Tutorial Proposal IEEE Congress on Evolutionary Computation 2025 (June 08-12, 2025, Hangzhou, China)

Title: Evolutionary Feature Reduction for Machine Learning

Introduction:

We are now in the era of big data, where vast amounts of high-dimensional data become ubiquitous in a variety of domains, such as social media, healthcare, and cybersecurity. When machine learning algorithms are applied to such high-dimensional data, they suffer from the curse of dimensionality, where the data becomes very sparse. Furthermore, the highdimensional data might contain redundant and/or irrelevant features that blur useful information from relevant features. Feature reduction can address the above issues by building a smaller but more informative feature set.

Feature selection and *feature construction* are two main approaches of feature reduction. *Feature selection* aims to select a small subset of original (relevant) features. *Feature construction* aims to create a small set of new high-level (informative) features based on the original feature set. Although both approaches are essential data pre-processing steps, they are challenging due to their large search spaces and complex interactions between features. While exhaustive searches are impractical due to their intensive computation cost, traditional heuristic searches require less computational resources but can be trapped at local optima. Recently, evolutionary computation (EC) has been widely applied to achieve feature reduction because of its potential global search ability. Existing EC-based feature reduction approaches successfully reduce the data dimensionality while still improve the classification performance as well as the interpretability of the built models.

This tutorial firstly describes a general framework of feature reduction followed by the applications of feature reduction in real-world scenarios. Then, we will show how EC techniques, particularly genetic algorithms, particle swarm optimisation, differential evolution, genetic programming, ant colony optimisation and evolutionary multi-objective optimisation (EMO) can be applied to address challenges in feature reduction. The effectiveness of EC-based feature reduction is illustrated through several applications including bioinformatics, image analysis and pattern classification, symbolic regression, and cybersecurity. The tutorial concludes with existing challenges for future research.

Outline of the tutorial:

1. Introduction to feature reduction

- a. What is feature reduction?
 - i. providing definitions of feature reduction
 - ii. feature selection vs feature construction
- b. Why is feature reduction necessary?
 - i. illustrating through real-world examples

- c. How to perform feature reduction?
 - i. providing a general framework of feature reduction
 - ii. avoid feature reduction bias
- d. How to category feature reduction approaches based on the fitness function?
 - i. advantages and disadvantages of filters/wrappers/embedded approaches
- **2. Feature selection:** this section reviews existing works based on how feature selection is represented in different EC algorithms (representations)
 - a. Graph-based representations:
 - i. mostly ant colony optimisation
 - b. Tree-based representations:
 - i. mostly genetic programming
 - c. Vector-based representations: most widely used representations
 - i. genetic algorithms
 - ii. particle swarm optimisation
 - iii. differential evolution
 - iv. EMO algorithms such as NSGAII and MOEA/D
- **3. Feature construction**: this section mostly reviews existing works applying genetic programming that is the most suitable EC algorithm for feature construction
 - a. Why genetic programming for feature construction?
 - b. Single-tree representations
 - c. Multi-tree representations
- **4. Hybridisation of feature selection and feature construction:** this section reviews an emerging topic in which a subset of original features is combined with a set of new high-level features.
- **5. Real-world applications of feature reduction:** this section illustrates several examples of real-world applications where feature reduction is successfully applied to boost the learning performance
 - a. Bioinformatics
 - b. Image analysis and pattern classification
 - c. Symbolic regression
 - d. Cybersecurity

6. Existing challenges

- a. Large search spaces mainly due to the representations
- b. Intensive computational cost
- c. Poor scalability especially when the number of instances is large
- d. Overfitting especially for feature construction

Length of the tutorials: Two hours

Level of the tutorial: Introductory

Organisers:

Dr Bach (Hoai) Nguyen

<u>Hoai.Bach.Nguyen@ecs.vuw.ac.nz</u> Centre for Data Science and Artificial Intelligence (CDSAI) & School of Engineering and Computer Science Victoria University of Wellington (VUW), Wellington 6140, New Zealand Homepage: https://people.wgtn.ac.nz/bach.nguyen/

A/Prof Ruwang Jiao

rwjiao@suda.edu.cn School of Future Science and Engineering Soochow University, China Homepage: https://web.suda.edu.cn/rwjiao/

Prof Bing Xue

Bing.Xue@ecs.vuw.ac.nz CDSAI & School of Engineering and Computer Science VUW, Wellington 6140, New Zealand Homepage: <u>https://people.wgtn.ac.nz/bing.xue/</u>

Prof Mengjie Zhang

Mengjie.Zhang@ecs.vuw.ac.nz CDSAI & School of Engineering and Computer Science VUW, Wellington 6140, New Zealand Homepage: <u>https://people.wgtn.ac.nz/mengjie.zhang/</u>

Biography of the organisers:

Dr Bach Nguyen is currently a Lecturer in Artificial Intelligence at the CDSAI & School of Engineering and Computer Science, Victoria University of Wellington (VUW). He has over 40 publications in fully refereed international journals and conferences. His research focuses mainly on evolutionary computation, machine learning, classification, feature selection, transfer learning, and multi-objective optimisation.

Dr Nguyen is currently the Vice-Chair of the IEEE CIS Data Mining and Big Data Analytics Technical Committee, the Vice-Chair of the IEEE Task Force on Evolutionary Feature Selection and Construction, and the Vice-Chair of the IEEE New Zealand Central Section.

