An Improved Ant Colony Optimization Algorithm for the Detection of SNP-SNP Interactions

Yingxia Sun[1], Junliang Shang[1,2(✉)], JinXing Liu[1], and Shengjun Li[1]

[1] School of Information Science and Engineering,
Qufu Normal University, Rizhao 276826, China
`{sunyingxia026,shangjunliangl10,qfnulsj}@163.com,`
`sdcavell@126.com`
[2] Institute of Network Computing, Qufu Normal University,
Rizhao 276826, China

Abstract. An increasing number of studies have found that one of the most important factors for emergence and development of complex diseases is the interactions between SNPs, that is to say, epistasis or epistatic interactions. Though many efforts have been made for the detection of SNP-SNP interactions, the algorithm of such studies is still ongoing due to the computational and statistical complexities. In this work, we proposed an algorithm IACO based on ant colony optimization and a novel introduced fitness function *Svalue*, which combined both Bayesian networks and mutual information, for detecting SNP-SNP interactions. Furthermore, a memory based strategy is also employed to improve the performance of IACO, which effectively avoids ignoring the optimal solutions that have already been identified. Experiments of IACO are performed on both simulation data sets and a real data set of age-related macular degeneration (AMD). Results show that IACO is promising in detecting SNP-SNP interactions, and might be an alternative to existing methods for inferring epistatic interactions. The software package is available online at http://www.bdmb-web.cn/index.php?m=content&c=index&a=show&catid=37&id=98.

Keywords: SNP-SNP interaction · Bayesian network · Mutual information · Ant colony · Optimization

1 Introduction

With the development of high-throughput sequencing technologies, it is universally acknowledged that SNP (single nucleotide polymorphism) is one of the most common forms of genetic variants in human genome, which usually affects complex diseases by their nonlinear interactions, namely, epistatic interactions or epistasis [1]. Currently, epistasis has caused the extensive concerns in exploring the pathogenic mechanism of non-Mendelian diseases, such as hypertension, diabetes, Alzheimer's disease and many others [2]. Although many efforts have been made for the detection of SNP-SNP interactions, the algorithm of such studies is still ongoing due to their computational

and statistical complexities, including the complexity of pathogenesis, the complexity of genetic models and the influence of environment factors.

Recently, a number of generic ant colony optimization (ACO) based methods have been proposed [1, 3–8] to detect the epistatic interactions. For instance, Christmas *et al.* [6] used generic ACO algorithm to identify the epistatic interactions in type 2 diabetes data. Results indicate that ACO algorithm is able to find statistically significant epistatic interactions. Wang *et al.* [4] proposed AntEpiSeeker based on ACO algorithm and designed two-stage optimization procedure for detection SNP-SNP interactions. Though it enhances the power of ACO algorithms, it is sensitive to main effect SNPs and it requires large amounts of ants over numerous iterations to obtain acceptable results. Shang *et al.* [1] proposed AntMiner based on ACO algorithm and incorporated heuristic information into ant colony optimization for epistasis detection, although it contributes to improving the computational efficiency and solution accuracy, it is also very time-costing. Jing *et al.* [5] proposed MACOED on the basis of ACO algorithm and designed a multi-objective method to detect SNP epistasis, which may be criticized for the complexity of multi-objective computation and the randomness of its searching strategy.

In this study, we proposed a method IACO based on ant colony optimization and a novel introduced fitness function *Svalue*, which combined both Bayesian networks and mutual information for detecting SNP-SNP interactions, effectively and efficiently evaluates how well the SNP-SNP combinations associates with the phenotype. In addition, a memory based strategy is also employed to improve the performance of IACO, which effectively avoids ignoring the optimal solutions that have already been identified and speed up the convergence of IACO algorithm. Experiments of IACO are performed on simulation data sets and a real data set of age-related macular degeneration (AMD). And we also compared IACO with some other representative methods, including AntEpiSeeker [4], AntMiner [1], and MACOED [5]. Results demonstrate that IACO outperforms others in inferring SNP-SNP interactions. Besides, the application of IACO on a real data set of age-related macular degeneration (AMD) may provide some new clues for the detection of AMD. IACO might be an alternative to existing methods for inferring epistatic interactions.

