Retrieving Records of Genealogy registers online using Genetic Algorithm based Adaptive Heuristic Search Technique

Govind Singh
M.Tech (Computer Science and Engineering)
Graphic Era Hill University,
Bhimtal Campus, Uttarakhand, India

Kamlesh Kumar
Assistant Professor, School of Computing
Department
Graphic Era Hill University,
Bhimtal Campus, Uttarakhand, India

ABSTRACT

If we talk about evolutionary computing, genetic algorithm is a main area to be research. It is a growing area of an Artificial Intelligence. This paper focuses on evolution of Genetic algorithm and its impact on genealogy database. GA is also better than conventional Artificial Intelligence because of its robustness. GA focuses on input/output pattern that is if we give some input as chromosomes, we produce some output as an offspring. GA is beneficial for us because they do not break easily even if the inputs changed slightly, or in presence of reasonable noise.

In this paper we intended on developing a computerized relational database on the basis of GA Heuristic Search technique. It will be easy to find out the relation in between the people. This paper will lead to a project for Hindu genealogy registers at Haridwar, Uttarakhand, India. This has been a subject of study for many years. People comes from outside for pilgrimage or for cremation of their dead kin and register their family detail there in a register maintained by Brahmin priests called 'Pandas'. Through this project we want the change this register system. As we are living in a technological world and computer is our basic need. So we want to maintain a computerized relational database which will be secure enough while accessing the data from it.

With the implementation of Genetic Algorithm, it will really be an easy task to find out the relation among people and to maintain the records.

Keywords

Genetic algorithm, Artificial Intelligence, Chromosomes, relational database, Heuristic search

1. INTRODUCTION

After the idea of evolutionary computing was proposed in near 1960s, John Holland and his students invented Genetic Algorithm as a heuristic search technique for natural selection and are able to find out proper solution in reasonable amount of time. GA is generally a biological process. All living organisms consist of cells and each cell contains same number of chromosomes, which consists of genes. There are several operators used in GA as inheritance, mutation, selection, and crossover. On the basis of these operators the new offspring is generated and after that we will maintain a family tree. This comparison will result to develop the relational database. If we face any problem, first of all we concern with the best solution to that problem. There are many solutions to a single problem. All these solutions are within a frame called Search Space.

The main objective of this paper is to collect, organize and maintain the family tree database with the help of genetic operators. So we have defined our paper in two parts. In the first part we focus on fundamental aspects of genetic algorithm such as fitness evaluation, operators, parameters etc. Haridwar is a holy place of Hindus. People come here for several purposes and register their family tree detail on the hand written register. Many people are also not aware about this so in second scenario we focus on maintaining and accessing the data from the computerized database. We want latest terminology in this field through this project and heuristic search technique will help us to improve our strategy.

2. ACCESS TO GENETIC ALGORITHM

First of all we focus on the evolution of genetic algorithm. How GA came into light? GA is inspired by Darwin's theory about evolution- "Survival of the Fittest" which says that Survival is a natural process resulting in the evolution of organisms. "In nature it means that the fittest animal is the one that survives". Solution to the problem is solved by genetic algorithm. There are many solutions to a problem called population. Solution from one population is taken to form a new population. According to their fitness new population is selected to reproduce the offspring and it is assumed that the new offspring will be better than its parents. In searching a large state-space it provides a quick solution than many other search strategies.

GA is based on an analogy with the genetic structure of chromosomes within a population of individuals using the following foundations:

- Individuals in a population compete for resources and mates.

- Those individuals most successful in each 'competition' will produce more offspring than those individuals that perform poorly.

- Genes from `good' individuals propagate throughout the population so that two good parents will sometimes produce offspring that are better than either parent.

- Thus each successive generation will become more suited to their environment.

So a basic genetic algorithm requires a genetic representation of the solution as well as a fitness function. GA then proceeds

to initialize a population of solution and applies the operators on them.

3. IMPLEMENTATION AND BUILDING BLOCKS OF GENETIC ALGORITHM

Solving a problem generally means looking for a solution, which is best among others. While solving any problem using genetic algorithm, it provides a better performance within the search space because of the flow of operators used to implement the problem. So based on the natural selection we have the following flow control strategy:

3.1. Initialization of genetic algorithm

3.2. Selection Operator

3.3. Crossover Operator

3.4. Mutation Operator

3.5. Termination

3.1. Initialization of Genetic Algorithm

Initial population generation is the first step towards genetic algorithm formation in which individual solutions are randomly generated. There are typically many solutions to a problem but we focus on to the optimal solution with lower complexity. There are n numbers of chromosomes; on the basis of these we evaluate the fitness function f(x).

