

## Research Article

# An Improved Opposition-Based Learning Particle Swarm Optimization for the Detection of SNP-SNP Interactions

Junliang Shang,<sup>1</sup> Yan Sun,<sup>1</sup> Shengjun Li,<sup>1</sup> Jin-Xing Liu,<sup>1,2</sup>  
Chun-Hou Zheng,<sup>3</sup> and Junying Zhang<sup>4</sup>

<sup>1</sup>School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China

<sup>2</sup>Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China

<sup>3</sup>College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui 230039, China

<sup>4</sup>School of Computer Science and Technology, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Jin-Xing Liu; [sdcavell@126.com](mailto:sdcavell@126.com)

Received 25 September 2014; Revised 30 December 2014; Accepted 2 January 2015

Academic Editor: Yuedong Yang

Copyright © 2015 Junliang Shang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

SNP-SNP interactions have been receiving increasing attention in understanding the mechanism underlying susceptibility to complex diseases. Though many works have been done for the detection of SNP-SNP interactions, the algorithmic development is still ongoing. In this study, an improved opposition-based learning particle swarm optimization (IOBLPSO) is proposed for the detection of SNP-SNP interactions. Highlights of IOBLPSO are the introduction of three strategies, namely, opposition-based learning, dynamic inertia weight, and a postprocedure. Opposition-based learning not only enhances the global explorative ability, but also avoids premature convergence. Dynamic inertia weight allows particles to cover a wider search space when the considered SNP is likely to be a random one and converges on promising regions of the search space while capturing a highly suspected SNP. The postprocedure is used to carry out a deep search in highly suspected SNP sets. Experiments of IOBLPSO are performed on both simulation data sets and a real data set of age-related macular degeneration, results of which demonstrate that IOBLPSO is promising in detecting SNP-SNP interactions. IOBLPSO might be an alternative to existing methods for detecting SNP-SNP interactions.

## 1. Introduction

There is an increasing interest in understanding the underlying genetic architecture of complex diseases, such as cancer, heart disease, diabetes, Crohn's disease, and many others, which represent the major part of current clinical diseases [1, 2]. Research of complex diseases is one of the hottest fields of bioinformatics and genome-wide association studies (GWAS) become routine strategies. With the methods of GWAS, hundreds of thousands of single nucleotide polymorphisms (SNPs) speculated to associate with complex diseases have been identified. Nevertheless, these SNPs have limited effects on predicting the phenotype, and a large fraction of genetic contributions to complex diseases remain unclear. Recent advances make it clear that besides rare SNPs not genotyped in GWAS, the "missing heritability" can be partly

explained by nonlinear interactive effects of multiple SNPs, namely, SNP-SNP interactions [3]. Detection of SNP-SNP interactions is therefore a compelling next step in GWAS.

In general, the detection of SNP-SNP interactions is a great challenge [4]. The first challenge is the intensive computational burden imposed by the enormous search space, which prohibits real applications of most existing methods, especially those exhaustive ones. For instance, search space of a 100,000-SNP data set with considered maximum order of 3 is an astronomical number  $\sum_{k=1}^3 C_{100,000}^k$ . The second challenge is the complexity of genetic architecture of a complex disease. Limited or even no prior knowledge available for a complex disease, such as the order and the effect magnitude of a SNP-SNP interaction, makes it difficult for the development of heuristic methods. The third evaluation measures that determine how well a SNP combination contributes to

the phenotype are limited. Evaluation measures should be efficient in computational cost and insensitive to both SNP combination order and dependency type. Though several evaluation measures have been widely used in the detection of SNP-SNP interactions, developing new evaluation measures that can effectively and efficiently capture SNP-SNP interactions is still a direction.

Though methodological and computational perplexities of the detection of SNP-SNP interactions have been well recognized, the algorithmic development is still ongoing. Exhaustive algorithms, for example, MDR [5], appear promising for small scale data sets. However, for large scale data sets, especially those for GWAS, the detection of SNP-SNP interactions becomes a *needles-in-a-haystack* problem and exhaustive algorithms lose their ability [6, 7]. Heuristic algorithms are popular since they can retain as many informative SNPs as possible while largely reducing computational complexity. For example, Jiang et al. formulated the detection of SNP-SNP interactions from the viewpoint of binary classification and designed *epiForest* on the basis of the gini importance given by the random forest to select a small set of candidate SNPs [8]. Zhang and Liu proposed a Bayesian partition approach BEAM to find groups of genotypes with large posterior probability [9]. Tang et al. introduced the concept of epistatic module and designed a Gibbs sampling approach *epiMODE* to detect such modules [10]. Wan et al. developed a SNP-SNP interaction detection method SNPRuler based on both predictive rule inference and two-stage design [11]. They also presented another method BOOST, which involves only Boolean values and allows the use of fast logic operations to obtain contingency tables [12]. Besides machine learning methods, entropy based methods are also applied to this field. Chanda et al. developed an interaction index based on entropy theory to prioritize interacting SNPs [13]. They also applied two entropy theoretic measures to three SNP-SNP interaction detection methods: AMBIENCE [14] with a phenotype associated information measure; KWII [15] with the coinformation measure detecting SNP-SNP interactions associated with the binary phenotype; and CHORUS [16] combining these two measures together to identify associations with quantitative traits.

Recently, many swarm intelligence based algorithms have been proposed for the detection of SNP-SNP interactions [17–31]. Among them, particle swarm optimization (PSO) appears promising and some related works have been reported [22–30]. Yang et al. [22] used the binary PSO with odds ratio as the fitness function (OR-BPSO) to evaluate the risk of breast cancer. Based on the OR-BPSO, Chang et al. [23] proposed the odds ratio-based discrete binary PSO (OR-DBPSO) for the detection of SNP-SNP interactions with the quantitative phenotype. Chuang et al. [24] proposed a chaotic PSO (CPSO) that identifies the best SNP combination for breast cancer association studies. For enhancing the reliability of the PSO in the identification of the best SNP-SNP interaction associated with breast cancer, they also developed an improved PSO (IPSO) [25] and proved that the IPSO is highly reliable than the OR-BPSO. More recently, they used the gauss chaotic map PSO (Gauss-PSO) to detect the best association with breast cancer [26]. Experimental

results revealed that the Gauss-PSO was able to identify higher difference values between cases and controls than both the PSO and the CPSO. Yang et al. [27] developed a double-bottom chaotic map PSO (DBM-PSO) that overcomes the respective disadvantages of the PSO and the CPSO. Then, DBM-PSO is successfully applied to determine gene-gene interactions based on *chi-square* test [28]. Hwang et al. [29] proposed a complementary-logic PSO (CLPSO) to increase the efficiency of significant model identification in case-control study. Wu et al. [30] applied PSO to analyze the SNP-SNP interactions associated with hypertension. However, these methods, almost all of which are developed by a group except the one proposed by Wu et al. [30], only focus on finding the best genotype-genotype of a SNP-SNP interaction among possible genotypes of SNP combinations, but not the SNP-SNP interactions among possible SNP combinations. Obviously, the limited sample size of SNP data affects their computational accuracies of fitness functions and hence hinders their further applications. Furthermore, these methods are experimented on very small scale data sets (<30 SNPs) of certain complex diseases, performance of which on various kinds of large scale data sets are still unclear.

In this study, we proposed an improved opposition-based learning particle swarm optimization (IOBLPSO) with mutual information as its fitness function to detect SNP-SNP interactions. IOBLPSO is the first PSO based method to find SNP-SNP interactions among possible SNP combinations. Highlights of IOBLPSO are the introduction of three strategies, that is, opposition-based learning (OBL), dynamic inertia weight, and a postprocedure. Among them, OBL is the core, which is presented in the stage of updating particle experiences and common knowledge of swarm, not only for enhancing the global explorative ability, but also for avoiding premature convergence. Dynamic inertia weight is computed before the stage of updating particle velocities to allow particles to cover a wider search space when the considered SNP is likely to be a random SNP and to converge on promising regions of the search space while capturing a highly suspected SNP. The postprocedure is used as the final stage for carrying out a deep search in highly suspected SNP sets. Experiments of IOBLPSO are performed on lots of simulation data sets under the evaluation measures of both detection power and computational complexity. Results demonstrate that IOBLPSO is promising for the detection of all simulation models of SNP-SNP interactions. IOBLPSO is also applied on data set of age-related macular degeneration (AMD). Results show the strength of IOBLPSO on real applications and capture important features of genetic architecture of AMD that have not been described previously, which provide new clues for biologists on the exploration of AMD associated SNPs. IOBLPSO might be an alternative to existing methods for the detection of SNP-SNP interactions.

