REVIEW PAPER
International Journal of Recent Trends in Engineering, Vol 2, No. 4, November 2009

Managing Uncertain Data using Multi Criteria Repeat Crossover Genetic Algorithm

Dr.G.RADHAMANI [1] , NIJU.P.JOSEPH [2]
[1] Dr. GRD College of Science, Coimbatore, Tamil Nadu, India
radhamanig@hotmail.com
[2] SCMS School of Engineering & Technology, Cochin, Kerala, India
pjniju@gmail.com

***Abstract*— Multi Criteria Genetic algorithms can produce good balance between their precision and their complexity. Systems use genetic algorithms for providing learning and adaptation capabilities from uncertain data set. A set of techniques discussed here, can be used to enhance searching and retrieving information from an existing fuzzy Knowledge Base (KB). The proposed genetic algorithm tries to be capable of generating solutions from both crisp and fuzzy valued data. The search algorithm transforms vague and uncertain data represented as fuzzy sets for accurate solution generation. Procedures are proposed which learns from the different available knowledge base for its application on uncertain data. In this paper, a repeat crossover Genetic Algorithm is formulated for its usage on uncertain data represented as fuzzy sets and it can be implemented for many real world applications.**

***Index Terms*— Uncertain data, efficient search, knowledge learning, Repeat Crossover, Genetic algorithms.**

I. INTRODUCTION

In most real-world problems, data have a certain degree of imprecision. Sometimes, this imprecision is small enough [5] so that it can be safely ignored. On other occasions, the uncertainty of the data can be modeled by a probability distribution (e.g., additive random noise). Lastly, there is a third kind of problems where the imprecision is significant, and a probability distribution is not a natural model.

For example, in [5], up to eight sources of information appropriately characterized by intervals were studied: plus-or-minus reports, significant digits, intermittent measurement, non-detects, censoring, data binning, missing data and gross ignorance.

As a further matter, there are ongoing researches about the use of Genetic Algorithm (GA) for modeling the interaction between variability and imprecision. The use of GA is to learn and evaluate Genetic Systems, and to advocate the use of specific reasoning methods and search functions to solve problems of uncertainty. In this work a comprehensive algorithm is proposed in order to learn from an existing Knowledge Base (KB) of imprecise data. The description includes some issues about the reasoning methods suitable for using vague data, bounds of the accuracy of a KB on vague data, and multi criteria genetic algorithms capable of optimizing a mix of crisp and fuzzy objectives.

II. RELATED WORK

The Use of imprecise data in fuzzy systems is analyzed in [1]. A Multi objective Genetic Fuzzy System with Imprecise Probability Fitness for Vague Data is discussed in [2]. Fuzzy-Pareto-Dominance driven multi objective genetic algorithm was given a focus in [3]. Predictive knowledge discovery by multi objective genetic fuzzy systems for estimating consumer behavior models was done in [4].Experimental uncertainty estimation and statistics for data having interval uncertainty is implemented in [5]. Mining uncertain data with multi objective genetic fuzzy systems to be applied in consumer behavior modeling was studied in [6]. Genetic learning of fuzzy rules based on low quality data was discussed in [7].

III. PROPOSED WORK

The Proposed work considers the uncertainty in occurrence of data in many applications. It develops a Multi criteria Repeat Crossover algorithm for managing Uncertain Data.

PROPOSED MULTICRITERIA REPEAT CROSSOVER GENETIC ALGORITHM (MRCGA):

Step1. (Initialization) Choose population size N based on multi criteria (Crispy and Fuzzy Input data), proper crossover probability $c\ p$ and mutation probability $m\ p$, respectively. Generate initial population P (0). Let the generation number t = 0.

Step2. (Crossover) Choose the parents for crossover from P (t) with probability $c\ p$. If the number of parents chosen is based on multi objectives, then in P (t), randomly match every two parents as a pair and use the proposed crossover operator c1 to each pair to generate two offspring. All these offspring constitute a set denoted by O.

