WANG Xubin

Website: wangxubin.site
Google Scholar:Xubin Wang

Email: wangxb19@mails.jlu.edu.cn
Github: github.com/wangxb96

Education

- **Jilin University** — Changchun, China
  MPhil - Computer Science and Technology — *September 2019 - June 2022*
  Advisor: *Prof. Xiangtao Li*

- **Tianjin Normal University** — Tianjin, China
  B.Eng. - Computer Science and Technology — *September 2015 - June 2019*

Experience

- **Beijing Normal Universsity** — Zhuhai, China
  Research Assistant at Institute of Artificial Intelligence and Future Networks — *July 2022 - Now*

Skills Summary

- **Programming Languages**: Python, Matlab, C/C++, SQL, Bash
- **Tools**: Git, Latex, Markdown, MySQL, SQLite
- **English Ability**: CET-4: 538, CET-6: 511 (*Passing Scores: 425*)
- **Languages**: Mandarin Chinese (native); Wu Chinese, Qingyuan Dialect (native); English

Publications

- [**U2**] **Wang, X.**, Wang, Y., Ma, Z., Wong, K. C., & Li, X. Evolving Ras Pathway Activation from Cancer Gene Expression Data using Nature-inspired Ensemble Optimization - Submitted to *TCBB*
- [**U1**] **Wang, X.**, Wang, Y., Ma, Z., Wong, K. C., & Li, X. Exhaustive Exploitation of Nature-inspired Computation for Cancer Screening in an Ensemble Manner - Submitted to *TCBB*
- [**C1**] Gao, H., Bian, C., **Wang, X.**, Li, X., & Wang, Y. (2022). Exploring Cancer Biomarker Genes from Gene Expression Data via Nature-inspired Multiobjective Optimization. *CCDC 2022*.
- [**J2**] Bian, C., **Wang, X.**, Su, Y., Wang, Y., Wong, K. C., & Li, X. (2022). scEFSC: Accurate Single-cell RNA-seq Data Analysis via Ensemble Consensus Clustering Based on Multiple Feature Selections. *Computational and Structural Biotechnology Journal*.
- [**J1**] **Wang, X.**, Wang, Y., Wong, K. C., & Li, X. (2022). A self-adaptive weighted differential evolution approach for large-scale feature selection. *Knowledge-Based Systems*, 235, 107633.

Projects

- **High-dimensional Data Processing Project (Feature Selection, Heuristic Optimization)**: We proposed a novel weighted differential evolution algorithm based on self-adaptive mechanism, named SaWDE, to solve large-scale feature selection problem. In SaWDE, a multi-population mechanism is employed, while an self-adaptive mechanism is proposed to select an appropriate strategy, based on which we design a weighting model to measure the importance of each feature. The experimental results on 24 high-dimensional datasets show that SaWDE can achieve excellent performance on both the training and test datasets compared to twelve state-of-the-art methods. Moreover, the SaWDE algorithm could reduce 75% to 80% of the features in many cases. **Code:** https://github.com/wangxb96/SaWDE
- **Implementation of a Book Recommender System (Collaborative Filtering Algorithm, Euclidean Distance, Pearson's Correlation Coefficient)**: In this graduation design, a user-based collaborative filtering algorithm is employed to calculate the user-to-user similarity and to recommend the products with higher final ratings based on the user rating matrix. Specifically, Euclidean distance and Pearson's correlation coefficient are used to calculate the similarity, and two recommendation lists are obtained to recommend products that are common to both lists. If there is no common item, the item with higher rating is recommended. This design was highly rated by three reviewers: the scores were 93, 90, and 93, which were the highest in the grade.
- **Precision Cancer Identification Program (GWO Optimization, Ensemble Learning)**: To address the challenges in cancer detection, we proposed a framework called EODE, which incorporates the grey wolf optimizer (GWO) to optimize feature subsets and then creates an optimized ensemble classifier in a collaborative manner. We conducted experiments on 35 datasets containing multiple types of cancer, and the experimental results showed the effectiveness of our algorithm compared to six benchmark machine learning algorithms (KNN, DT, ANN, SVM, DISCR and NB), six state-of-the-art ensemble algorithms (RF, ADABOOST, RUSBOOST, SUBSPACE, TOTALBOOST and LPBOOST), seven nature-inspired methods (ACO, CS, DE, GA, GWO, PSO and ABC), and four nature-inspired ensemble methods. **Code:** https://github.com/wangxb96/EODE

- **Imbalanced Data Processing Project (Imbalanced Ensemble Learning, ACO Optimization)**: To address the challenge of unbalanced data, a model named hybrid sampling nature-inspired optimization ensemble (HSNOE) is developed, which preprocesses the original cancer gene expression data by a hybrid sampling method and then integrates the ant colony optimization (ACO) to optimize the feature subsets and the generation of an optimized ensemble in a collaborative manner. Experiments based on Pan-cancer dataset from 33 different cancer types across The Cancer Genome Atlas (TCGA) PanCanAtlas project demonstrate that HSNOE outperforms state-of-the-art benchmark methods, achieving an AUROC of nearly 94% on the test set. **Code:** https://github.com/wangxb96/HSNOE
- **Time-series Prediction Project (Feature Selection, RNN)**: The time-series data are screened out of redundant and irrelevant features using feature selection methods, then the data are normalized using normalization approaches, on which the LSTM is used for prediction, while short data complementation and interval sampling are used to improve the model robustness. We finally achieve a prediction accuracy of more than 90%.

Honors and Awards

- The First Prize Scholarship × 2
- The Second Prize Scholarship × 1
- Merit Student × 3
- Postgraduate Academic Scholarship × 1

Practical Activities

- **Teaching Assistant for Postgraduate Examination** Changchun, China
  Instructed students in the field of data structures and operating systems. *Aug. 2019 - Dec. 2019*
- **Teaching Assistant for Undergraduate Students** Changchun, China
  Organized events & instructed experiments in the class Introduction to Artificial Intelligence. Oct. 2020 - Nov. 2020