## 2. Methods

**2.1. Particle Swarm Optimization (PSO).** The PSO, proposed by Kennedy and Eberhart [32], is a member of the family of swarm intelligence algorithms, which mimic the collective

behaviors of organisms based on information sharing, like ants and birds, which can jointly perform many complex tasks though each individual is very limited in its capability. The PSO is a stylized representation of the movement of birds (viewed as particles) in a flock, where each particle uses its own experience and the common knowledge gained by the entire swarm to find an optimal position [29].

In PSO, the position of a particle represents a possible solution. In each generation, the position of each particle is adjusted according to its updated velocity and is estimated by a fitness function for providing a good search direction. Whether the velocity of each particle is updated depends on three variables: its previous velocity, its individual experience, and the common knowledge of the swarm. Specifically, the individual experience of each particle is updated while fitness value of its current position is higher than that of its previous experience; the common knowledge of the swarm is updated by the one of individual experiences of all particles with the highest fitness value while such value is higher than that of their previous common knowledge. This feedback strategy leads the swarm to gradually converge to an optimal solution [25–29].

Owing to its high capability and good generality in solving complex problems, the PSO has become a widely adopted swarm intelligence algorithm. However, it still has several defects, for example, premature convergence, stagnation phenomenon, and slow convergence speed in the later evolution period, which imply that the PSO should be further improved, especially for a specific complex problem, for example, the detection of SNP-SNP interactions. In general, the PSO consists of 4 stages: (1) initializing particles, (2) evaluating particles using fitness function, (3) updating particle velocities and positions, and (4) updating particle experiences and common knowledge of swarm. These stages are detailed and described in the following section.

**2.2. IOBLPSO: An Improved Opposition-Based Learning Particle Swarm Optimization for the Detection of SNP-SNP Interactions.** The flowchart of IOBLPSO is shown in Figure 1, where its highlights are with grey background. Below we describe IOBLPSO in detail from 6 stages.

**(1) Mapping SNPs and Initializing Particles.** At present, the popular way of mapping SNPs is to collect them as a matrix, where a row represents genotypes of an individual and a column represents a SNP. Genotypes of a SNP are coded as  $\{0, 1, 2\}$ , corresponding to homozygous common genotype (e.g., AA, BB), heterozygous genotype (e.g., Aa, aA, Bb, bB), and homozygous minor genotype (e.g., aa, bb). The label of an individual is a binary phenotype being either 0 (control) or 1 (case).

Based on above numerical mapping, the position of the  $p_{th}$  particle at iteration  $t$  can be represented as  $\text{Position}_t(p) = (\text{SNP}_{p1}^t, \dots, \text{SNP}_{pk}^t, \dots, \text{SNP}_{pK}^t)$ , where  $p \in \{1, 2, \dots, P\}$ ,  $k \in \{1, 2, \dots, K\}$ ,  $t \in \{1, 2, \dots, T\}$ ,  $P$  is the number of particles,  $K$  is the considered order of SNP-SNP interactions,  $T$  is the number of iterations,  $\text{SNP}_{pk}^t$  is the index of the selected  $k_{th}$  SNP of the  $p_{th}$  particle at iteration  $t$ ,  $\text{SNP}_{pk}^t \in \{1, 2, \dots, M\}$ ,

and  $M$  is the number of SNPs in the data set. The velocity of the  $p_{th}$  particle at iteration  $t$  is represented as  $\text{Velocity}_t(p) = (v_{p1}^t, \dots, v_{pk}^t, \dots, v_{pK}^t)$ , where  $v_{pk}^t$  is the velocity of SNP  $\text{SNP}_{pk}^t$  and  $v_{pk}^t \in [1 - M, M - 1]$ . Similarly, the individual experience of the  $p_{th}$  particle, that is, the position of the  $p_{th}$  particle with the highest fitness value until iteration  $t$ , can be denoted as  $pbest_t(p) = (p\text{SNP}_{p1}^t, \dots, p\text{SNP}_{pk}^t, \dots, p\text{SNP}_{pK}^t)$ , and the common knowledge of swarm, that is, the best position of all particles with the highest fitness value until iteration  $t$ , is denoted as  $gbest_t = (g\text{SNP}_1^t, \dots, g\text{SNP}_k^t, \dots, g\text{SNP}_K^t)$ .

Before the first iteration,  $\text{Position}_1(p)$ ,  $\text{Velocity}_1(p)$ ,  $pbest_1(p)$ , and  $gbest_1$  are randomly initialized in their respective domains.

**(2) Updating Dynamic Inertia Weight.** Inertia weight is used to control the impact of the previous velocity of a particle on its current velocity. A large inertia weight facilitates the global exploration and thus enables the method to execute a search over various regions, while a small inertia weight facilitates the local exploitation, which searches a promising region [27]. In order to effectively balance the global exploration and the local exploitation, a dynamic inertia weight is introduced to IOBLPSO, which can be defined as

$$W_{pk}^t = \frac{\max(\text{count}_t) - \text{count}_t [p\text{SNP}_{pk}^t]}{\max(\text{count}_t) - \min(\text{count}_t)}, \quad (1)$$

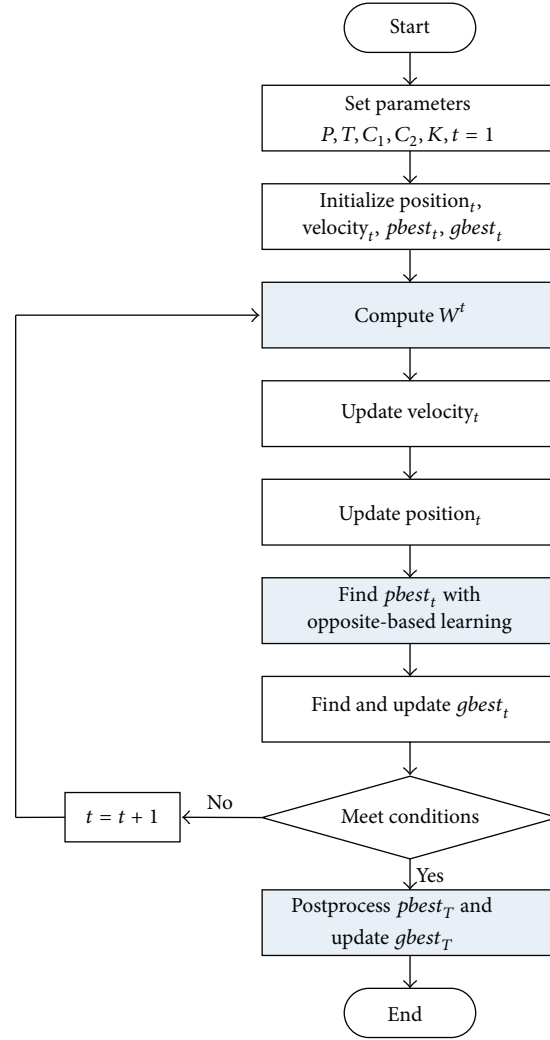
where  $\text{count}_t = (\text{ct}_1^t, \dots, \text{ct}_m^t, \dots, \text{ct}_M^t)$  and  $\text{ct}_m^t$  is a counter that counts the number of SNP  $m$  presented in  $pbest$  from iteration 1 to iteration  $t$ . This strategy allows particles to cover a wider search space while the considered SNP is likely to be a random SNP and to converge on promising regions of the search space while capturing a highly suspected SNP.

