A MULTI-OBJECTIVE METHOD FOR DISCOVERY OF PATHWAYS IN PROTEIN-PROTEIN INTERACTION NETWORKS

NGUYEN Hoai Anh
Faculty of Information Technology
Le Quy Don Technical University, Ha Noi, Viet Nam
Emails: nguyenhoaianh@yahoo.com

***Abstract*—Biological pathways such as metabolic or signaling ones play an important role in understanding cell activities and evolution. A cost-effective method to discover such pathways is analyzing accumulated information about protein-protein interactions, which are usually given in forms of undirected networks or graphs. Previous findings show that orienting protein interactions can improve pathway discovery. However, assigning orientation for protein interactions is a combinatorial optimization problem which has been proved to be NP-hard, making it critical to develop efficient algorithms.**

In the previous report [1], we solved the problem of finding pathways in the protein-protein interaction networks using genetic algorithms. Our proposed algorithm was also compared with the ones of other authors and proved to be better. However, our algorithm gave out only one goal, which was maximizing the weight function of the interactive network, while ignoring the number of standard pathways in the network. Through experiments, we have realized that weight functions with high value do not always give the high number of standard pathways. As a result, this report introduces multi-objective generic algorithms to deal with two targets at the same time: Maximizing the weight and function, as well as calculating the number of standard pathways in the interaction network.

To extend the previous report, we first studied the mathematical model of the interaction network orientation problem with multi-objective characteristic. Based on such model, we designed a multi-objective genetic algorithm to find the solutions for the problem. Many experimental runs were carried out on the yeasts protein interaction network data set to find the best answer. Results from Multi-Objective Genetic Algorithm (MOGA), when compared with those of our previous proposed GA algorithms (SOGA – Single-Objective Genetic Algorithm), turn out to better solve the problem.

***Keywords*-Multi–objective; genetic algorithms; protein; interaction network;**

I. INTRODUCTION

PPI databases have been the source of interaction information in biological cells. This kind of databases is usually large since data is aggregated over time from the experimental findings. So the discovery of new knowledge from the database has become a challenge for computational biology. Note that edges representing PPIs have been experimentally defined and tested. Certainly reconstructing biological pathways (or biological networks) has attracted a lot of attentions: the reconstruction of regulatory networks [2, 3], the analysis of metabolic networks [4, 5], and the discovery of signaling networks and pathways [6, 7]. However, directionality of interactions in networks has not been investigated thoroughly, while direction is necessary to know how information is moved from one to another. The orientation of the signaling network is more difficult than the regulatory and metabolic networks, due to the lack of orientation information. For example, ChIP-chip and ChIP-Seq used to analyze protein interactions with DNA: identify transcription factors which regulate genes, studies of microRNAs and performed upstream of genes [8, 9, 10]. Metabolic networks are simulated by using knowledge of the order of genes and enzymes [11]. In contrast, it is a fact that PPI data is almost always undirected; therefore the problem of orienting interaction edges for signal transmission in signaling networks is costly. Typical works in this area can be found in [12, 13, 14]. This demonstrates the attraction of finding an efficient algorithm for edge-orientation in PPI networks.

For an overview, authors in [12] stated the problem of orienting edges in the protein interaction network as an optimization problem and proved that this problem is NP hard. Then they presented a random orientation (plus local search) algorithm (ROLS) to perform edge orientation and evaluated calculated results with the data from biological experiments in order to determine if the path found is consistent with the experimental or not. The results were also compared with several algorithms proposed in [15, 16]. In evaluating the algorithm results, the authors found out 37 standard pathways that had been tested through biological experiments. But there were still paths that did not appear in the standard set and such interactions could not occur in biological experiments, even though the objective function values of these pathways were high. Formerly, relatively few methods have been developed to clearly solve the edge orientation problem. In [17], the authors defined the maximum tree orientation (MTO) problem, which focused on reachability.

