**MTHSA-DHEI: Multitasking Harmony Search Algorithm for Detecting High-*Order* Epistatic Interactions**

Shouheng Tuo 1,2,3\*, Chao Li1,2,3, Fan Liu1,2,3, Aimin Li4,Lang He1,2,3**,** Zong Woo Geem5, JunLiang Shang6, Haiyan Liu1,2,3, YanLing Zhu1,2,3, TianRui Chen1,2,3, ZengYu Feng1,2,3

1. School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, China
2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an 710121, Shaanxi, China
3. Xi'an Key Laboratory of Big Data and Intelligent Computing, Xi'an 710121, Shaanxi, China
4. School of Computer Science and Engineering, Xi’an University of Technology, Xi’an, Shaanxi, 710048, China
5. Department of Energy IT, Gachon University, Seongnam 13120, Korea
6. School of Computer Science, Qufu Normal University, Rizhao 276826, China

Corresponding author: E-mail: [tuo\_sh@126.com](mailto:tuo_sh@126.com), [tuo\_sh@xupt.edu.cn](mailto:tuo_sh@xupt.edu.cn) (S.T)

## The harmony search algorithm

The harmony search algorithm (HSA) is a new swarm intelligence optimization algorithm inspired by the process of improvising a musical harmony in a music orchestra (Geem, et al., 2001,2008) [1][2]. To solve an optimization model with *k* decision variables, a harmony represents a vector that consists of values of the *k* decision variables. Harmony memory (HM), which is composed of harmonies, is similar to the population in genetic algorithm (GA). The number of harmonies in HM is called HMS. An HM is a matrix of order HMS×*k* or an augmented matrix of order HMS × (*k* + 1) [18], as follows:



where (i=1, 2,…, HMS) denotes the i-th harmony in HM and  is the value of the objective function of .

HSA aims to optimize the harmony in HM by adjusting the pitch.

### 1.1 The steps of the standard harmony search algorithm

**Step 1. Initializing the parameters.**

The parameters include harmony memory consideration rate (HMCR), pitch adjustment rate (PAR), harmony memory size (HMS), pitch fret width (*fw*) and terminal condition (i.e., the maximum number of objective function evaluated (MaxFEs)). HMCR is primarily used to balance the exploration power and exploitation power of the algorithm. PAR and *fw* are used for local pitch adjustment.

**Step 2. Initializing the harmony memory (HM) and calculating the fitness value of each harmony.**

|  |
| --- |
| **For** i=1: HMS  ;  **For** j=1: ***k***  **Do**  ;  **While** ()  ;  **End**  // calculating the fitness value of harmony    **End** |



**Step 3. Improvising a new** **harmony.**

|  |
| --- |
| **For i=1🡪N**  **If rand(0,1) < HMCR**  🡨 randomly choose from  with probability HMCR.  **If** rand (0,1)<PAR  .  **EndIf**  **Else**  🡨 select a value from the feasible search space randomly with probability 1-HMCR  **EndIf**  **EndFor** |

**Step 4. Updating the HM.**

**If**  is better than the worst harmony  in the HM

 🡨 

**end**

**Step 5. Checking the stopping condition.**

If the stopping criterion is met, the computation is terminated. Otherwise, Steps (3) and (4) are repeated.

### 1.2 An Example for introducing the proposed algorithm



**Figure S**1. An example that a new solution (SNP combination) generated based on the harmony HM2 and four elite harmony sets of task-2 is transferred to task-1, where task-2 aims for detecting 3-order epistatic interactions and task-1 aims to detect 2-order or 4-order epistatic interactions.

