Evolutionary Computation for Feature Selection and Feature Construction

Bing Xue and Mengjie Zhang
Evolutionary Computation Research Group
Victoria University of Wellington, New Zealand

Bing.Xue@ecs.vuw.ac.nz and Mengjie.Zhang@ecs.vuw.ac.nz

http://gecco-2021.sigevo.org/

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07…$15.00
https://doi.org/10.1145/3449726.3461415

Instructors

- **Bing Xue** is a currently a Professor in Computer Science and Program Director of Science in School of Engineering and Computer Science at VUW. She has over 300 papers published in fully refereed international journals and conferences and her research focuses mainly on evolutionary computation, feature selection, feature construction, machine learning, classification, symbolic regression, evolving deep neural networks, image analysis, transfer learning, multi-objective machine learning. Dr Xue is currently the founding Chair of IEEE Computational Intelligence Society (CIS) Task Force on Evolutionary Feature Selection and Construction, Vice-Chair of IEEE CIS Task Force on Transfer Learning & Transfer Optimization, Vice-Chair of IEEE CIS Task Force on Evolutionary Deep Learning and Applications. She is also served as associate editor of several international journals, such as IEEE Transactions on AI, IEEE Computational Intelligence Magazine and IEEE Transactions on Evolutionary Computation.

- **Mengjie Zhang** is a Fellow of Royal Society of NZ, a Fellow of IEEE, and an IEEE Distinguished Lecturer. He is Professor of Computer Science at the School of Engineering and Computer Science, Victoria University of Wellington, New Zealand. His research is mainly focused on evolutionary computation, particularly genetic programming, evolutionary deep and transfer learning, image analysis, feature selection and reduction, and evolutionary scheduling and combinatorial optimisation. He has published over 600 academic papers in refereed international journals and conferences. He is currently an associate editor for over ten international journals (e.g. IEEE TEVC, ECJ, ACM TELO, IEEE TCYB, and IEEE TETCI). He has been serving as a steering committee member and a program committee member for over eighty international conferences. He is a reviewer of research grants for many countries/regions (e.g. Canada, Portugal, Spain, Germany, UK, Netherland, Austria, Mexico, Czech, Italy, HK, Australia, NZ).

Outline

- Feature Selection and Feature Construction
- Evolutionary Computation (EC) for Feature Selection and Feature Construction
- Evolutionary Feature Selection Methods
- Evolutionary Feature Construction Methods
- Issues and Challenges

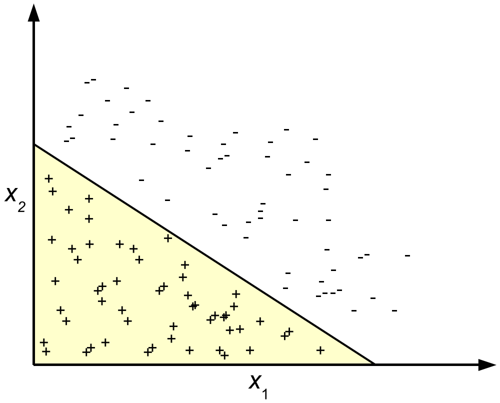
Data set (Classification) — Example 1

Credit card application:

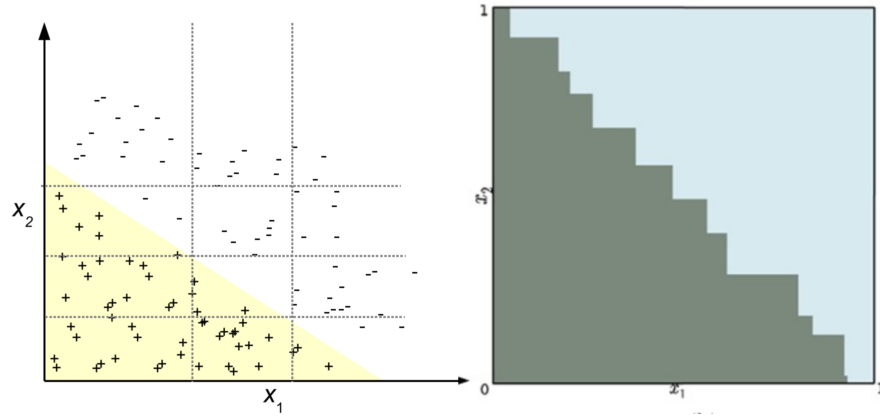
- 7 applicants (examples/instances/observations)
- 2 classes: Approve, Reject
- 3 features/variables/attributes

	Job	Saving	Family	Class
Applicant 1	true	high	single	**Approve**
Applicant 2	false	high	couple	**Approve**
Applicant 3	true	low	couple	**Reject**
Applicant 4	true	low	couple	**Approve**
Applicant 5	true	high	children	**Reject**
Applicant 6	false	low	single	**Reject**
Applicant 7	true	high	single	**Approve**

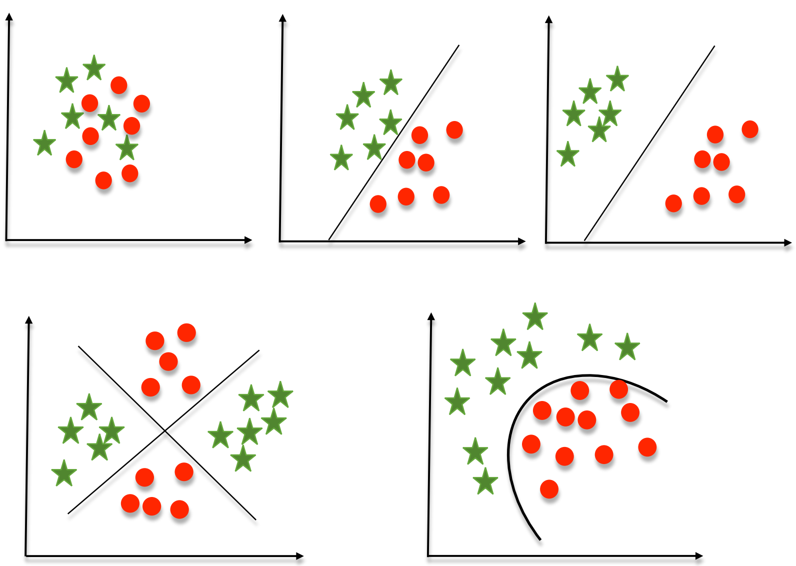
What is a good feature?



What is a good feature?

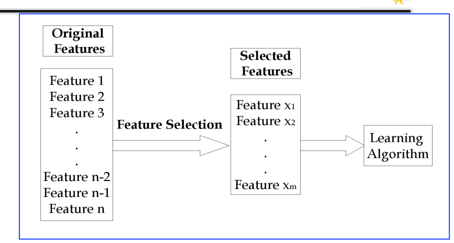


What is a good feature?

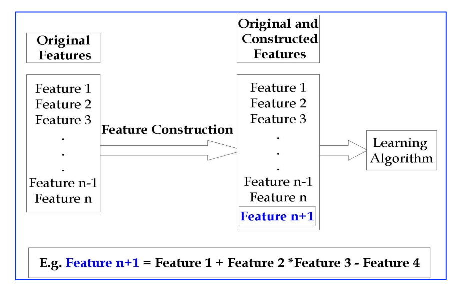


Feature Selection and Feature Construction

- Feature selection aims to pick a subset of relevant features to achieve similar or better classification performance than using all features.



- Feature construction is to construct new high-level features using original features to improve the classification performance.



Increasingly Popular Topic

Popular Articles | Latest Published Articles

A Fast And Elitist Multiobjective Genetic Algorithm: NSGA-II — NSGAII
K. Deb; A. Pratap; S. Agarwal; T. Meyarivan

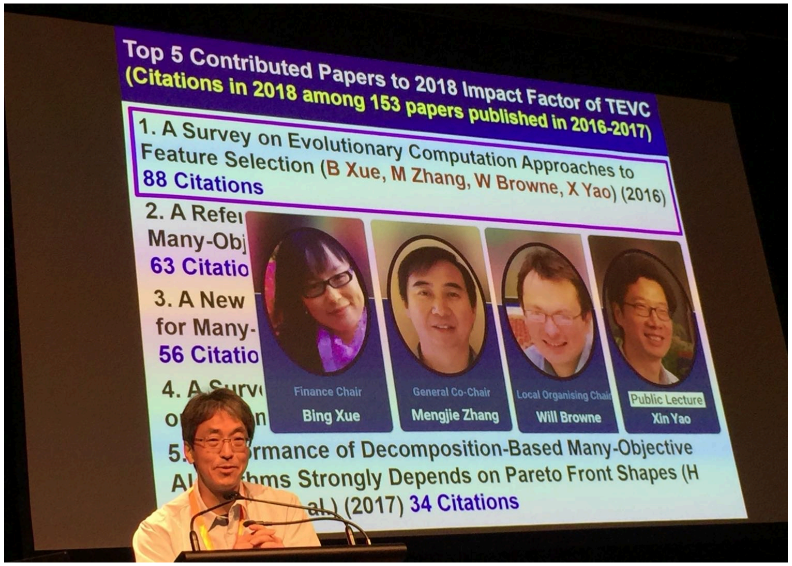
An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints — NSGAIII
Kalyanmoy Deb; Himanshu Jain

MOEA/D: A Multiobjective Evolutionary Algorithm Based On Decomposition — MOEA/D
Qingfu Zhang; Hui Li

A Survey On Evolutionary Computation Approaches To Feature Selection
Bing Xue; Mengjie Zhang; Will N. Browne; Xin Yao

A Coevolutionary Framework For Constrained Multiobjective Optimization Problems
Ye Tian; Tao Zhang; Jianhua Xiao; Xingyi Zhang; Yaochu Jin

Most Contributed Paper to 2018 IF of TEVC



Why Feature Selection ?

- "Curse of the dimensionality"
  - Large number of features: 100s, 1000s, even millions
- Not all features are useful (relevant)
- Redundant or irrelevant features may reduce the performance (e.g. classification accuracy)
- Costly: time, memory, and money
- Feature selection
  - to select a small subset of relevant features from the original large set of features in order to maintain or even improve the performance

Why Feature Construction?

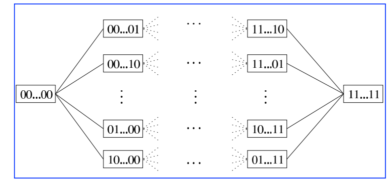
- The quality of input features can drastically affect the learning performance.
- Even if the quality of the original features is good, transformations might be required to make them usable for certain types of classifiers.
- Feature construction does not add to the cost of extracting (measuring) original features; it only carries computational cost.
- In some cases, feature construction can lead to dimensionality reduction or implicit feature selection.

What can FS/FC do ?

- Reduce the dimensionality (No. of features)
- Improve the (classification) performance
- Simplify the learnt model
- Speed up the processing time
- Help visualisation
- Improve interpretability and explainablity
- Reduce the cost, e.g. save memory
- and ?