In the framework of this paper, a different approach is used to solve the problem of edge orienting outlined in [12]; in particular we designed a multi objective genetic algorithm (MOGA) for it. Genetic Algorithm (GA) is one of popular and successful computational models in the field

978-1-4799-5431-5/14/$31.00 ©2014 IEEE

of intelligent computing [18], especially for dealing with NP-hard problems. Along with other intelligent computing techniques such as fuzzy computing, neural networks, multi-agent systems, genetic algorithms develop more and more strongly and are widely applied in different fields of life. The design of our algorithm takes into account conflicting elements in PPI networks in order to reduce unnecessary edges, thus greatly improve computing speed. Results show that our algorithms found a good solution for this problem.

The structure of our paper consists of 6 sections: Section 1 introduces the problem, Section 2 gives general knowledge of the problem and the genetic algorithm, Section 3 introduces mathematical model of the interaction network orientation problem with multi-objective characteristic, Section 4 describes in detail the MOGA designed to solve the problem posed, Section 5 make an assessment of the results achieved by MOGA. The final part is the paper conclusion.

II. BACKGROUND

A. Problem of orienting edges in protein interaction networks PPIs

Proteins are involved in most of biological processes in cells; however, instead of operating independently, they interact with other proteins or macromolecules such as DNA and RNA. They together form a complex network of interactions to perform biological functions. Along with experimental studies, the database of information about protein interactions (PPI) is also formed and developed over time. This database is constantly updated and added with new elements of protein interactions announced by researchers around the world. An example is given in Figure 1 where the graph shows a part of the network of protein interactions in yeast created by *Cytospase* software. From the graph, we can see that the protein interaction network of an organism can be represented by an undirected graph in which each vertex denoted is a protein and each edge represents an interaction of PPIs network. This interactive network contains signaling pathways that comes from a protein source through transformation to transmit biological information to a specific target protein. The signaling pathways verified by experiment are gathered into a database to serve for the interpretation of biological problems. The discovery of the signaling pathway in protein interaction networks are still performed by scientists. The problem here is the need to have a certain method to reconstruct the known signal pathways from the undirected protein interaction networks, then analyse such pathways to make predictions new signaling pathways for purposes of biological studies such as understanding disease signals, creating new drugs to treat diseases caused by the deviation from the signal pathway.

This is a difficult problem because there are many paths that can link two proteins in the interaction network. However, we can establish assumptions to simplify the problem. It is likely that biological responses are controlled by reasonably short signaling cascades, so we should only search for length-bounded paths. So far, pathways in signaling databases such as KEGG and the Science Signaling Database of Cell Signaling

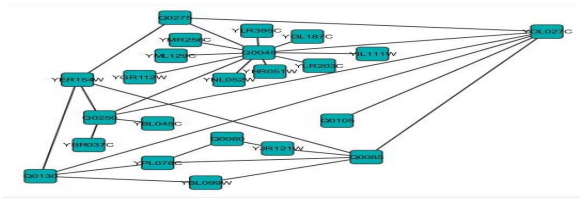


Fig. 1. A part of the protein interaction network of yeast includes 23 proteins and 30 interactions. According to database BIOGRID network 2.0.51 of yeast, its PPIs have 5570 proteins and 140849 interactions [18].

on average contain only five edges between a target and its closest source [12]. The goal of the problem is to extract the signal pathways of length k from source to target that are highy reliable. We can formally express a PPI network by a weighted undirected graph $G = (V, E)$, where V is the set of vertices of the graph labeled by names of proteins, E is the set of edges of the graph describing interactions between proteins. We have $u, v \in V$, edge $e = (u, v) \in E$ if and only if u, v can interact with each other. We define $S \subseteq V$ as the set of source vertices of paths and $T \subseteq V$ as the set of target vertices of paths. A path has a maximum length of at most k between pairs of sources - target in form of $\langle s_i, t_i \rangle$, where $s_i \in S \subseteq V$ and $t_i \in T \subseteq V$. The goal is to orient the edge $e = (u, v) \in E$ from u to v or from v to u so that the weight of the path from source to target with the maximum length k is the largest. Each path has the form $p = (v_1, v_2), (v_2, v_3), , (v_l, v_{l+1})$ and $l \leq k$ for some pairs $\langle s_i, t_i \rangle$. A path is said to be satisfied in the orientation graph if and only if every edge (v_j, v_{j+1}) has its orientation path from v_j to v_{j+1} in the network.

