Automatic Threshold Selections by exploration and exploitation of optimization algorithm in Record Deduplication

[1]K.Deepa, [2]Dr.C.Vivek, [3]S.Palanivel Rajan,

[1]Professor, Sri Ramakrishna Engineering College, Coimbatore, Tamilnadu, India

E-mail ID : deepa.senthil@srec.ac.in

[2]Associate Professor, Department of Electronics and Communication Engineering,

M.Kumarasamy College of Engineering, Karur, Tamilnadu, India

Email ID : vivekc.phd@gmail.com

[3]Assistant Professor, Department of Electronics and Communication Engineering,

M.Kumarasamy College of Engineering, Karur, Tamilnadu, India

Email ID : palanivelrajanme@gmail.com

ABSTRACT

A deduplication process uses similarity function to identify the two entries are duplicate or not by setting the threshold. This threshold setting is an important issue to achieve more accuracy and it relies more on human intervention. Swarm Intelligence algorithm such as PSO and ABC have been used for automatic detection of threshold to find the duplicate records. Though the algorithms performed well there is still an insufficiency regarding the solution search equation, which is used to generate new candidate solutions based on the information of previous solutions. The proposed work addressed two problems: first to find the optimal equation using Genetic Algorithm(GA) and next it adopts an modified Artificial Bee Colony (ABC) to get the optimal threshold to detect the duplicate records more accurately and also it reduces human intervention. CORA dataset is considered to analyze the proposed algorithm.

INDEX TERMS

Genetic Algorithm, Modified ABC, Similarity metrics, Cosine Similarity, Levenshtein Distance

1. INTRODUCTION

Knowledge Discovery in Databases (KDD) is the process of identifying valid, useful, and understandable patterns from large datasets [20]. Data mining is the core of the KDD process. It is a computational process of discovering patterns in large datasets [10]. Data cleaning deals with identifying and removing errors and inconsistencies from data to improve the quality of data. Data cleaning plays a important task in the process of data mining. The fundamental component of data cleaning is usually termed as duplicate record identification that is the process of identifying the record pairs indicating the same entity (duplicate records) [14]. Multiple versions of the same record are often accumulated when databases are constructed from multiple sources. The task of detecting these different versions is known as record deduplication [28]. All records that have exactly or approximately the same data in one or more fields are identified as duplicates. The problem of identifying syntactically different records that describe unique entities is denoted by all terms such as record linkage, duplicate detection and more [30]. It is essential to enrich the quality of data through data cleaning methods. Numerous data cleaning techniques are being employed for diverse purposes. Similarities among records and fields are identified using 'Similarity Functions. Duplicate elimination functions' are employed to identify if two or more records signify the same real world object [25].

The recent researches have proposed many methods for the deduplication purposes with many distinct features by their own. The proposed work addressed two problems: first one to find the optimal equation using GA and the next to get an optimal threshold using modified ABC. Similarity metrics is used to calculate similarity values among the fields and each such value is called as evidence. GA combines different parts of evidence to find a duplicate record detection function [11]. This enables to identify whether two entries in a repository are duplicates or not. Since duplicate detection process is a time consuming process, the aim is to recommend a method that finds a proper combination of the best pieces of evidence, thus yielding a function that maximizes performance for training purposes [27]. Each expression requires the optimal threshold value in order to classify the duplicate and non duplicate entries [8]. Normally, thresholds are set by the users based on the requirement of specific applications and the optimal thresholds are set up by either minimizing or maximizing an objective function with respect to the values of the thresholds [22]. In order to find the optimal threshold and to reduce the human intervention, the proposed work uses an intelligence algorithm, modified ABC. The rest of the paper is organized as; the section 2 gives a review of related works regarding deduplication. Section 3 gives basics concepts of GA ,ABC, Exploration and Exploitation 4 describes the proposed approach; Section 5 shows the results and discussion about the proposed approach and in the 6th section we conclude our research work.