**(3) Evaluating Particles Using Fitness Function.** Fitness function of the IOBLPSO plays an important role on deciding which SNP combination is the SNP-SNP interaction and measuring how much the effect of a captured SNP-SNP interaction to the phenotype is. In the IOBLPSO, mutual information is applied as its fitness function, since it is well developed and can measure multivariate dependence without complex modeling. Mutual information has been widely used as a promising measure for feature selection and here is defined as

$$MI(X; Y) = H(X) + H(Y) - H(X, Y), \quad (2)$$

where  $H(X)$  is the entropy of  $X$ ;  $X$ , representing a SNP combination, is the general expression of  $\text{Position}_t(p)$ ,  $pbest_t(p)$ , and  $gbest_t$ ;  $H(Y)$  is the entropy of the phenotype  $Y$ ;  $H(X, Y)$  is the joint entropy of both  $X$  and  $Y$ . It is clear that higher mutual information value, namely, fitness value, indicates stronger association between the phenotype and the SNP combination.

**(4) Updating Particle Velocities and Positions.** IOBLPSO executes a search for SNP-SNP interactions by continuously



$P$ : number of particles  
 $T$ : number of iterations  
 $C_1$ : acceleration factor of the individual particle experience  
 $C_2$ : acceleration factor of the common knowledge  
 $K$ : order of considered SNP-SNP interactions

$t$ : current iteration  
 Position: positions of particles  
 Velocity: velocities of particles  
 $pbest$ : individual particle experience  
 $gbest$ : common knowledge of swarm  
 $W$ : inertia weight

FIGURE 1: The flowchart of IOBLPSO. Three components with grey background are highlights of IOBLPSO.

updating particle velocities and particle positions in all iterations. The velocity of  $\text{SNP}_{pk}^t$  is updated using the following equations:

$$\begin{aligned} \tilde{v}_{pk}^{t+1} = & W_{pk}^t \cdot v_{pk}^t + C_1 \cdot r_1 \cdot (p\text{SNP}_{pk}^t - \text{SNP}_{pk}^t) \\ & + C_2 \cdot r_2 \cdot (g\text{SNP}_k^t - \text{SNP}_{pk}^t), \end{aligned} \quad (3)$$

$$v_{pk}^{t+1} = \begin{cases} \tilde{v}_{pk}^{t+1} & \tilde{v}_{pk}^{t+1} \in [1 - M, M - 1] \\ \text{rand}(1 - M, M - 1) & \tilde{v}_{pk}^{t+1} \notin [1 - M, M - 1], \end{cases}$$

where  $C_1$  and  $C_2$ , controlling how far a particle moves in a single iteration, are acceleration factors and  $r_1$  and  $r_2$  are random values in  $(0, 1)$ . To obtain a valid velocity, a random value is sampled in  $[1 - M, M - 1]$  while  $\tilde{v}_{pk}^{t+1}$  exceeds its domain. Based on  $v_{pk}^{t+1}$ , the position of  $\text{SNP}_{pk}^t$  can be updated by the following two equations:

$$\begin{aligned} \overline{\text{SNP}}_{pk}^{t+1} = & \text{SNP}_{pk}^t + v_{pk}^{t+1}, \\ \text{SNP}_{pk}^{t+1} = & \begin{cases} \text{int}(\overline{\text{SNP}}_{pk}^{t+1}) & \overline{\text{SNP}}_{pk}^{t+1} \in [1, M] \\ \text{int}(\text{rand}(1, M)) & \overline{\text{SNP}}_{pk}^{t+1} \notin [1, M]. \end{cases} \end{aligned} \quad (4)$$



Because of  $\text{SNP}_{pk}^{t+1}$  being a SNP index, an integer between 1 and  $M$  is randomly sampled if  $\overline{\text{SNP}_{pk}^{t+1}}$  exceeds its domain. Such random sampling strategies on updating of both  $v_{pk}^{t+1}$  and  $\text{SNP}_{pk}^{t+1}$  help to increase the diversity of the search, the more possibility of jumping out local optima and getting into global optima.

(5) *Updating Particle Experiences and Common Knowledge of Swarm.* Another strategy introduced to IOBLPSO is the OBL. The basic principle of OBL is the consideration of a solution and its corresponding opposite solution simultaneously to approximate the global optima [33]. In the IOBLPSO, if the solution is  $\text{Position}_t(p)$ , its corresponding opposite solution can be defined as

$$\text{Position}'_t(p) = 1 + M - \text{Position}_t(p). \quad (5)$$

By comparing fitness values of  $\text{Position}_t(p)$ ,  $\text{Position}'_t(p)$ , and  $pbest_t(p)$ , the individual experience of the  $p_{th}$  particle at iteration  $t+1$ , that is,  $pbest_{t+1}(p)$ , is updated to the one among them with highest fitness value, which can be written as

$$pbest_{t+1}(p) = \begin{cases} \text{Position}_t(p) & MI(\text{Position}_t(p); Y) = \text{Val} \\ \text{Position}'_t(p) & MI(\text{Position}'_t(p); Y) = \text{Val} \\ pbest_t(p) & MI(pbest_t(p); Y) = \text{Val}, \end{cases} \quad (6)$$

where  $\text{Val} = \max(MI(\text{Position}_t(p); Y), MI(\text{Position}'_t(p); Y), MI(pbest_t(p); Y))$ . From this equation, it can be seen that the employed OBL strategy facilitates IOBLPSO not only expanding the search space and enhancing the global explorative ability, but also accelerating the convergence and avoiding premature convergence.

Similarly, whether the common knowledge of the swarm at iteration  $t+1$ , for example,  $gbest_{t+1}$ , is updated or maintained as  $gbest_t$  depends on fitness values of individual experiences of all particles at iteration  $t+1$  and can be defined as

$$gbest_{t+1} = \begin{cases} pbest_{t+1}(p) & MI(pbest_{t+1}(p); Y) > MI(gbest_t; Y) \\ gbest_t & MI(pbest_{t+1}(p); Y) \leq MI(gbest_t; Y). \end{cases} \quad (7)$$

(6) *Deep Searching with a Postprocedure.* A postprocedure is provided when completing the iteration process to carry out a deep search of SNP-SNP interactions in a highly suspected SNP sets. First, all SNPs are descending sorted according to their counters in  $\text{count}_T$ , and the specified number of top SNPs (By default, 10) are selected into the highly suspected SNP sets. Second, IOBLPSO conducts an exhaustive search within the highly suspected SNP sets to determine whether fitness value of one or more SNP combinations is higher than that of  $gbest_T$ . If indeed detected,  $gbest_T$  is updated by the best one among them.  $gbest_T$  is therefore the final result of IOBLPSO.

### 3. Results and Discussion

3.1. *Simulation Data.* Six commonly used models of SNP-SNP interactions with their orders being equal to 2 (i.e.,  $K = 2$ ) are exemplified for the study [7, 9, 10, 34, 35]. Model 1 and Model 2 are models displaying both marginal effects and interactive effects, and others show no marginal effects but interactive effects. Specifically, the penetrance in Model 1 increases only when both SNPs have at least one minor allele [9, 10]; Model 2 assumes that the minor allele in one SNP has the marginal effect; however the effect is inversed while minor alleles in both SNPs are present [9]; Model 3 and Model 4 are directly cited from the reference [35]; Model 5 is a ZZ model [34]; and Model 6 is an XOR model [35]. Model 3~Model 6 are exemplified here since they provide a high degree of complexity to challenge ability of a method in detecting SNP-SNP interactions [7]. For each model, 50 data sets are generated by the simulator EpiSIM [36], each containing 2000 cases and 2000 controls genotyped with 100 SNPs. For each data set, random SNPs are set independently with MAFs chosen from [0.05, 0.5] uniformly and detailed parameters of ground-truth SNPs are recorded in Figure 2, where ground-truth SNPs refer to the causative SNPs that truly associated with the phenotype, in other words, the SNPs in models added into the simulation data sets.

3.2. *Evaluation Measure.* Detection power is one of the generally used evaluation measures in the field of the detection of SNP-SNP interactions, and various forms of detection power have been proposed depending on what is desired to measure [4, 7, 9–11, 21, 31, 37]. In this study, two types of detection power are introduced, namely, Power 1 and Power 2.