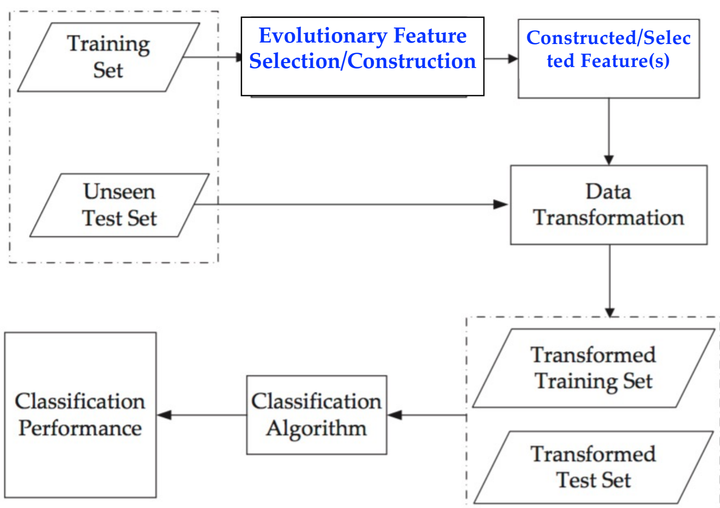
Challenges in FS and FC

- Large search space: 2^n possible feature subsets
  - 1990: n < 20
  - 1998: n <= 50
  - 2007: n ≈ 100s
  - Now: 1000s, 1 000 000s



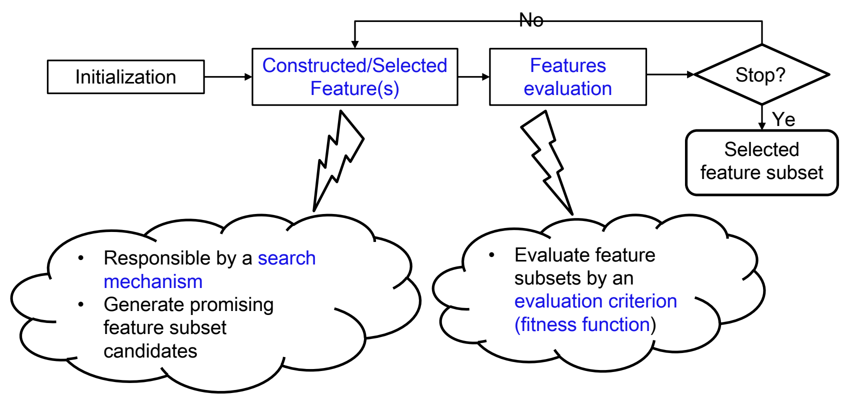
- Feature interaction
  - Relevant features may become redundant
  - Weakly relevant or irrelevant features may become highly useful
- Slow processing time, or even not possible
- Multi-objective Problems

General FS/FC System



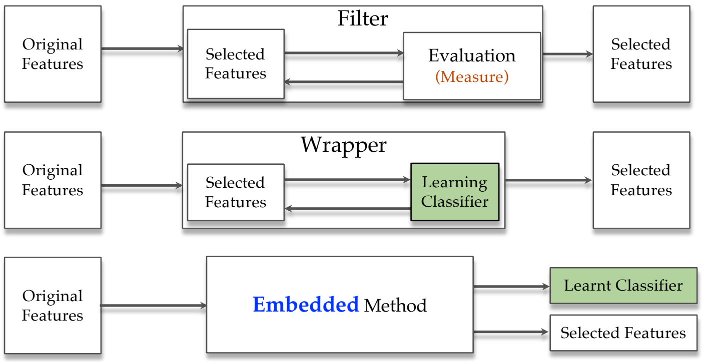
Feature FS/FC Process

- On training set:



Feature Selection Approaches

- Based on Evaluation ——— learning algorithm
  - Three categories: Filter, Wrapper, Embedded
  - Hybrid (Combined): commonly wrapper + filter



Feature Selection Approaches

- Generally:

	Classification Accuracy	Computational Cost	Generality (different classifiers)
Filter	Low	Low	High
Embedded	Medium	Medium	Medium
Wrapper	High	High	Low

Feature Selection

Feature Selection Approaches

- Conventional approaches
  - The Relief algorithm
  - The FOCUS algorithm
  - Sequential forward/backward floating selection
  - Statistical feature selection methods
  - Sparsity based feature selection methods
- Evolutionary Computation (EC) based approaches

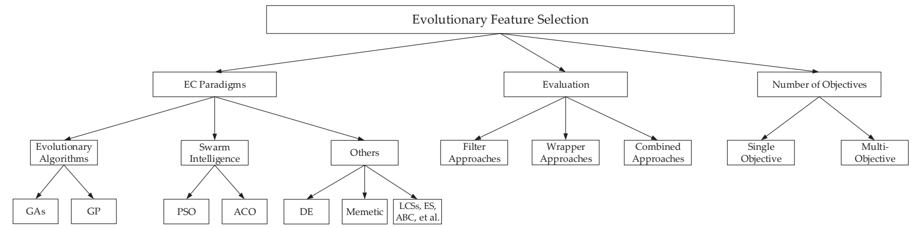
Li, J., Guo, R., Liu, C., & Liu, H. (2019, July). Adaptive unsupervised feature selection on attributed networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 92-100).
Feng, Chao, Chao Qian, and Ke Tang. "Unsupervised feature selection by pareto optimization." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 3534-3541. 2019.
Tang, Chang, Xinwang Liu, Xinzhong Zhu, Jian Xiong, Miaomiao Li, Jingyuan Xia, Xiangke Wang, and Lizhe Wang. "Feature selective projection with low-rank embedding and dual Laplacian regularization." IEEE Transactions on Knowledge and Data Engineering (2019).
Li, Yun, Tao Li, and Huan Liu. "Recent advances in feature selection and its applications." Knowledge and Information Systems 53.3 (2017): 551-577.
Cheng, Kewei, Jundong Li, and Huan Liu. "FeatureMiner: a tool for interactive feature selection." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
Zhai, Yiteng, Yew-Soon Ong, and Ivor W. Tsang. "The Emerging" Big Dimensionality"." IEEE Computational Intelligence Magazine 9.3 (2014): 14-26.
Gui, J., Sun, Z., Ji, S., Tao, D., & Tan, T. Feature selection based on structured sparsity: A comprehensive study. IEEE transactions on neural networks and learning systems, 28(7), (2017): 1490-1507.
Zhai, Yiteng, Yew-Soon Ong, and Ivor W. Tsang. "Making trillion correlations feasible in feature grouping and selection." IEEE transactions on pattern analysis and machine intelligence 38.12 (2016): 2472-2486.

Why Evolutionary Computation ?

- Don't need domain knowledge
- Don’t make any assumption
  - e.g. differentiable, linearity, separability, equality
- Easy to handle constraints
- EC can simultaneously build model structures and optimise parameters
- Population based search is particularly suitable for multi-objective optimisation

EC for Feature Selection

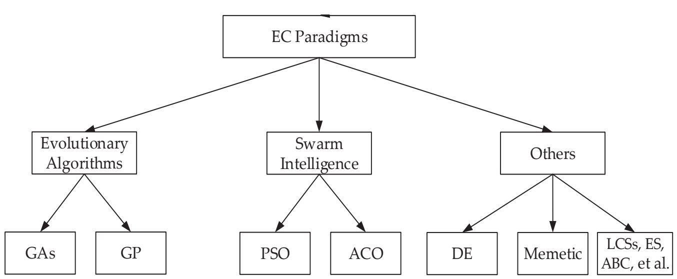
- EC Paradigms
- Evaluation
- Number of Objectives



Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." IEEE Transactions on Evolutionary Computation20, no. 4 (2016): 606-626.

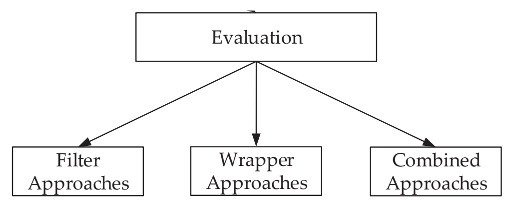
EC for Feature Selection

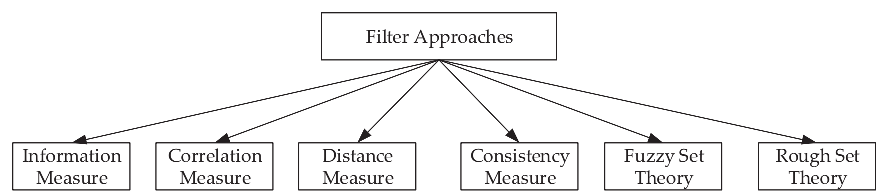
- Genetic algorithms (GAs), Genetic programming (GP)
- Particle swarm optimisation (PSO), ant colony optimisation(ACO)
- Differential evolution (DE), memetic algorithms, learning classifier systems (LCSs)



Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." IEEE Transactions on Evolutionary Computation20, no. 4 (2016): 606-626.

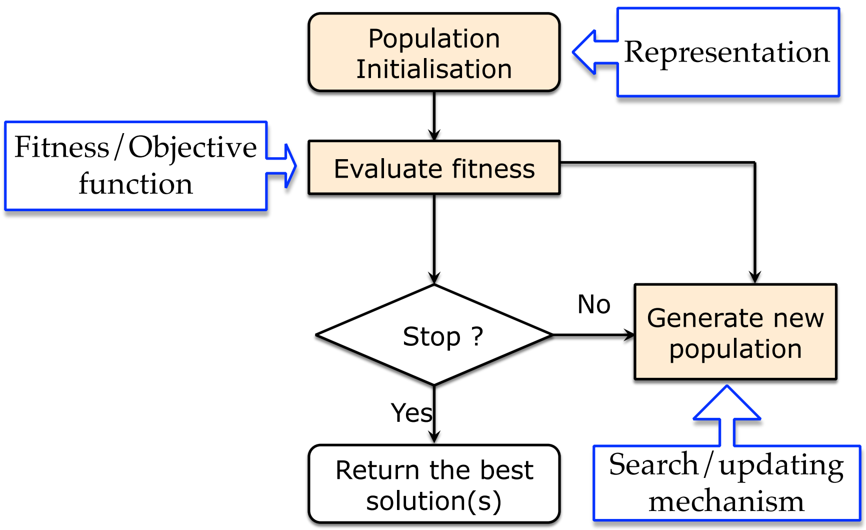
EC for Feature Selection





Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." IEEE Transactions on Evolutionary Computation20, no. 4 (2016): 606-626.

