GENETIC-BASED NUTRITION RECOMMENDATION MODEL

[1]Fayoumi, S.A.A., [1]A.A. Hegazy and [2]M.A. Belal

[1]College of Computing and Information Technology,
Arab Academy for Science and Maritime Science Technology, Cairo, Egypt
[2]College of Computers and Information, Helwan University, Cairo, Egypt

ABSTRACT

Evolutionary computing is the collective name for a range of problem-solving techniques based on principles of biological evolution, such as natural selection and genetic inheritance. These techniques are being widely applied to a variety of problems in many vital fields. Also, Evolutionary Algorithms (EA) which applied the principles of Evolutionary computations, such as genetic algorithm, particle swarm, ant colony and bees algorithm and so on play an important role in decision making process. EAs serve a lot of fields which can affect our life directly, such as medicine, engineering, transportations, communications. One of these vital fields is Nutrition which can be viewed from several points of view as medical, physical, social, environmental and psychological point of view. This study, presents a proposed model that shows how evolutionary computing generally and genetic algorithm specifically-as a powerful algorithm of evolutionary algorithms-can be used to recommend an appropriate nutrition style in a medical and physical sides only to each person according to his/her personal and medical measurements.

Keywords: Evolutionary Computations, Evolutionary Algorithms (EA), Genetic Algorithm (GA), Nutrition, Body Status Index (BSI)

1. INTRODUCTION

Evolutionary Computation (EC) is a subcategory of artificial intelligence that includes continuous and combinatorial optimization problems. It uses iterative progress, such as growth or development in a population. This population is then selected in a guided random search using parallel processing to achieve the desired end. Such processes are often inspired by biological mechanisms of evolution (Clune *et al.*, 2009).

Many problems nowadays can be solved in theory by intelligently searching through many possible solutions using evolutionary computation. EC algorithms can be considered global optimization methods with a met heuristic or stochastic character and are mostly applied for black box problems often in the context of expensive optimization.

An Evolutionary Algorithm (EA) is a subset of computation, a generic populations based metaheuristic optimization algorithm and searching techniques. An EA uses mechanisms stimulated by evolution, such as selection, reproduction, mutation and recombination. Candidate solutions to the optimization problem play the role of individuals in a population and the fitness function determines the quality of the solutions. Evolution of the population then takes place after the repeated application of the above operators. Artificial Evolution (AE) describes a process involving individual evolutionary algorithms; EAs are individual

Corresponding Author: Fayoumi, S.A.A., College of Computing and Information Technology,
Arab Academy for Science and Maritime Science Technology, Cairo, Egypt

components that participate in an AE. Evolutionary algorithms often perform well approximating solutions to all types of problems because they ideally do not make any assumption about the underlying fitness landscape. The contribution is to apply EA on one of rich fields of research which is Nutrition. Nutrition process is a complicated process because many factors affect this process directly. These factors are as describes and as shown in **Fig. 1**:

- Medical: It concerns if the person suffers from any recent or historical diseases
- Physical: It concerns for person's measurements such as weight, height and so on
- Psychological: It concerns of person's emotional status such as depressed, dispirit, motivated or loved
- Social: It studies nutrition habits according to the traditions and social level of living
- Environmental: If the person lives in hot, cold or moderate countries

Nutrition models that were made, were based on many techniques such as neural network, fuzzy logic,and expert systems. But No nutrition model-so far-was built based on any of Evolutionary Algorithms such as Ant Colony, Genetics, Particle Swarm and Bees Algorithm and so on. In this search, the model is focused on the physical and medical sides in nutrition process. The chosen evolutionary algorithm to recommend a suitable nutrition style to any person is Genetic Algorithm, which will be explained in details in the next section.