Power 1 [4, 7, 9–11, 21, 31] is defined as the proportion of data sets in which all ground-truth SNPs are detected with no false positives, which can be written as

$$\text{Power 1} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (8)$$

where  $N$  is the number of data sets with the same parameter settings (here,  $N = 50$ ), and  $x_i \in \{0, 1\}$  is the detection tag; that is, if 2 ground-truth SNPs in data set  $i$  are detected with no false positives,  $x_i = 1$ ; otherwise,  $x_i = 0$ . Though Power 1 seems not practical since false positives are inevitable for any statistical tests and fewer false positives result in larger false negatives, we still introduce it because it is advantageous in practical applications and might be of interest to biologists due to false positives implying wasted experimental effort to validate the results.

Sometimes, allowing some small Type-I error rate is more reasonable; thus Power 2 [4, 7, 21] is introduced here, which is defined as an average proportion of ground-truth SNPs in the top 2 detected SNPs, and can be written as

$$\text{Power 2} = \frac{1}{2 \cdot N} \sum_{i=1}^N y_i, \quad (9)$$

where  $y_i$  is the number of ground-truth SNPs in the top 2 SNPs identified in data set  $i$ .

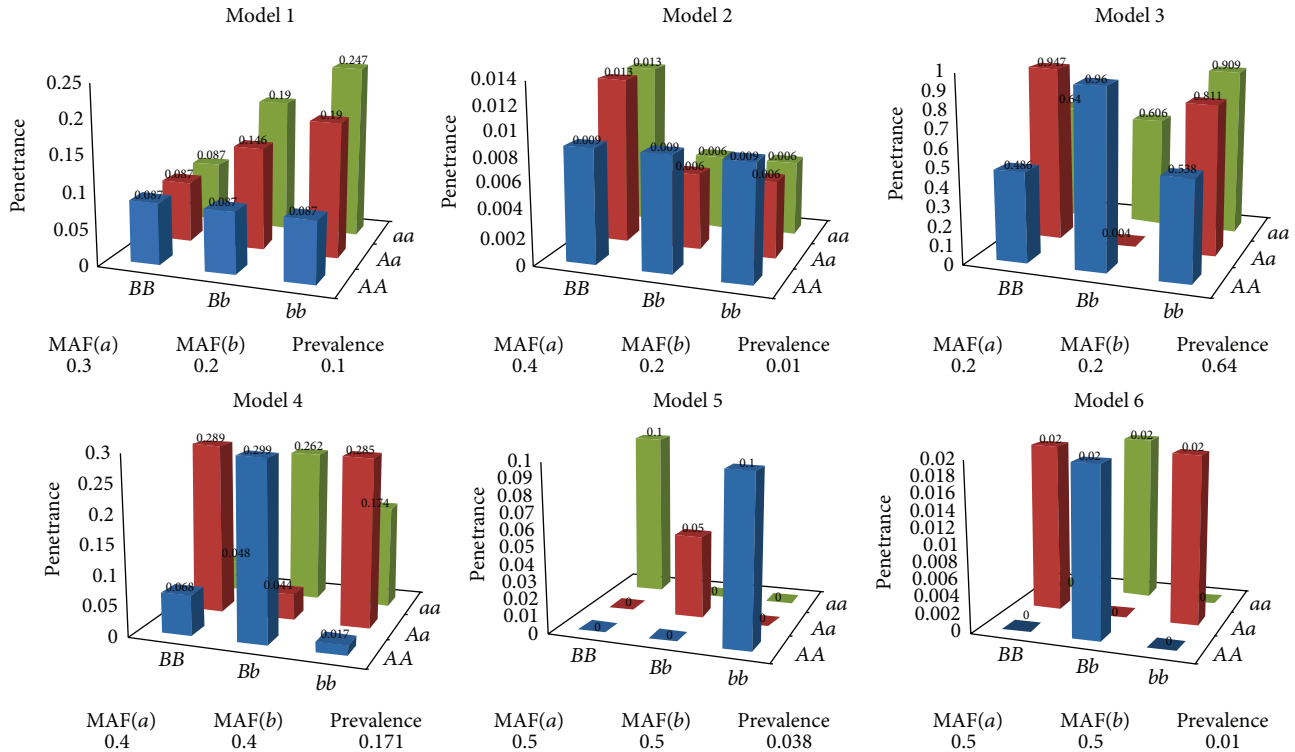


FIGURE 2: The simulation models of SNP-SNP interactions. In the figure, penetrance is the probability of the occurrence of a disease given a particular genotype; prevalence is the proportion of individuals that occur a disease; MAF( $a$ ) and MAF( $b$ ) are, respectively, minor allele frequencies of  $a$  and  $b$ .

Computational complexity is also considered. We measure running time in the same computational environment to assess realistic applicability of compared methods.

**3.3. Performance of IOBLPSO on Simulation Data.** To demonstrate the validity of IOBLPSO, its detection power is evaluated by comparison with several typical SNP-SNP interaction detection methods, that is, BOOST [12], AntEpiSeeker [20], SNPRuler [11], and TEAM [38]. These machine learning methods are recently proposed, claimed to facilitate large scale data sets, and their packages are online freely available [7]. Besides these methods, two modified PSO methods for SNP-SNP interaction detection, namely, DBM-PSO [27] focusing on finding the best genotype-genotype of a SNP-SNP interaction among possible genotypes of SNP combinations and the PSO focusing on finding SNP-SNP interactions among possible SNP combinations, are also compared.

In the study, parameters of each method are generally set as default. Only a few are changed according to suggestions in order to balance result accuracy and computational cost. For BOOST, interaction threshold is set to 10, that is, results of BOOST are the SNP-SNP interactions whose likelihood ratio test statistic values  $>10$  with 4 degrees of freedom. For AntEpiSeeker, the numbers of ants and iterations is set to 500 and 10, respectively. For TEAM, permutation number is set to 100. For a fair comparison, parameter settings of PSO based methods are the same. Specifically, the number of particles  $P$  and the number of iterations  $T$  are respective set to 100 and

100; both acceleration factors  $C_1$  and  $C_2$  are set to 2 [39]; the inertia weight  $W$  is set to 0.65. It is believed that performance of IOBLPSO mainly depends on parameters ( $P, T$ ). Hence we further examine the influence of these parameters on detection power with (25, 100), (50, 100), (100, 25), (100, 50), and (100, 100).

Detection power of compared methods on simulation data sets is reported in Figure 3. Detection power of IOBLPSO and the PSO with different numbers of particles is shown in Figure 4, and that with different numbers of iterations is shown in Figure 5. The average running time of the methods on simulation data sets is recorded in Table 1. From Figures 3, 4, and 5 and Table 1, we have the following observations.

It is seen that IOBLPSO outperforms compared methods on all cases regardless of models, the numbers of particles, and iterations. Specifically, no matter, according to Power 1 or Power 2, detection power of IOBLPSO on all models and ( $P, T$ ) settings is comparable and sometimes superior to that of compared methods, which might be the result of introducing three effective strategies into IOBLPSO: OBL expanding the search space and enhancing the global explorative ability, dynamic inertia weight guiding the particles to more promising regions, and postprocedure carrying out a deep search in highly suspected SNP sets; with the numbers of particles or iterations grow, detection power of both IOBLPSO and the PSO increase quickly, especially IOBLPSO; IOBLPSO identifies almost all ground-truth SNPs

TABLE 1: Average running time (seconds) of compared methods on simulation data sets. Experiments are conducted with Intel Xeon 2.00 GHz CPUs and 6 GB of RAM running Microsoft Windows XP Professional x64 Edition 2003 Service Pack 2 for computational complexity analysis.