A General Approach



GAs for Feature Selection

- Over 25 years ago, first EC techniques
  - Filter, Wrapper, Single Objective, Multi-objective
- Representation
  - Binary string

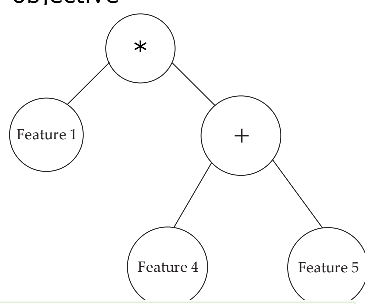
1	0	1	1	0	0	1	0	0	1	1	0	0

- Search mechanisms
  - Genetic operators
- Multi-objective feature selection
- <u>Scalability issue</u>

R. Leardi, R. Boggia, and M. Terrile, "Genetic algorithms as a strategy for feature selection," Journal of Chemometrics, vol. 6, no. 5, pp. 267–281, 1992.
Ishibuchi, Hisao, and Tomoharu Nakashima. "Multi-objective pattern and feature selection by a genetic algorithm." In GECCO, pp. 1069-1076. 2000.
Z. Zhu, Y.-S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," Pattern Recognition, vol. 40, no. 11,pp. 3236–3248, 2007.
W. Sheng, X. Liu, and M. Fairhurst, "A niching memetic algorithm for simultaneous clustering and feature selection," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 7, pp. 868–879, 2008.
Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." IEEE Transactions on Evolutionary Computation, 20, no. 4 (2016): 606-626.
Tao, Zhou, Lu Huiling, Wang Wenwen, and Yong Xia. "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling." Applied Soft Computing 75 (2019): 323-332.

GP for Feature Selection

- Implicit feature selection
  - Filter, Wrapper, Single Objective, Multi-objective
- Embedded feature selection
- Feature construction
- <u>Computationally expensive</u>



M. G. Smith and L. Bull, "Genetic programming with a genetic algorithm for feature construction and selection," Genetic Programming and Evolvable Machines, vol. 6, no. 3, pp. 265–281, 2005.
B. Tran, B. Xue, M. Zhang. "Class Dependent Multiple Feature Construction Using Genetic Programming for High-Dimensional Data". Proceedings of the 30th AI2017,LNCS. Vol. 10400. Springer. Melbourne, Australia, August 19-20th, 2017. pp. 182-194.
Viegas, Felipe, et al. "A Genetic Programming approach for feature selection in highly dimensional skewed data." Neurocomputing 273 (2018): 554-569.
Nag, Kaustuv, and Nikhil R. Pal. "A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification." IEEE transactions on cybernetics 46.2 (2016): 499-510.
Tran, Binh, Bing Xue, and Mengjie Zhang. "Genetic programming for multiple-feature construction on high-dimensional classification." Pattern Recognition 93 (2019): 404-417.
Nag, Kaustuv, and Nikhil R. Pal. "Feature Extraction and Selection for Parsimonious Classifiers with Multiobjective Genetic Programming." IEEE Transactions on Evolutionary Computation (2019).

PSO for Feature Selection

- Very popular in recent years
  - Filter, Wrapper, Single Objective, Multi-objective
- Representation, continuous PSO vs Binary PSO
- Search mechanism
- Fitness function
- <u>Scalability</u>

0.7	0.12	0.84	0.69	0.25	0.06	0.92	0.45	0.36	0.80	0.67	0.30	0.41

1	0	1	1	0	0	1	0	0	1	1	0	0

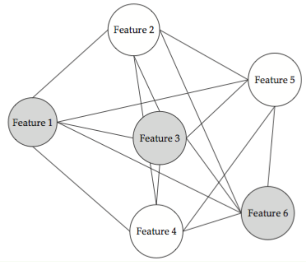
E. K. Tang, P. Suganthan, and X. Yao, "Feature selection for microarray data using least squares SVM and particle swarm optimization," in IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 1–8, 2005.
Zhang, Yong, Dun-wei Gong, and Jian Cheng. "Multi-objective particle swarm optimization approach for cost-based feature selection in classification." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 14.1 (2017): 64-75.
Gu, Shenkai, Ran Cheng, and Yaochu Jin. "Feature selection for high-dimensional classification using a competitive swarm optimizer." *Soft Computing* 22.3 (2018): 811-822..
Hafiz, Faizal, et al. "A two-dimensional (2-D) learning framework for Particle Swarm based feature selection." Pattern Recognition. 76 (2018): 416-433.
Binh Tran and Bing Xue and Mengjie Zhang."A New Representation in PSO for Discretisation-Based Feature Selection", IEEE Transactions on Cybernetics, vol. 48, no. 6, pp.1733-1746, 2018
Xue, Y., Xue, B., & Zhang, M.. Self-adaptive particle swarm optimization for large-scale feature selection in classification. ACM Transactions on Knowledge Discovery from Data (TKDD), 13(5), 1-27. 2019
Tran, Binh, Bing Xue, and Mengjie Zhang. "Adaptive multi-subswarm optimisation for feature selection on high-dimensional classification." In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 481-489. 2019.
Bach Hoai Nguyen, Bing Xue, Mengjie Zhang, and FengyuZhou "A survey on swarm intelligence approaches to feature selection in data mining", Swarm and Evolutionary Computation, vol. 54, num , pp 100663:1-30. , May 2020. (doi.org/10.1016/j.swevo.2020.100663)

ACO for Feature Selection

- Start from around 2003
  - Filter, Wrapper, Single Objective, Multi-objective

- Representation
- Search mechanism
- Filter approaches

- Scalability



S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," Neurocomputing, 2014.
C.-K. Zhang and H. Hu, "Feature selection using the hybrid of ant colony optimization and mutual information for the forecaster," in International Conference on Machine Learning and Cybernetics, vol. 3, pp. 1728–1732, 2005.
L. Ke, Z. Feng, and Z. Ren, "An efficient ant colony optimization approach to attribute reduction in rough set theory," Pattern Recognition Letters, vol. 29, no. 9, pp. 1351–1357, 2008.
Moradi, Parham, and Mehrdad Rostami. "Integration of graph clustering with ant colony optimization for feature selection." Knowledge-Based Systems 84 (2015): 144-161.
Tabakhi, Sina, and Parham Moradi. "Relevance–redundancy feature selection based on ant colony optimization." Pattern recognition 48, no. 9 (2015): 2798-2811.
Sara, V. J., S. Belina, and K. Kalaiselvi. "Ant colony optimization (ACO) based feature selection and extreme learning machine (ELM) for chronic kidney disease detection." International Journal of Advanced Studies of Scientific Research 4, no. 1 (2019).
Paniri, M., Dowlatshahi, M.B. and Nezamabadi-pour, H., 2020. MLACO: A multi-label feature selection algorithm based on ant colony optimization. Knowledge-Based Systems, 192, p.105285.
Bach Hoai Nguyen, Bing Xue, Mengjie Zhang, and FengyuZhou "A survey on swarm intelligence approaches to feature selection in data mining", Swarm and Evolutionary Computation, vol. 54, num , pp 100663:1-30. , May 2020. (doi.org/10.1016/j.swevo.2020.100663)

DE, LCSs, and Memetic

- DE: since 2008
  - potential for large-scale
- LCSs:
  - implicit feature selection
  - embedded feature selection
- memetic:
  - population search + local search
  - Wrapper + filter

Hancer, Emrah, Bing Xue, and Mengjie Zhang. "Differential evolution for filter feature selection based on information theory and feature ranking." Knowledge-Based Systems 140 (2018): 103-119.
Z. Li, Z. Shang, B. Qu, and J. Liang, "Feature selection based on manifold-learning with dynamic constraint handling differential evolution," in IEEE Congress on Evolutionary Computation (CEC), pp. 332–337, 2014.
Hoai Bach Nguyen, Bing Xue, Hisao Ishibuchi, Peter Andreae, and Mengjie Zhang. "Multiple Reference Points MOEA/D for Feature Selection". Proceedings of 2017 Genetic and Evolutionary Computation Conference (GECCO 2017) Companion. ACM Press. Berlin, German, 15 - 19 July 2017.pp 157-158.
I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1424 –1437, 2004.
Z. Zhu, S. Jia, and Z. Ji, "Towards a memetic feature selection paradigm [application notes]," IEEE Computational Intelligence Magazine, vol. 5, no. 2, pp. 41–53, 2010.
Y. Wen and H. Xu, "A cooperative coevolution-based pittsburgh learning classifier system embedded with memetic feature selection," in IEEE Congress on Evolutionary Computation, pp. 2415–2422, 2011.
Zhang, Yong, Dun-wei Gong, Xiao-zhi Gao, Tian Tian, and Xiao-yan Sun. "Binary differential evolution with self-learning for multi-objective feature selection." Information Sciences 507 (2020): 67-85.

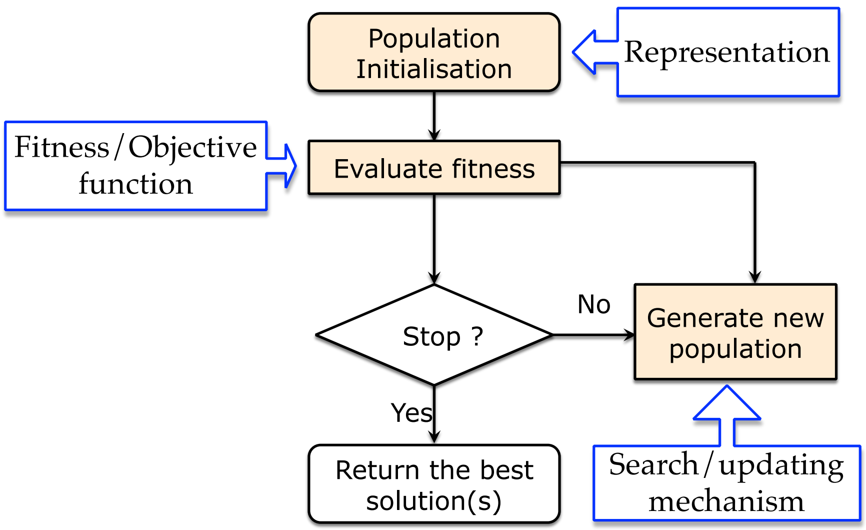
Related Areas (Applications)

- Biological and biomedical tasks
  - gene analysis, biomarker detection, cancer classification, and disease diagnosis
- Image and signal processing
  - image analysis, face recognition, human action recognition, EEG brain-computer-interface, speaker recognition, handwritten digit recognition, personal identification, and music instrument recognition.
- Network/web service
  - Web service composition and development, network security, and email spam detection.
- Business and financial problems
  - Financial crisis, credit card issuing in bank systems, and customer churn prediction.
- Others
  - power system optimisation, weed recognition in agriculture, melting point prediction in chemistry, and weather prediction.

Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." IEEE Transactions on Evolutionary Computation20, no. 4 (2016): 606-626.

Feature Selection

A General Approach



Variable-length PSO for FS

- Comprehensive Learning PSO

Dimension:	1	2	3	4	5	1	2	3	...	L
Position:	0.8	0.3	0.9	0.4	0.1	0.6	0.2	0.7	...	0.5
Velocity:	0.1	0.2	0.5	0.4	0.1	0.2	0.4	0.3	...	0.2
Exemplar:	8	7	5	6	2	3	7	8	...	1

Learning Probability (Pc) = 0.25 Renew Exemplar Count = 3

Fig. 1. Representation of a VLPSO particle with length L.

