#051

BICA VPS 2023

Novel Feature Selection Method Based on Slime Mold Network Formation Behavior.

Chenyang Yan[1]. [1]Ningbo City College of Vocational Technology, Ningbo Zhejiang ,310000, China yanchenyang@nbcc.edu.cn

First Author

SUMMARY

- Inspired by the adaptive network formation behavior of slime mold, we propose a novel feature selection algorithm (SMFS).
- The results on several benchmark datasets demonstrate the efficiency and effectiveness of the SMFS method.

INTRODUCTION

Some non-neuronal organisms, despite having no intelligence [1]. We are particularly fascinated by the work of Tero and Nakagaki [2], who showed the remarkable ability of Physarum to design adaptive networks. They set up an experiment with a plate containing 36 food sources that corresponded to the geo-graphic locations of cities in the Tokyo area. They simulated the geographic features by varying the intensity of light, since Physarum tends to avoid bright light. They placed a small plasmodium of Physarum at the location of Tokyo and observed its behavior. Initially, the Physarum occupied most of the available space on the plate, but then it refined its network by thinning out some connections and leaving behind a subset network. This network not only resembled the Tokyo rail system visually, but also had similar efficiency, fault tolerance, and cost as the networks designed by human engineers.

APPROACH

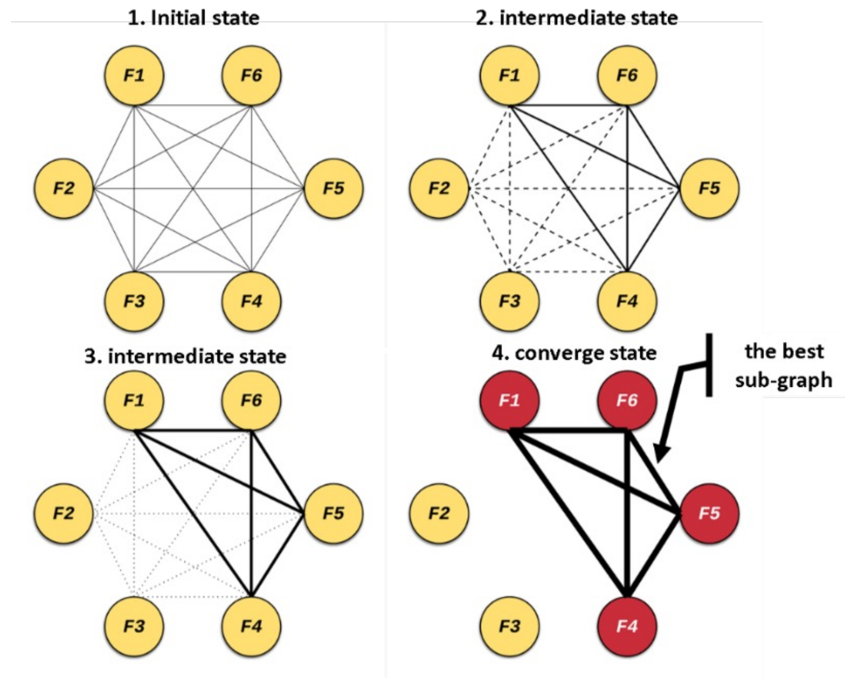

Fig. 1. Imitate the network formation behavior of slime mold to solve the feature selection problem

METHODS

Inspired by Tero and Nakagaki's work, we transform the feature selection problem into an optimal subgraph problem and construct an algorithm based on the rules of the slime mold to solve the problem. In a nutshell, the candidate features are considered vertices in a complete undirected graph. These vertices are also the virtual source. The edges of the graph are the initial tubes. The virtual slime plasmodia flows and forages through these tubes according to specific rules. In the end, the morphology of the slime tubes will converge to form a complete sub-graph. All vertices on the subgraph are the selected features .We call this model SMFS, a Slime Mold inspired Feature Selection algorithm.