2 Methods

2.1 Mathematical Description of the Problem

In genome-wide association study (GWAS) [9], SNPs are bi-allelic labels with major and minor alleles being denoted by capital letters (e.g., A) and lowercase letters (e.g., a) respectively. There are three genotypes for each SNP: homozygous common genotype (AA), heterozygous genotype (Aa), and homozygous minor genotype (aa) [1]. The most common way of mapping SNPs is to collect them as a matrix, where a column represents a SNP and a row represents the genotypes of a sample. Furthermore, the elements 1, 2, 3 in the matrix on behalf of AA, Aa and aa respectively. In addition, we generally called the case group and the control group of each individual as a sample and used a matrix to collect them. Rows of the matrix are equivalent to the rows of SNP

matrix. The label of a sample is either 1 (case) or 2 (control) [10]. In mathematics, Genome-wide SNP interaction research is described as a study of multiple SNP combinations for high-dimensional and small sample size data, which is used to greatly predict the phenotype of the sample.

2.2   Traditional Ant Colony Optimization

Swarm intelligence algorithm [1, 3–8, 10–12] is an optimization algorithm to solve the problems of large-scale data, which can handle some discrete data optimization problems through simulation of animal behavior. As a new swarm intelligence algorithm, ACO takes inspiration from the foraging behavior of some ant species [1], which has already been used to solve the problems of traveling salesman, graph coloring and others.

In the progress of foraging, ants can communicate with each other by secreting pheromones on the ground, but pheromones will gradually evaporate as time passes. The subsequent ants choose the path according to pheromones and tend to choose the paths with higher pheromones [1]. With the passage of time almost all ants choose a nearest path from a food source to their nest.

In this study, m ants are represented by $\{m_1, m_2 \cdots m_m\}$ respectively, each of which is used to detect a K-SNP interaction at each iteration, where K is a user-specified order of epistatic interactions.

The ants choose the next position by probability function (PF). PF is a probability that ant k choose the next position j from the position i at iteration t, and is denoted as $p_k^{ij}(t)$. The PF of traditional ant colony optimization is defined as

$$P_k^{ij}(t) = \begin{cases} \dfrac{\tau_{ij}(t)^{\alpha}\eta_{ij}^{\beta}}{\sum_{u\in U_k(t)}\tau_{iu}(t)^{\alpha}\eta_{iu}^{\beta}} & i \in U_k(t) \\ 0 & otherwise \end{cases} \tag{1}$$

where $\tau_{ij}(t)$ is the pheromones of position i to position j at iteration t. η_{ij} is the heuristic information of position i to position j. α and β are controlling importance of pheromones and heuristic information respectively. $U_k(t)$ is a set of positions that are not selected by ant k at iteration t.

Pheromones will evaporate in the process of movement. Pheromones of position i to position j at iteration $t + 1$ are updated according to

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \tag{2}$$

where ρ is a user-defined evaporation coefficient and $\Delta\tau_{ij}(t)$ is an increment of pheromones from position i to position j at iteration t. The $\Delta\tau_{ij}(t)$ is defined as

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t) \tag{3}$$

where m is the number of ants and $\Delta\tau_{ij}^{k}(t)$ is the legacy of pheromones from position i to position j for ant k at iteration t, which is defined as

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \dfrac{Q}{S_k(t)} & f\ ant\ k\ via\ the\ path\ of\ ij\ at\ iteration\ t \\ 0 & otherwise \end{cases} \tag{4}$$

where Q is a user-defined positive constant and $S_k(t)$ is the path length for ant k at iteration t.

2.3 Evaluation Measure Svalue

2.3.1 Mutual Information

Information entropy is used to measure the characteristic of uncertainty, higher uncertainty implies the higher information entropy. Mutual information [12, 13] is a measure based on information entropy, which can measure the dependence of SNP-SNP combinations and phenotype. The IACO employs mutual information to measure how much the relevance of SNP-SNP interactions to the phenotype. The formula of mutual information can be written as

$$MI(S; Y) = H(S) + H(Y) - H(S, Y) \tag{5}$$

where $H(S)$ is the entropy of S, $H(Y)$ is the entropy of Y, and $H(S,Y)$ is the joint entropy of both S and Y, S is a position of a particle indicating a SNP combination, Y is the phenotype.