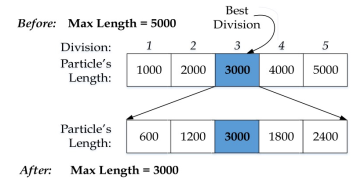


Fig. 3. Example of length changing in a swarm with five divisions.

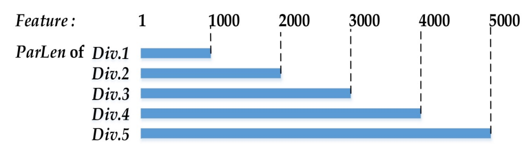


Fig. 2. Example of population division for a problem with 5000 features and the number of division is 5.

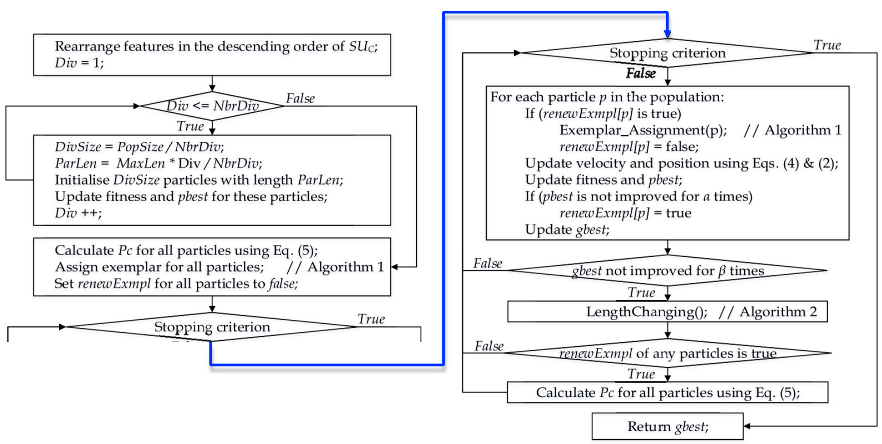
Feature Ranking

$$\mathrm{SU}(F, C) = \left[\frac{\mathrm{IG}(F|C)}{H(F) + H(C)}\right]$$

$$\mathrm{IG}(F|C) = H(F) - H(F|C)$$

Binh Tran and Bing Xue and Mengjie Zhang. "Variable-Length Particle Swarm Optimisation for Feature Selection on High-Dimensional Classification", IEEE Transactions on Evolutionary Computation, Vol. 23, pp 473-487, 2019. pp 473-487.

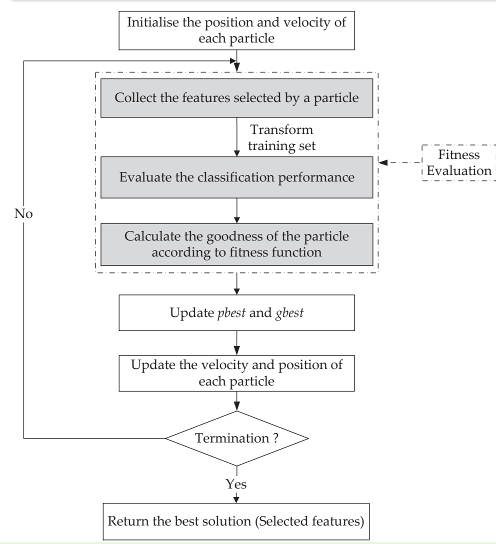
Variable-length PSO for FS



$$\text{fitness} = (\gamma \cdot \text{accuracy} + (1 - \gamma) \cdot \text{distance}).$$

Binh Tran and Bing Xue and Mengjie Zhang. "Variable-Length Particle Swarm Optimisation for Feature Selection on High-Dimensional Classification", IEEE Transactions on Evolutionary Computation, Vol. 23, pp 473-487, 2019. pp 473-487.

PSO for FS: initialisation and updating

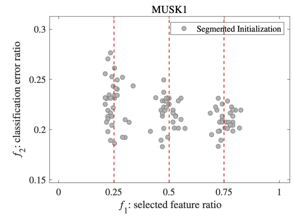


- Initialisation:
  - Forward selection
  - Backward selection
  - Mixture of both
- Updating:
  - Consider the number of features in the pest and gbest updating

Bing Xue, Mengjie Zhang, Will N. Browne."Particle Swarm Optimisation for Feature Selection in Classification: Novel Initialisation and Updating Mechanisms". Applied Soft Computing. Vol 18, PP. 261--276, 2014

Initialization in Multi-objective FS

- Segmented initialization mechanism respectively generates three sub-populations whose solutions are randomly located around the forward, middle and backward areas
  - areas with a small, medium or large number of selected features respectively.



- Offspring modification:
  - find out all the duplicated solutions
  - Modify duplicated solutions to become unique ones
    - by each flipping one or two dimensions (each dimension corresponds to one original feature), according to the analysis of common features in the first nondominated front

Hang Xu, Bing Xue, and Mengjie Zhang. "Segmented Initialization and Offspring Modification in Evolutionary Algorithms for Bi-objective Feature Selection". Proceedings of 2020 Genetic and Evolutionary Computation Conference (GECCO 2020). ACM Press. Cancun, Mexico. July 8th-12th 2020, 9pp

Initialization in Multi-objective FS

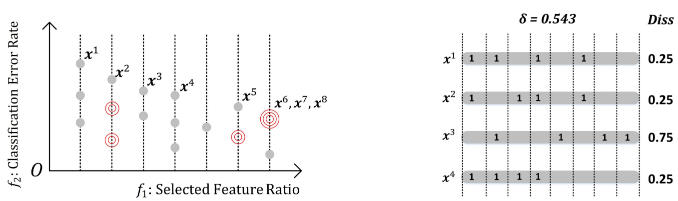
IGD and HV show that segmented initialization mechanism and offspring modification mechanism each contributed positively to the success of the new plug-in MOEAs, while combining them together contributed the most.

No.	Dataset Names	Features	Instances	Classes
1	Climate	18	540	2
2	Statlog_German	24	1000	2
3	Breast_Cancer	30	569	2
4	Connectionist_Bench_Sonar	60	208	2
5	Mice_Protein_Expression	77	1077	8
6	Hill_Valley	100	606	2
7	MUSK1	166	476	2
8	Semeion_Handwritten_Digit	256	1593	10
9	Arrhythmia	278	452	16
10	LSVT_Voice_Rehabilitation	310	126	2
11	Madelon_Train_Validation	500	2600	2
12	ISOLET5	617	1559	26
13	Multiple_Features	649	2000	10
14	SRBCT	2308	83	4
15	Leukemia1	5327	72	3
16	DLBCL	5469	77	2
17	Brain1	5920	90	5
18	Leukemia	7070	72	2

No.	NSGA-II	SIOM-NSGAII	MOEA/HD	SIOM-MOEAHD	HypE	SIOM-HypE
1	8.8843e-01 ± 7.89e-03	†8.8913e-01 ± 7.07e-03	8.8859e-01 ± 7.44e-03	†8.8940e-01 ± 7.43e-03	8.8964e-01 ± 7.39e-03	†8.9046e-01 ± 6.96e-03
2	7.2368e-01 ± 1.01e-02	†7.2578e-01 ± 6.31e-03	7.2476e-01 ± 5.55e-03	†7.2747e-01 ± 8.03e-03	7.2244e-01 ± 7.98e-03	†7.2394e-01 ± 8.55e-03
3	9.3936e-01 ± 6.29e-03	†9.3930e-01 ± 5.60e-03	9.3751e-01 ± 5.22e-03	†9.3850e-01 ± 6.05e-03	9.2626e-01 ± 1.46e-02	9.3848e-01 ± 5.96e-03
4	7.8557e-01 ± 2.21e-02	†7.9769e-01 ± 2.09e-02	8.0020e-01 ± 2.14e-02	†7.9177e-01 ± 2.34e-02	7.6884e-01 ± 2.71e-02	7.9011e-01 ± 2.29e-02
5	7.2290e-01 ± 2.53e-02	7.3768e-01 ± 1.51e-02	5.9972e-01 ± 2.14e-01	7.3325e-01 ± 1.90e-02	7.2896e-01 ± 1.56e-02	7.3943e-01 ± 1.27e-02
6	5.9247e-01 ± 2.46e-02	6.3030e-01 ± 1.22e-02	6.2509e-01 ± 1.43e-02	†6.3098e-01 ± 1.22e-02	6.1941e-01 ± 1.46e-02	†6.2524e-01 ± 1.14e-02
7	8.3315e-01 ± 2.35e-02	9.0255e-01 ± 1.56e-02	8.6995e-01 ± 2.41e-02	8.9757e-01 ± 1.52e-02	8.3748e-01 ± 2.50e-02	8.7525e-01 ± 1.83e-02
8	7.2989e-01 ± 1.55e-02	8.0715e-01 ± 1.26e-02	7.5261e-01 ± 1.40e-02	8.1625e-01 ± 1.35e-02	7.1856e-01 ± 1.24e-02	7.6815e-01 ± 1.13e-02
9	6.6089e-01 ± 1.67e-02	6.9266e-01 ± 1.45e-02	5.0177e-01 ± 2.26e-02	6.6032e-01 ± 4.83e-02	6.4381e-01 ± 2.50e-02	6.6781e-01 ± 1.60e-02
10	8.2075e-01 ± 3.08e-02	9.2341e-01 ± 2.14e-02	8.4570e-01 ± 2.55e-02	9.2429e-01 ± 2.91e-02	8.2964e-01 ± 2.07e-02	8.9074e-01 ± 2.24e-02
11	5.8352e-01 ± 1.51e-02	8.1875e-01 ± 3.38e-02	5.7165e-01 ± 1.53e-02	7.5211e-01 ± 3.35e-02	5.9267e-01 ± 1.84e-02	7.3684e-01 ± 2.93e-02
12	6.9352e-01 ± 1.51e-02	8.0820e-01 ± 1.00e-02	6.8977e-01 ± 1.20e-02	8.0297e-01 ± 1.27e-02	6.9876e-01 ± 1.08e-02	7.8173e-01 ± 1.23e-02
13	7.8757e-01 ± 1.13e-02	9.0081e-01 ± 1.02e-02	7.8274e-01 ± 1.18e-02	8.8876e-01 ± 1.18e-02	7.6248e-01 ± 1.12e-02	8.6388e-01 ± 1.21e-02
14	2.9424e-01 ± 2.14e-03	5.6236e-01 ± 1.74e-01	2.5868e-01 ± 2.29e-03	3.6473e-01 ± 8.43e-02	2.9455e-01 ± 2.38e-03	5.7069e-01 ± 1.70e-01
15	5.3839e-01 ± 1.76e-02	7.1387e-01 ± 1.97e-02	5.2521e-01 ± 1.84e-02	6.9892e-01 ± 2.46e-02	5.2619e-01 ± 2.03e-02	7.0910e-01 ± 2.72e-02
16	6.0251e-01 ± 2.09e-02	7.9549e-01 ± 3.58e-02	5.8483e-01 ± 1.50e-02	7.8223e-01 ± 2.46e-02	5.9509e-01 ± 1.98e-02	7.9343e-01 ± 3.57e-02
17	4.8237e-01 ± 2.88e-03	6.3834e-01 ± 8.24e-03	4.6436e-01 ± 8.95e-03	6.2643e-01 ± 7.22e-03	4.7869e-01 ± 3.90e-03	6.3286e-01 ± 6.19e-03
18	5.4917e-01 ± 1.72e-02	7.2973e-01 ± 2.82e-02	5.2839e-01 ± 2.06e-02	7.1279e-01 ± 3.65e-02	5.3623e-01 ± 1.94e-02	7.2762e-01 ± 3.28e-02