All vertices and edges in the graph have weights which are denoted $w(v)$ and $w(e)$ respectively. Because all vertices $v \in V$ are involved in the signaling pathway (i.e., the graph does not have any isolation peak) so here we assign 1 to every weight of the vertices. Edge weight is a value in the range $[0, 1]$, which is based on the probability of each protein interaction. The value of the weight is typical for reliability in the presence of an edge or the involvement of a protein in the path, and the weight of the path is the probability of a protein interaction in path calculated by the formula

$$w(p) = \prod_{v \in p} w(v) * \prod_{e \in p} w(e) \tag{1}$$

Thus, the goal of the problem is to maximize the total weight of the satisfied paths; or in other words, to optimize the objective function

$$\sum_{p \in P} Is(p) * w(p) \tag{2}$$

Where P is the set of paths between sources and targets with lengths of at most k. $w(p)$ is the path weight. $Is(p)$ is a function of only two values 0 or 1. $Is(p) = 0$ if path p is not satisfied, $Is(p) = 1$ if path p is satisfied.

B. Overview on Genetic Algorithms

GA is one of development tendencies in evolutionary computation. It was researched, developed, and applied since the

last century in search, optimization and machine learning. The exploitation of the evolution principle as a heuristics has made the genetic algorithm an effective approach for the optimization problem (with acceptable solutions) without using traditional conditions (continuous or differentiable) as prerequisites.

One of the important characteristics of GA is the usage of a set (or *population*) of solutions. The search is done parallel on multiple points that can interact with each other according to natural evolution principles. In the context of using genetic algorithms, we can use the concept of "*individual*" in equivalence with the notion of "*solution*". The basic steps of a genetic algorithm are described as follows:

- ***Step 1***: $t = 0$; Initialize $pop(t) = \{x_1, x_2, \ldots, x_N\}$, N is the population size.
- ***Step 2***: Evaluate $pop(t)$.
- ***Step 3***: Create the mating pool $MP = se\{pop(t)\}$ with se is the selection operator.
- ***Step 4***: Define $pop'(t)$=$cr\{MP\}$, with cr is the crossover operator.
- ***Step 5***: Define $pop''(t)$=$mu\{pop'(t)\}$, with mu is mutation operator.
- ***Step 6***: Evaluate $pop''(t)$
- ***Step 7***: Define $pop(t+1)$=$pop''(t)$ and set $t = t + 1$
- ***Step 8***: Go back Step 3, if the stopping criterion is not satisfied.

1) Individual representation: This is one of the important tasks in designing genetic algorithms, deciding the application of evolutionary operators. One of the traditional representations of genetic algorithms is binary representation. With this, each individual in the population is represented as a sequence of bits 0 and 1, also known as chromosomes. Each chromosome represents a parameter of individual components.

2) Selection operator: The selection of individuals can be done when we need a number of individuals to produce the next generation. Each individual has an adaptive value (fitness). This value is used to determine which individual to choose. The selection method used in this paper is tournament selection. This method bases on the fitness function value to choose individuals.

3) Crossover operator: crossover operator is applied to generate new children individuals from parent individuals with the best traits inherited from their parents. In the search context, the crossover operator performs a search around the area of the solution represented by individual parents.

4) Mutation operator: Similar to crossover operator, mutation operator is used to simulate biological mutations. The result of mutations often generates new individuals which are different from their parents. The purpose of mutation operator is to expand searching areas out of local ones.

III. INTRODUCTION OF THE INTERACTION NETWORK ORIENTATION PROBLEM WITH A MULTI-OBJECTIVE CHARACTERISTIC

From testing results of genetic algorithm in [1], we found that not in any case higher objective function value leads to the number of high standard pathway. In other words, higher total weight of satisfied paths did not always result in the number of high standard pathway. In fact, both objectives are important. So, there should be a way to simultaneously satisfy both objectives. In this section, we give a mathematical model of the interaction network orientation problem with multi-objective characteristic.