The entropy and the joint entropy are defined as

$$H(S) = -\sum_{j_1=1}^{3} \cdots \sum_{j_K=1}^{3} \left(p\left(s_{j_1}, \cdots, s_{j_K}\right) \cdot \log p\left(s_{j_1}, \cdots, s_{j_K}\right) \right) \tag{6}$$

$$H(Y) = -\sum_{j=0}^{1} \left(p\left(y_j\right) \cdot \log p\left(y_j\right) \right) \tag{7}$$

$$H(S, Y) = -\sum_{j_1=1}^{3} \cdots \sum_{j_K=1}^{3} \sum_{j=0}^{1} \left(p\left(s_{j_1}, \cdots, s_{j_K}, y_j\right) \cdot \log p\left(s_{j_1}, \cdots, s_{j_K}, y_j\right) \right) \tag{8}$$

where S is the genotype of a SNP coded as {1, 2, 3}, corresponding to homozygous common genotype, heterozygous genotype, and homozygous minor genotype, y is the label of a sample coded as {0, 1}, corresponding to control and case. $p(\cdot)$ is the probability distribution function.

Apparently, the higher the mutual information score is, the stronger the association between the SNP subset and the disease.

2.3.2   Bayesian Network

In recent years, with the rapidly development of machine learning, there are a lot of researches devoting to some statistical learning methods. For instance, Bayesian network [14, 15]. Many measures in Bayesian networks are used to evaluate the correlation between SNP-SNP combinations and disease. In this work, we adopt *BN* to measure the correlation between disease and different SNP combinations. Specifically, in the GWAS *BN*, the genotypes and phenotypes are denoted as a set of nodes, and their conditional dependences are denoted as a set of edges. On the basis of previous studies [15] we choose the *BN* score as follows:

$$BN = \sum_{i=1}^{I} \left(\sum_{b=1}^{r_i+1} \log(b) - \sum_{j=1}^{J} \sum_{d=1}^{r_{ij}} \log(d) \right) \tag{9}$$

where *I* is the combinatorial number of SNP nodes with different values (if *l*-SNP nodes are connected to disease node *y*, the number of SNP nodes' combinations is 3^l as the possible value of a SNP node is 0, 1, or 2), *J* is the state number of disease node *y* two for all samples), r_i is the number of cases with SNP nodes taking the i_{th} combination, and r_{ij} is the number of cases where the disease node takes the j_{th} state and its parents take the i_{th} combination [5]. In this measure, lower BN value implies greater relevance of SNP-SNP combinations and phenotype.

2.3.3   Svalue

In previous studies, most of studies used single objective evaluation method to evaluate the correlation between SNP-SNP interactions and phenotype. However, many evaluation methods have different effects on the disease in different model due to the complexity of some disease. In this work, we adopt a new evaluation measure *Svalue*, which is based on Bayesian network and mutual information. As mentioned before, the higher the mutual information score and the lower the BN score, the stronger the association between the SNP subsets and the disease. But we can't evaluate the pros and cons of different SNP-SNP combinations due to the different assessment of two evaluation measures and the different order of magnitude for two evaluation functions. So, we proposed a new evaluation measure as follows:

$$Svalue(A) = \theta \frac{MI}{BN} \tag{10}$$

where *MI* and *BN* are the mutual information score and *BN* score respectively. θ is a user-specified constant. The ability to identify between SNP-SNP combinations and phenotype will be better when the *Svalue* score is larger.