2. REVIEW OF RELATED WORK

A structure to enhance the duplicate detection of records with the aid of trainable measures of textual similarity was proposed by Mikhail Bilenko et al [6]. They utilized a learnable text distance functions for each database field, and illustrated that such measures are skilled enough to adapt to the precise notion of similarity that was suitable for the field's domain. An extended modification of learnable string edit distance and a vector-space based measure that utilizes a Support Vector Machine (SVM) for training were the two learnable text similarity measures proposed that were appropriate for the task. Extensive experimental evaluation on a variety of data sets demonstrated that the framework is capable of enhancing detection accuracy over conventional methodologies. Mois´es G. de Carvalho et al. [9-13] have proposed a genetic programming approach to record deduplication that combines several different pieces

4515 | P a g e
O c t o b e r   2 0 1 6
w w w . c i r w o r l d . c o m

of evidence taken out from the data content to find a deduplication function that is able to identify whether two entries in a repository are duplicates or not [21]. This is due to the fact that clean and duplicate-free repositories not only allow the retrieval of quality information but also lead to more concise data and to potential savings in computational time and resources to process this data [24]. Moreover, the suggested functions are computationally less demanding since they use less evidence.

Juliana B. dos Santos et al. [3] proposed a method for estimating the quality of a similarity function that does not require human intervention. Quality is measured in terms of recall and precision calculated at several different thresholds. Based on the results of the proposed estimation process and the requirements of a specific application, a user is able to choose a suitable threshold value. The estimation process relies on a clustering and the choice of a similarity threshold based on the silhouette coefficient which is an internal quality measure for clusters [16]. An extensive set of experiments on artificial and real datasets demonstrates the effectiveness of the proposed approach. ABC based methodology presented by Kumbhar et at.[4-8] maximizes the accuracy and minimizes the number of connections of an ANN by concurrently evolving the synaptic weights, the ANN's architecture and the transfer functions of each neuron [18]. The methodology is then tested with several pattern recognition.   Samanta et al.[9] proposed algorithm to search out the optimal combinations of different operating parameters for three widely used Non-Traditional Machining (NTM) processes i.e., electrochemical machining, electrochemical discharge machining and electrochemical micromachining processes using ABC. The single and multi-objective optimization problems are solved using this algorithm for the considered NTM processes [19]. Guopu Zhu and Sam Kwong [2] proposed the modified ABC algorithm called gbest-guided ABC (GABC) algorithm by incorporating the information of global best (gbest)solution into the solution search equation to improve the exploitation which is inspired by PSO. MariemGzara and AbdelbassetEssabri [15] projected the most parallel evolutionary algorithm for single and multi-objective optimization motivated by the need for the reduction of computation time and resolution of larger problems.. In multi-objective optimization problem, more exploration of the search space is required to obtain the whole or the best approximation of the Pareto Optimal Front [17]. They presented a new clustering based parallel multi-objective evolutionary algorithm that balances exploration and exploitation of the search space.

3. MATERIALS AND METHODS

3.1 Genetic Algorithm

Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. They represent an intelligent exploitation of a random search used to solve optimization problems.ad they exploit historical information to direct the search into the region of better performance within the search space [26.] In this algorithm, a population of strings (called chromosomes) which encode candidate solutions (called individuals) to an optimization problem, progress toward better solutions. Usually, solutions are represented in binary as strings of 0's and 1's, but other encodings are also possible. The evolution generally starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every solution in the population is calculated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and randomly mutated) to form a new population which is used in the next iteration of the algorithm [29]. Commonly, the algorithm terminates when either a highest number of generations has been produced, or a satisfactory fitness level has been reached for the population.