3.2. Selection Operator

Selection is a base process for any of the problem domain. This is the first operator applied on population. The problem is how to select the chromosomes. We select two parent chromosomes from a population and evaluate a fitness function, where fitter solutions (evaluated from fitness function) are more likely to be selected. Several methods are used to select the optimal solution such as Rank selection, Tournament selection, Boltzmann selection etc. Sometimes the optimal solution also creates the problem. So we go for the worst solution at that time.
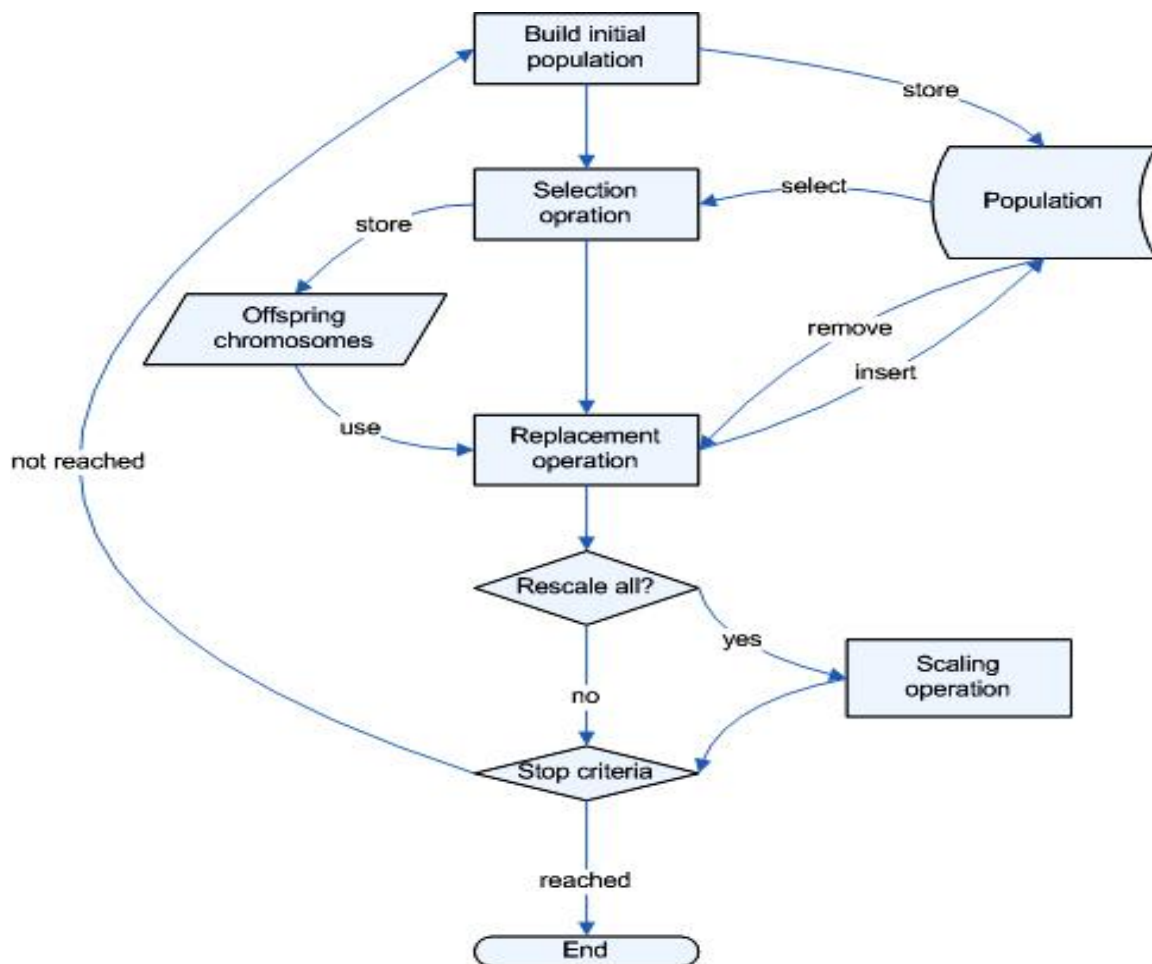