2.4   Improved Ant Colony Optimization (IACO)

2.4.1   Path Selection Strategy

IACO algorithm is based on the generic ant colony algorithm (ACO). And in this work, we use a matrix to store the information of pheromones, which number of columns is

all SNPs data sets. In order to treat each SNP equally, the pheromones of all SNPs is set to 1 in the initial time of the algorithm. Then ants can independently choose the SNPs on the basis of pheromones and heuristic information. The ant-decision rule is ant k select SNP i at iteration t, and it is denoted as $P_k^i(t)$. The PF of IACO is defined as

$$P_k^i(t) = \begin{cases} R & q \le q_0 \\ S & q > q_0 \end{cases} \tag{11}$$

where q is a number generated randomly with uniform distribution in (0,1), and q_0 is a user-defined threshold to control the rate of convergence and to avoid falling into locally optimal solution. R and S are the rules for ants to select SNPs, described as

$$S = \begin{cases} 1 & i = rand\,(U_k(t)) \\ 0 & otherwise \end{cases} \tag{12}$$

$$R = \begin{cases} \dfrac{\tau_i(t)^\alpha \eta_i^\beta}{\sum\limits_{u \in U_k(t)} \tau_u(t)^\alpha \eta_u^\beta} & i \in U_k(t) \\ 0 & otherwise \end{cases} \tag{13}$$

where $\tau_i(t)$ is the pheromones of SNP i at iteration t, and η_i is the heuristic information of SNP i.

2.4.2 Pheromone Updating Strategy

Each ant can find a K-order SNP-SNP combination at the completion of each iteration, but the connection degree of different SNP-SNP combinations and disease is distinct. Therefore we adopt a new pheromone update function to change the pheromones of different SNP-SNP combinations respectively, and described as

$$\tau_i(t+1) = (1 - \rho)\tau_i(t) + \Delta\tau_i(t) \tag{14}$$

where $\Delta\tau_i(t)$ is the variation of pheromones, which is described as

$$\Delta\tau_i(t) = \sum_{k=1}^{m} \Delta\tau_i^k(t) \tag{15}$$

$$\Delta\tau_i^k(t) = \begin{cases} Svalue(A) & k \in M_i(t) \\ 0 & otherwise \end{cases} \tag{16}$$

In those equations, $M_i(t)$ is a set of ants who select SNP i at iteration t, A is the SNP-SNP combinations which are identified by ant k at iteration i, and $Svalue(A)$ is the $Svalue$ of A.

2.4.3 Memory Based Strategy

In this work, we rank all SNP-SNP combinations that identified by ants at iteration t with descending *Svalue*, and the combinations are described as $A_1, A_2, \cdots A_m$, where their *Svalues* are $(1, Svalue(A_1)), (2, Svalue(A_2)), \cdots (m, Svalue(A_m))$. Then we should find an inflection point f and hold the points before the inflection point as the optimal solutions of this iteration, which is defined as

$$f = arc \max_{g=3}^{m} \left(\left(Svalue(A_g) - Svalue(A_{g-1}) \right) - \left(Svalue(A_{g-1}) - Svalue(A_{g-2}) \right) \right) \quad (17)$$

In order to save the optimal solutions that have already been identified, speed up the convergence and reduce the calculation of *Svalue* for some optimal solutions that have already been calculated, we introduce a memory strategy, which put the optimal solutions to the sets of suspected solutions. Otherwise, we design the memory strategy to remain the optimal solutions to compare with the identified solutions in the current iteration and get final suspected solutions. All suspected solutions in the last iteration will be fed into the next stage of IACO to process [5].