3.2 Simple Genetic Algorithm procedure

- ✓ Choose the initial population of individuals
- ✓ Evaluate the fitness of each individual in that population
- ✓ Repeat on this generation until termination (time limit, sufficient fitness achieved, etc.):
- ✓ Select the best-fit individuals for reproduction
- ✓ Breed new individuals through crossover and mutation operations to give birth to offspring
- ✓ Evaluate the individual fitness of new individuals
- ✓ Replace least-fit population with new individuals

3.3 Artificial Bee Colony (ABC)

ABC is a swarm based meta-heuristic algorithm that was introduced by [23] Karaboga in 2005 for optimizing numerical problems [1]. The intelligent foraging behavior of honey bees inspired Karaboga.The algorithm is very simple and robust. It classifies the colony of artificial bees into three kinds employed bees, onlookers, and scouts. Employed bees are recognized by their association with a particular food source which they are currently exploiting or are "employed" at. The information about this particular source is carried out with them and shared with onlookers. Onlooker bees are those bees that are waiting on the dance area in the hive for the information to be shared by the employed bees about their food sources, and then make a choice to choose a food source. A bee carrying out random search is called a scout

The steps involved in this algorithm are given below.

Initialization Phase
REPEAT
       Employed Bees Phase
        Onlooker Bees Phase
        Scout Bees Phase
        Memorizing of the best solution achieved so far
UNTIL(Cycle=Maximum Number or a Maximum CPU time)

In the initialization phase, the ABC algorithm generates a randomly distributed initial food source positions of S solutions, where S denotes the size of the employed bees or onlooker bees. Each solution xi (i= 1,2,. . ., SN) is a N-dimensional vector where N is the number of optimization parameters. Each nectar amount fit_i is then evaluated. In the ABC algorithm, nectar amount is the value of benchmark function.

In the employed bees' phase, each employed bee finds a new food source *vi* in the neighborhood of its current source x_i. To calculate the new food source the following expression is used.

$$v_{ij} = x_{ij} + \varphi_{ij}\ (x_{ij} - x_{kj})\qquad\qquad(1)$$

where $k \in (1, 2, . . . , SN\_$ and $j \in (1, 2, . . . , D)$ are randomly chosen indexes, and $k \neq i$. φ_{ij} is a random number between$[-1, 1]$. The employed bee then compares the new one against the current solution and memorizes the better one by means of a greedy selection mechanism.

In the onlooker bees' phase, each onlooker chooses a food source with a probability which is related to the nectar amount (fitness) of a food source shared by employed bees. To calculate the probability, the following expression is used.

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n}\qquad\qquad(2)$$

In the scout bee phase, if a food source cannot be improved through a predetermined cycles, called "limit", it is removed from the population. The employed bee of that food source then becomes scout and the scout bee must now find a new random food source position. These steps are repeated through a predetermined number of cycles, called Maximum Cycle Number or until a termination criterion is satisfied.

3.4 Exploration and Exploitation

The difficulties in the modeling of optimization algorithm are the balancing of exploration and exploitation. The main objective of the optimization algorithms is to produce an optimal solution. The optimal solution defined by an optimization algorithm after the ending criteria may be the best that can be produced, but it is not always the best for the problem. Thus, exploration and exploitation scenarios are introduced to the optimization algorithm. The exploration problem deals with the convergence of the unlikely solutions to a single group and finding the best from it. On the other hand, the exploitation of the problem deals with exploring the top solution from the likely solutions.ABC algorithm is selected to improve their performance by including the exploitation and exploration problem.

4. PROPOSED APPROACH ON DUPLICATE RECORD DETECTION.

Step 1: Similarity Computation for all pair of records

Similarity functions compute the similarity of each field with other record field and assign a similarity value for each field. The similarity metrics used in the proposed work are Levenshtein distance and cosine similarity.

Levenshtein distance: The chosen name fields of the records are "record 1" and "Record 2". The "Levenshtein distance" is computed by calculating the minimum number of operations that has to be made to transform one string to the other, usually these operations are: replace, insert or deletion of a character. The levenshtein distance between the records is finding out by considering the record as a whole.

Cosine similarity: The cosine similarity between the two records name field "Record 1" and"Record 2" are calculated as follows: First, the dimension of both strings are obtained by taking the union of two stringelements in the record 1 and "record 2" as (word1, word2, …….wordN) and then the frequency of occurrence vectors of the two elements are calculated i.e. "record 1" = (<vector value1>, <vector value2>,……<>) and "record 2"= (<vector value1>, <vector value2>,……<>) . Finally we obtain the dot Product and magnitude of both strings.