Step3. (Repeat Crossover) Choose population O .Match parents based on higher sorted fitness function order into a set and use the proposed crossover operator c2 to generate set of off springs. All these offspring constitute a set denoted by s.

Step4. (Local Search) For each offspring generated by crossover, the proposed local search scheme is used to it to generate an improved offspring. All these improved offspring constitute a set denote by 1 *s*.

Step5. (Mutation) Select the parents for mutation from set 1 *s* with probability *m p*. For each chosen parent, the proposed mutation operator is used to it to generate a new offspring. These new offspring constitute a set denoted by 2*s*.

Step6. (Selection) Select the best *N* individuals among all the generated set G as the next generation population *P* (*t* +1), let *t* = *t* +1.

Step7. (Termination) If termination conditions hold, then stop, and keep the best solution obtained as the approximate global optimal solution of the problem; otherwise, go to step 2.

IV. INVENTORY MANAGEMENT USING MRCGA

The Proposed Genetic Algorithm MRCGA is implemented in inventory stock management where there are uncertainties in stock management .The randomly generated initial chromosome is created by having the stock levels within the lower limit and the upper limit for all the contributors of the supply chain, factory and the supply centers. As known, chromosome is constituted by genes which defines the length of the chromosomes. The stock level of each member of the chromosome is referred as gene of the chromosome. Hence for *n* length supply chain, the chromosome length is also *n*. As we are using only three members of the chain, the length of the chromosome *n* is 3, i.e. 3 genes. Each gene of the chromosome is representing the amount of stock that is in excess or in shortage.

A. Chromosome Representation

These kinds of chromosomes are generated for the genetic operation. Initially, only two chromosomes will be generated and from the next generation a single random chromosome value will be generated. The chromosomes thus generated is then applied to find its number of occurrences in the database content by using a *Select count*( ) function The function will give the number of occurrences of the particular amount of stock level for the three members *Nc* that are going to be used further in the fitness function.

B. Fitness Function

Fitness functions ensure that the evolution is toward optimization by calculating the fitness value for each individual in the population. The fitness value evaluates the performance of each individual in the population:

$$f(k) = \log(1 - (Nc / Np)) \quad , k = 1, 2, 3\ldots m$$

Where, *N c* is the number of counts that occurs throughout the period and *N p* is the total number of inventory values obtained after clustering.
m is the total number of chromosomes for which the fitness function is calculated.
The fitness function is carried out for each chromosome and the chromosomes are sorted on the basis of the result of the fitness function. Then the chromosomes are subjected for the genetic operation crossover and mutation.

C. Cross Over & Repeat Crossover

As far as the crossover operation is concerned, repeat crossover is used in this study. The first two chromosomes in the mating pool are selected for crossover operation. The genes that are right of the cross over point in the two chromosomes are swapped and hence the cross over operation is done. After the crossover operation two new chromosomes are obtained. These chromosomes are subjected to repeat crossover process.

D. Mutation

The newly obtained chromosomes from the repeat crossover operation are then pushed for mutation. By performing the mutation, a new chromosome will be generated. This is done by a random generation of two points and then performing swaps between both the genes.
The mutation operation provides new chromosomes that do not resemble the initially generated chromosomes. After obtaining the new chromosome, another random chromosome will be generated. Then again the process repeats for a particular number of iteration while the two chromosomes that are going to be subjected for the process is decided by the result of the fitness function.
Each number of iteration will give a best chromosome and this is will be considered to find an optimal solution for the inventory control. When the number of iterations is increased then the obtained solution moves very closer to the accurate solution. More the number of iterations results in more accurate optimal solution. Eventually with the help of the Genetic algorithm, the best stock level to be maintained in the members of the supply chain could be predicted from the past records and so that the loss due to the holding of excess stock level and shortage level can be reduced in the upcoming days.