## disease models

### 8 EINME models

### Table S1. Eight EINME models. The 3rd column denotes whether the model satisfies the Hardy-Weinberg equilibrium (HWE). In column 4–column 8, the values represent the prediction accuracy from k-order (k=1, 2,…,5) epistatic interaction.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | **k-order** | **HWE** | **1-order(sd)** | **2-order(sd)** | **3-order(sd)** | **4-order(sd)** | **5-order(sd)** | **tar.gz link** |
| EINME-1 | 3 | No | .502(.001) | .511(.007) | .886(.023) |  |  | [threewayBests](http://discovery.dartmouth.edu/model_free_data/evolution/threeway/threewayBests.tar.gz) |
| EINME-2 | 3 | Yes | .504(.002) | .509(.003) | .680(.024) |  |  | [HWthreewayBests](http://discovery.dartmouth.edu/model_free_data/evolution/HWthreeway/HWthreewayBests.tar.gz) |
| EINME-3 | 4 | No | .502(.001) | .510(.003) | - | .897(.018) |  | [fourwayBests](http://discovery.dartmouth.edu/model_free_data/evolution/fourway/fourwayBests.tar.gz) |
| EINME-4 | 4 | Yes | .507(.003) | .513(.003) | - | .673(.009) |  | [HWfourwayBests](http://discovery.dartmouth.edu/model_free_data/evolution/HWfourway/HWfourwayBests.tar.gz) |
| EINME-5 | 4 | No | .501(.000) | .504(.001) | .518(.003) | .567(.010) |  | [fourwayNoLowBests](http://discovery.dartmouth.edu/model_free_data/evolution/fourwayNoLow/fourwayNoLowBests.tar.gz) |
| EINME-6 | 5 | No | .502(.001) | .510(.002) | - | - | .895(.009) | [fivewayBests](http://discovery.dartmouth.edu/model_free_data/evolution/fiveway/fivewayBests.tar.gz) |
| EINME-7 | 5 | Yes | .511(.003) | .518(.003) | - | - | .693(.008) | [HWfivewayBests](http://discovery.dartmouth.edu/model_free_data/evolution/HWfiveway/HWfivewayBests.tar.gz) |
| EINME-8 | 5 | No | .503(.001) | .508(.001) | .518(.002) | .543(.004) | .690(.008) | [fivewayNoLowBests](http://discovery.dartmouth.edu/model_free_data/evolution/fivewayNoLow/fivewayNoLowBests.tar.gz) |

The eight datasets are generated by Himmelstein et al, 2011[3], which disables the discovery of disease-causing models for certain existing heuristic methods due to the lack of clues of causative SNP markers.

### 12 EIME models

### Table S2. The parameters and the values of penetrance of 12 EIME models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model type** | **EIME** | **order** | **Heritability(H2)** | **MAF** | **Heterogeneity proportion** |
| Additive model | EIME -1 | 5 | 0.1 | 0.1 | 1.0 |
| EIME -2 | 5 | 0.1 | 0.25 | 1.0 |
| EIME -3 | 5 | 0.1 | 0.5 | 1.0 |
| EIME -4 | 5 | 0.1 | 0.2 | 1.0 |
| Threshold model | EIME -5 | 5 | 0.1 | 0.1 | 1.0 |
| EIME -6 | 5 | 0.25 | 0.1 | 1.0 |
| EIME -7 | 5 | 0.5 | 0.1 | 1.0 |
| EIME -8 | 5 | 0.1 | 0.2 | 1.0 |
| Multiplicative  model | EIME -9 | 4 | 0.005 | 0.1 | 1.0 |
| EIME -10 | 4 | 0.005 | 0.2 | 1.0 |
| EIME -11 | 4 | 0.005 | 0.4 | 1.0 |
| EIME -12 | 4 | 0.004 | 0.05 | 1.0 |

**H2** denotes the genetic heritability. **MAF** represents the minor allele frequencies.

**The datasets are generated using GAMETES software [4].**

## Experimental results

**Table S3.** Experimental results of proposed algorithm for 8 EINME datasets with 100SNPs, 1k SNPs, 10k SNPs and 50k SNPs