Step 2: Generate list of evidence

In this approach each pair of evidence <attribute, similarity function> that represents the use of a specific similarity function over the values of a specific attribute found in the data being analyzed(Moises G. de Carvalhoet al., 2011). For example, if we want to deduplicate a database table with the attributes (e.g., name, address, and city) using a specific similarity function (e.g., the Levenshtein function), generate the following list of evidence: a <name,

Levenshtein >, b <address, Levenshtein >, c <city, Levenshtein >. Develop a set of expressions by using these evidences with the simple mathematical functions (+,*,-,/,)

Step 3: Optimized expression

A set of such expression are supplied as input to GA to find best among the supplied inputs which is capable of providing better solution for the problem. The optimization algorithm ABC find the optimal threshold for each expression. The detailed algorithm is explained below.

Population

Initialize the population with user provided individuals. Here, set of expressions is considered as an initial population which is shown below.

Initial Population

(a+b)+(c+d)
(a+b)*(c+d)

In the above population, a, b, c, d correspond to evidence defined on the attributes name, address, Phone number and category respectively.

Fitness

The fitness value is a value generated from the fitness function which is one of the most important components in this process. If the fitness function is badly chosen then it will surely fail to find the best expression. In this approach, we have used F1 metric as the fitness function and can be calculated as

Precision (P): It is defined as the fraction of identified duplicate pairs that are correct.

$$Precision\ (P) = \frac{Number\ of\ true\ duplicate\ records\ identified}{Total\ number\ of\ duplicate\ records\ identified}$$

Recall (R): It is the fraction of actual duplicate pairs that are identified correctly in the input dataset.

$$Recall\ (R) = \frac{Number\ of\ true\ duplicate\ records\ identified}{Total\ number\ of\ duplicate\ records\ present\ in\ the\ dataset}$$

Recall can be seen as a measure of completeness, whereas precision is a measure of exactness or fidelity.

F1-measure (F): It gives equal weight to both precision and recall and it is the harmonic mean of precision and recall. The traditional F-measure or balanced F-score is computed as

$$F = \frac{(2 * R * P)}{(R + P)}$$

Likewise find the fitness value for each expression in the population based on threshold value. Since F1-value varies with different threshold, it is necessary to choose an optimized threshold to classify the dataset as duplicates and non-duplicates accurately.

4.1 Optimal threshold using modified ABC

In modified ABC, a population starts with a random set of threshold (particle) on each expression. This set is considered as a employeed bees. Find the fitness for all bees. This set passed on to onlooker-bee and scout bee and selects the best threshold value on each such expression which classify the set of records as duplicate and non-duplicate. The steps are given below.

Modified ABC

Step1. **Select** input

Step2. **Create** random population

Step3. **Initialize** employed bees

Step4. **Generate** random e_i values

$$E = [e_1, e_2, ..., e_n]$$

4518 | P a g e
O c t o b e r 2016
w w w . c i r w o r l d . c o m

Step5. **Calculate** fitness

Step6. **Calculate** corresponding $e_{i\ values}$

$$e_i \rightarrow [+,-,*,....]::[e_{i-1}, e_{i+1},....,e_n] \Rightarrow E_i^{correspond}$$

Step7. **Find**e_{best}

Step8. **Select** e_{best} as onlookers

$$[e_{best} \leftarrow v_i = v^0 + \varphi(v^0 - v_k)] \rightarrow updated\ to \rightarrow o_i$$

Step9. **Apply** operator update process on o_i

Step10. **Calculate** fitness of o_i

Step11. Select $o_{best.}$

If new_fitness>old_fitness

Replace o_{best}

Step12. **Repeat** step4 to step11 until stopping criteria

Step13. **If** abandon cycle

Introduce scout bee s_i

Step14. Apply operator update process on s_i

Step15. $S_i \rightarrow$ bee colony

Step16. **Repeat** step4 to step11 until stopping criteria

Step17.stop

Select the best n expressions having high fitness value and apply crossover and mutation to generate new set of population. Repeat the process until termination criteria is reached. Once the optimal expression has obtained during the training phase, the duplicate detection of testing datasets is done with the help of the same expression.