2. GENETIC ALGORITHM

Genetic Algorithms are search Algorithms based on the mechanisms of natural selection and genetics (Bacardit, 2004). They have been applied to search, optimization and machine learning problems with great success. Genetic Algorithms explore the search space by using a population of solutions, instead of a single point. This Population is evaluated and then potentially improved by the mechanisms of selection, crossover and mutation. One of the abilities of genetic algorithms is to keep a good balance between exploration of the search space and exploitation of the best found solutions. This equilibrium allows exploring large search space efficiently, tending to avoid local minima. Genetic algorithms can also be applied to a wide range of domains, since it does not require many assumptions on the data model. They can also work with different representations allowing even wider applicability. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, natural selection and recombination (or crossover).
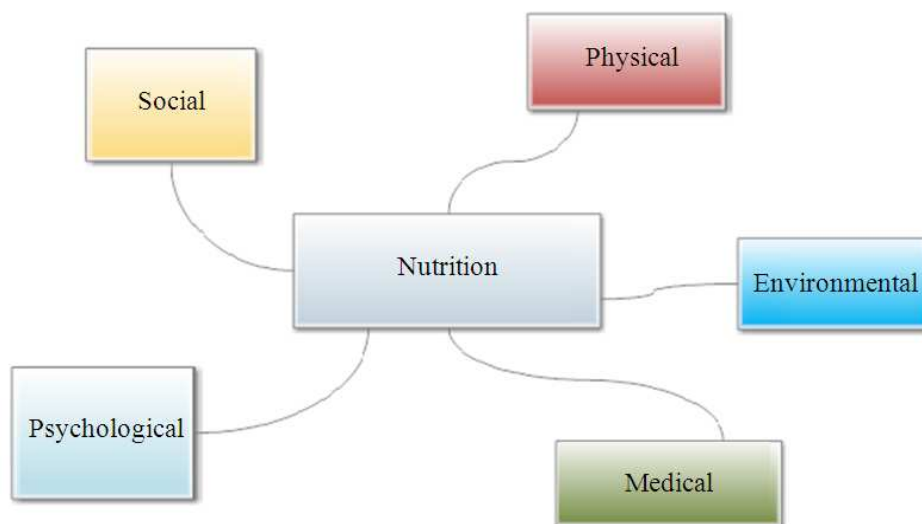

Fig. 1. Factors that affect nutrition process

Genetic algorithms are typically implemented as a computer simulation in which a population of abstract representations (called chromosomes) of candidate solutions (called individuals) to an optimization problem evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but different encodings are also possible. The evolution starts from a population of completely random individuals and happens in generations. In each generation, the fitness of the whole population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), modified (mutated or recombined) to form a new population, which becomes current in the next iteration of the algorithm.

Nature has always been able to solve one kind of task: Survival and adaptation to the environment. Charles Darwin proposed the concept of natural selection: The strongest individuals of a population (the better adapted to the environment) are the ones that survive. This process by itself has one problem: If the strong individuals dominate the weak completely, they will take over the population until all individuals are equal, which stops the adaptation process. Therefore the concept of diversity is necessary. Mendel discovered that parents transmit their biological information to the offspring in the reproduction process. All this information necessary to define an individual is codified at cellular level in a structure called chromosome. Parts of the chromosome codify hair color, height. New individuals are created by mixing the genetic information of the parents in a process called crossover. Therefore, new individuals are a mix of the information of the parents.

However, this does not account for the problem stated above, that is, the crossover of two identical individuals' means producing two identical offspring. A mechanism that introduces diversity is necessary: Mutation. Mutation can be defined as some small mistakes introduced during the mixing process of crossover. Thanks to mutation, new information that did not exist in the parents is introduced. Sometimes this change creates worse individuals, but sometimes it creates better ones, who are the next step in the evolutionary process.

Genetic Programming, This paradigm is often described as a variation of genetic algorithms rather than a mainstream evolutionary computation paradigm in itself. Individuals being evolved in this paradigm are various kinds of computer programs, consisting not only of data structures but also of functions (or operations) applied to those data structures. These programs are usually represented using trees (Holland, 1975). Genetic Algorithms, This is the most popular paradigm of evolutionary computation. Genetic algorithms emphasize crossover as the main exploratory search operator and consider mutation as a minor operator, typically applied with a very low probability. In early ("classic") genetic algorithms individuals were represented by binary strings, but nowadays more elaborate representations, such as real-valued strings, are also used (Holland, 1975; Goldberg, 1989).