Dr Nguyen co-chaired of IEEE Symposium on Computational Intelligence in Data Mining in IEEE Symposium on Computational Intelligence (SSCI) 2021, 2022. He was the organiser of the Special Session on Evolutionary Feature Selection, Construction, and Extraction in IEEE Congress on Evolutionary Computation (CEC) in 2021, 2022, 2023, and 2024. He also organized the Special Session on Evolutionary Transfer Learning and Domain Adaptation in SSCI 2021 and 2022. He delivered a Turorial on Evolutionary Feature Reduction in IEEE CEC 2021, IEEE WCCI 2022, IEEE CEC 2023, and IEEE WCCI 2024.

Dr Nguyen has been serving as a program committee member for over 10 international conferences including AAAI, IJCAI, IEEE CEC, GECCO, and IEEE SSCI. He has been serving as a reviewer for over 10 international journals including IEEE Transactions on Evolutionary Computation and IEEE Transactions on Cybernetics.

A/Prof Ruwang Jiao is currently an associate professor at Soochow University, China. He was a research fellow in artificial intelligence at Victoria University of Wellington, New Zealand. He received the Humboldt Research Fellowship and the JSPS International Research Fellowship in 2024. His research mainly focuses on feature selection, dimensionality reduction, and evolutionary multi-objective learning.

Dr Jiao currently serves as the Chair of the IEEE Taskforce on Evolutionary Computation for Feature Selection and Construction, and a member the Technical Committee on Evolutionary Computation of the IEEE Intelligent Computing Society. He co-chaired the special session on EMOML at IEEE CEC 2023 and IEEE WCCI 2024 and delivered a tutorial at IEEE SSCI 2023. He is an Associate Editor for the Journal of Intelligent & Fuzzy Systems and a young editor for Intelligent Marine Technology and Systems. He has been invited to serve as a PC member for international conferences such as ECAI, IEEE ICDM, GECCO, IEEE CEC, IEEE SSCI, and Evostar. He has published over 30 papers in esteemed journals and conferences, including IEEE TEVC, ECJ (MIT Press), IEEE TCYB, and IEEE TAP.

Prof Bing Xue is currently a Professor and Deputy Head of School in School of Engineering and Computer Science at VUW. She is also Deputy Director of the CDSAI and Fellow of Engineering of New Zealand. She has over 400 papers published in fully refereed international journals and conferences and her research focuses mainly on evolutionary computation, machine learning, classification, symbolic regression, feature selection, evolving deep neural networks, image analysis, transfer learning, multi-objective machine learning.

Prof Xue is currently a Member of the IEEE Computational Intelligence Society AdCom, a Member of the ACM SIGEVO Executive Committee, and the Vice-Chair of IEEE Task Force on Evolutionary Feature Selection and Construction. She has also served as an Associate Editor of 8 international journal articles including IEEE Transactions on Evolutionary Computation, IEEE Computational Intelligence Magazine, and IEEE Transactions on Artificial Intelligence.

Prof Xue has been organising many international conferences, such as General Chair of PRICAI 2025, Conference Chair of IEEE CEC 2024 and EuroGP 2024. She has also been the organiser of the Special Session on Evolutionary Feature Selection and Construction in IEEE Congress on Evolutionary Computation (CEC) 2015 - 2020. She was a chair for a number of international conferences including the Chair of Women@GECCO 2018 and a co-Chair of the Evolutionary Machine Learning Track for GECCO 2019-2022. She is the Lead Chair of IEEE Symposium on Computational Intelligence in Feature Analysis, Selection, and Learning in Image and Pattern Recognition (FASLIP) at SSCI 2016-2022, a Program Co-Chair of the 7th International Conference on Soft Computing and Pattern Recognition (SoCPaR2015), a Program Chair of the 31st Australasian Joint Conference on Artificial Intelligence (AI 2018), Finance Chair for 2019 IEEE Congress on Evolutionary Computation (CEC), Tutorial Co-Chair of 2022 IEEE WCCI, and Conference Chair of 2024 IEEE CEC.

Prof Mengjie Zhang is a Fellow of Royal Society of New Zealand, a Fellow of Engineering New Zealand, a Fellow of IEEE, an IEEE Distinguished Lecturer, currently Professor of Computer Science at Victoria University of Wellington, where he heads the interdisciplinary

Evolutionary Computation and Machine Learning Research Group. He is the Director of the Centre for Data Science and Artificial Intelligence at the University.

His research is mainly focused on AI, machine learning and big data, particularly in evolutionary learning and optimisation, feature selection/construction and big dimensionality reduction, computer vision and image analysis, scheduling and combinatorial optimisation, classification with unbalanced data and missing data, and evolutionary deep learning and transfer learning. Prof Zhang has published over 900 research papers in refereed international journals and conferences. He has been serving as an associated editor for over ten international journals including IEEE Transactions on Evolutionary Computation, IEEE Transactions on Cybernetics, the Evolutionary Computation Journal (MIT Press), and involving many major AI and EC conferences as a chair. He received the "EvoStar/SPECIES Award for Outstanding Contribution to Evolutionary Computation in Europe" in 2023. Since 2007, he has been listed as a top five (currently No. 3) world genetic programming researchers by the GP bibliography (http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/index.html). He is also a Clarivate Highly Cited Researcher in the field of Computer Science -- 2023.

Prof Zhang is currently the Chair for IEEE CIS Awards Committee. He is also a past Chair of the IEEE CIS Intelligent Systems Applications Technical Committee, the Emergent Technologies Technical Committee and the Evolutionary Computation Technical Committee, a past Chair for IEEE CIS PubsCom Strategic Planning subcommittee, and the founding chair of the IEEE Computational Intelligence Chapter in New Zealand.