Fig 1: Block diagram of Genetic Algorithm formation

3.3. Crossover operator

After we select the best generated population, crossover operator takes place. When two organisms mate their genes; the resultant offspring may end up having half the genes from one parent and half from the other. This process is called as Crossover or Recombination. The purpose of crossover is only we want to produce a better chromosome than its parents.
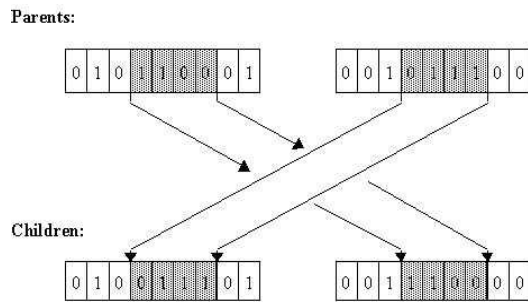
Fig 2: Two point crossover of Binary strings

Crossover operators are of many types. As One point crossover, two point crossover, Uniform crossover, Arithmetic and heuristic crossover. Generally we prefer single point crossover because it is easy to cross only with one point in the string.

3.4. Mutation operator

After the crossover, mutation takes place. This operator maintains the genetic diversity from one generation of population to the next.
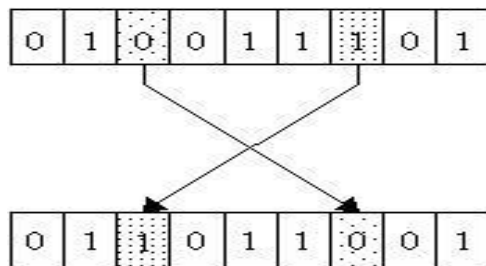

Fig 3: Mutation in binary bits

It is a random alteration of a gene, or genes, in a chromosome. This leads a new gene values being added to the gene pool. This operator helps to prevent the population from stagnating at any local optima.

Mutation is of many types as Flip bit, Boundary, Non uniform, Uniform and Gaussian. In the genetic algorithm, mutation involves selecting a random point, or points, in the chromosome and substituting the value found at that point with an appropriate value that has been randomly generated.

3.5. Termination

When all the above operators are applied to a population, a new offspring is generated and that acts as a parent for new generation. This flow continues until best individual is good enough. Genetic algorithms are very easy to implement but difficult to analyze its behavior. With the help of above operators applied on it, it is very convenient to use GA.

4. APPLICATIONS OF GENETIC ALGORITHM

Genetic algorithm is a field used in every areas related to computing, electronics, human life and many others. One of the major applications is in Evolutionary algorithm itself. It is useful in developing family tree. Traditional methods of search and optimization are too slow in finding a solution in a very complex search space, even implemented in supercomputers. Genetic algorithm provides a platform for many applications. Moreover it is useful is some areas of Nonlinear problems. There are several operators used in GA, so it provides a solution to be very robust from its parent solutions. Here are some applications as follows:

4.1. Maintaining genealogy database

The first and most important application of GA is in making genealogical database. This paper is also made to focus on this particular topic. In our daily life we have many relatives of ours. To remember the relation with all of them is a very uncomfortable job for anyone, so relational database will show all the relevant details and relationship with a person.

4.2. In scheduling applications

Scheduling is a very useful process in any organization and if we schedule any time table or any database then to do this by hand is very time consuming process. GA has proved very useful in automating and streamlining the process.

4.3. Optimized telecommunication routing

Many times we have noticed that our mobile networks, LAN, internet, fax and all are not working properly. But GA provides us a portable platform in such conditions. It allows a dynamic and anticipatory routing of circuits for telecommunication networks. GAs is being developed to optimize placement and routing of cell towers for best coverage and ease of switching.

4.4. Encryption and code breaking

It is necessary to encrypt the data for security purpose. If there is much complexity in breaking the encrypted code, with the help of GA it is very convenient to decrypt the code. So GA may be used in large areas in computer field for protecting cyber-crimes.

4.5. Game playing

If we are familiar with computer games, then we feel difficulty in playing the game against computer because it uses GA on it. If that GA will be applied on a human brain, it also will work like a super computer. GA is a very powerful tool for inconsistency of data. If we have a fuzzy data, GA can solve the data in a little amount of time.