The following flowchart illustrates the basic steps in a genetic algorithm **Fig. 2.**

2.1. Basic Mechanisms of Genetic Algorithms (Holland, 1975)

2.1.1. Individual

A candidate solution to the problem we are solving.

2.2.2. Chromosome

The codification of an individual. Usually individuals, unlike in nature, are codified using a single chromosome.

2.2.3. Gene

Each of the atomic values of a chromosome.

2.2.4. Fitness Function

The function that indicates the degree of adaptation of an individual to the environment where it lives. That is, how good is the individual in solving the problem.

2.2.5. Parent selection

The process that chooses the most fitted individuals to the environment to produce offspring. This process uses the value given by the fitness function to each individual to decide which the most fit individuals are.

There are many selection algorithms, some of them choose individuals based on their proportion of fitness value over the whole population, other methods are rank based and only take into account if an individual is better than another, not how much better it is.

2.2.6. Crossover

A process inspired in natural reproduction. Parents mix their chromosomes to create the offspring. Usually there is some probability (pc) of a candidate parent producing offspring. The mix can be performed in several ways. The most classical one, the one point crossover which is called single crossover as shown in **Fig. 3**, it chooses randomly a cut point in the crossover and creates offspring by mixing the contents to the left of

this point from one parent with the contents to the right of the point from the other parent. On the other hand, Multi crossover **Fig. 4**, it is as the same as single crossover but it chooses multi cut points.
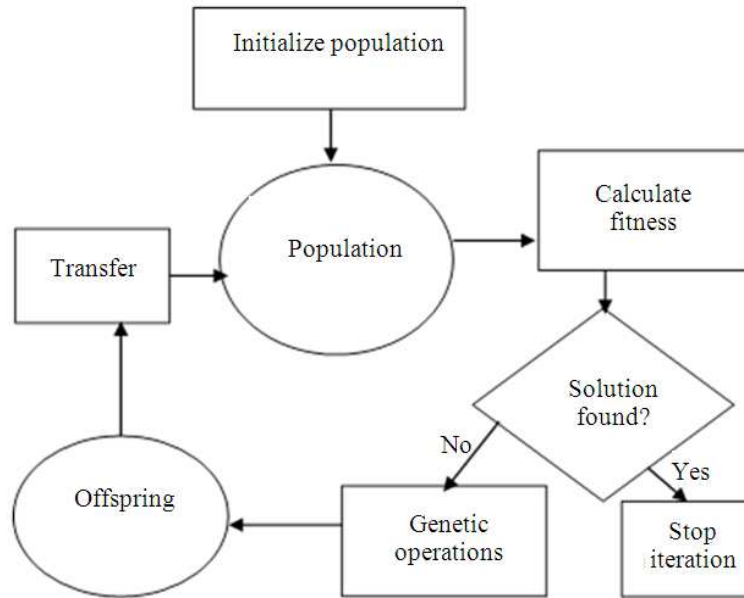

