Tutorial Proposal
IEEE Congress on Evolutionary Computation 2023
(02 - 05 July 2023, Chicago, US. CEC 2023)

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 high-dimensional 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
Hoai.Bach.Nguyen@ecs.vuw.ac.nz
School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6140, New Zealand
Homepage: https://people.wgtn.ac.nz/bach.nguyen/

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

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

Dr Bach Nguyen is currently a Lecturer in Artificial Intelligence at School of Engineering and Computer Science, Victoria University of Wellington (VUW). He has over 30 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 Chair of the IEEE Task Force on Evolutionary Feature Selection and Construction, and the Chair of Young Professionals Affinity Group of 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 in CEC 2021 and WCCI 2022. He also organized the Special Session on Evolutionary Transfer Learning and Domain Adaptation in SSCI 2021 and 2022. He delivered a Tutorial on Evolutionary Feature Reduction in CEC 2021 and WCCI 2022.

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.

Prof Bing Xue is currently a Professor and Deputy Head of School in School of Engineering and Computer Science at VUW. She has over 200 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 the Chair of the Evolutionary Computational Technical Committee (IEEE CIS), the Chair of IEEE Task Force on Evolutionary Deep Learning and Applications, the Vice-Chair of IEEE Task Force on Evolutionary Feature Selection and Construction, and the Vice-Chair of IEEE CIS Task Force on Transfer Learning & Transfer Optimization.

Prof Xue is the organiser of the special session on Evolutionary Feature Selection and Construction in IEEE Congress on Evolutionary Computation (CEC) 2015 - 2020. Prof Xue has been 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.

She is an Editor of the IEEE Computational Intelligence Society Newsletter, an Associate Editor or Member of the Editorial Board for seven international journals, including IEEE Transactions of Evolutionary Computation, IEEE Computational Intelligence Magazine, and ACM Transactions on Evolutionary Learning and Optimisation.

Prof Mengjie Zhang is a Fellow of Royal Society of New Zealand, a Fellow of IEEE, a Panel Member of the Marsden Fund (New Zealand Government Funding), and currently Professor of Computer Science at Victoria University of Wellington, where he heads the interdisciplinary Evolutionary Computation Research Group. He is a member of the University Academic Board, a member of the University Postgraduate Scholarships Committee, Associate Dean (Research and Innovation) in the Faculty of Engineering, and Chair of the Research Committee of the Faculty of Engineering and School of Engineering and Computer Science.

His research is mainly focused on evolutionary computation, particularly genetic programming, particle swarm optimisation and learning classifier systems with application areas of feature selection/construction and dimensionality reduction, computer vision and image processing, evolutionary deep learning and transfer learning, job shop scheduling, multi-objective optimisation, and clustering and classification with unbalanced and missing data. He is also interested in data mining, machine learning, and web information extraction. Prof Zhang has published over 500 research papers in refereed international journals and conferences in these areas.

He has been serving as an associated editor or editorial board member for over 10 international journals including IEEE Transactions on Evolutionary Computation, IEEE Transactions on Cybernetics, the Evolutionary Computation Journal (MIT Press), ACM Transactions on Evolutionary Learning and Optimisation, Genetic Programming and Evolvable Machines (Springer), IEEE Transactions on Emergent Topics in Computational Intelligence, Applied Soft Computing, and Engineering Applications of Artificial Intelligence, and as a reviewer of over 30 international journals. He has been a major chair for 8 international conferences. He has also been serving as a steering committee member and a program committee member for over 80 international conferences including all major conferences in evolutionary computation.

Since 2007, he has been listed in the top five world genetic programming researchers by the GP bibliography (http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/index.html).

He is the Tutorial Chair for GECCO 2014, an AIS-BIO Track Chair for GECCO 2016, an EML Track Chair for GECCO 2017, and a GP Track Chair for GECCO 2020. Since 2012, he has been co-chairing several parts of IEEE CEC, SSCI, and EvoIASP/EvoApplications conference (he has been involving major EC conferences such as GECCO, CEC, EvoStar, SEAL). Since 2014, he has been co-organising and co-chairing the special session on evolutionary feature selection and construction at IEEE CEC and SEAL, and also delivered a keynote/plenary talk for IEEE CEC 2018, IEEE ICAVSS 2018, DOCSA 2019, IES 2017 and Chinese National Conference on AI in Law 2017.

Prof Zhang was the Chair of the IEEE CIS Intelligent Systems Applications, the IEEE CIS Emergent Technologies Technical Committee, and the IEEE CIS Evolutionary Computation Technical Committee; a Vice-Chair of the IEEE CIS Task Force on Evolutionary Feature Selection and Construction, the IEEE CIS Task Force on Evolutionary Computer Vision and Image Processing, and the IEEE CIS Task Force on Evolutionary Deep Learning and Applications; and also the founding chair of the IEEE Computational Intelligence Chapter in New Zealand.