With the multi-objective model, the interaction network orientation problem also can formally express a PPI network by a weighted undirected graph $G = (V, E)$, and use the fully definition similar to the single-objective models. The first objective of the problem is optimized the objective function (2).

$$f_1 = \sum_{p \in P} Is(p) * w(p) \tag{3}$$

To identify the second objective, we use a directed graph $G' = (V', E')$ to express the gold standard network, where V' is the set of vertices of the graph labeled by names of proteins, E' is the set of edges of the graph describing interactions between proteins. We have $u', v' \in V'$, edge $e' = (u', v') \in E'$ if and only if u' can interact with v', edge $e' = (v', u') \in E'$ if and only if v' can interact with u'. Since then, we have defined standard pathway in the graph G after orientation.

A path p in graph G after orientation is called a standard pathway when p satisfies the following bound: p satisfied orientation of the graph G, length of p is equal to k (where k is the maximum length from source to target), p has at most $k - 2$ consecutive edges in G'.

Thus, the second objective of the problem is to maximize number of standard paths; or in other words, to optimize the objective function

$$f_2 = \sum_{p \in P} Is(p) * standard(p) \tag{4}$$

Where P is the set of paths between sources and targets with lengths of at most k. $Is(p)$ is defined in (2). $standard(p)$ is a function of only two values 0 or 1. $standard(p) = 0$ if path p is not standard path, $standard(p) = 1$ if path p is standard path.

The interaction network orientation problem becomes the optimization of the two objective functions f_1 given by (4) and f_2 given by (5).

IV. METHODOLOGY OF MULTI OBJECTIVE GENETIC ALGORITHM (MOGA)

The idea of designing multi objective genetic algorithms (MOGA) to solve the edge orientation problem starts with a randomly initialized population (population P) of individuals in which the number of individuals of the population is a constant natural number n, each individual is represented by the sequence of the chromosomes. Population will be evolved over many generations, in the evolution of each generation, some individual is picked up randomly and local search is performed for the purpose of increasing the convergence speed. The set of non-dominated individuals of each generation is kept for the next population. After the evolution process completed,

TABLE I
FIVE INDIVIDUALS WITH THE VALUES OF THE TWO OBJECTIVE FUNCTIONS.

Individual	f_1	f_2
A1	7485	56
A2	7996	52
A3	6926	53
A4	8050	36
A5	8033	36

the set of non-dominated individuals in the population will eventually be the selected path of the problem.

Representation of individuals in MOGA is quite similar in [1]. In the following part, we will discuss the design of operators and evaluation of individuals in MOGA.

A. Evaluation of individuals

The individual evaluation involves calculating the fitness value f_1 and f_2. Given the values of f_1, f_2, the problem is to identify in two individuals $A1$ and $A2$, which individual is better? Based on [20] we introduced the concept of better individuals: A individual $A1$ is said to dominate (or better) the other individual $A1$; if both conditions i and ii are true: (i) The individual $A1$ is no worse than $A2$ in both objectives; (ii) The individual $A1$ is strictly better than $A2$ in at least one objective.

Thus, when comparing two individuals, there are three possibilities that can be the outcome of the dominance check between two individuals $A1$ and $A2$. i.e. (i) individuals $A1$ dominates (or better) individuals $A2$, (ii) individuals $A2$ dominates (or better) individuals $A1$, or (iii) individuals $A1$ and $A2$ do not dominate each other.

For greater clarity, we consider an example consisting of 5 individuals with the values of the two objective functions are given in Table I. According to the definition stated, individuals $A4$ dominates (or better) individuals $A5$, individuals $A1$ dominates (or better) individuals $A3$, individuals $A1$ and $A2$ do not dominate each other.

B. The operators

1) Selection operator: with MOGA we need to create a mating pool by the mean of selection. However, the selection of operator in MOGA is carried out randomly with two individuals in the current population.

2) Crossover operator: In MOGA, we also use a two-point crossover operator. The crossover operation calls for two index points to be selected on the parent bit-strings. Everything between the two points is swapped between the parent organisms, rendering two child organisms.

3) Mutation operator: In MOGA, we also use bit inversion: randomly select a bit and change its state to the opposite state.

C. The set of non-dominated individuals

For a given finite set of individuals, we can perform all possible pair comparisons and find which individual dominates which and which individuals are non-dominated with respect to each other. Lastly, we hope to have a set of individuals, any two of which do not dominate each other. This set also has another property. For any individual outside of this set, we can always find an individual in this set which will dominate the former. Thus, this particular set has a property of dominating all other individuals which do not belong to this set. In other words, this means that the individuals of this set are better compared to the rest of individuals. This set is called the non-dominated set for the given set of individuals. In five individuals given in Table V, individual $A1, A2$ and $A4$ constitutes the non-dominated set of the given set of five individuals. Thus, we define a set of non-dominated individuals as follows.