3 Results and Discussion

3.1 Simulation and Real Data Sets

We exemplify 4 benchmark models of SNP-SNP interactions for the experiments [14, 16–19], parameters are set as follows, penetrance is the probability of the occurrence of a disease given a particular genotype, and penetrance functions of four models is reported in Fig. 1; MAF(α)is the minor allele frequency of α. *AA*, *Aa* and *aa* are homozygous common genotype, heterozygous genotype, and homozygous minor genotype; and prevalence is the proportion of samples that occur a disease. Specifically, Model 1 is a model that display both marginal effects and interactive effect, the penetrance of which increases only when both SNPs have at least one minor allele [14, 16], the MAF(a), MAF(b) and the prevalence of Model 1 are 0.300, 0.200 and 0.100 respectively; Model 2 is a model showing both marginal effects and interactive effect, the additional minor allele at each locus of which does not further increase the penetrance [14], the MAF(a), MAF(b) and the prevalence of Model 2 are 0.400, 0.400 and 0.050 respectively; Model 3 is also a model displaying both marginal effects and interactive effect, which assumes that the minor allele in one SNP has the marginal effect, however, the marginal effect is inversed while minor alleles in both SNPs are present [14], the MAF(a), MAF(b) and the prevalence of Model 3 are 0.400, 0.200 and 0.010 respectively; Model 4 is a model that shows only interactive effects, which is directly cited from the reference [18]. Model 4 is exemplified here since it provides a high degree of complexity to challenge ability of a method in detecting SNP-SNP interactions, the MAF(a), MAF(b) and the prevalence of Model 4 are 0.400, 0.400 and 0.171 respectively. For each model, 50 data sets are simulated by *epi*SIM [11], each containing 4000 samples with the ratio of cases and controls being 1. For each data set, random SNPs are set with their MAFs chosen from $[0.05, 0.5]$ uniformly.
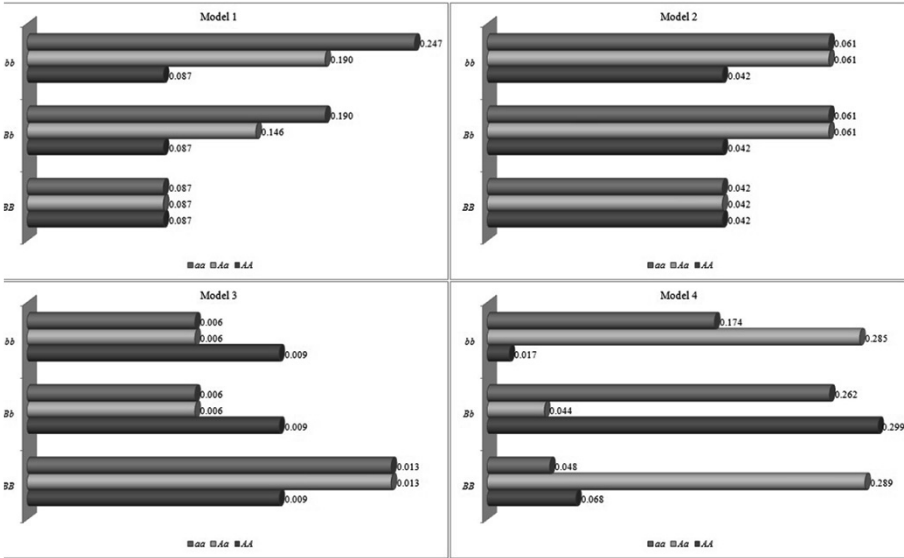
Fig. 1. Penetrance functions of four models

A real data set of age-related macular degeneration (AMD) is used for testing the practical ability of IACO. AMD, refers to pathological changes in the central area of the retina, is the most important cause of irreversible visual loss in elderly populations, and is considered as a complex disease having multiple SNP-SNP interactions [16]. The AMD data set contains 103611 SNPs genotyped with 96 cases and 50 controls, which has been widely used as a benchmark data set in the field of testing methods of detecting SNP-SNP interactions [1, 10, 12–14, 16, 19, 20].

3.2 Experiments on Simulation Data Sets

Three methods for detection SNP-SNP interactions are compared with IACO, which are AntEpiSeeker, AntMiner and MACOED. In order to equitable comparison, we set the same parameters for four methods. The number of ants m and the number of iterations T are set to 500 and 10 respectively; the dimension K for IACO is set to 2; the heuristic information η is set to 1; the controlling importance of pheromones α and heuristic information β are both set to 1; the threshold for control the rate of convergence q_0 is set to 0.6; the evaporation coefficient ρ is set to 0.2; the constant θ is set to 10000.