Methods	BOOST	AntEpiSeeker	SNPRuler	TEAM	DBM-PSO	ISO	IOBLPSO
Running time	0.36	1146.60	1.56	13.14	68.73	13.40	20.75

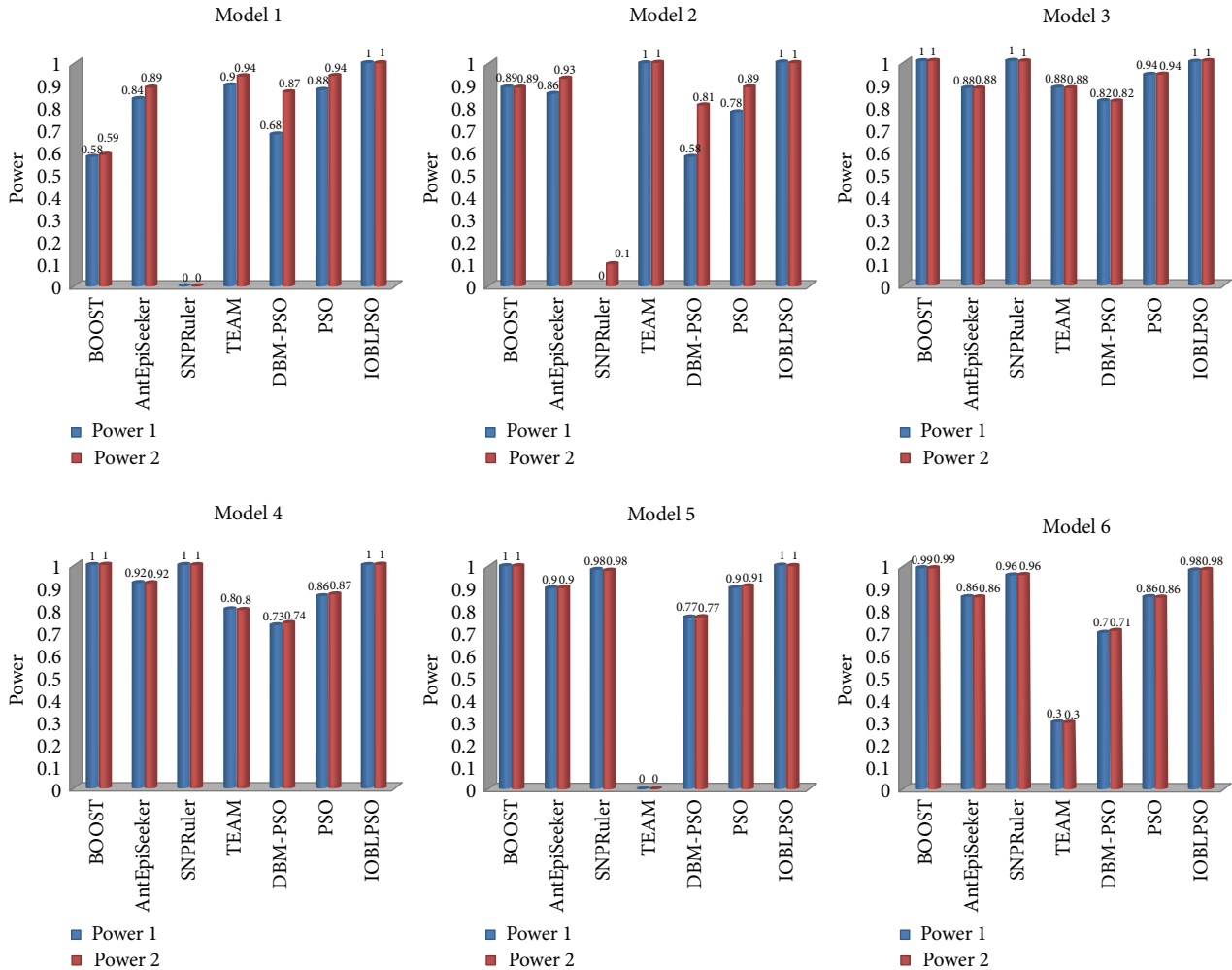


FIGURE 3: Detection power of compared methods on simulation data sets.

on all models with the parameter setting (100, 100), even with (25, 100) or (100, 25); IOBLPSO has perfect detection power on Model 1 and Model 2; that is to say, compared with other PSO based methods, IOBLPSO needs less particles and/or iterations to obtain higher detection power, implying that IOBLPSO can handle large scale data sets for GWAS and its scalability is better than others; for Model 1 and Model 2, Power 1 and Power 2 of IOBLPSO reach a prefect level, Power 1 and Power 2 of other methods have different values since these two models display not only interaction effects but also marginal effects, leading to compared methods sometimes only identifying several ground-truth SNPs, but not SNP-SNP interactions; similarly, for each method on Model 3~Model 6, Power 1 and Power 2 of each compared method are almost always equal because single ground-truth SNPs show no main effects; in terms of computational complexity,

though IOBLPSO is not the fast one among all compared methods, it can finish the work at affordable time costs; more importantly, its time costs can be estimated and controlled by setting the numbers of particles and iterations freely under the premise of ensuring sufficient accuracy.

3.4. Application to Real AMD Data. In the study, potential of IOBLPSO can also be verified by analyzing a real AMD data set [40], which contains 103.611 SNPs genotyped with 96 cases and 50 controls. AMD, which 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 whereby multiple SNP-SNP interactions interact with environmental factors to the disease [4, 10]. We run IOBLPSO on AMD data set 20 times with different combinations of the number of particles

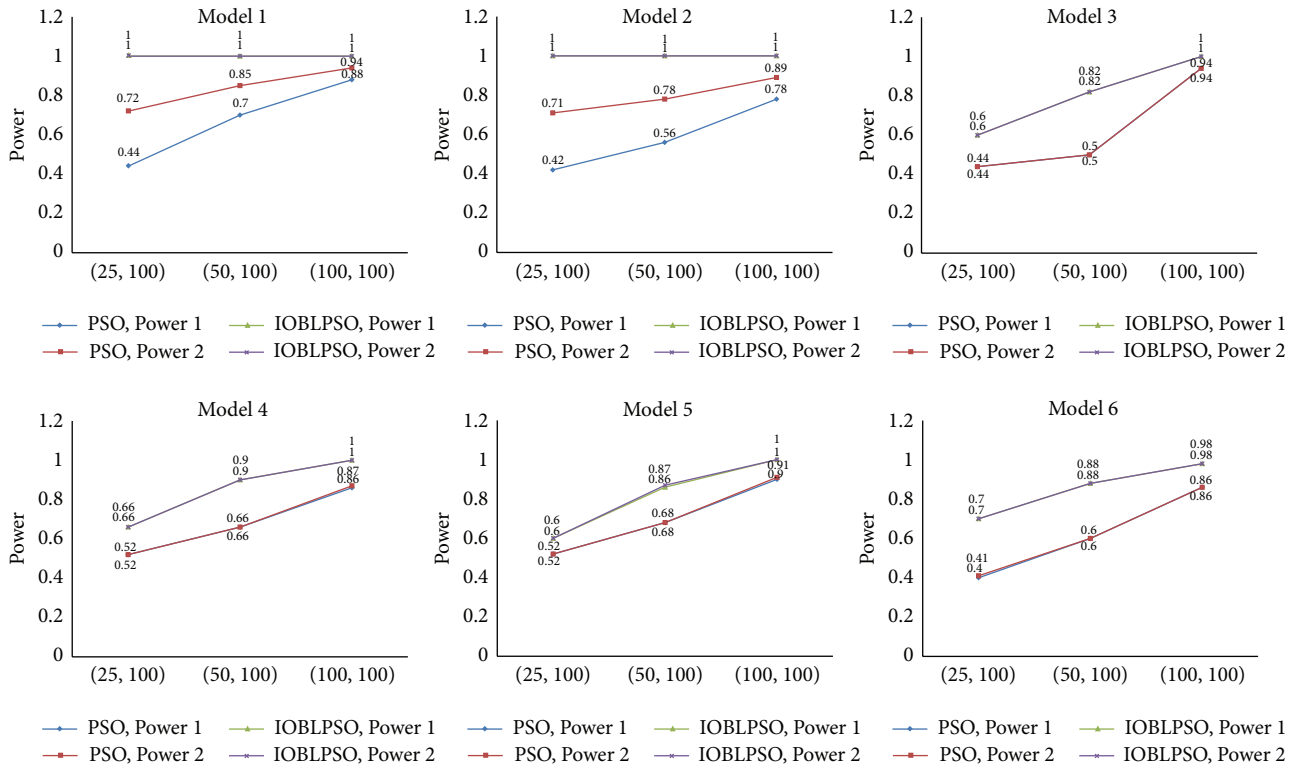


FIGURE 4: Detection power of IOBLPSO and the PSO with different numbers of particles. The numbers of particles are set to 25, 50, and 100, while the number of iterations is equal to 100.