V. CASE STUDY

A. Prepare data

1) Yeast PPIs interaction network: In this paper we use the database of yeast PPIs taken from database BioGRID (`http://thebiogrid.org/download.php`), this is an on-line database of genetic interactions of organisms on a large scale. As mentioned above, this database is updated over time basing on new researches and findings by experiments from biologists. Therefore, for ease of comparison between our results and those of existing algorithms, we use the same database version 2.0.51 BioGrid with the authors [12]. This database is a two-dimensional data table that has 140849 lines; each line contains interactive information of a pair of proteins. We are interested in information about experiment types which are used to detect interactions, because we will combine this information with the table of confidence scores for each type of experiment to determine weights for each interaction edge [19].

The edge weight depends on two factors: First, the reliability of the experiment type; Second, the number of separate experiments that have such interactions. In essence, the edge weight of an edge $(Pro1, Pro2)$ is the probability of interacting protein pairs $Pro1$ and $Pro2$ which is calculated using the formula

$$P(interact(Pro1, Pro2)) = 1 - \prod_{i \in I_{Pro1,Pro2}} (1 - c(i)) \quad (5)$$

where i is a member of the set $I_{Pro1,Pro2}$, which contains all separate interactive experiments from the database of PPIs, and $c(i)$ the reliability of experiment type i.

After determining weights for protein interactions in (3) we get a data sheet of the protein interaction pairs and weights of the interactions. This is the input data of the algorithm.

2) Gold standard pathways: To confirm that the orientations produced by our algorithms not only achieve good objective function value but also produce biologically meaningful results, we compared the PPI network of yeast that it oriented by our algorithm with all yeast signaling pathways from the Science Signaling Database of Cell Signaling. The database focuses all the signal path has been verified by experiment called the gold standard pathway. Collect all the gold standard pathway is called the gold standard network. To evaluate our algorithm, we compared the overlap of the individual pathways

in the set of pathways found by the algorithm with the gold standard network. We see that, the gold standard network is much smaller compared with the complete interaction network, containing 76 proteins and 122 interactions.

3) The source – target protein pairs: The algorithm inputs will use a set $S \subseteq V$ that includes experimentally proven source proteins in a path and a set $T \subseteq V$ that contains proteins where signaling pathway ends. List of source - target pairs of is determined basing on the standard pathway taken from [19].

B. Testing scenario

1) Testing scenario of Single Objective Genetic Algorithms (SOGA): In [1], we have testing scenario of SOGA, follow as: First, we use the Depth First Search algorithm for a set of paths from source to target, then generating a set of conflicted edges. The yeast PPIs database gives us 993 conflicted edges. After that, SOGA is used to find the best orientation setting for conflicted edges. We conducted the test run many times to compare results obtained by SOGA designed by us with the results of the random orientation algorithm plus local search, called ROLS, (Note that in [12], ROLSs performance was shown better than that of the algorithms MIN-SAT, MAX-CSP and MTO). With each loop, we find an outstanding individual and keep it for the next generation.

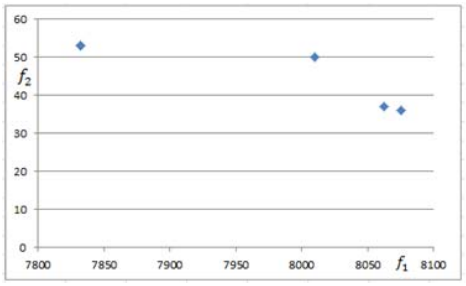
The SOGA test run is planned as follows: Set the initial population of 100 individuals, each individual has n chromosomes (which equals to the total number of conflicted edges in the set of conflicted edges). Input parameters for genetic algorithms include: total generation number of 50, crossover probability of 0.9 and mutation probability of 0.001. To ensure the same experimental conditions, we also run tests 20 times and take the highest value of the objective values (like in [12]), for each run populations and individuals are initialized randomly.