5. RESULTS AND DISCUSSION

This section shows the results of the experiments and discussed about the performance of the proposed work.

5.1 Dataset Description

The proposed approach used CORA Dataset which is commonly employed for evaluating duplicate record detection approaches.

Cora Bibiliographic: This dataset contains 864 entries including 112 duplicates, that were that were taken from riddle repository . Attributes used are author names, year, title, venue and other information.

5.2 Results

Experiment 1**:** Fig.1 shows the F1-measure of the proposed work using two similarity measures such as Levenshtein distance and Cosine Similarity on Cora dataset.

4519 | P a g e
O c t o b e r   2 0 1 6
w w w . c i r w o r l d . c o m

I S S N   2 3 2 1 - 8 0 7 X
V o l u m e   1 2   N u m b e r 1 1
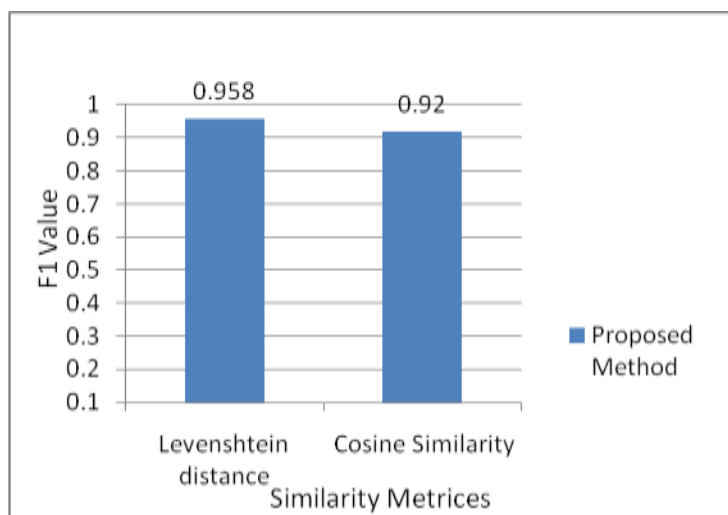J o u r n a l   o f   A d v a n c e s   i n   C h e m i s t r y

Fig.1: F1 values Vs Similarity metrics on Cora

Experiment 2: Fig.2 shows the precision and recall values for Cora dataset. It can be observed, that the proposed duplicate detection approach is very efficient by achieving high precision and recall.
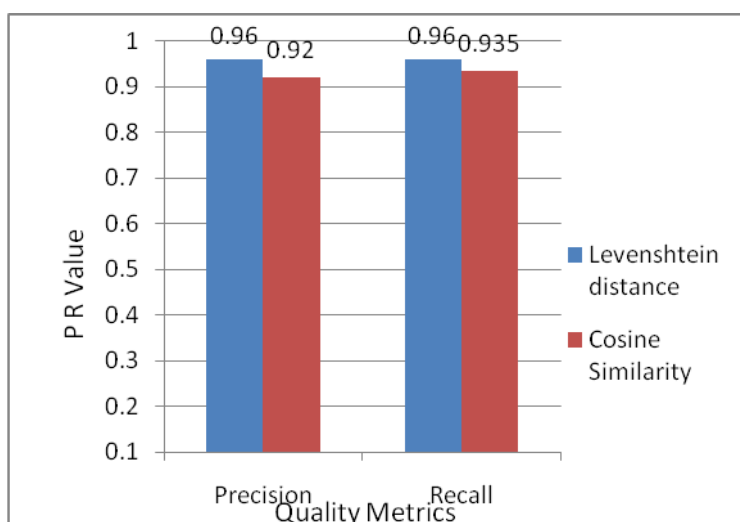

Fig.2:. Evaluation metric obtained for the proposed approach on Cora

6. CONCLUSION

The duplication has been one of the most emerging techniques for data redundancy and duplication. The duplication creates lots of problems in the information retrieval system. We have provided GA and modified ABC to detect the duplicate records and to find the optimal threshold. The experimentation of the proposed algorithms showed significant results. We used the cora dataset to evaluate the performance of the algorithm.