Hang Xu, Bing Xue, and Mengjie Zhang. "Segmented Initialization and Offspring Modification in Evolutionary Algorithms for Bi-objective Feature Selection". Proceedings of 2020 Genetic and Evolutionary Computation Conference (GECCO 2020). ACM Press. Cancun, Mexico. July 8th-12th 2020, 9pp

Duplication Analysis-Based MOFS

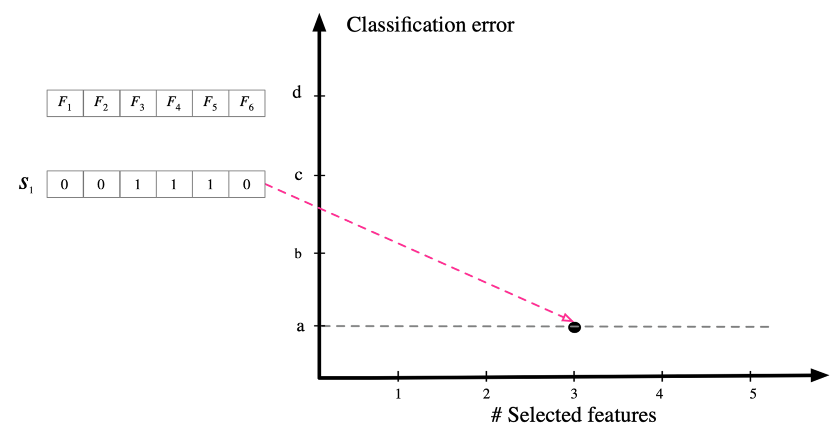


- The reproduction process is modified to improve the quality of offspring;
- A duplication analysis method is proposed to filter out the redundant solutions
- A diversity-based selection method is adopted to further select the reserved solutions

Hang Xu, Bing Xue, and Mengjie Zhang. "Segmented Initialization and Offspring Modification in Evolutionary Algorithms for Bi-objective Feature Selection". Proceedings of 2020 Genetic and Evolutionary Computation Conference (GECCO 2020). ACM Press. Cancun, Mexico. July 8th-12th 2020, 9pp
Xu, H., Xue, B., & Zhang, M. (2020). A duplication analysis based evolutionary algorithm for bi-objective feature selection. IEEE Transactions on Evolutionary Computation.

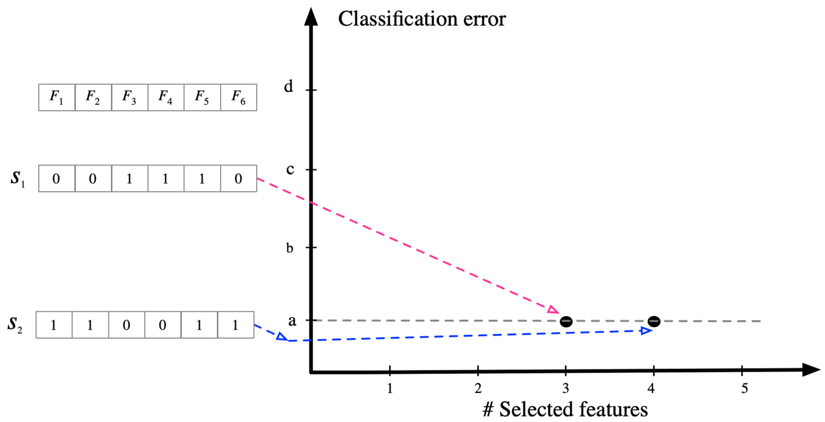
Evolutionary Multimodal Optimisation for FS

The goal is to find **multiple optimal feature subsets**



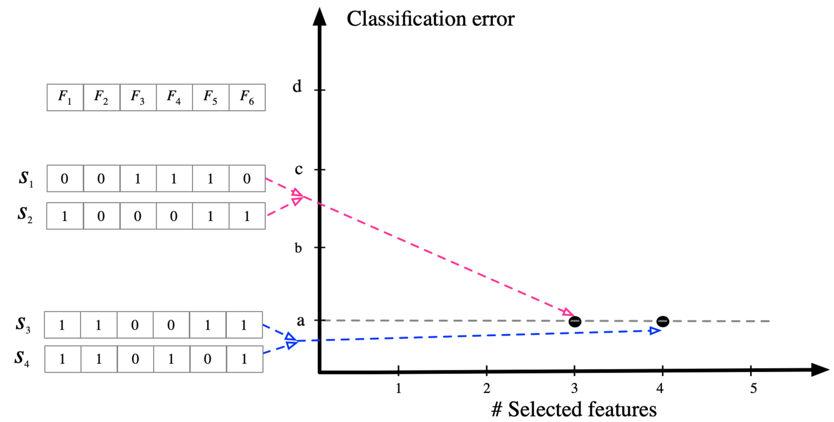
Evolutionary Multimodal Optimisation for FS

The goal is to find **multiple optimal feature subsets**



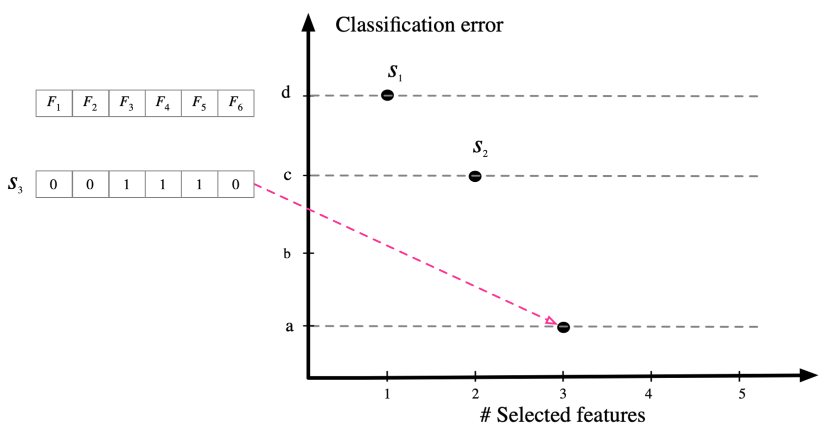
Evolutionary Multimodal Optimisation for FS

The goal is to find **multiple optimal feature subsets**



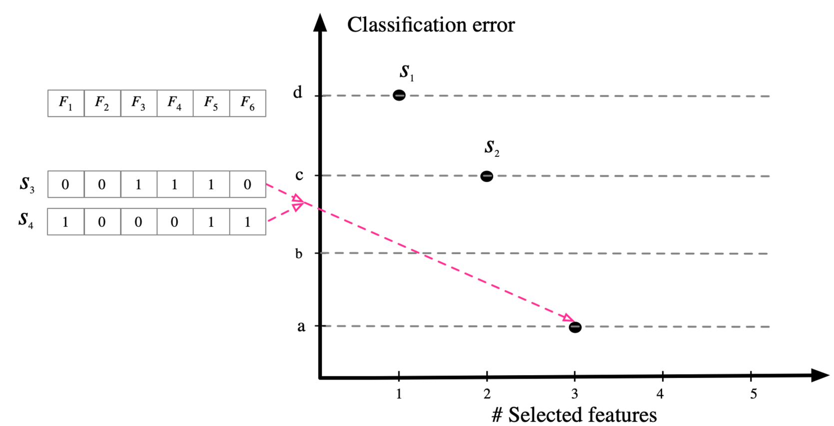
Evolutionary Multimodal Optimisation for FS

The goal is to find **multiple optimal feature subsets**

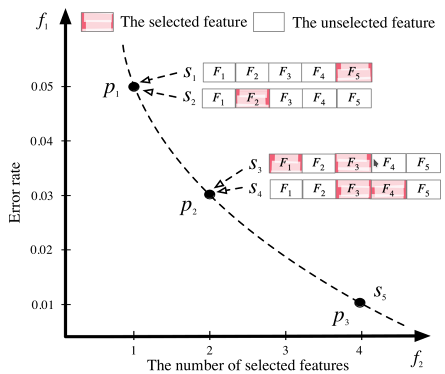


Evolutionary Multimodal Optimisation for FS

The goal is to find **multiple optimal feature subsets**



Evolutionary Multimodal Optimisation for FS

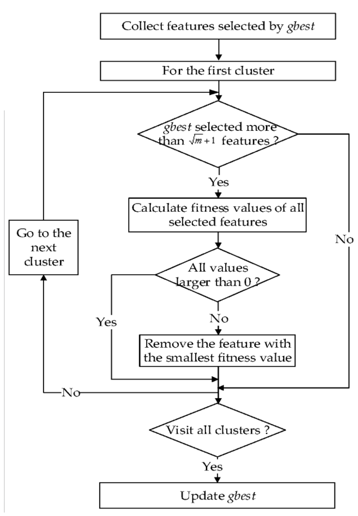


Why Multimodal Optimisation for FS?

- In Breast Cancer Wisconsin (Original) Data Set, there are 699 instances, 9 features and 2 classes.
- Using subset [**F1**, F2, F7] or [F2, **F3**, F7] with KNN can achieve the **same 97.81%** classification accuracy.
- F1 is '**Clump Thickness**' and F3 is '**Uniformity of Cell Shape**'. Obviously, the first feature is easier to be collected than the third one.