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Models** | **Number of SNPs** | **1st Power** | **2nd Power** | **3nd Power** | **FEs** | **runtime (s)** | **Precision** | **Recall** | **F1-score** | **FDR-6** |
| EINME-1 | 100 | 100 | 100 | 100 | 1002 | 1.2 | 100% | 100% | 100% | 0 |
| EINME-2 | 100 | 100 | 100 | 100 | 995 | 1.2 | 100% | 100% | 100% | 0 |
| EINME-3 | 100 | 100 | 100 | 100 | 2601 | 4.0 | 100% | 100% | 100% | 0 |
| EINME-4 | 100 | 100 | 100 | 100 | 3315 | 5.3 | 100% | 100% | 100% | 0 |
| EINME-5 | 100 | 97 | 97 | 97 | 5100 | 10.6 | 100% | 97% | 98% | 0 |
| EINME-6 | 100 | 100 | 100 | 100 | 3800 | 6.9 | 100% | 100% | 100% | 0 |
| EINME-7 | 100 | 95 | 95 | 95 | 5838 | 12.9 | 100% | 95% | 97% | 0 |
| EINME-8 | 100 | 53 | 53 | 53 | 18362 | 32.7 | 100% | 53% | 69% | 0 |
| EINME-1 | 1000 | 100 | 100 | 100 | 8105 | 9.8 | 100% | 100% | 100% | 0 |
| EINME-2 | 1000 | 100 | 100 | 100 | 8693 | 10.2 | 100% | 100% | 100% | 0 |
| EINME-3 | 1000 | 100 | 100 | 100 | 13958 | 18.1 | 100% | 100% | 100% | 0 |
| EINME-4 | 1000 | 100 | 100 | 100 | 19292 | 26.8 | 100% | 100% | 100% | 0 |
| EINME-5 | 1000 | 93 | 93 | 93 | 83365 | 203.6 | 100% | 93% | 96% | 0 |
| EINME-6 | 1000 | 100 | 100 | 100 | 14229 | 20.2 | 100% | 100% | 100% | 0 |
| EINME-7 | 1000 | 95 | 95 | 95 | 32701 | 269.3 | 100% | 95% | 97% | 0 |
| EINME-8 | 1000 | 37 | 37 | 37 | 282899 | 526.6 | 100% | 37% | 54% | 0 |
| EINME-1 | 10000 | 100 | 100 | 100 | 12722 | 13.5 | 100% | 100% | 100% | 0 |
| EINME-2 | 10000 | 100 | 100 | 100 | 15051 | 18.3 | 100% | 100% | 100% | 0 |
| EINME-3 | 10000 | 100 | 100 | 100 | 23639 | 30.6 | 100% | 100% | 100% | 0 |
| EINME-4 | 10000 | 100 | 100 | 100 | 28969 | 42.1 | 100% | 100% | 100% | 0 |
| EINME-5 | 10000 | 87 | 87 | 87 | 141556 | 271.2 | 100% | 87% | 93% | 0 |
| EINME-6 | 10000 | 100 | 100 | 100 | 27876 | 36.4 | 100% | 100% | 100% | 0 |
| EINME-7 | 10000 | 71 | 71 | 71 | 142607 | 310.9 | 100% | 71% | 83% | 0 |
| EINME-8 | 10000 | 11 | 11 | 11 | 797659 | 3237.3 | 100% | 11% | 19.8% | 0 |
| EINME-1 | 50000 | 97 | 97 | 97 | 47096 | 71.1 | 100% | 97% | 98% | 0 |
| EINME-2 | 50000 | 93 | 93 | 93 | 98025 | 146.7 | 100% | 93% | 96% | 0 |
| EINME-3 | 50000 | 100 | 100 | 100 | 43690 | 70.0 | 100% | 100% | 100% | 0 |
| EINME-4 | 50000 | 100 | 100 | 100 | 59922 | 96.0 | 100% | 100% | 100% | 0 |
| EINME-5 | 50000 | 34 | 34 | 34 | 441725 | 739.5 | 100% | 34% | 51% | 0 |
| EINME-6 | 50000 | 100 | 100 | 100 | 29880 | 52.9 | 100% | 100% | 100% | 0 |
| EINME-7 | 50000 | 55 | 55 | 55 | 316726 | 880.5 | 100% | 55% | 71% | 0 |
| EINME-8 | 50000 | 2 | 2 | 2 | 960747 | 8153.4 | 100% | 2% | 3.9% | 0 |

**Table S4.** Experimental results of proposed algorithm for 12 EIME datasets with 100SNPs, 1k SNPs, 10k SNPs and 50k SNPs