REFERENCES

1. Bitton, D., and DeWitt (1983). Duplicate Record Elimination in Large Data Files. *ACM Transactions on Database Systems,* 8(2),255-265.

2. P.Sasikumar, C.Vivek, P.Jayakrishnan , ''Key-Management Systems in Vehicular Ad-Hoc Networks'', International Journal of Computer Applications, 2010; 10 , 23-28.

3. Kumbhar, P.Y., and Krishnan, P.S.(2011). Use of Artificial Bee Colony (ABC) Algorithm in Artificial Neural Network Synthesis. *International Journal of Advanced Engineering Sciences and Technologies*, 11(1), 162-171.

4. Shriram K. Vasudevan, C.Vivek, S.Srivathsan ''An Intelligent Boxing Application through Augmented Reality for two users–Human Computer Interaction Attempt'', Indian Journal of Science and Technology, 2014; 8, 1-7.

5. Mariem Gzara & Abdelbasset Essabri (2011). Balanced Explore- Exploit Clustering based Distributed Evolutionary Algorithm for Multiobjective Optimisation, International transaction on Studies in informatics and control,20, 2, 97- 106.

6. C.Vivek, S.Audithan, ''Texture Classification by Shearlet Band Signatures'', Asian Journal of Scientific Research, 2014; 7, 94-101.

7. Shriram K. Vasudevan, C.Vivek ''An Intelligent Attempt to Export Files into Cloud in Handheld Devices through Gesture Recognition'', Indian Journal of Science and Technology, 2015; 8, 1-8.

8. Guopu Zhu & Sam Kwong (2010). Gbestguided artificial bee colony algorithm for numerical function optimization, Applied Mathematics and Computation,217, 3166-3173.

9. S.Vijayprasath, S.Palanivel Rajan, R.Sukanesh, "Experimental Explorations on EOG Signal Processing for Realtime Applications in LabVIEW", Proceedings of the IEEE, International Conference on Advanced Communication Control and Computing Technologies, INSPEC Accession Number: 13037966, ISBN No.: 978-1-4673-2045-0, pp. 67-70, 2012.

10. Juliana B. dos Santos, Carlos A. Heuser, Viviane P. Moreira and Leandro K. Wives(2011). Automatic threshold estimation for data matching applications. *Information Sciences*, 181(13), 2685-2699.

11. R.Sukanesh, S.Palanivel Rajan, "Cellular Phone based Biomedical System for Health Care", Proceedings of the IEEE, International Conference on Communication Control and Computing Technologies, INSPEC Accession Number: 11745436, ISBN No.: 978-1- 4244-7769-2, pp. 550-553, 2010.

12. C.Vivek, S.Audithan ''Robust Analysis of the Rock Texture Image Based on the Boosting Classifier with Gabor Wavelet Features'', Journal of Theoretical and Applied Information Technology, 2014, 69, 562-570.

13. S.Palanivel Rajan, K.Sheik Davood, "Performance Evaluation on Automatic Follicles Detection in the Ovary", International Journal of Applied Engineering Research, ISSN No.: 0973-4562, Vol. 10, Issue 55, pp. 1-5, 2015.

14. S.Palanivel Rajan, T.Dinesh, "Analysis of Human Brain Disorders for Effectual Hippocampus Surveillance", International Journal of Modern Sciences and Engineering Technology, ISSN 2349-3755, Vol. No.: 2, Issue No.: 2, pages: 38-45, 2015.

15. S.Palanivel Rajan, S.Vijayprasath, "Performance Analysis on Web based Traffic Control for DDoS Attacks", International Journal of Engineering Research and General Sciences, ISSN: 2091 - 2730, Volume 3, Issue 1, 2015.

16. S.Vijayprasath, S.Palanivel Rajan, "Performance Investigation of an Implicit Instrumentation Tool for Deadened Patients Using Common Eye Developments as a Paradigm", International Journal of Applied Engineering Research, ISSN No.: 0973-4562, Vol. 10, Issue No.1, pp. 925-929, 2015.