Peng Wang, Bing Xue, Jing Liang and Mengjie Zhang. "Improved Crowding Distance in Multi-objective Optimization for Feature Selection in Classification". Proceeding of the 23th European Conference on Applications of Evolutionary Computation (EvoApplications 2021). Lecture Notes in Computer Science. Vol. , Leipzig, Germany. Seville, Spain, 7-9 April 2021
Peng Wang, Bing Xue, Jing Liang and Mengjie Zhang. " A Grid-dominance based Multi-objective Algorithmfor Feature Selection in Classification." IEEE Congress on Evolutionary Computation (CEC 2021). Krakow, Poland, 28 June - 1 July 2021, 8pp

PSO FS: with backward elimination



- Filter measure based on mutual information in backward elimination:

$$f'(s_i) = \frac{1}{x_i}(Rel(s_i) - \frac{1}{|s|-1}Red(s_i))$$

- x_i: the position value in the ith dimension
- By adding $\frac{1}{x_i}$, $f'(s_i)$ ensures that if two features has the same $f(s)$ value, the one with a smaller position value (i.e. smaller probability) will be removed
- s_i is removed only when $f'(s_i) < 0$ and $f'(s_i)$ is the smallest value

Accurate Wrapper, Global

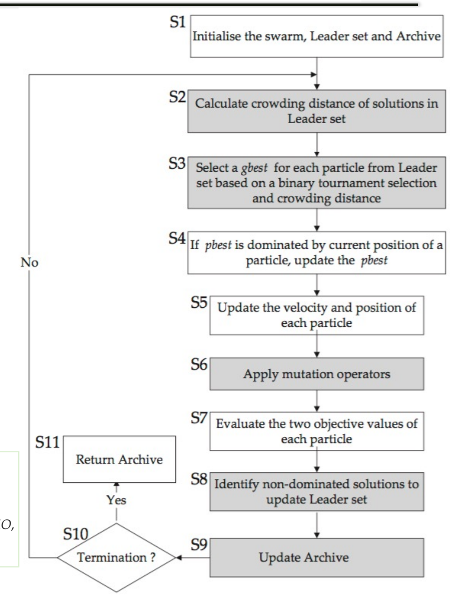
+

Fast Filter, Local

Bach Hoai Nguyen, Bing Xue, Ivy Liu and Mengjie Zhang."Filter based Backward Elimination in Wrapper based PSO for Feature Selection in Classification", Proceedings of 201 IEEE Congress on Evolutionary Computation. Beijing, China. 6-11 July, 2014. IEEE Press. PP.3111-- 3118. 2015

Multi-objective PSO for FS

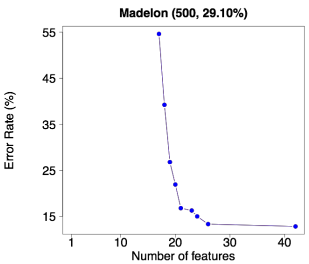
- Introduce and develop the first multi-objective PSO approach to feature selection
  - Simultaneously minimise the number of features and the error rate
  - **>800 citations since 2013**



Bing Xue, Mengjie Zhang, Will Browne. "Particle swarm optimization for feature selection in classification: A multi-objective approach, IEEE Transactions on Cybernetics, vol. 43, no. 6, pp. 1656-1671, 2013.
M. R. Sierra and C. A. C. Coello, "Improving PSO-based multi-objective optimization using crowding, mutation and epsilon-dominance", *Proc. EMO*, pp. 505-519, 2005

Multi-objective PSO for FS

- Simultaneously minimise the number of features and the error rate



T-Test on Hypervolume Ratios on Training Accuracy

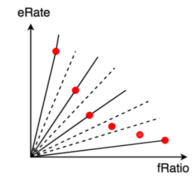
Dataset	Wine		Australian		Zoo		Vehicle		German		WBCD	
	NS	CMD	NS	CMD	NS	CMD	NS	CMD	NS	CMD	NS	CMD
NSPSOFS		+		+		+		+		+		+
CMDPSOFS	-		-		-		-		-		-	
NSGAII	-	=	-	=	-	-	-	-	-	=	-	=
SPEA2	-	=	-	=	-	=	-	=	-	=	-	=
PAES	-	=	-	=	-	-	-	-	-	-	-	=

Dataset	Lung		Ionosphere		Hillvalley		Musk1		Madelon		Isolet5	
	NS	CMD	NS	CMD	NS	CMD	NS	CMD	NS	CMD	NS	CMD
NSPSOFS		+		+		+		+		+		+
CMDPSOFS	-		-		-		-		-		-	
NSGAII	-	-	-	-	-	+	-	+	=	+	+	+
SPEA2	-	-	-	-	=	+	-	+	=	+	+	+
PAES	-	-	-	-	-	-	-	-	-	-	-	-

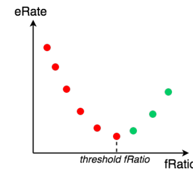
Bing Xue, Mengjie Zhang, Will Browne. "Particle swarm optimization for feature selection in classification: A multi-objective approach, IEEE Transactions on Cybernetics, vol. 43, no. 6, pp. 1656-1671, 2013.

MOEA/D for Feature Selection

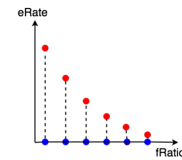
- Multiple Reference Points based Decomposition for Multi-objective FS



(a) Weight vector for FS



(b) Not always conflicting



Multiple reference points in MOEA/D.

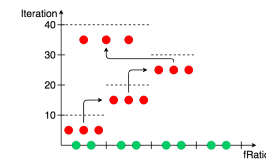
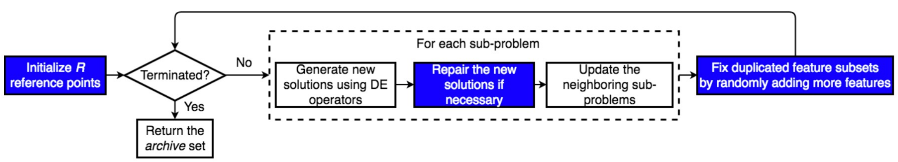


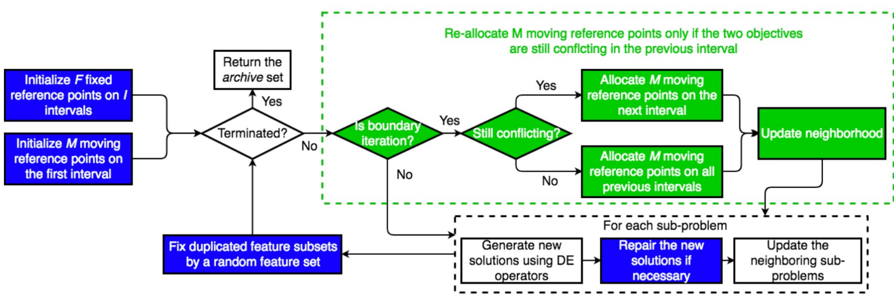
Figure 5.4: Dynamic reference points example: *fixed* points are green, *moving* points are red, dashed line shows the interval that *moving* points are located in the corresponding iterations.

Hoai Bach Nguyen, Bing Xue, Hisao Ishibuchi, Peter Andreae, and Mengjie Zhang. "Multiple Reference Points MOEA/D for Feature Selection". Proceedings of 2017 Genetic and Evolutionary Computation Conference (GECCO 2017) Companion. ACM Press. Berlin, German, 15 - 19 July 2017.pp 157-158
Bach Hoai Nguyen, Bing Xue, Peter Andreae, Hisao Ishibuchi, Mengjie Zhang. "Multiple Reference Points based Decomposition for Multi-objective Feature Selection in Classification: Static and Dynamic Mechanisms", IEEE Transactions on Evolutionary Computation, vol. 24, no. 1, pp.170 - 184, 2019

MOEA/D for Feature Selection



(a) Static multiple reference points strategy (MOEA/D-STAT).

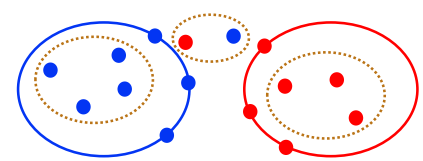


(b) Dynamic multiple reference points strategy (MOEA/D-DYN).

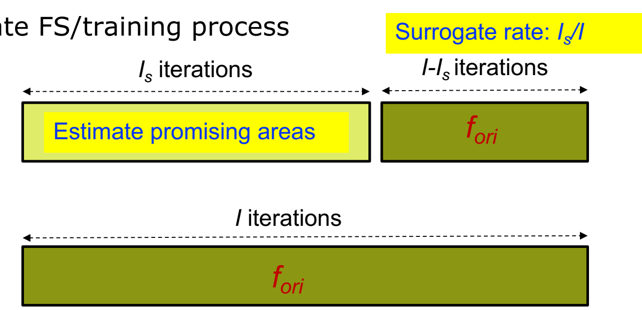
Hoai Bach Nguyen, Bing Xue, Hisao Ishibuchi, Peter Andreae, and Mengjie Zhang. "Multiple Reference Points MOEA/D for Feature Selection". Proceedings of 2017 Genetic and Evolutionary Computation Conference (GECCO 2017) Companion. ACM Press. Berlin, German, 15 - 19 July 2017.pp 157-158
Bach Hoai Nguyen, Bing Xue, Peter Andreae, Hisao Ishibuchi, Mengjie Zhang. "Multiple Reference Points based Decomposition for Multi-objective Feature Selection in Classification: Static and Dynamic Mechanisms", IEEE Transactions on Evolutionary Computation, vol. 24, no. 1, pp.170 - 184, 2019

Surrogate Training Set/ FS Processing

- Original fitness function (f_{ori}) vs Surrogate fitness function (f_{sur})
- Surrogate training set



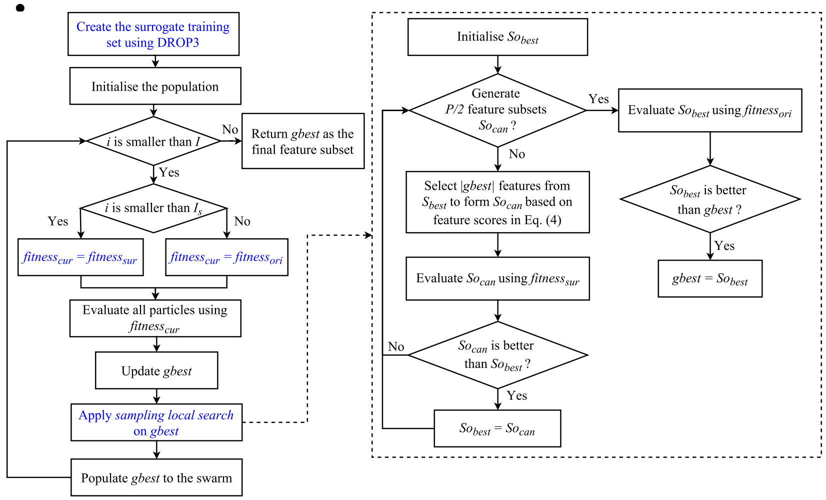
- Surrogate FS/training process



Hoai Bach Nguyen, Bing Xue, and Peter Andreae. "PSO with Surrogate Models for Feature Selection: Static and Dynamic Clustering-based Methods", Memetic Computing, vol. 10, pp. 291-300, 2018

Surrogate Training Set/ FS Processing

•



Hoai Bach Nguyen, Bing Xue, and Peter Andreae. "PSO with Surrogate Models for Feature Selection: Static and Dynamic Clustering-based Methods", Memetic Computing, vol. 10, pp. 291-300, 2018