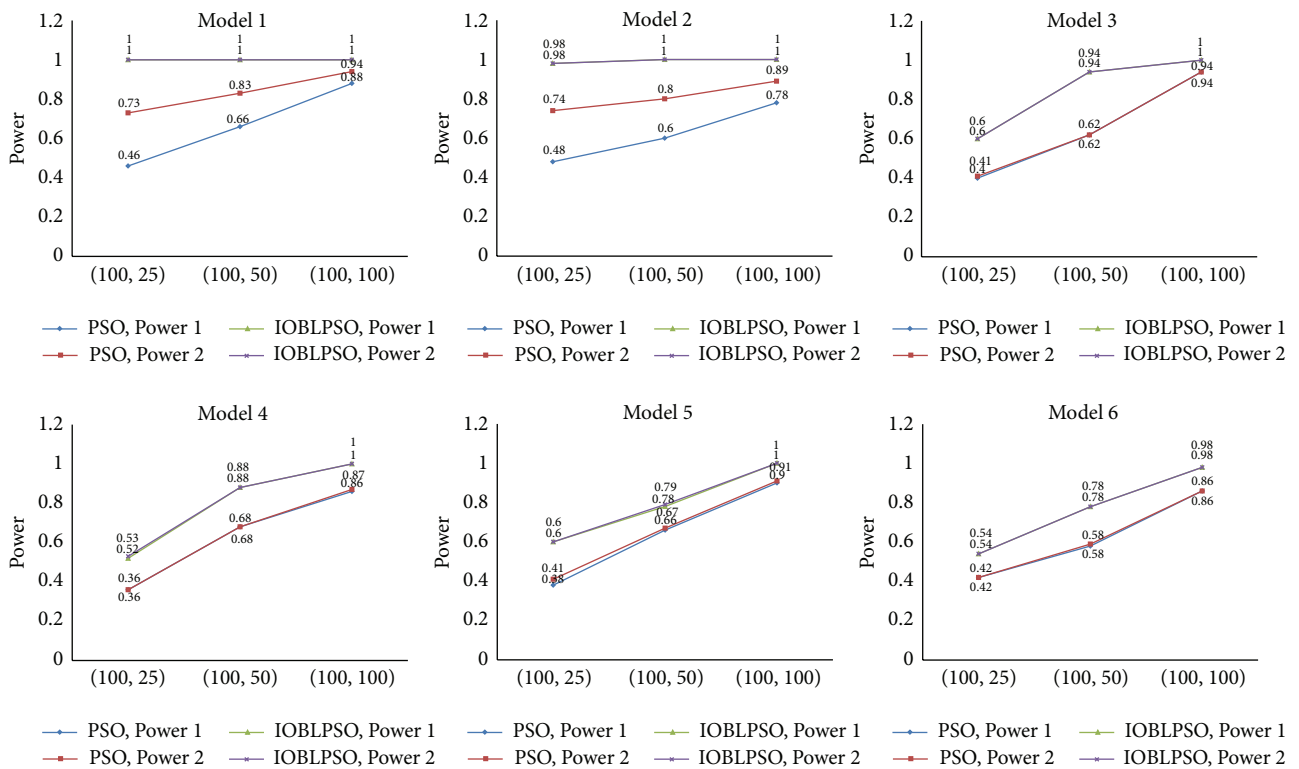


FIGURE 5: Detection power of IOBLPSO and the PSO with different numbers of iterations. The numbers of iterations are set to 25, 50, and 100, while the number of particles is equal to 100.



TABLE 2: Detected SNP-SNP interactions associated with AMD.  $P$  values of detected SNP-SNP interactions before Bonferroni correction, as well as their linkage disequilibrium (LD) correlation coefficients  $r^2$ , are also recorded. The SNPs in different SNP-SNP interactions have low LD.

SNP	Gene	Chromosome	Times	Mutual information value		$P$ value	LD ( $r^2$ )
				Individual	Interaction		
rs380390	<i>CFH</i>	1	1	<b>0.1412</b>	<b>0.2955</b>	$5.0006e - 08$	0.0089
rs1374431	<i>N/A</i>	2		0.0198			
rs380390	<i>CFH</i>	1	2	<b>0.1412</b>	0.2949	$5.7995e - 08$	0.0015
rs2402053	<i>N/A</i>	7		0.0476			
rs1329428	<i>CFH</i>	1	2	0.1218	0.2853	$3.0012e - 08$	0.0019
rs9328536	<i>MED27</i>	9		0.0563			
rs380390	<i>CFH</i>	1	2	<b>0.1412</b>	0.2777	$1.4359e - 07$	0.0050
rs10512174	<i>ISCA1</i>	9		0.0844			
rs380390	<i>CFH</i>	1	2	<b>0.1412</b>	0.2775	$1.9451e - 07$	0.0018
rs718263	<i>NCALD</i>	8		0.0471			
rs380390	<i>CFH</i>	1	1	<b>0.1412</b>	0.2760	$1.9798e - 07$	0.0019
rs223607	<i>N/A</i>	6		0.0079			
rs380390	<i>CFH</i>	1	8	<b>0.1412</b>	0.2752	<b><math>1.1381e - 09</math></b>	0.0031
rs1363688	<i>N/A</i>	5		0.0949			
rs380390	<i>CFH</i>	1	1	<b>0.1412</b>	0.2678	$6.4999e - 07$	0.0013
rs210758	<i>N/A</i>	4		0.0131			
rs380390	<i>CFH</i>	1	1	<b>0.1412</b>	0.2641	$2.4273e - 07$	0.0040
rs10507949	<i>N/A</i>	13		0.0948			

$P$  (10000, 20000) and the number of iterations  $T$  (500, 1000), each running 5 times. The order of SNP-SNP interactions  $K$  is set to 2 since the small sample size of 146 individuals is insufficient for secure detection of any higher order SNP-SNP interactions. Both acceleration factors  $C_1$  and  $C_2$  are set to 2. Detected SNP-SNP interactions associated with AMD are listed in Table 2, where their mutual information values of individual SNPs and SNP-SNP interactions are recorded.

It has been widely accepted that rs380390 and rs1329428 are believed to be significantly associated with AMD [10]. These two SNPs are in an intron of the *CFH* gene in chromosome 1. There are biologically plausible mechanisms for the involvement of *CFH* in AMD and at least 100 mutations in *CFH* have been proven to increase the risk of AMD and other disorders. *CFH* is a regulator that activates the alternative pathway of the complement cascade, the mutations in which can lead to an imbalance in normal homeostasis of the complement system. This phenomenon is thought to account for substantial tissue damage in AMD [41]. In the IOBLPSO, these two SNPs are detected as members of SNP-SNP interactions, especially the rs380390. Almost all SNP-SNP interactions include rs380390, since it has the strongest main effect, leading to its combinations with other SNPs displaying strong interaction effects. This phenomenon indicates that IOBLPSO is sensitive to those SNPs displaying strong main effects.

The SNP-SNP interaction (rs380390, rs1374431), also reported by [42, 43], has the strongest interaction effect. Rs1374431 is located in a noncoding region between genes *LOC644301* and *KIAA1715*. *KIAA1715* is usually found in adult brain regions. Although no evidences were reported with this gene related to AMD, it may be a plausible candidate gene

associated with AMD [42, 43]. Another SNP-SNP interaction (rs380390, rs2402053) has the second highest mutual information value. The SNP rs2402053 is in the intergenic region between genes *TFEC* and *TES* in chromosome 7q31 [44]. It is worth noting that mutations in some genes on 7q31-7q32 are revealed in patients with retinal disorders [45]. Therefore, rs2402053 may be a new genetic factor contributing to the underlying mechanism of AMD [46–50].

It is interesting that the SNP-SNP interaction (rs380390, rs1363688) was successfully detected 8 times by the IOBLPSO and by other methods [4, 21, 51], though it has moderate interaction effect. However, in terms of  $P$  value, the interaction (rs380390, rs1363688) is the most statistically significant one among all detected SNP-SNP interactions, which might be the reason of it being frequently detected. This fact implies that IOBLPSO is capable of capturing SNP-SNP interactions with statistically significant  $P$  values, though its fitness function is the mutual information. The SNP-SNP interaction (rs1329428, rs9328536) [52, 53], rs10507949 [4], and rs10512174 [51] also have been identified in AMD association studies, but their functions are still unclear. Other SNPs, that is, rs210758, rs223607, and rs718263, are the first time being identified. Further studies with the use of large-scale case-control samples are needed to confirm whether these SNPs have true associations with AMD. We hope that, from these results, some clues could be provided for the exploration of causative factors of AMD.