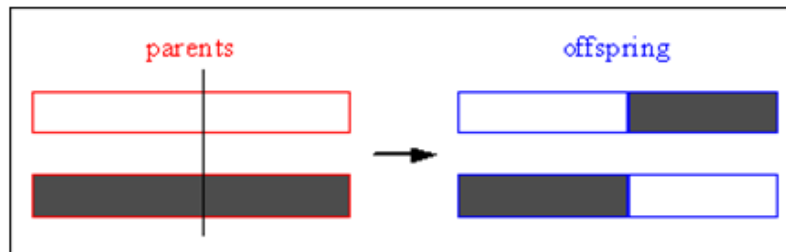
Fig. 2. Basic steps in genetic algorithms
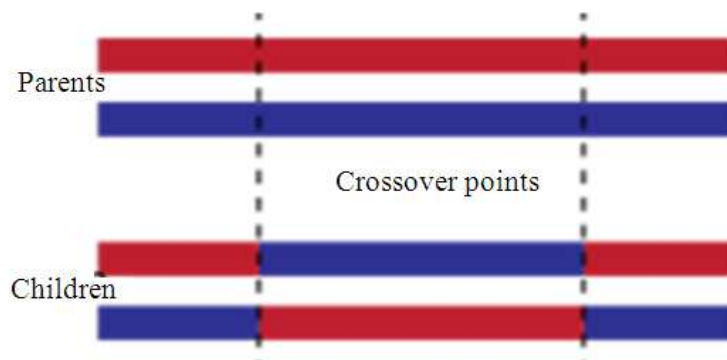

Fig. 3. Single crossover

Fig. 4. Multi crossover

2.2.7. Mutation

The alteration of the genetic material of an individual. Like crossover, this operator is controlled by a certain probability, usually gene-wise. For binary representations, the most typical mutation is to flip a gene, changing from 1 to 0 or from 0 to 1.

2.2.8. Replacement

The process that given an original population and an offspring's population merges them to create the population for next iterations. The most classical approach is to use only the offspring population and the best individual of the original population.

3. THE PROPOSED MODEL

The proposed model aims to how to recommend accurate nutrition advices and promote food supplementary for any person according to his/her health measurements using genetic algorithm **Fig. 5.**

The model permits each user to enter two kinds of inputs personal and medical inputs. Personal inputs will enter to Body Status Index (BSI) model to calculate the Body Mass Index (BMI) for each person to recognize if he/she is under weight, normal, overweight, or obese and also it calculates the recommended calories per day for each case. Moreover, the GA model depends on the medical data, it has a historical data repository that contains two

data sets. First data set is called Personal and Medical Data (PMD) which contains real data for 300 cases which are collected manually by retrospective audit review **Fig. 6**. These data are being split into two parts first part which concerns with personal data such as: Age, gender, if the person smokes or not, his/her type effort (muscles based or mental based) and for females, the system asks for pregnancy and breast feeding. Second part asks for medical measures such as blood sugar, blood pressure, uric acid, cholesterol, triglycerides and hemoglobin. The sample (300 cases) was from 20 to 60 years old.

Furthermore, the second data set is an output data set which is called Recommended Nutrition Data (RND) **Fig. 7**. This data set was built after medical consultations from many medical doctor and nutrition specialists for each case from 300 cases and this data set can be considered as a reference data set that measures how the proposed system succeeded. The Recommendations are:

- Quantity of sugar and salt
- Quantity of water
- How many minutes of exercise
- Quantity of protein (red and white meat and vegetarian protein
- Quantity of vegetables and fruits
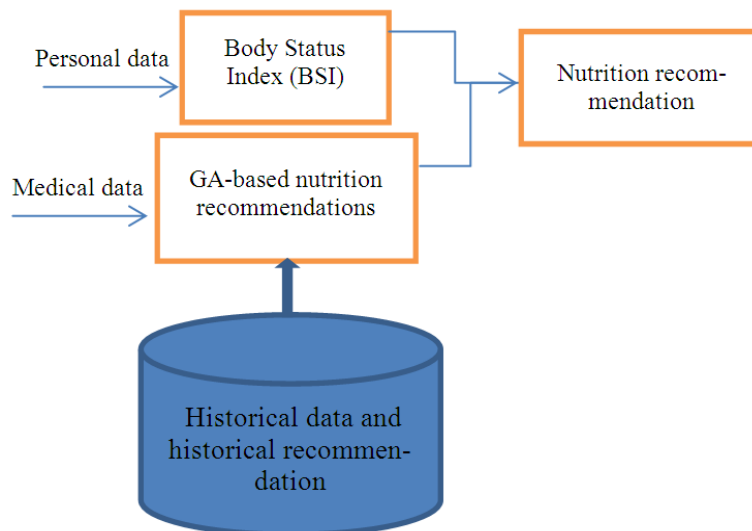- And vitamins supplementary such as: Vitamin B12, Iron and Folic Acid

Fig. 5. A proposed model for the proposed system

Fig. 6. Input data set

Fig. 7. Output data set

3.1. Chromosome Representation

The proposed chromosome is considered as a vector of 16 values from discrete numbers which mapped for real object such as tea spoon-moderate dish-tablet, medium piece and cup and so on.