2) Testing scenario of MOGA: Similar, MOGA test run is planned as follows: Set the initial population of 100 individuals, each individual has n chromosomes (which equals to the total number of conflicted edges in the set of conflicted edges). Input parameters for genetic algorithms include: total generation number of 50, crossover probability of 0.9 and mutation probability of 0.001. We run tests 10 times, for each run populations and individuals are initialized randomly. Then, results of MOGA will be compared to SOGA to demonstrate its advantages.

C. Results and analysis

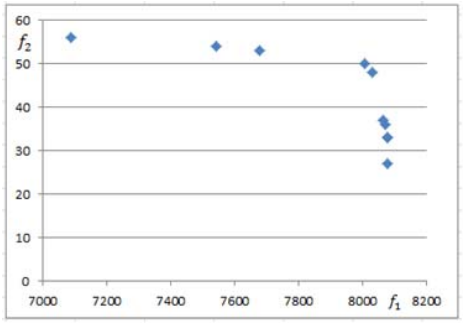
The results obtained in each run is the set of non-dominated individuals. The number of elements in this set is not fixed. Figure 2 to figure 11 describes the elements in the set of non-dominated individuals that has found in 10 times run tests.

From ten non-dominated sets of ten time runs, we find the unique non-dominated set (figure 12). Comparing the objective function of SOGA [1] and MOGA (figure 12), we see that in all cases, MOGA gave out at least one individual better



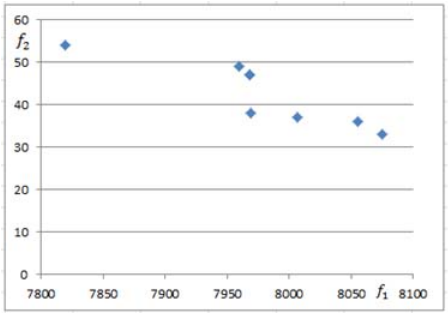
f_1	f_2
7832.373	53
7832.373	53
8009.546	50
8075.230	36
8062.096	37

Fig. 2. The set of non-dominated individuals with seed = 1.



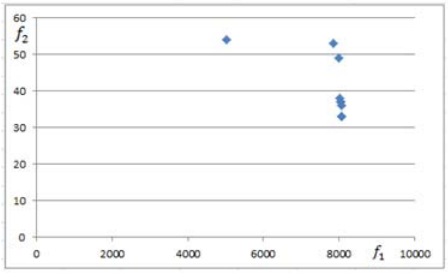
f_1	f_2
7542.315	54
8005.542	50
7676.396	53
7087.653	56
8031.165	48
8075.798	33
8075.798	33
8077.725	27
8070.053	36
8063.494	37

Fig. 3. The set of non-dominated individuals with seed = 2.



f_1	f_2
7819.360	54
7967.814	47
7967.814	47
8006.771	37
7969.266	38
8074.683	33
8055.032	36
7959.451	49

Fig. 4. The set of non-dominated individuals with seed = 3.

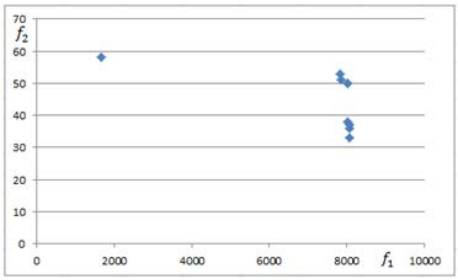


f_1	f_2
5031.326	54
7861.686	53
8010.51	38
8055.639	37
8055.639	37
8077.623	33
8077.623	33
7987.731	49
8071.362	36

Fig. 5. The set of non-dominated individuals with seed = 4.

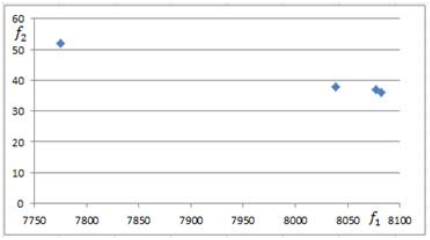
individuals of SOGA in both objectives. Comparative results are given in Table II.