RESULTS

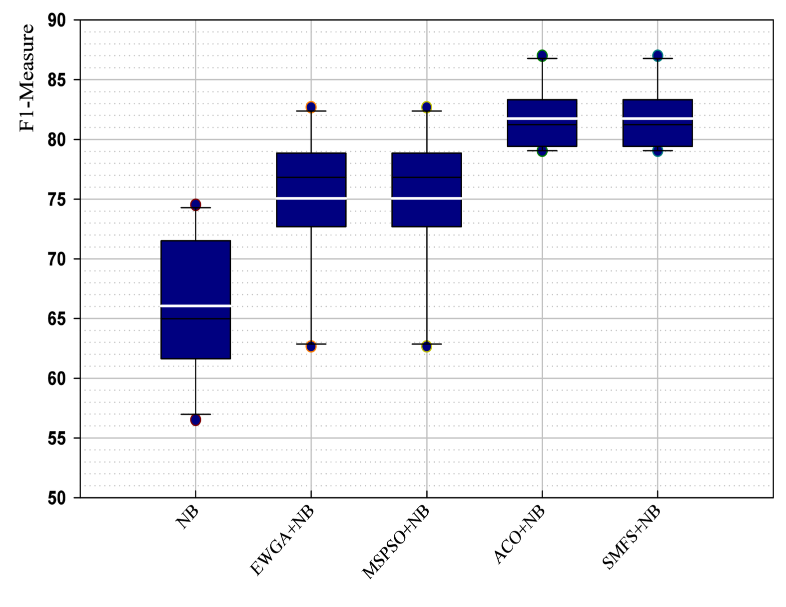

Fig2 F1-Measure scores of different algorithms in 10-fold cross-validation. The SMFS (feature importance / Multinomial Naive Bayes) is our proposed method, the Multinomial Naive Bayes is the benchmark algorithm, and the EWGA, MSPSO, and ACO are the comparison algorithms. The solid line shows the mean of the scores.

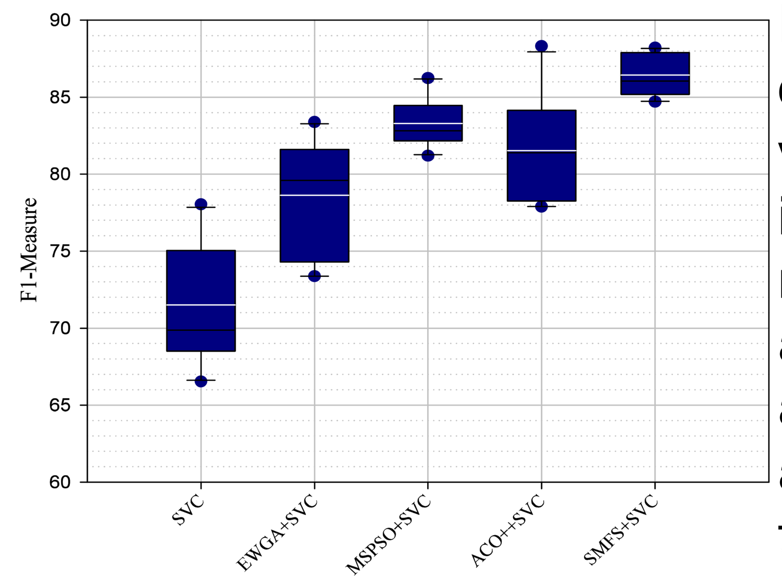

Fig3 F1-Measure scores obtained by different algorithms in 10-fold cross-validation. The SMFS (feature importance /SVC) is our proposed method, the SVC is the benchmark algorithm, and the EWGA, MSPSO, and ACO are the comparison algorithms. The solid line represents the mean of the scores.

ANALYSIS

Figure 2 and 3 show a box diagram of the F1-Measure scores obtained by the SMFS variants using different short-term return calculation methods (L2-regularized tf-idf score or feature importance) and different classifiers (L2-regularized L2-loss SVC and Multinomial Naive Bayes), comparison algorithms (SVC, NB without feature selection, EWGA, MSPSO, and ACO) in 10 cross-check validations. It can be observed that regardless of the classifier used, the F1-Measure obtained by SMFS is far better than the benchmark algorithm and EWGA and slightly better than ACO and MSPSO.

DISCUSSION

SMFS trades time performance for the ability to select better feature sets. In practice, feature selection is often part of data preprocessing, and final classification performance is often more important than runtime, so the performance of SMFS can be considered to be superior to MSPOS to some extent.

Combined with the above test results, SMFS is superior to the benchmark algorithm, EWGA, and ACO algorithm and slightly better than the MSPSO algorithm in the application scenario that is not sensitive to the algorithm run time.

CONCLUSIONS

1. The SMFS algorithm can select a proper subset of features and obtain good classification performance on the test dataset.
2. SMFS is just a preliminary exploration of the application of slime mold net-work formation mechanism. SMFS's availability and reliability cannot be compared with the state-of-the-art feature selection algorithms.

ACKNOWLEDGMENTS

This work was supported in part by the National (China) Educational Technology Research and Planning Project [Grant Number 166243001].