In order to evaluate the performance of the algorithm, we introduced power as evaluate criteria, which is defined as the proportion of 50 data sets that all connected SNPs are recognized. Power of these four compared methods on simulation data sets is reported in Fig. 2. Apparently, power of IACO is much higher than that of AntE-piSeeker; power of IACO is much higher than that of MACOED except the Model 4; power of IACO is much higher than that of AntMiner except the Model1 and Model2,
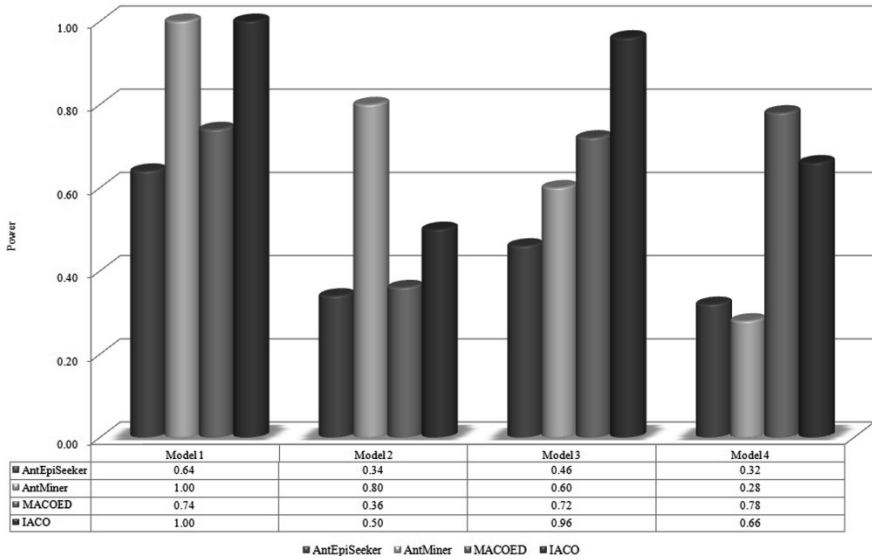
	Model 1	Model 2	Model 3	Model 4
■ AntEpiSeeker	0.64	0.34	0.46	0.32
■ AntMiner	1.00	0.80	0.60	0.28
■ MACOED	0.74	0.36	0.72	0.78
■ IACO	1.00	0.50	0.96	0.66

■ AntEpiSeeker　■ AntMiner　■ MACOED　■ IACO

Fig. 2. Power of compared methods on simulation data sets.

on account of AntMiner incorporate heuristic information into ACO for SNP-SNP interaction, but the running time of AntMiner is significantly higher than IACO; the performance of AntEpiSeeker is worse on all models because it focus on identify the models of pathogenic SNPs; the MACOED introduce logistic regression to evaluate SNP-SNP combinations, which is unstable for different parameter settings, and the most major problem of it is the searching strategy becomes stochastic search. Specifically, the computer time of AntEpiSeeker, MACOED and IACO is on an order of magnitude, immensely lower than AntMiner. With comprehensive consideration, IACO outperforms others in detection power for large datasets and it has superb stability. In the future research, we will consider introducing heuristic information into ACO to improve the power of our methods.

3.3 Application to Real Data Set

The IACO is applied on AMD data set four times with parameter settings (m, T) being $(10000, 500)$, $(10000, 1000)$, $(20000, 500)$ and $(20000, 1000)$. In total, there are captured SNP-SNP interactions that might be associated with the AMD are reported in Table 1 with descending *Svalue*.

Obviously, all reported SNP-SNP interactions contains either rs380390 or rs1329428 except the fourth line SNPs, of which reason is that these two SNPs have strongest main effects among all tested SNPs, and these two SNPs have already proved to be significantly associated with AMD [4]. All of them in the gene of *CFH*, which is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies [21].

Table 1. Top 11 captured SNP-SNP interactions associated with AMD. *CFH*: complement factor H. *CRSP8*: cofactor required for Sp1 transcriptional. *JMJD2C*: jumonji domain containing 2C. *ISCA1*: HESB like domain containing 2. *N/A*: no gene is available. Chr: Chromosome.