E. Experimental Results

The optimization of inventory control in supply chain management based on genetic algorithm is analyzed with the help of MATLAB. The stock levels for the three different members of the supply chain, Production Center, Supplier1 and Supplier 2 are generated using the MATLAB script and this generated data set is used for evaluating the performance of the genetic algorithm.

Some sample set of data used in the implementation is given in table 1. Some 15sets of data are given in the table 1 and these are assumed as the records of the past period. The two initial chromosomes are generated at the beginning of the genetic algorithm. These initial

Chromosomes are subjected for the genetic operators, Crossover and Mutation. The resultant chromosome thus obtained is again processed with repeat crossover and mutation so that it moves towards the best chromosome after the each iterative execution. Hence at the end of the execution of 'n' iterations, best chromosome '-591 -329 269' is obtained.

While applying the genetic algorithm with the past records, it can be decided that controlling this resultant chromosome is sufficient to reduce the loss either due to the holding of excess stocks or due to the shortage of stocks. Hence it obtains a stock level that is a better prediction for the inventory optimization in supply chain management.

TABLE1.
A SAMPLE OF DATA SETS HAVING STOCK LEVELS

Sl. No	Production Center	Supplier 1	Supplier 2
1	-591	-329	269
2	-479	-796	-548
3	-591	-329	269
4	494	392	285
5	-591	-329	269
6	372	573	-345
7	999	-934	108
8	146	118	532
9	-591	-329	269
10	-591	-329	269
11	-746	721	-677
12	792	-456	837
13	-550	-634	158
14	-550	-634	158
15	611	-295	-443

V. CONCLUSION AND FUTURE RESEARCH

When the data is imperfect and uncertain, an algorithm to manage such data is proposed. This algorithm tries to repeatedly crossover and mutate the uncertain data in inventory stock management .This leads to better solutions in many iterations .The algorithm proposed here helps in arriving at solutions which can process uncertain data. This algorithm can be implemented in uncertain data based applications which have both crisp and fuzzy data inputs. Better insight gained by this technique will help to harness its power for various applications. In future research, the algorithm can be improved in the areas of speed, accuracy and consistency over a larger uncertain input data environment.

REFERENCES

[1] L. Sanchez and I. Couso , "Advocating the use of imprecisely observed data in genetic fuzzy systems," IEEE Transactions on Fuzzy Systems. Volume 15, Issue 4 Page(s):551 - 562, Aug. 2007.

[2] L. Sanchez, I. Couso, and J.Casillas, "A Multi objective Genetic Fuzzy System with Imprecise Probability Fitness for Vague Data," in Proc. of the 2006 IEEE International Conference on Evolutionary Fuzzy Systems, Ambleside, UK, pp. 131–137, 2006.

[3] M. Koeppen, K. Franke, and B. Nickolay, "Fuzzy-Pareto-Dominance driven multiobjective genetic algorithm." in Proc. 10th International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey, pp. 450–453, 2003.

[4] J. Casillas, O. Delgado, and F. Mart´ýnez-L´opez, "Predictive knowledge discovery by multi objective genetic fuzzy systems for estimating consumer behavior models," Proceedings of the 4th International Conference in Fuzzy Logic and Technology (EUSFLAT 2005), Barcelona, Spain, 272-278,2005.

[5] S.Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, L. Ginsburg," Experimental uncertainty estimation and statistics for data having interval uncertainty", Technical Report, SAND2007-0939, Sandia National Laboratories, 2007.

[6] J.Casillas"Mining uncertain data with Multi objective genetic fuzzy systems to be applied in consumer behavior modeling ". Expert Systems with Applications: An International Journal Volume 36, Issue 2, March 2009.

[7] Sanchez., I. Couso, J. Casillas, "Genetic learning of fuzzy rules based on low quality data", Fuzzy Sets and Systems ,2009.

[8] J. Teich, "Pareto-front exploration with uncertain objectives." in EMO, 2001, pp. 314–328.

[9] P. Limbourg, "Multi-objective optimization of problems with epistemic uncertainty." in EMO, 2005, pp. 413–427.

ACEEE