Effect of the surrogate training set

Dataset	NF	Training Results AveTrain ± Std	T	Test Results AveTest ± Std	T	Time (ms)
Vehicle						
All	18.0	89.44 ± 0.00	-	83.07 ± 0.00	-	
OriPSO	4.8	91.41 ± 0.69	=	84.66 ± 1.29	-	113187.62
SurPSO	5.1	91.43 ± 0.69		85.23 ± 0.82		34294.93
German						
All	24.0	83.14 ± 0.00	-	65.33 ± 0.00	-	
OriPSO	5.6	78.70 ± 0.52	-	67.63 ± 1.90	=	229110.30
SurPSO	6.4	84.33 ± 0.52		68.35 ± 0.75		60600.90
Ionosphere						
All	34.0	90.24 ± 0.00	-	86.67 ± 0.00	-	
OriPSO	4.8	95.27 ± 0.81	+	88.48 ± 2.12	=	33564.40
SurPSO	4.0	94.64 ± 0.81		88.38 ± 2.61		9903.90
Lung						
All	56.0	86.36 ± 0.00	-	80.00 ± 0.00	=	
OriPSO	5.1	99.89 ± 1.73	+	79.50 ± 4.44	=	707.38
SurPSO	4.4	99.20 ± 1.73		79.50 ± 5.45		451.75
Sonar						
All	60.0	88.97 ± 0.00	-	84.13 ± 0.00	+	
OriPSO	14.8	95.55 ± 1.78	+	81.94 ± 3.32	=	23605.65
SurPSO	12.4	93.55 ± 1.78		82.58 ± 3.14		7291.62
Movementlibras						
All	90.0	98.52 ± 0.00	-	95.06 ± 0.00	-	
OriPSO	9.2	98.54 ± 0.16	=	95.33 ± 0.41	=	105780.20
SurPSO	9.4	98.63 ± 0.16		95.29 ± 0.32		41165.30
Plant						
All	64.0	99.55 ± 0.00	+	99.10 ± 0.00	+	
OriPSO	3.0	99.24 ± 0.04	=	98.67 ± 0.03	=	1756442.18
SurPSO	3.1	99.26 ± 0.04		98.68 ± 0.05		566479.72
Hillvalley						
All	100.0	79.83 ± 0.00	-	59.07 ± 0.00	-	
OriPSO	24.8	81.54 ± 0.83	=	58.85 ± 1.88	-	1557361.75
SurPSO	22.3	81.16 ± 0.83		59.92 ± 1.46		443232.00
LSVT						
All	310.0	79.55 ± 0.00	-	55.26 ± 0.00	-	
OriPSO	27.9	85.45 ± 4.72	=	65.53 ± 1.53	=	18240.05
SurPSO	27.4	85.11 ± 4.72		65.07 ± 4.89		4708.20
Multiple Features						
All	649.0	99.49 ± 0.00	-	98.57 ± 0.00	-	
OriPSO	118.2	99.66 ± 0.05	=	99.04 ± 0.10	=	7203332.90
SurPSO	143.6	99.65 ± 0.05		99.05 ± 0.12		2068439.52

- On all datasets SurPSO maintains or improve the testing accuracy while reducing 70% computation time than OriPSO

Information Theory Feature Selection

$$Fitness = Rel - Red$$

where,

$$Rel = \sum_{x_i \epsilon X} I(x_i; c)$$

$$Red = \sum_{x_i, x_j \epsilon X} I(x_i; x_j)$$

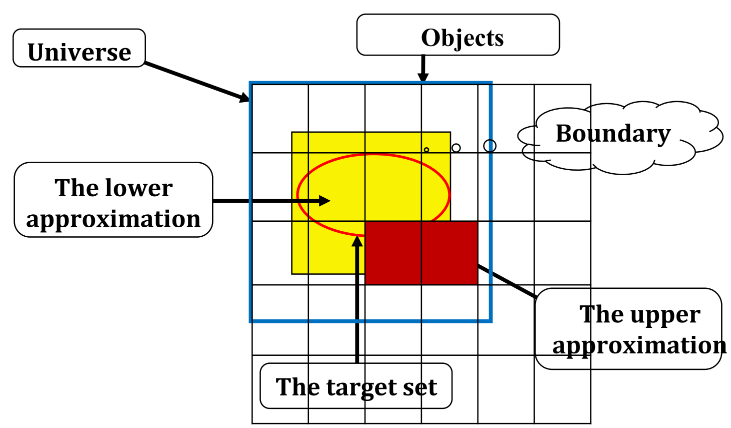
$\boldsymbol{X}$ is the selected feature subset
$\boldsymbol{x_i, x_j}$: single feature in X
$\boldsymbol{c}$ is the class labels
$\boldsymbol{Rel}$: relevance between X and c
$\boldsymbol{Red}$: redundancy in X

- **Relevance**:
  - Classification performance
  - The relevance (mutual information) between each selected feature and the class labels
- **Redundancy**:
  - Number of features
  - The relevance (mutual information) between the selected features

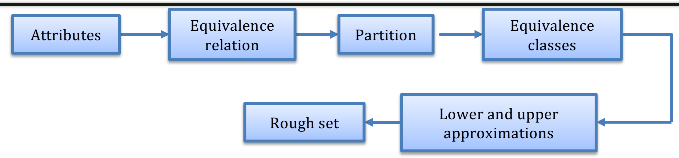
Liam Cervante, Bing Xue and Mengjie Zhang."Binary Particle Swarm Optimisation for Feature Selection: A Filter Based Approach". Proceedings of 2012 IEEE World Congress on Computational Intelligence/ IEEE Congress on Evolutionary Computation (WCCI/CEC 2012). Brisbane, Australian. 10-17 June, 2012. IEEE Press. pp. 881-888.
Peng, Hanchuan, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." IEEE Transactions on pattern analysis and machine intelligence27.8 (2005): 1226-1238.

Filter FS based on Rough Set

- Promote rough set theory for feature selection
  - Others': mainly < 200 features
  - Ours: more than 600 features



Filter FS based on Rough Set



$$Fitness = \frac{\sum_{i=1}^{n} |\underline{apr}P\ U_i|}{|U|} + \frac{\sum_{x \in \{EqC\}} \frac{|x|}{|U|}}{|EqC|}$$

- U is the universe or the **whole dataset**
- U_i is **one class** in the dataset
- EqC: *equivalence classes*
- $\underline{apr}P$ is the **lower approximation** in **probabilistic** rough set theory
- A parameter **α** to relax the definition of $\underline{apr}P$

Bing Xue, Liam Cervante, Lin Shang, Will Browne and Mengjie Zhang."Binary PSO and rough set theory for feature selection: a multi-objective filter based approach". International Journal of Computational Intelligence and Applications (IJCIA), Vol. 13, No. 2 (2014). pp. 1450009(1-34)
Liam Cervante, Bing Xue, Lin Shang, Mengjie Zhang. "A Multi-Objective Feature Selection Approach Based on Binary PSO and Rough Set Theory". Proceedings of the 13th European Conference on Evolutionary Computation in Combinatorial Optimisation (EvoCOP 2013). Lecture Notes in Computer Science. Vol. 7832. Vienna, Austria. 3-5 April 2013. pp. 25-36
Liam Cervante, Bing Xue, Lin Shang and Mengjie Zhang. "A Dimension Reduction Approach to Classification Based on Particle Swarm Optimisation and Rough Set Theory". Proceedings of the 25th Australasian Joint Conference on Artificial Intelligence. Lecture Notes in Artificial Intelligence.Vol. 7691. Springer. Sydney, Australia, December 2012. pp. 313-325

Feature Subset Selection based on Ranking

- MRFS based on mutual information
  - Relief
  - Fisher Score

$$Fit_{mirf}(S) = \max\left(\underbrace{\sum_{x_k \in S} NI(x_k; y)}_{relevance} - \beta\left(\underbrace{\sum_{x_k \in S} NRelief_{order}(x_k) + NFisher_{order}(x_k)}_{ranking}\right)\right)$$

$$NI(x_k; y) = \frac{I(x_k; y)}{\sqrt{\sum_{m=1}^{M} I(x_m; y)^2}} \qquad NFisher_{order}(x_k) = \frac{Fisher_{order}(x_k)}{p * \sum_{m=1}^{M} Fisher_{order}(x_m)^2}$$

$$NI(x_k; x_l) = \frac{I(x_k; x_l)}{\sqrt{\sum_{m=1}^{M-1} \sum_{j=m+1}^{M} I(x_m; x_j)^2}} \qquad NRelief_{order}(x_k) = \frac{Relief_{order}(x_k)}{p * \sum_{m=1}^{M} Relief_{order}(x_k)^2}$$

Hancer, Emrah, Bing Xue, and Mengjie Zhang. "Differential evolution for filter feature selection based on information theory and feature ranking." Knowledge-Based Systems 140 (2018): 103-119.

Information Theory Feature Selection

- Information theory in evolutionary feature selection
  - Fast algorithm — mutual information
  - New measures, evaluate multiple features
  - Evolutionary multi-objective filter feature selection

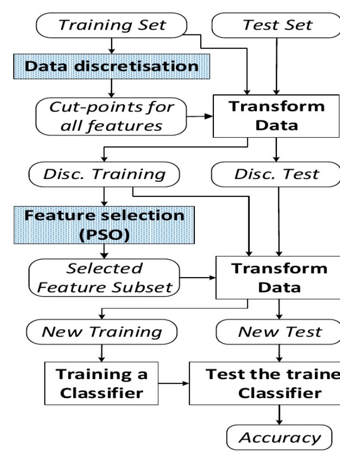
F-MI	0.05	0.05	0.05	0.06	0.07	0.09	0.15	0.18
F-E	2.88	97.7	8.64	27.95	9.85	256.57	2.96	236.42
F-RS	2.07	2485.61	8.21	55.3	14.81	1372.93	0.69	928.25
F-PRS	2.86	2766.29	8.28	38.36	9.95	1827.06	0.68	911.3
W-SVM	24.41	5143.18	53.28	270.64	118.37	2441.21	5.4	10937.87
W-5NN	6.12	9311.59	18.89	264.51	72.72	4095.07	1.68	1936.67
W-DT	5.19	189.43	10.53	43.15	47.87	244.55	3.82	529.7
W-NB	13.46	304.08	15.89	150.37	19.42	377.24	4.13	706.23

Bing Xue, Liam Cervante, Lin Shang, Will Browne, Mengjie Zhang. “A Multi-Objective Particle Swarm Optimisation for Filter Based Feature Selection in Classification Problems". Connection Science. Vol. 24, No. 2-3, pp. 91-116, 2012.
Bing Xue, Liam Cervante, Lin Shang, Will N. Browne, Mengjie Zhang. “Evolutionary Algorithms and Information Theory for Filter Based Feature Selection in Classification". International Journal on Artificial Intelligence Tools. Vol. 22, Issue 04, August 2013. pp. 1350024 -- 1 - 31. DOI: 10.1142/S0218213013500243.
Bing Xue, Mengjie Zhang and Will Browne. "A Comprehensive Comparison on Feature Selection Approaches to Classification". International Journal of Computational Intelligence and Applications (IJCIA). Vol. 14, No. 2. 2015. pp. 1550008 (1-23).