Further, MOGA was able to find up to 59 paths that match the criterion (The best of SOGA is 50). This demonstrates the application of multi-objective algorithm to the interaction network orientation problem is reasonable.



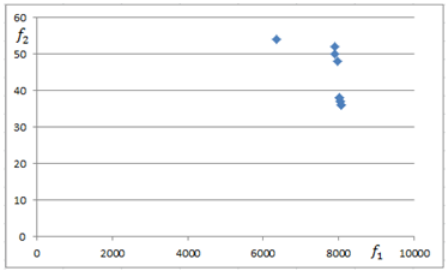
f_1	f_2
1673.485	58
7838.398	53
8012.254	50
7851.33	51
8012.254	50
8069.817	36
8022.994	38
8067.917	37
8074.062	33

Fig. 6. The set of non-dominated individuals with seed = 5.



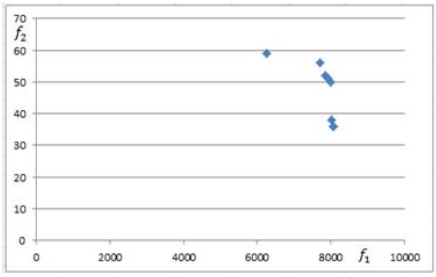
f_1	f_2
7775.127	52
8038.840	38
8076.883	37
8082.172	36

Fig. 7. The set of non-dominated individuals with seed = 6.



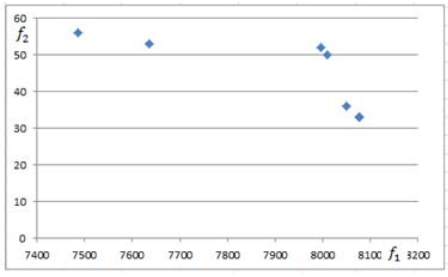
f_1	f_2
7905.958	50
7962.282	48
6346.404	54
7900.648	52
8017.423	38
8071.398	36
8071.398	36
8052.379	37
8052.117	37
8017.042	38

Fig. 8. The set of non-dominated individuals with seed = 7.



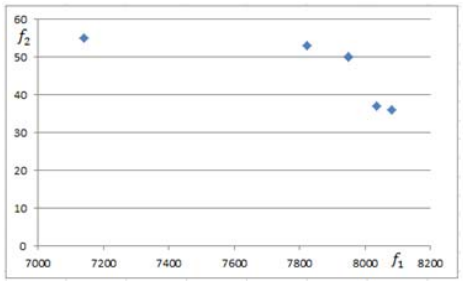
f_1	f_2
7839.902	52
8006.949	50
7702.833	56
7912.047	51
6261.101	59
8023.963	38
8077.728	36
8077.728	36

Fig. 9. The set of non-dominated individuals with seed = 8.



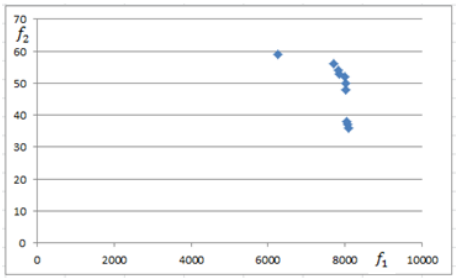
f_1	f_2
7485.607	56
7996.813	52
8009.032	50
7636.397	53
8050.337	36
8076.984	33
8076.984	33
8076.265	33

Fig. 10. The set of non-dominated individuals with seed = 9.



f_1	f_2
7821.860	53
7949.186	50
7949.186	50
7142.005	55
8081.267	36
8034.378	37

Fig. 11. The set of non-dominated individuals with seed = 10.



f_1	f_2
8081.267	36
8076.883	37
8038.84	38
8031.165	48
8012.254	50
7996.813	52
7861.686	53
7819.36	54
7702.833	56
6261.101	59

Fig. 12. The set of unique non-dominated individuals

TABLE II
COMPARING THE OBJECTIVE FUNCTION (OBF) AND NUMBER OF STANDAR PATHS (NOSP) OF SOGA AND MOGA

No	OBF of SOGA	OBF of MOGA	NoSP of SOGA	NoSP of MOGA
1	8027	8081	35	36
2	8022	8076	29	37
3	8016	8038	29	38
4	7985	8031	13	48
5	7983	8012	37	50
6	7982	7996	38	52
7	7982	7861	35	53
8	7968	7819	35	54
9	7956	7702	39	56
10	7933	6261	35	59
MEAN	**7985.4**	**7787.7**	**32.5**	**48.3**