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Models** | **Number of SNPs** | **1st Power** | **2nd Power** | **3nd Power** | **FEs** | **runtime (s)** | **Precision** | **Recall** | **F1-score** | **FDR-6** |
| EIME-1 | 100 | 100 | 100 | **100** | 2195 | 3.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-2 | 100 | 100 | 100 | **100** | 2331 | 3.7 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-3 | 100 | 100 | 100 | **100** | 2567 | 4.2 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-4 | 100 | 100 | 100 | **100** | 2261 | 3.4 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-5 | 100 | 100 | 100 | **0** | 2700 | 4.3 | / | 0.0% | / | 0 |
| EIME-6 | 100 | 100 | 100 | **100** | 2391 | 3.7 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-7 | 100 | 100 | 100 | **100** | 2173 | 3.2 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-8 | 100 | 100 | 100 | **1** | 3831 | 6.7 | 100.0% | 1.0% | 2.0% | 0 |
| EIME-9 | 100 | 100 | 100 | **0** | 1379 | 1.7 | / | 0.0% | / | 0 |
| EIME-10 | 100 | 100 | 100 | **0** | 1286 | 1.5 | / | 0.0% | / | 0 |
| EIME-11 | 100 | 100 | 100 | **100** | 1259 | 1.5 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-12 | 100 | 100 | 100 | **100** | 1242 | 1.4 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-1 | 1000 | 100 | 100 | **100** | 13106 | 17.0 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-2 | 1000 | 100 | 100 | **100** | 12931 | 17.7 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-3 | 1000 | 100 | 100 | **100** | 14619 | 22.5 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-4 | 1000 | 100 | 100 | **100** | 13142 | 18.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-5 | 1000 | 100 | 100 | **0** | 16264 | 26.0 | / | 0.0% | / | 0 |
| EIME-6 | 1000 | 100 | 100 | **100** | 12906 | 17.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-7 | 1000 | 100 | 100 | **100** | 12506 | 16.4 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-8 | 1000 | 100 | 100 | **1** | 27796 | 56.3 | 100.0% | 1.0% | 2.0% | 0 |
| EIME-9 | 1000 | 100 | 100 | **0** | 2823 | 5.5 | / | 0.0% | / | 0 |
| EIME-10 | 1000 | 100 | 100 | **0** | 2694 | 5.3 | / | 0.0% | / | 0 |
| EIME-11 | 1000 | 100 | 100 | **100** | 2690 | 5.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-12 | 1000 | 100 | 100 | **100** | 2559 | 4.8 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-1 | 10000 | 100 | 100 | 100 | 18378 | 27.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-2 | 10000 | 100 | 100 | 100 | 18080 | 26.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-3 | 10000 | 100 | 100 | 100 | 20787 | 33.8 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-4 | 10000 | 100 | 100 | 100 | 18436 | 27.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-5 | 10000 | 100 | 100 | 0 | 24061 | 40.7 | / | 0.0% | / | 0 |
| EIME-6 | 10000 | 100 | 100 | 100 | 20166 | 31.1 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-7 | 10000 | 100 | 100 | 100 | 18189 | 25.6 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-8 | 10000 | 100 | 100 | 1 | 40976 | 89.8 | 100.0% | 1.0% | 2.0% | 0 |
| EIME-9 | 10000 | 99 | 99 | 0 | 14488 | 23.2 | / | 0.0% | / | 0 |
| EIME-10 | 10000 | 95 | 98 | 0 | 13711 | 21.4 | / | 0.0% | / | 0 |
| EIME-11 | 10000 | 100 | 100 | 100 | 12372 | 17.2 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-12 | 10000 | 100 | 100 | 100 | 13254 | 18.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-1 | 50000 | 100 | 100 | 100 | 35431 | 50.5 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-2 | 50000 | 100 | 100 | 100 | 37271 | 53.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-3 | 50000 | 100 | 100 | 100 | 41014 | 63.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-4 | 50000 | 100 | 100 | 100 | 35711 | 50.1 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-5 | 50000 | 98 | 98 | 0 | 46837 | 77.8 | / | 0.0% | / | 0 |
| EIME-6 | 50000 | 100 | 100 | 100 | 36569 | 52.3 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-7 | 50000 | 100 | 100 | 100 | 35661 | 49.4 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-8 | 50000 | 99 | 99 | 1 | 82610 | 167.7 | 100.0% | 1.0% | 2.0% | 0 |
| EIME-9 | 50000 | 91 | 91 | 0 | 20907 | 25.4 | / | 0.0% | / | 0 |
| EIME-10 | 50000 | 89 | 89 | 0 | 20163 | 23.6 | / | 0.0% | / | 0 |
| EIME-11 | 50000 | 100 | 100 | 100 | 19946 | 22.9 | 100.0% | 100.0% | 100.0% | 0 |
| EIME-12 | 50000 | 100 | 100 | 100 | 15591 | 16.7 | 100.0% | 100.0% | 100.0% | 0 |

## References

[1]. Z.W. Geem, J. Kim, G. Loganathan, Music-inspired optimization algorithm harmony search, Simulation 76, 60-68(2001).

[2]. Z, W. Geem. Novel derivative of harmony search algorithm for discrete design variables. Applied Mathematics & Computation 199(1), 223-230(2008).

[3] Himmelstein et al. Evolving hard problems: Generating human genetics datasets with a complex etiology. BioData Mining 4, 21(2011). doi:10.1186/1756-0381-4-21. <http://discovery.dartmouth.edu/model_free_data/>.

[4] Urbanowicz, R. J., Kiralis, J., Sinnott-Armstrong, N. A., Heberling, T., Fisher, J. M. & Moore, J. H. GAMETES: a fast, direct algorithm

for generating pure, strict, epistatic models with random architectures. ***BioData mining*** 2012; 5:1–14.