5. GENEALOGY DATABASE DESIGN OVERVIEW

Through this paper I want to develop a genealogy application for retrieval of data for a holy place Haridwar in Uttarakhand. I will show the relationship in between individuals. This application will work as when the name of any person will be entered in the database, it will show all the family detail for that person because in Haridwar, panda (authorized person) stores the hand written data about every person visiting to Haridwar in a register. This is a very fuzzy process that creates inconsistency. So we want to put whole the database in a computer for security and ease to access purpose.

For any database the most basic requirement is that it models the intended real world situation accurately. There should not be many restrictions on the database. As if we want to develop a genealogical database then there may be any number of children in a family. If we provide restrictions in such a database, then our database will be flawed because there is no such restriction in human life. A good model should be able to cope with all the modern methods of producing children,

including artificial insemination of sperm, artificial implantation of the fertilized egg in the genetic mother or in a surrogate, and even cloning. Multiple spouses, at the same time or sequentially, should not cause any database problem.

In this paper we focus on Human Based Genetic Algorithm that is all about maintaining a family tree. All the operators

used for human interface. There is one extra operator that is initialization which allows a HBGA to start with an empty population.
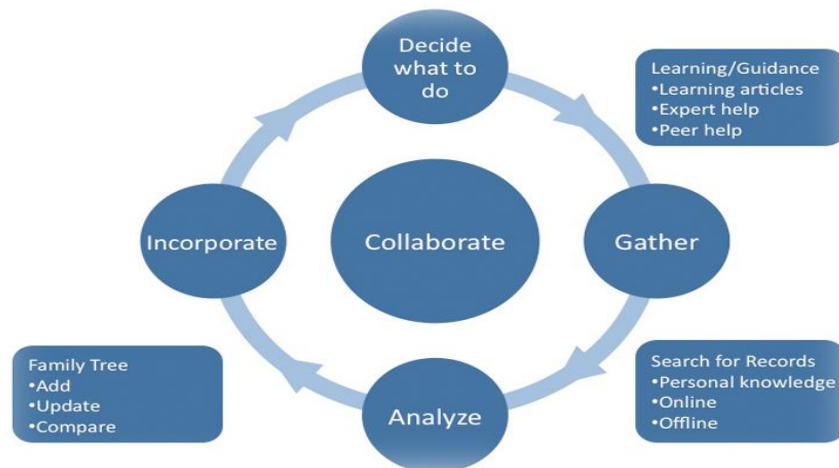

Fig 4: Genealogical workflow diagram for Family-Tree

For preparing such a database we will have to go through the tools used for the software. It is a very frustrating job to schedule the database; we will build the table as per the need. As in above shown cycle of genealogical workflow we have to collaborate whole the knowledge to maintain a family tree.

6. PROBLEM STATEMENT IN GENEALOGICAL DATABASE

A database is a collection of records stored in a computer systematically and a computer program used to manage and query a database is known as Database Management System. No any system is made with zero numbers of shortcomings. Our paper is concerned with distributed database in which we will have to focus on the data which is geographically scattered. The problem with this type of database is that if two persons are working on their own project related to relational database. We will have to try and avoid duplications of whole databases or even individual records and to use the hyper link mechanism to co-reference data held at other sites. The advantage of linking databases is that each researcher can continue to explore their own area and amend and update the information under their domain of responsibility without the permission or active cooperation of the others involved.

Genealogical database on the web should not have copies of one data to another. The main thing to note is that the data size will be too large if we concern with only one website to manage all the data so we will have to link our database with more than one web page that will low the workload.

7. BASIC INFORMATION RETRIEVAL SYSTEM

As we concern with maintaining a genealogical database so retrieving the knowledge from the database is a very tricky job. To access the information we will make our database easy to access and platform independent. We may find our data from the database by our name, age, state, city, town etc. Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on metadata or on full-text (or other content-based) indexing. When a user

will enter any query into the system, the process will begin simultaneously.

