

# Identifying SNP Interactions with Multiple Phenotypes using Fish-Agent Logic Regression

Li Wei<sup>1,2</sup> and Shaojie Jiao<sup>1,2</sup>

<sup>1</sup>Department of Computer Science and Engineering,  
University of Connecticut,  
Storrs, CT 06269, U.S.A.

<sup>2</sup>Joint Lab for Pervasive Healthcare Systems,  
University of Connecticut - Xi'an Jiaotong University  
[li.wei@engr.uconn.edu](mailto:li.wei@engr.uconn.edu)

**Abstract.** As single nucleotide polymorphisms (SNPs) and their interactions are considered to predispose to human disease, it is crucial to identify the interacting SNPs that influence disease susceptibility in disease association studies. Logic Regression (LR) approach is proposed to test the association between the candidate SNPs and the case / control status, in which Logic Tree (LT) is used to model the relationships among selected SNPs by boolean operators (“AND” or “OR”). Monte Carlo Markov Chain (MCMC) has been incorporated into original LR method and then a fish-swarm framework is designed to optimal the MCMC search results and speed up the convergence process. Although those approaches has been applied successfully, however, symbiosis among diseases which has shown in clinical research may cause a loss of statistical power. Some non-causal SNPs may indicate false positive because they truly associate with another disease related to the one we focus on. It is therefore better to detect the genotype-phenotype association on multiple related phenotypes. In this article, we design a novel fish-agent swarm framework to handle multiple phenotypes. In this framework, we create a number of agents for each of the phenotypes. Each agent contains a LT as a solution, while two behaviors, “Follow” and “Exclude”, are defined on agent to enhance the convergence and increase the power, separately. In each iteration, an agent selects one behavior to gain a better solution by modifying the current LT, where the probability of selecting a behavior is computed by the following factors: the current best solutions for the same and related phenotypes, distances between the agent itself and the one holds the best solution in the solution space, and a prior distribution on phenotypes and a posterior distribution on all candidate SNPs. We finally show the performance of our approach on several simulation experiments.

**Keywords:** Disease association studies, Logic regression, Fish-swarm algorithm, Multiple phenotypes

**Acknowledgments.** Research is supported by Joint Lab Foundation for Undergraduate Innovation [110118].