#### 4. Conclusions

Detection of SNP-SNP interactions is believed to be important in understanding underlying mechanism of complex

diseases. In this study, we proposed an improved opposition-based learning particle swarm optimization, or IOBLPSO, to detect SNP-SNP interactions. To the best of our knowledge, IOBLPSO is the first PSO based method to detect SNP-SNP interactions among possible SNP combinations. Highlights of IOBLPSO are the introduction of three strategies: OBL, dynamic inertia weight, and a postprocedure. Among them, OBL is the core, which is presented in the stage of updating particle experiences and common knowledge of swarm, not only for enhancing the global explorative ability, but also for avoiding premature convergence. Dynamic inertia weight is computed before the stage of updating particle velocities to allow particles to cover a wider search space while the considered SNP is likely to be a random SNP and to converge on promising regions of the search space while capturing a highly suspected SNP. The postprocedure is introduced as the final stage for carrying out a deep search in highly suspected SNP sets. Experiments of IOBLPSO are performed on lots of simulation data sets under the evaluation measures of detection power and computational complexity. Results demonstrate that IOBLPSO is promising for the detection of all simulation models of SNP-SNP interactions. IOBLPSO is also applied on a real AMD data set, results of which not only show the strength of IOBLPSO on real applications, but also capture important features of genetic architecture of AMD that have not been described previously. These features might provide new clues for biologists on the exploration of AMD associated genetic factors.

IOBLPSO might be an alternative to existing methods for detecting SNP-SNP interactions and has several merits. First, IOBLPSO is easy to be implemented, and its time costs can be estimated and controlled. Second, OBL and other two strategies help to improve the performance of IOBLPSO. Third, mutual information is effective in measuring SNP-SNP interactions. Fourth, compared with other methods, IOBLPSO needs less particles and/or iterations to obtain higher detection power, implying that IOBLPSO can handle large scale data sets for GWAS and its scalability is better than others. Though IOBLPSO is a beneficial exploration in the detection of SNP-SNP interactions, it still has several limitations; for example, multiple SNP-SNP interactions in a data set are not considered simultaneously; IOBLPSO is sensitive to those SNPs that display strong main effects. Furthermore, recent advancements in sequencing technology have enabled the sequencing of the whole-exome or even whole-genome of a cohort. The rare or de novo mutations resulting from these experiments should be considered. For example, Wu et al. recently proposed a bioinformatics method called SPRING for prioritizing candidate mutations [54]. It is therefore interesting to consider the problem of interactive effects of such de novo mutations. Limitations of IOBLPSO, as well as this new research hotspot, will inspire us to continue working in the future.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported by the Scientific Research Reward Foundation for Excellent Young and Middle-age Scientists of Shandong Province (BS2014DX004), the Science and Technology Planning Project of Qufu Normal University (xkj201410), the Scientific Research Foundation of Qufu Normal University (BSQD20130119), the Shandong Provincial Natural Science Foundation (ZR2013FL016), the China Postdoctoral Science Foundation Funded Project (2014M560264), the Shenzhen Municipal Science and Technology Innovation Council (JCYJ20140417172417174), the Award Foundation Project of Excellent Young Scientists in Shandong Province (BS2014DX005), the Project of Shandong Province Higher Educational Science and Technology Program (J13LN31), the Scientific Research Foundation of Qufu Normal University (XJ201226), and the Innovation and Entrepreneurship Training Project for College Students of Qufu Normal University (2014A096).

## References

- [1] L. R. Cardon and J. I. Bell, "Association study designs for complex diseases," *Nature Reviews Genetics*, vol. 2, no. 2, pp. 91–99, 2001.
- [2] N. Risch and K. Merikangas, "The future of genetic studies of complex human diseases," *Science*, vol. 273, no. 5281, pp. 1516–1517, 1996.
- [3] B. Maher, "The case of the missing heritability," *Nature*, vol. 456, no. 7218, pp. 18–21, 2008.
- [4] J. Shang, J. Zhang, Y. Sun, X. Dai, and Y. Zhang, "EpiMiner: a three-stage co-information based method for detecting and visualizing epistatic interactions," *Digital Signal Processing*, vol. 24, pp. 1–13, 2014.
- [5] M. D. Ritchie, L. W. Hahn, N. Roodi et al., "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer," *The American Journal of Human Genetics*, vol. 69, no. 1, pp. 138–147, 2001.
- [6] J. H. Moore, F. W. Asselbergs, and S. M. Williams, "Bioinformatics challenges for genome-wide association studies," *Bioinformatics*, vol. 26, no. 4, pp. 445–455, 2010.
- [7] J. Shang, J. Zhang, Y. Sun, D. Liu, D. Ye, and Y. Yin, "Performance analysis of novel methods for detecting epistasis," *BMC Bioinformatics*, vol. 12, article 475, 2011.
- [8] R. Jiang, W. Tang, X. Wu, and W. Fu, "A random forest approach to the detection of epistatic interactions in case-control studies," *BMC Bioinformatics*, vol. 10, no. 1, article S65, 2009.
- [9] Y. Zhang and J. S. Liu, "Bayesian inference of epistatic interactions in case-control studies," *Nature Genetics*, vol. 39, no. 9, pp. 1167–1173, 2007.
- [10] W. Tang, X. Wu, R. Jiang, and Y. Li, "Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy," *PLoS Genetics*, vol. 5, no. 5, Article ID e1000464, 2009.
- [11] X. Wan, C. Yang, Q. Yang, H. Xue, N. L. S. Tang, and W. Yu, "Predictive rule inference for epistatic interaction detection in genome-wide association studies," *Bioinformatics*, vol. 26, no. 1, pp. 30–37, 2010.
- [12] X. Wan, C. Yang, Q. Yang et al., "BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control