In our database there are mainly three things to be noticed:

- Representation of the document

- Representation of user's queries

- Algorithm used to match user with appropriate document

We have to make our project so comfortable that anyone can access it from anywhere but the changes will be made by the authorized person only. In Uttarakhand most of the people are from rural areas so we will put all the requirements of rural population as well as urban population so that it will be easy to understand the relation in between people.
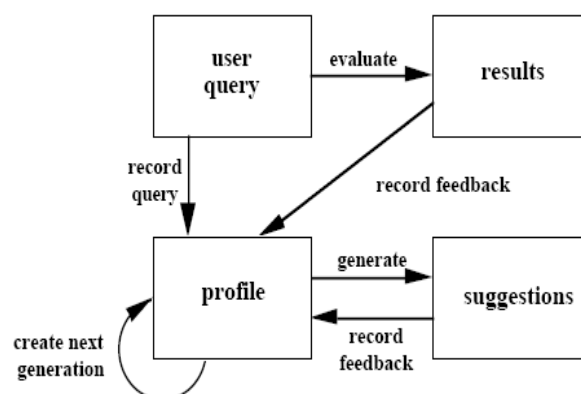

Fig 5: Personalizing Information Retrieval system

Our database will work as the above system. Document representation should be easy that it will find the solution to the query very easily. The algorithm used will match the user requirement with the document representation accurately so no collision may occur in the system as in traditional system.

8. CONCLUSION AND FUTURE WORK

In this paper we concern with how to form a genetic diversity with the help of genetic algorithm and compare the cycle of genetic algorithm to the actual population. The second thing to be notice in this paper we concern with is how the population grows and how to get an accurate offspring from its parents and how we will use new offspring to generate the better child. All the chromosomes are in the binary format so we may encrypt our data for better security using genetic algorithm. Genetic algorithm is much more than only producing the offspring from its parents. In artificial creativity, automated design, bioinformatics, developing polygenetic trees etc. there is a very critical use of GA.

In our future work we concern with one of the application that is maintaining genealogical database using GA tools. We will develop a database for the Hindu population go to Haridwar in Uttarakhand. People come there for pilgrimage or any other purpose and register their name in a hand written database. Through this paper we only want to put all the data in a computerized database that will be accessible from anywhere and will reduce inconsistency and collision of the data. Information retrieval for a specific population will also be focused through this paper. Researchers can also proceeds in this area by developing a website for the population through GA.

9. ACKNOWLEDGEMENT

10. REFERENCES

[1] Erick Cantú-Paz "A Survey of Parallel Genetic Algorithms", Department of Computer Science and Illinois Genetic Algorithms Laboratory University of Illinois at Urbana-Champaign

[2] David Beasley, David R. Bull, Ralph R. Martin "An overview of Genetic algorithm"

[3] Behrouz Minaei-Bidgoli1, William F. Punch III "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System", Michigan State University 2340 Engineering Building

[4] Praveen Pathak, Michael Gordon, Weiguo Fan "Effective Information Retrieval using Genetic Algorithms based Matching Functions Adaptation", University of Michigan Business School

[5] "15 Real-World Uses of Genetic Algorithms", http://brainz.org/15-real-world-applications-genetic-algorithms/

[6] "Genetic algorithm overview" http://www.stumptown.com/diss/chapter1.html

[7] "List of genetic algorithm applications" http://en.wikipedia.org/wiki/List_of_genetic_algorithm_applications

[8] Petter Øgland, Department of Informatics, University of Oslo, "Implementing continuous improvement using genetic algorithms"

[9] Brain Larson, "Microsoft SQL server 2005 Reporting services 2005"

[10] "The concept of Hindusm" form http://www.hinduwebsite.com/hinduism/concepts.asp

[11] "Genetic algorithm" from Wikipedia, the free encyclopedia

[12] Basic description from "Introduction to Genetic Algorithms" http://www.obitko.com/tutorials/genetic-algorithms/ga-basic-description.php

[13] Lawrence "David" Davis Tica Associates, "Genetic algorithm and their applications"

[14] Haridwar, Hindu genealogy registers at Haridwar http://en.wikipedia.org/wiki/Haridwar

[15] M.Sanderson, "Test Collection Based Evaluation of Information Retrieval Systems".

[16] Brian Tompsett, Department of Computer Science, Problems of distributed Genealogical Databases

[17] Mahanim Omar, Adam Baharum, Yahya Abu Hasan, School of Mathematical Sciences, Universiti Sains Malaysia, "A JOB-SHOP SCHEDULING PROBLEM (JSSP) USING GENETIC ALGORITHM (GA)".

[18] Keshavamurthy B. N , Asad Mohammed Khan & Durga Toshniwal, Department of Electronics & Computer Engineering, Indian Institute of Technology, Roorkee, Uttarakhand, India, "Improved Genetic Algorithm Based Classification".

[19] Barry Leadbeater, "Genealogy database design overview", http://www.familyhistorysa.info/.