The genes which are not represented as tablets or ampules-such as all kinds of vitamins-have 5 values, very low, low, normal, high and very high respectively, but vitamins are represented by 4 values as a maximum according to medical consultation.

Table 1. Recommendation user representation

Recommendation	Corresponding user representation
Sugar, salt and fats	Tea spoon
Vitamin B12	Ampules
Iron, folic acid and calcuim	Tablets
Vegetables and fruits	By one
Water	Glass
Carbohydrates and Vegetarian proteins	Medium plate
Red and white meat	Hand size piece.
Exercise	Minutes

The values in order:

Sugar:	0-10-15-20-25 g
Salt:	0-5-10-15-20 g
Vitamin B12:	0-1000-2000-3000 microgram
Iron:	0-17-34-51 mg.
Folic Acid:	0-500-1000-1500 microgram
Water:	from 2 to 6 liter.
Exercise:	from 15 to 60 min increased by 15 min.
Red Meat:	0-50-100-150-200 g
White Meat:	0-50-100-150-200 g
Vegetables:	0-100-200- 300-400 g
Fruits:	0- 50-100-150-200 g
Fats:	0-15-30-45-60 g
Calcium:	0-800-1600-2400 mg
Carbohydrates:	0-50-100-150- 200 g
Vegetarian Proteins:	0- 50- 100- 150- 200 grams.

The above values are for the purpose of the using genetic algorithm. The output of the model will be as shown in **Table 1**.

3.2 Genetic Steps

3.2.1. First Step: Generation

Population Size = 100 chromosomes as initial population selected randomly with equal probability for each value. Then, the genetic cycle was performed 1000 times including selection, crossover and mutation.

3.3. Second Step: Evolution of generations

3.3.1. Selection

The process of selection is based on the genetic operator called fitness proportionate selection or it known as Roulette Wheel selection. It is being used in genetic algorithm for selecting potentially useful solutions for recombination. Here, the fitness function which will be explained and discuss in the last of this section, assigns a fitness to possible solutions or chromosomes. This fitness level is used to associate a probability of selection with each individual chromosome.

3.4. Crossover

3.4.1. Single Crossover

A single index was put randomly. All data beyond that index in either chromosome is swapped between the two parent chromosomes. The resulting chromosomes are the children (offspring).

3.5. Multi Crossover

It is also called two points crossover. Two indexes were put randomly, this is because of the chromosome size is large that contains 16 value and population size is 100 which can be considered as a small quantity. So, multi crossover allows swapping as much as possible to collect or combine good gene from the same chromosome.

In "result section" in this thesis, we will show which one was better than the other. Is single crossover or multi crossover?

3.6. Mutation

The system tended to mutate 10% for all genes with equal probability to be increased or decreased by one step according to step size for each gene.

3.7. Fitness Function

Because of there are many different fields with different measures, we put weights for all fields according to its importance to unify the final format of fitness function which is as follows Equation 1:

$$F_{ci} = \min_{j=0}^{j=300} \left(\sum_{k=0}^{k=16} |O_{jk} - C_{ik}| + \sum_{k=0}^{k=13} |I_{jk} - test_k| \right)$$

When:
Ci = The Chromosome number
Ii = The input record number
Test = The test input measurements
Oi = The Output recommendation

All genes are weighted by multiplying each gene in certain weight according to its importance.

