zyn S., Yasar C., Liao T., "Artificial bee colony algorithm with dynamic population size to combined economic and emission dispatch problem", Electrical Power and Energy Systems, vol.45, pp.144-153, 2014.
- Li G., Peifeng N., Xiao X., "Development and investigation of efficient artificial bee colony algorithm for numerical function optimization", Applied soft computing, vol.12, no. 1, pp. 320-332, 2012.
- Brajevic I., Tuba M., "An upgraded artificial bee colony (ABC) algorithm for constrained optimization problems", Journal of Intelligent Manufacturing, vol. 24, no. 4, pp.729-740, 2013.
- Karaboga D., Basturk B., "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm", Journal of global optimization vol.39, no. 3, pp.459-471, 2007.
- Babaeizadeh, S., and A. Rohanin, "Modified Artificial Bee Colony Algorithm with Chaotic Search Method for Constrained Optimization Problems", Proceeding of 2nd International Science Postgraduate Conference 2014 (ISPC2014), Faculty of Science, Universiti Teknologi Malaysia, (2014a).
- Babaeizadeh, S., and A. Rohanin, "A Modified Artificial Bee Colony Algorithm for Constrained Optimization Problems", Journal of Convergence Information Technology, Vol.9, No.6, 2014b, pp. 151- 163.
- Subotic M., "Artificial bee colony algorithm with multiple onlookers for constrained optimization problems", In Proceeding of the European Computing Conference, pp. 251-256. 2011.
- Bacanin N., Tuba M., "Artificial bee colony (ABC) algorithm for constrained optimization improved with genetic operators", Studies in Informatics and Control, vol. 21, no. 2, pp. 137-146, 2012.
- Akay B., Karaboga D., "Artificial bee colony algorithm for large-scale problems and engineering design optimization", Journal of Intelligent Manufacturing, vol. 23, no. 4, pp. 1001-1014, 2012.
- Karaboga D., Akay B., "A modified artificial bee colony (ABC) algorithm for constrained optimization problems", Applied Soft Computing, vol. 11, no. 3, pp.3021-3031, 2011.
- Mezura-Montes E., Cetina-Domínguez O., "Empirical analysis of a modified artificial bee colony for constrained numerical optimization", Applied Mathematics and Computation, vol. 218, no. 22 pp.10943-10973, 2012.
- Stanarevic N., Tuba M., Bacanin N. , "Modified artificial bee colony algorithm for constrained problems optimization", International journal of mathematical models and methods in applied sciences, vol.5, no. 3, pp. 644-651, 2011.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

15. Tsai H. C., "Integrating the artificial bee colony and bees algorithm to face constrained optimization problems", Information Sciences, vol. 258, pp. 80-93, 2014.
16. Uma. S, Suguna.J, "Human Interaction Pattern Mining Using Enhanced Principal Component Analysis", International Journal of Informative & Futuristic Research, ISSN: 2347-1697, vol.2, no.7,pp. 2279-2289, March 2015.
17. Salamin, H.; Vinciarelli, A., "Automatic Role Recognition in Multiparty Conversations: An Approach Based on Turn Organization, Prosody, and Conditional Random Fields," Multimedia, IEEE Transactions on , vol.14, no.2, pp.338,345, April 2012
18. Kleinbaum, David G., et al., "Applied Regression Analysis and Multivariable methods", 3rd edition, (Pacific Grove, Ca.:Brooks/Cole Publishing Company, 1998).
19. Dervis Karaboga , Bahriye Basturk., "Artificial Bee colony Optimization Algorithm for solving Constrained Optimization Problems", IFSA 2007, LNAI 4529, pp.789-798, 2007.
20. Uma. S, Suguna.J, "Temporal Semantic Analysis Based Human Interaction Pattern Mining using Partial Ancestral Graph", Research Journal of Applied Sciences, Engineering and Technology, vol. 8, no. 12, pp. 1487-1491, 2014
21. Christian Becker and Zhiwen Yu, "Tree-Based Mining for Discovering Patterns f Human Interactions in Meetings", 'Proc.Eigth IEEE Int'l Conf.Knowledge Discovery and Data Mining, pp 107-115, 2012.

BIOGRAPHY

¹S.Uma received master's degree in Information Technology from Bharathidasan University, Trichy in 2004 and M.Phil degree in Computer Science from Bharathidasan University, Trichy in 2007. She is currently an Assistant Professor in School of Computer Science, C.M.S.College of Science & Commerce (Autonomous), Coimbatore, Tamil Nadu. Her research interests are Data Mining and Image Processing.

²Dr. J. Suguna received the master's degree in mathematics from Annamalai University, Chidambaram in 1988 and the Ph.D. degree in computer science from the Bharathiar University, Coimbatore in 2009. She is currently an Associate Professor with the Department of Computer Science and Controller of Examinations,Vellalar College for Women (Autonomous), Erode, Tamil Nadu. Her research interests are AI, Data Mining, Text Mining and Image Processing. She is the author or co-author of over 20 publications in journals, conference proceedings and book chapters. She has presented a paper in an International Conference held at Cincinnati University, Cincinnati, Ohio, USA. She has produced over 15 M.Phil. Scholars in Computer Science.