17. S.Palanivel Rajan, "A Survey on Detecting Methods of Driver Drowsiness", International Journal of Innovations in Engineering and Management, ISSN: 2319-3344, Vol. 4, Issue No. 1, pp. 1-5, March 2015.

18. C.Vivek, S.Palanivel Rajan, "Z-TCAM: An Efficient Memory Architecture Based TCAM", Asian Journal of Information Technology, ISSN No.: 1682-3915, Vol. No.: 15, Issue : 3, pp. 448-454, 2016.

19. S.Palanivel Rajan, R.Sukanesh, "Experimental Studies on Intelligent, Wearable and Automated Wireless Mobile Tele-Alert System for Continuous Cardiac Surveillance", Journal of Applied Research and Technology, ISSN No.: 1665–6423, Vol. No. 11, Issue No.: 1, pp.133 - 143, 2013.

20. S.Palanivel Rajan, R.Sukanesh, S.Vijayprasath, "Analysis and Effective Implementation of Mobile Based Tele-Alert System for Enhancing Remote Health-Care Scenario", HealthMED Journal, ISSN No. : 1840-2291, Vol. No. 6, Issue No. 7, pp. 2370–2377, 2012.

21. S.Palanivel Rajan, R.Sukanesh, S.Vijayprasath, "Design and Development of Mobile Based Smart Tele-Health Care System for Remote Patients", European Journal of Scientific Research, ISSN No.: 1450-216X/1450-202X, Vol. No. 70, Issue 1, pp. 148-158, 2012.

22. Mikhail Bilenko and Raymond J. Mooney(2003). Adaptive Duplicate Detection Using Learnable String Similarity Measures. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 39-48.

23. Moises G. de Carvalho, Alberto H. F. Laender, Marcos Andre Goncalves and Altigran S. da Silva(2011). A Genetic Programming Approach to Record Deduplication. *IEEE Transaction on Knowledge and Data Engineering*,pp 399-412.

24. Monge A.E and Elkan C.P.(1996). The field matching problem: Algorithms and applications. *SIGMOD workshop on research issues on knowledge discovery and data mining*, 267-270.

25. S.Palanivel Rajan, "Review and Investigations on Future Research Directions of Mobile Based Telecare System for Cardiac Surveillance", Journal of Applied Research and Technology, ISSN No.: 1665–6423, Vol. 13, Issue 4, pp. 454-460, 2015.

26. S.Palanivel Rajan, R.Sukanesh, "Viable Investigations and Real Time Recitation of Enhanced ECG Based Cardiac Tele-Monitoring System for Home-Care Applications: A Systematic Evaluation", Telemedicine and e-Health Journal, ISSN: 1530-5627, Vol. No.: 19, Issue No.: 4, pp. 278-286, 2013.

27. S.Palanivel Rajan, "A Significant and Vital Glance on "Stress and Fitness Monitoring Embedded on a Modern Telematics Platform", Telemedicine and e-Health Journal, ISSN: 1530-5627 (Online ISSN: 1556-3669), Vol. No.: 20, Issue No.: 8, pages: 757-758, 2014.

28. S.Palanivel Rajan, R.Sukanesh, T.Kamalesshwaren, R.Prasaanth, P.Thangaperumal, "Performance Analysis of Mobile Phone Radiation Minimization through Characteristic Impedance Measurement", International Journal of Computer Science Issues, ISSN No.: 1694-0814, Volume 9, Issue 2, No. 3, pp. 540-544, March 2012.

29. Samanta, S. and Chakraborty, S.(2011). Parametric Optimization of Some Non-traditional Machining Processes using Artificial Bee Colony Algorithm. *Engineering Applications of Artificial Intelligence*,24(6), 946-957.

30. S.Vijayprasath, R.Sukanesh, S.Palanivel Rajan, "Certain Investigations on Eye State Analysis for the Development of a Non-Intrusive Fatigue Detection System", European Journal of Scientific Research, ISSN No.: 1450-216X/1450-202X, Vol. 93, No 3, pp.342-350, 2012.

4522 | P a g e
O c t o b e r   2 0 1 6
w w w . c i r w o r l d . c o m