4. METHODOLOGY

The methodology in this proposed model is when the user enters his inputs, the genetic algorithm generate many solutions (chromosomes) from the output dataset randomly and apply the fitness function on it. The chromosome is in the same order of the output dataset.
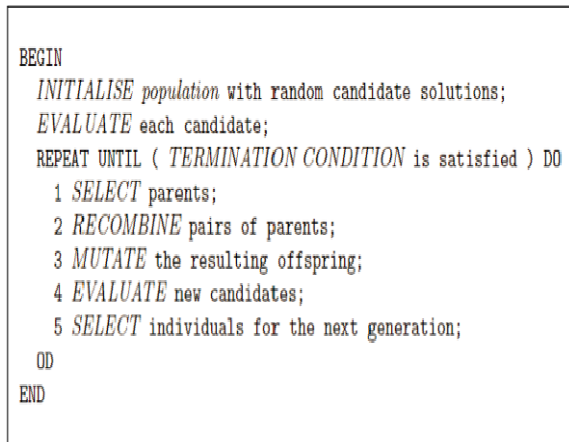
```
BEGIN
   INITIALISE population with random candidate solutions;
   EVALUATE each candidate;
   REPEAT UNTIL ( TERMINATION CONDITION is satisfied ) DO
      1 SELECT parents;
      2 RECOMBINE pairs of parents;
      3 MUTATE the resulting offspring;
      4 EVALUATE new candidates;
      5 SELECT individuals for the next generation;
   OD
END
```

Fig. 8. Cycle of genetic algorithm

Then the genetic algorithm applied its steps on these chromosomes until reach the best solution **Fig. 8**. Here, the termination condition is the number of generation when was mentioned before. Also, the fitness function will be executed in the step of Evaluate each candidate.

5. RESULTS

The model passed on three phases of testing and at each phase, the results were tested and recorded to evaluate the accuracy and efficiency of such model.

These three phases are:

- Phase One: Normal selection with no genetic algorithms
- Phase Two: Applying Genetic Algorithm with Single Crossover
- Phase Three: Applying Genetic Algorithm with Multi Crossover

One of many cases that was used to test the model is taken here as an example.

A 34 years old female and her measurements are:

Blood sugar	=	300,
Blood Pressure	=	150/ 110,
Uric Acid	=	6.7,
Cholesterol	=	210,
Triglyceride	=	65,
Hemoglobin	=	10.9,
Smoking	=	No,
Type of work	=	Mental based,
Pregnant	=	Yes and
Breast Feed	=	No.

The standard recommendation for this case and as shown in **Table 2** was given from nutrition consultant.

Table 2. Standard nutrition recommendation

Recommendation	Consultation
Sugar	Zero gm
Salt	Zero gm
Vitmain B12	2000 mg
Iron	34 mg
Folic	1000 microgram
White Meat	150 gm
Red Meat	50 gm
Carbohydrates	50 gm
Water	4 liters
Exercise	15 min
Vegetarian Proteins	100 gm
Vegetables	300 gm
Fruits	150 gm
Need Insulin	Yes
Fats	Zero
Calcuim	1600 mg

Table 3. Results of normal selection with no GA

Recommendation	Consultation
Sugar	5 gm
Salt	5 gm
Vitmain B12	0 mg
Iron	51 mg
Folic	1000 microgram
White Meat	150 gm
Red Meat	150 gm
Carbohydrates	100 gm
Water	3 liters
Exercise	15 min
Vegetarian Proteins	100 gm
Vegetables	400 gm
Fruits	150 gm
Need Insulin	Yes
Fats	60
Calcuim	1600 mg

Table 4. Results of applying GA with single cross over

Recommendation	Consultation
Sugar	Zero gm
Salt	10 gm
Vitmain B12	2000 mg
Iron	34 mg
Folic	1000 microgram
White Meat	150 gm
Red Meat	150 gm
Carbohydrates	50 gm
Water	5 liters
Exercise	60 min
Vegetarian proteins	100 gm
Vegetables	300 gm
Fruits	100 gm
Need Insulin	Yes
Fats	Zero
Calcuim	1600 mg