- studies," *The American Journal of Human Genetics*, vol. 87, no. 3, pp. 325–340, 2010.
- [13] P. Chanda, L. Sucheston, A. Zhang, and M. Ramanathan, "The interaction index, a novel information-theoretic metric for prioritizing interacting genetic variations and environmental factors," *European Journal of Human Genetics*, vol. 17, no. 10, pp. 1274–1286, 2009.
- [14] P. Chanda, L. Sucheston, A. Zhang et al., "AMBIENCE: a novel approach and efficient algorithm for identifying informative genetic and environmental associations with complex phenotypes," *Genetics*, vol. 180, no. 2, pp. 1191–1210, 2008.
- [15] P. Chanda, A. Zhang, D. Brazeau et al., "Information-theoretic metrics for visualizing gene-environment interactions," *The American Journal of Human Genetics*, vol. 81, no. 5, pp. 939–963, 2007.
- [16] P. Chanda, L. Sucheston, S. Liu, A. Zhang, and M. Ramanathan, "Information-theoretic gene-gene and gene-environment interaction analysis of quantitative traits," *BMC Genomics*, vol. 10, article 509, 2009.
- [17] C. S. Greene, J. M. Gilmore, J. Kiralis, P. C. Andrews, and J. H. Moore, "Optimal use of expert knowledge in ant colony optimization for the analysis of epistasis in human disease," in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, pp. 92–103, Springer, Berlin, Germany, 2009.
- [18] C. S. Greene, B. C. White, and J. H. Moore, "Ant colony optimization for genome-wide genetic analysis," in *Ant Colony Optimization and Swarm Intelligence*, pp. 37–47, Springer, Berlin, Germany, 2008.
- [19] R. Rekaya and K. Robbins, "Ant colony algorithm for analysis of gene interaction in high-dimensional association data," *Revista Brasileira de Zootecnia*, vol. 38, no. 1, pp. 93–97, 2009.
- [20] Y. Wang, X. Liu, K. Robbins, and R. Rekaya, "AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm," *BMC Research Notes*, vol. 3, article 117, 2010.
- [21] J. Shang, J. Zhang, X. Lei, Y. Zhang, and B. Chen, "Incorporating heuristic information into ant colony optimization for epistasis detection," *Genes and Genomics*, vol. 34, no. 3, pp. 321–327, 2012.
- [22] C.-H. Yang, H.-W. Chang, Y.-H. Cheng, and L.-Y. Chuang, "Novel generating protective single nucleotide polymorphism barcode for breast cancer using particle swarm optimization," *Cancer Epidemiology*, vol. 33, no. 2, pp. 147–154, 2009.
- [23] H.-W. Chang, C.-H. Yang, C.-H. Ho, C.-H. Wen, and L.-Y. Chuang, "Generating SNP barcode to evaluate SNP-SNP interaction of disease by particle swarm optimization," *Computational Biology and Chemistry*, vol. 33, no. 1, pp. 114–119, 2009.
- [24] L.-Y. Chuang, H.-W. Chang, M.-C. Lin, and C.-H. Yang, "Chaotic particle swarm optimization for detecting SNP-SNP interactions for CXCL12-related genes in breast cancer prevention," *European Journal of Cancer Prevention*, vol. 21, no. 4, pp. 336–342, 2012.
- [25] L.-Y. Chuang, Y.-D. Lin, H.-W. Chang, and C.-H. Yang, "An improved PSO algorithm for generating protective SNP barcodes in breast cancer," *PLoS ONE*, vol. 7, no. 5, Article ID e37018, 2012.
- [26] L.-Y. Chuang, Y.-D. Lin, H.-W. Chang, and C.-H. Yang, "SNP-SNP interaction using Gauss chaotic map particle swarm optimization to detect susceptibility to breast cancer," in *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14)*, pp. 2548–2554, Waikoloa, Hawaii, USA, January 2014.
- [27] C.-H. Yang, S.-W. Tsai, L.-Y. Chuang, and C.-H. Yang, "An improved particle swarm optimization with double-bottom chaotic maps for numerical optimization," *Applied Mathematics and Computation*, vol. 219, no. 1, pp. 260–279, 2012.
- [28] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, and H.-W. Chang, "Double-bottom chaotic map particle swarm optimization based on chi-square test to determine gene-gene interactions," *BioMed Research International*, vol. 2014, Article ID 172049, 10 pages, 2014.
- [29] M.-L. Hwang, Y.-D. Lin, L.-Y. Chuang, and C.-H. Yang, "Determination of the SNP-SNP interaction between breast cancer related genes to analyze the disease susceptibility," *International Journal of Machine Learning and Computing*, vol. 4, no. 5, pp. 468–473, 2014.
- [30] S.-J. Wu, L.-Y. Chuang, Y.-D. Lin et al., "Particle swarm optimization algorithm for analyzing SNP-SNP interaction of renin-angiotensin system genes against hypertension," *Molecular Biology Reports*, vol. 40, no. 7, pp. 4227–4233, 2013.
- [31] M. Aflakparast, H. Salimi, A. Gerami, M.-P. Dubé, S. Visweswaran, and A. Masoudi-Nejad, "Cuckoo search epistasis: a new method for exploring significant genetic interactions," *Heredity*, vol. 112, no. 6, pp. 666–674, 2014.
- [32] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, December 1995.
- [33] H. R. Tizhoosh, "Opposition-based learning: a new scheme for machine intelligence," in *International Conference on Computational Intelligence for Modelling, Control and Automation, International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA/IAWTIC '05)*, vol. 1, pp. 695–701, Vienna, Austria, 2005.
- [34] W. N. Frankel and N. J. Schork, "Who's afraid of epistasis?" *Nature genetics*, vol. 14, no. 4, pp. 371–373, 1996.
- [35] W. Li and J. Reich, "A complete enumeration and classification of two-locus disease models," *Human Heredity*, vol. 50, no. 6, pp. 334–349, 2000.
- [36] J. Shang, J. Zhang, X. Lei, W. Zhao, and Y. Dong, "EpiSIM: simulation of multiple epistasis, linkage disequilibrium patterns and haplotype blocks for genome-wide interaction analysis," *Genes & Genomics*, vol. 35, no. 3, pp. 305–316, 2013.
- [37] X. Jiang, R. E. Neapolitan, M. M. Barmada, S. Visweswaran, and G. F. Cooper, "A fast algorithm for learning epistatic genomic relationships," in *Proceedings of the AMIA Annual Symposium*, p. 341, 2010.
- [38] X. Zhang, S. Huang, F. Zou, and W. Wang, "TEAM: efficient two-locus epistasis tests in human genome-wide association study," *Bioinformatics*, vol. 26, no. 12, pp. i217–i227, 2010.
- [39] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 240–255, 2004.
- [40] R. J. Klein, C. Zeiss, E. Y. Chew et al., "Complement factor H polymorphism in age-related macular degeneration," *Science*, vol. 308, no. 5720, pp. 385–389, 2005.
- [41] M. K. M. Adams, J. A. Simpson, A. J. Richardson et al., "Can genetic associations change with age? CFH and age-related macular degeneration," *Human Molecular Genetics*, vol. 21, no. 23, pp. 5229–5236, 2012.
- [42] B. Han, M. Park, and X.-W. Chen, "DASSO-MB: detection of epistatic interactions in genome-wide association studies using Markov blankets," in *Proceedings of the IEEE International*

- Conference on Bioinformatics and Biomedicine (BIBM '09)*, pp. 148–153, November 2009.
- [43] B. Han, M. Park, and X.-W. Chen, “A Markov blanket-based method for detecting causal SNPs in GWAS,” *BMC Bioinformatics*, vol. 11, no. 3, article S5, 2010.
- [44] E. S. Tobias, A. F. L. Hurlstone, E. MacKenzie, R. Mcfarlane, and D. M. Black, “The *TES* gene at 7q31.1 is methylated in tumours and encodes a novel growth-suppressing LIM domain protein,” *Oncogene*, vol. 20, no. 22, pp. 2844–2853, 2001.
- [45] S. J. Bowne, L. S. Sullivan, S. H. Blanton et al., “Mutations in the inosine monophosphate dehydrogenase 1 gene (*IMPDH1*) cause the RP10 form of autosomal dominant retinitis pigmentosa,” *Human Molecular Genetics*, vol. 11, no. 5, pp. 559–568, 2002.
- [46] B. Han, X.-W. Chen, Z. Talebizadeh, and H. Xu, “Genetic studies of complex human diseases: characterizing SNP-disease associations using Bayesian networks,” *BMC Systems Biology*, vol. 6, no. 3, article S14, 2012.
- [47] B. Han, X.-W. Chen, and Z. Talebizadeh, “FEPI-MB: identifying SNPs-disease association using a Markov Blanket-based approach,” *BMC Bioinformatics*, vol. 12, p. S3, 2011.
- [48] B. Han, X.-W. Chen, and Z. Talebizadeh, “A fast markov blankets method for epistatic interactions detection in genome-wide association studies,” in *Proceedings of the 9th International Workshop on Data Mining in Bioinformatics (BIOKDD '10)*, 2010.
- [49] S. Lee, M.-S. Kwon, I.-S. Huh, and T. Park, “CUDA-LR: CUDA-accelerated logistic regression analysis tool for gene-gene interaction for genome-wide association study,” in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '11)*, pp. 691–695, November 2011.
- [50] J. Fontanarosa and Y. Dai, “A block-based evolutionary optimization strategy to investigate gene-gene interactions in genetic association studies,” in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '10)*, pp. 330–335, Hong Kong, December 2010.
- [51] X. Guo, Y. Meng, N. Yu, and Y. Pan, “Cloud computing for detecting high-order genome-wide epistatic interaction via dynamic clustering,” *BMC Bioinformatics*, vol. 15, no. 1, article 102, 2014.
- [52] M.-S. Kwon, M. Park, and T. Park, “IGENT: efficient entropy based algorithm for genome-wide gene-gene interaction analysis,” *BMC Medical Genomics*, vol. 7, article S6, 2014.
- [53] K. Kim, M.-S. Kwon, S. Y. Lee, J. Namkung, M. D. Li, and T. Park, “GxG-Viztool: a program for visualizing gene-gene interactions in genetic association analysis,” in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '12)*, pp. 838–843, 2012.
- [54] J. Wu, Y. Li, and R. Jiang, “Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies,” *PLoS Genetics*, vol. 10, no. 3, Article ID e1004237, 2014.