VI. CONCLUSION

In this paper, we propose the multi objective genetic algorithm (MOGA) design for problem of orienting protein interaction network. This is a challenging problem for computational biology. We present a method to perform populations individuals that fit the problem, especially our designs take into account conflicting elements for solution representation, thus greatly improve computing speed. Results show that our algorithm properly settles this problem. As evidence of the correctness of our algorithm, we find that our algorithms has reconstructed many known signaling pathways, which is significant in biological research. In the future, we seek to improve the efficiency of the algorithm.

REFERENCES

[1] Nguyen Hoai Anh, Vu Cong Long, Tu Minh Phuong and Bui Thu Lam, *A Genetic-based approach for discovering pathways in protein-protein interaction networks*. InProceedings of SoCPaR2013.

[2] E. Segal, M. Shapira, A. Regev, D. Peer, D. Botstein, D. Koller, and N. Friedman, *Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data*. Nat. Genet. 2003.

[3] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano, *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context*. BMC Bioinformatics. 2006.

[4] Junji Kitagawa, and Hitoshi Iba, *Identifying Metabolic Pathways and Gene Regulation Networks with Evolutionary Algorithms*. Chapter 12. Evolution Computation in Bioinformatic. 2003.

[5] E. Fischer, and U.Sauer, *Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism*. Nat. Genet. 2005.

[6] J. Scott, T. Ideker, R.M. Karp, and R. Sharan, *Efficient algorithms for detecting signaling pathways in protein interaction networks*. J. Comput. Biol. 2006.

[7] G. Bebek, and J. Yang, *PathFinder: mining signal transduction pathway segments from protein-protein interaction networks*. BMC Bioinformatics. 2007.

[8] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. Kim, and R.P. Koche, *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells*. Nature. 2007.

[9] B.P. Lewis, C.B. Burge, , and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets*. Cell. 2005.

[10] X. Xie, J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, , and M. Kellis, *Systematic discovery of regulatory motifs in human promoters and 3 UTRs by comparison of several mammals*. Nature. 2005.

[11] S.J. Cox, S.S. Levanon, G.N. Bennett, and K. San, *Genetically constrained metabolic flux analysis*. Metab. Eng. 2005.

[12] Anthony Gitter, Judith Klein-Seetharaman, Anupam Gupta, and Ziv Bar-Joseph, *Discovering Pathways by Orienting Edges in Protein Interaction Networks*. Nucleic Acids Research, Vol. 39, No. 4. 2011.

[13] Jinghua Gu, Bradley, Jianhua Xuan, Chen Wang, and Li Chen, *Detecting aberrant signal transduction pathways from high-throughput data using GIST algorithm*. Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012.

[14] Dima Blokh, Danny Segev, and Roded Sharan, *Approximation Algorithms and Hardness Results for Shortest Path Based Graph Orientations*. Springer Berlin Heidelberg, 2012.

[15] R. Kohli, R. Krishnamurti, and P. Mirchandani, *The minimum satisfiability problem*. SIAM J. Discret. Math, 1994.

[16] M. Charikar, K. Makarychev, and Y. Makarychev, *Near-optimal algorithms for maximum constraint satisfaction problems*. ACM Trans. Alg, 2009.

[17] A. Medvedovsky, V. Bafna, U. Zwick, and R. Sharan, *An algorithm for orienting graphs based on cause-effect pairs and its applications to orienting protein networks*. InProceedings of the 8th international workshop on Algorithms in Bioinformatics, Karlsruhe, Germany, 2008.

[18] T. Back, *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, 1996.

[19] Anthony Gitter, Judith Klein-Seetharaman, Anupam Gupta, and Ziv Bar-Joseph, *Supporting Information, Discovering Pathways by Orienting Edges in Protein Interaction Networks, .* download from http://sb.cs.cmu.edu/OrientEdges/. [Accessed 10/8/2013]

[20] Kaylyanmoy Deb, *Multi - Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons, Ltd, 2001.