Table 5. Results of applying GA with multicrossover

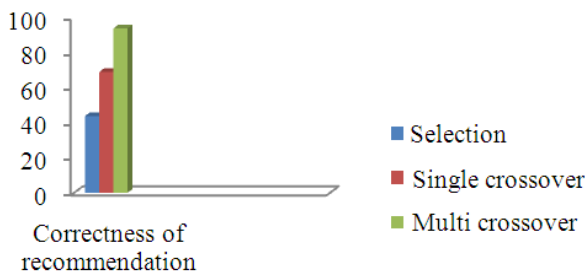| Recommendation	Consultation
Sugar	0 gm
Salt	5 gm
Vitmain B12	2000 mg
Iron	34 mg
Folic	1000 microgram
White Meat	150 gm
Red Meat	50 gm
Carbohydrates	100 gm
Water	4 liters
Exercise	15 minutes
Vegetarian Proteins	100 gm
Vegetables	300 gm
Fruits	150 gm
Need Insulin	Yes
Fats	Zero
Calcium	800 mg

Fig. 9. Performance of the model over three different phases

4.1. Results after Applying Phase One

In **Table 3**, there are 7 recommendations were accurate in comparison with **Table 2.**

4.2. Results after Applying Phase Two

In comparison with **Table 2 and 4** shows that the model reached a higher performance than results which were obtained by phase one. In this phase, 11 recommendations were accurate.

4.3. Results after Applying Phase Three

In comparison with **Table 2 and 5** shows that the model reached a much higher performance than results which were obtained by phase Two. In this phase, 13 recommendations were accurate.

The next chart illustrates the performance of the model over the three different phases in nutrition recommendation **Fig. 9.**

6. CHALLENGES

This model faced some challenges:

- Very few evolutionary algorithms are applied before in nutrition field and genetics was never used before in such field
- No previous experiments using genetics to be used as a reference to measure the correctness of the model which makes us to ask medical and nutrition experts to measure the efficiency of this contribution
- After a long time of search to find a data set that contains all used measurements together, there is no available datasets via internet contain all needed inputs. So, the model input dataset are collected manually from real cases in life from medical laboratories and the recommendation data set is also collected from medical and nutrition specialists
- The need to have a user friendly interface while keeping a good performance made us implement the GA using C++ and then implement an extra layer of the user interface using Java. (using multithreading to allow communication between C++ for GA program and Java for user interface

7. CONCLUSION

Evolutionary algorithms are problem solving techniques that are inspired from the biological changes of living cells. In this paper, evolutionary algorithms are used to suggest a nutrition life style to people, according to their medical history. Different experiments with various input values were run and evaluated. The more genetic- based steps, the higher the performance was. The system used a carefully collected dataset from real cases and evolutionary algorithms are proved to be powerful problem solving techniques. But, there are no models for nutrition used genetics to compare with this model to evaluate which is better. As said before, many factors are affect the nutrition process such as environmental, psychological, social, physical, medical, etc. so, as a future work, we can combine all this factors together and make them in a unified model using genetics to cover all possible factors that affect the person's nutrition process.

8. REFERENCES

Bacardit, J., 2004. Pittsburgh genetic-based machine learning in the data mining era: Representations, generalization and run-time. PhD Thesis, University of Ramon Llull, Barcelona.

Clune, J., B.E. Beckmann, C. Ofria and R.T. Pennock, 2009. Evolving coordinated quadruped gaits with the hyperneat generative encoding. Proceedings of the IEEE Congress on Evolutionary Computing Special Section on Evolutionary Robotics, May 18-21, IEEE Xplore Press, Trondheim, pp: 2764-2771. DOI: 10.1109/CEC.2009.4983289

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. 1st Edn., Addison-Wesley, Reading, ISBN-10: 0201157675, pp: 412.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. 1st Edn., University of Michigan Press, Ann Arbor, ISBN-10: 0472084607, pp: 183.