SNP 1				SNP 2				*Svalue*	Times
Index	Name	Gene	Chr	Index	Name	Gene	Chr		
43748	rs380390	*CFH*	1	80178	rs1363688	*N/A*	5	39.1602	4
43748	rs380390	*CFH*	1	57476	rs2402053	*N/A*	7	37.7208	4
43748	rs380390	*CFH*	1	63879	rs1374431	*N/A*	2	37.1915	3
13485	rs1368863	*N/A*	11	19405	rs1740752	*N/A*	10	37.1603	3
54108	rs1329428	*CFH*	1	79546	rs7467596	*CRSP8*	9	36.8086	1
54108	rs1329428	*CFH*	1	31604	rs9328536	*CRSP8*	9	36.8086	1
43748	rs380390	*CFH*	1	97535	rs2224762	*JMJD2C*	9	36.3767	1
43748	rs380390	*CFH*	1	75884	rs2380684	*N/A*	2	35.9957	2
43748	rs380390	*CFH*	1	10645	rs10512174	*ISCA1*	9	35.0884	1
43748	rs380390	*CFH*	1	86692	rs10488386	*N/A*	7	33.6912	3
43748	rs380390	*CFH*	1	19405	rs1740752	*N/A*	10	33.5013	1

Besides, the SNPs in the second column of the table might be identified due to they combine with rs380390 or rs1329428, since strong main effects of these two SNPs leads to their combinations with other SNPs almost displaying strong interactive effect. This also indicates that IACO is sensitive to those SNPs displaying strong main effects. It is worth noting that our methods find that the combinations of rs1368863 and rs1740752 also displaying strong interactive effect. These SNPs in the N/A gene, which are located on chromosome 11 and 10 in noncoding region, may induced AMD by gene mutation or induced diseases by changing the other disease-causing genes. The above observation may provide some new clues for the detection of AMD.

4   Conclusions

Epistasis has caused the extensive concern in exploring the mechanism underlying susceptibility to complex diseases. In this work, we proposed an algorithm IACO based on ant colony optimization and a novel introduced fitness function *Svalue*, which combined both Bayesian networks and mutual information, for detecting SNP-SNP interactions. Furthermore, a memory based strategy is also employed to improve the performance of IACO, which effectively avoids ignoring the optimal solutions that have already been identified. Experiments of IACO are performed on simulation data sets and a real data set of age-related macular degeneration (AMD). And we also compared IACO with some other representative methods, including AntEpiSeeker [4], AntMiner [1], and MACOED [5]. Results demonstrate that IACO outperforms others in detecting SNP-SNP interactions. Besides, the application of IACO on a real data set of age-related macular degeneration (AMD) may provide some new clues for the detection of AMD. The Matlab version of IACO software package is available online at http://www.bdmb-web.cn/index.php?m=content&c=index&a=show&catid=37&id=98.

Obviously, there are many advantages for IACO to identify epistatic interactions. First of all, IACO proposed a new fitness function *Svalue* based on Bayesian networks and mutual information, which unifies the different assessment of two evaluation measures, greatly improve the processing ability of the optimal solutions. In the next place, it adopted a memory based strategy to dispose the optimal solution, which rank *Svalue* of the suspected solutions with the *Svalue* of the SNP-SNP combinations identified at current iteration. One merit of memory based strategy is the optimal solutions in last iteration is used for the current iteration, which avoid ignoring the optimal solution that have already been found, speed up the convergence and reduce the calculation of *Svalue* for some optimal solutions that have already been calculated. What's more, we use a dynamic pheromone updating strategy to update the pheromones of SNPs. SNPs of bigger associated with disease will update more pheromones than that poor relevance of disease, which effectively improve the convergence speed of the algorithm.

Although IACO showed great performance for detection of epistatic interactions on both simulation and real data sets, there still remain some limitations. For instance, we set the heuristic information η to 1, which may degrade the performance of IACO. In future studies, we will consider introducing the heuristic information to IACO to deal with GWAS dataset. IACO is a continuous research project and we will continue improving in the future.

References

1. Shang, J., Zhang, J., Lei, X., Zhang, Y., Chen, B.: Incorporating heuristic information into ant colony optimization for epistasis detection. Genes Genomics **34**(3), 321–327 (2012)
2. Maher, B.: The case of the missing heritability. Nature **456**(7218), 18–21 (2008)
3. Rekaya, R., Robbins, K.: Ant colony algorithm for analysis of gene interaction in high-dimensional association data. Revista Brasileira de Zootecnia **38**(SPE), 93–97 (2009)
4. Wang, Y., Liu, X., Robbins, K., Rekaya, R.: AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm. BMC Res. Notes **3**(1), 117 (2010)
5. Jing, P., Shen, H.: MACOED: a multi-objective ant colony optimization algorithm for SNP epistasis detection in genome-wide association studies. Bioinformatics **31**, 634–641 (2014). btu702
6. Christmas, J., Keedwell, E., Frayling, T.M., Perry, J.R.: Ant colony optimisation to identify genetic variant association with type 2 diabetes. Inf. Sci. **181**(9), 1609–1622 (2011)

7. Greene, C.S., White, B.C., Moore, J.H.: Ant colony optimization for genome-wide genetic analysis. In: Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T., Winfield, A.F. (eds.) ANTS 2008. LNCS, vol. 5217, pp. 37–47. Springer, Heidelberg (2008)

8. Greene, C.S., Gilmore, J.M., Kiralis, J., Andrews, P.C., Moore, J.H.: Optimal use of expert knowledge in ant colony optimization for the analysis of epistasis in human disease. In: Pizzuti, C., Ritchie, M.D., Giacobini, M. (eds.) EvoBIO 2009. LNCS, vol. 5483, pp. 92–103. Springer, Heidelberg (2009)

9. Li, P., Guo, M., Wang, C., Liu, X., Zou, Q.: An overview of SNP interactions in genome-wide association studies. Briefings Funct. Genomics **14**, 143–145 (2014). elu036

10. Shang, J., Zhang, J., Sun, Y., Zhang, Y.: EpiMiner: a three-stage co-information based method for detecting and visualizing epistatic interactions. Digit. Sig. Process. **24**, 1–13 (2014)

11. Shang, J., Zhang, J., Lei, X., Zhao, W., Dong, Y.: EpiSIM: simulation of multiple epistasis, linkage disequilibrium patterns and haplotype blocks for genome-wide interaction analysis. Genes Genom. **35**, 1–12 (2013)

12. Ma, C., Shang, J., Li, S., Sun, Y.: Detection of SNP-SNP interaction based on the generalized particle swarm optimization algorithm. In: 2014 8th International Conference on Systems Biology (ISB), 2014, pp. 151–155. IEEE (2014)

13. Shang, J., Sun, Y., Fang, Y., Li, S., Liu, J.-X., Zhang, Y.: Hypergraph supervised search for inferring multiple epistatic interactions with different orders. In: Huang, D.-S., Jo, K.-H., Hussain, A. (eds.) ICIC 2015. LNCS, vol. 9226, pp. 623–633. Springer, Heidelberg (2015)

14. Zhang, Y., Liu, J.S.: Bayesian inference of epistatic interactions in case-control studies. Nat. Genet. **39**(9), 1167–1173 (2007)

15. Han, B., Chen, X.-w., Talebizadeh, Z., Xu, H.: Genetic studies of complex human diseases: Characterizing SNP-disease associations using Bayesian networks. BMC Syst. Biol. **6**(Suppl 3), S14 (2012)

16. Tang, W., Wu, X., Jiang, R., Li, Y.: Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. PLoS Genet. **5**(5), e1000464 (2009)

17. Frankel, W.N., Schork, N.J.: Who's afraid of epistasis? Nat. Genet. **14**(4), 371–373 (1996)

18. Li, W., Reich, J.: A complete enumeration and classification of two-locus disease models. Hum. Hered. **50**(6), 334–349 (2000)

19. Shang, J., Zhang, J., Sun, Y., Liu, D., Ye, D., Yin, Y.: Performance analysis of novel methods for detecting epistasis. BMC Bioinform. **12**(1), 475 (2011)

20. Shang, J., Sun, Y., Li, S., Liu, J.-X., Zheng, C.-H., Zhang, J.: An improved opposition-based learning particle swarm optimization for the detection of SNP-SNP interactions. BioMed Res. Int. **2015**, 524821 (2015)

21. Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.-Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T.: Complement factor H polymorphism in age-related macular degeneration. Science **308**(5720), 385–389 (2005)