Feature Selection Though Data Discretisation

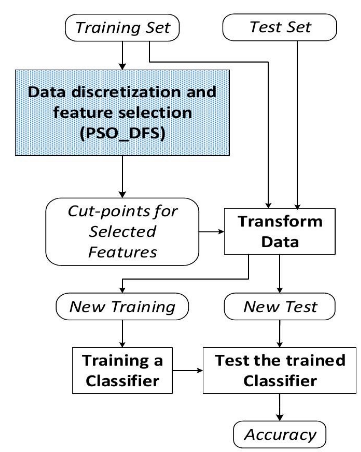
Bare-Bone Particle Swarm Optimisation

Two-stage (PSO-FS)



Proposed

- One-stage (PSO-DFS)



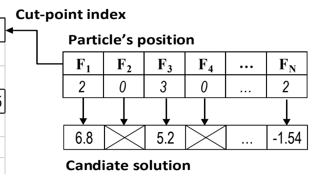
Binh Tran Ngan, Mengjie Zhang, Bing Xue. "Bare-Bone Particle Swarm Optimisation for Simultaneously Discretising and Selecting Features For High-Dimensional Classification". Proceedings of the 19th European Conference on the Applications of Evolutionary Computation (EvoApplications 2016, EvoIASP 2016). Lecture Notes in Computer Science. Vol. 9597. Porto, Portugal, March 30 - April 1, 2016. pp. 701-718
Binh Tran and Bing Xue and Mengjie Zhang."A New Representation in PSO for Discretisation-Based Feature Selection", IEEE Transactions on Cybernetics, vol. 48, no. 6, pp.1733-1746, 2018

FS Though Data Discretisation

- Utilising information from MLDP

Potential cut-point table

	#C	C_1	C_2	C_3	C_4
F_1	2	5.25	6.8		
F_2	3	10.7	24.9	50.2	
F_3	4	-0.45	0.67	5.2	20.5
F_4	1	-32.5			
...					
F_N	3	-3.72	-1.54	6.55	



- Hybrid fitness function

$$fitness = (\gamma \cdot balanced\_accuracy + (1-\gamma) \cdot distance)$$

where

$$distance = \frac{1}{1 + exp^{-5(D_b - D_w)}}$$

$$D_b = \frac{1}{M} \sum_{i=1}^{M} \min_{\{j|j \neq i, class(I_i) \neq class(I_j)\}} Dis(I_i, I_j)$$

$$D_w = \frac{1}{M} \sum_{i=1}^{M} \max_{\{j|j \neq i, class(I_i) = class(I_j)\}} Dis(I_i, I_j)$$

Binh Tran and Bing Xue and Mengjie Zhang."A New Representation in PSO for Discretisation-Based Feature Selection", IEEE Transactions on Cybernetics, vol. 48, no. 6, pp.1733-1746, 2018

FS Though Data Discretisation

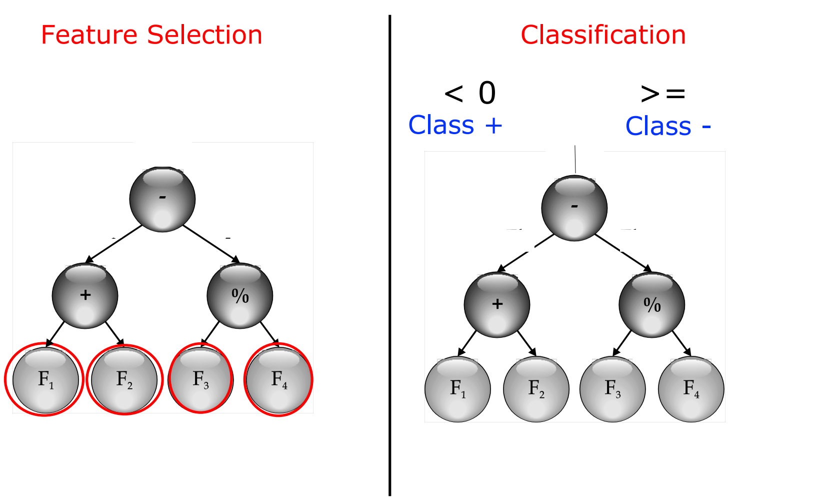
Table 3.2: Experimental results

Dataset	Method	Ave-Size	Best	Mean±StdDev	T_{RLS}	T_{CLS}
SRBCT	Full	2308.0	87.08		–	–
	PSO	915.0	99.17	96.56 ± 1.55	=	–
	PSO-RG	606.2	99.17	95.60 ± 1.66	=	–
	PSO-RLS	545.1	99.17	96.08 ± 1.73		–
	PSO-CLS	59.7	100.00	99.97 ± 0.15		
DLBCL	Full	5469.0	83.00		–	–
	PSO	2279.9	95.83	93.61 ± 2.19	=	+
	PSO-RG	305.5	93.33	85.92 ± 4.06	–	–
	PSO-RLS	1417.4	97.33	93.72 ± 1.79		+
	PSO-CLS	47.4	96.67	90.86 ± 3.19		
9Tumor	Full	5726.0	36.67		–	–
	PSO	2564.2	65.00	51.22 ± 5.23	+	=
	PSO-RG	1894.3	60.00	50.22 ± 4.54	+	=
	PSO-RLS	1352.0	58.33	48.39 ± 4.88		–
	PSO-CLS	46.7	60.00	51.39 ± 4.22		
Leukemia1	Full	5327.0	79.72		–	–
	PSO	2143.8	95.56	93.88 ± 1.47	=	–
	PSO-RG	786.1	94.31	89.63 ± 2.96	–	–
	PSO-RLS	1534.9	95.56	93.45 ± 1.71		–
	PSO-CLS	31.9	95.42	94.84 ± 1.16		
Brain1	Full	5920.0	72.08		–	–
	PSO	2481.6	80.00	75.89 ± 1.68	=	=
	PSO-RG	519.7	77.08	72.00 ± 3.13	–	–
	PSO-RLS	1549.0	77.50	75.00 ± 1.80		–
	PSO-CLS	1081.5	82.50	76.78 ± 2.09		
Leukemia2	Full	11225.0	89.44		–	–
	PSO	4577.7	93.89	92.07 ± 1.40	=	–
	PSO-RG	1116.6	95.00	90.72 ± 2.59	=	–
	PSO-RLS	3426.5	93.89	91.72 ± 1.46		–
	PSO-CLS	53.7	98.33	95.56 ± 1.68		
Brain2	Full	10367.0	62.50		–	–
	PSO	4249.9	81.25	75.83 ± 2.99	=	+
	PSO-RG	654.9	85.00	73.74 ± 4.95	=	=
	PSO-RLS	3099.0	82.50	75.35 ± 3.16		+
	PSO-CLS	2647.7	78.75	73.47 ± 2.82		
Prostate	Full	10509.0	85.33		=	–
	PSO	4603.1	88.17	85.04 ± 1.59	=	–
	PSO-RG	873.2	89.33	84.97 ± 2.55	=	–
	PSO-RLS	2690.3	89.17	85.79 ± 1.49		–
	PSO-CLS	2670.3	91.17	86.98 ± 1.76		
11Tumor	Full	12533.0	71.42		–	–
	PSO	5588.9	87.67	84.26 ± 1.35	=	–
	PSO-RG	2108.4	86.82	83.84 ± 2.25	=	–
	PSO-RLS	3163.9	87.77	84.19 ± 1.47		–
	PSO-CLS	266.8	90.72	87.51 ± 1.73		
Lung	Full	12600.0	78.05		–	–
	PSO	5353.3	84.73	83.18 ± 0.77	=	–
	PSO-RG	887.3	84.72	82.13 ± 1.78	–	–
	PSO-RLS	3453.9	86.87	83.50 ± 1.16		–
	PSO-CLS	311.6	96.43	90.78 ± 2.61		

Table 3.3: Comparison with traditional methods

Dataset	Method	Size	Training			Test		
			Best	Mean	S_{Tr}	Best	Mean	S_{Te}
SRBCT	Full	2,308.0		83.35	–		87.08	–
	LFS	6.1		98.19	–		88.75	–
	CFS	80.9		100.00	=		100.00	=
	PSO-CLS	59.7	100.00	100.00		100.00	99.97	
DLBCL	Full	5,469.0		81.71	–		83.00	–
	LFS	4.0		98.24	–		74.00	–
	CFS	58.0		99.22	–		91.67	=
	PSO-CLS	47.4	100.00	100.00		96.67	90.86	
9Tumor	Full	5,726.0		33.44	–		36.67	–
	LFS	12.6		82.39	–		41.67	–
	CFS	38.0		90.71	–		53.33	+
	PSO-CLS	46.7	97.78	97.78		60.00	51.39	
Leukemia1	Full	5,327.0		79.77	–		79.72	–
	LFS	4.8		99.17	–		81.39	–
	CFS	56.0		100.00	=		93.19	–
	PSO-CLS	31.9	100.00	100.00		95.42	94.84	
Brain1	Full	5,920.0		65.07	–		72.08	–
	LFS	9.9		89.13	–		59.17	–
	CFS	115.4		99.93	–		79.58	+
	PSO-CLS	1081.5	100.00	99.96		82.50	76.78	
Leukemia2	Full	11,225.0		88.82	–		89.44	–
	LFS	4.3		99.08	–		90.00	–
	CFS	79.0		100.00	=		98.89	+
	PSO-CLS	53.7	100.00	100.00		98.33	95.56	
Brain2	Full	10,367.0		63.52	–		62.50	–
	LFS	5.6		98.80	+		53.33	–
	CFS	63.4		100.00	+		71.25	–
	PSO-CLS	2647.7	99.20	98.55		78.75	73.47	
Prostate	Full	10,509.0		82.08	–		85.33	–
	LFS	4.9		82.44	–		73.17	–
	CFS	51.6		98.12	–		90.17	+
	PSO-CLS	2670.3	98.92	98.64		91.17	86.98	
Lung	Full	12,600.0		71.59	–		78.05	–
	LFS	12.2		95.12	–		80.55	–
	CFS	NA		NA			NA	
	PSO-CLS	311.6	99.11	99.02		96.43	90.78	
11Tumor	Full	12,533.0		71.01	–		71.42	–
	LFS	14.3		79.96	–		61.71	–
	CFS	NA		NA			NA	
	PSO-CLS	266.8	100.00	100.00		90.72	87.51	

GP for Embedded Feature Selection



Soha Ahmed, Mengjie Zhang, Lifeng Peng and Bing Xue. "Genetic Programming for Measuring Peptide Detectability". Proceedings of the 10th International Conference on Simulated Evolution and Learning (SEAL 2014). Lecture Notes in Computer Science. Vol. 8886. Dunedin, New Zealand. December 15-18, 2014. pp. 593-604

GP for Embeded Feature Selection

- Existing feature selection metrics have some biases class imbalance problems
- Each terminal node (leaf) consists of a ("basic") feature selection metric, which returns a set of features considered highly discriminative by such metric
- The set operations may be union (∪), intersection (∩), set difference (□), and so on

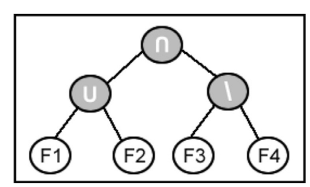


Fig. 2. Hypothetical individual under our modeling strategy: $(\cap (\cup f_1 f_2) (\setminus f_3 f_4)) = (f_1 \cup f_2) \cap (f_3 \setminus f_4)$.

Viegas, Felipe, Leonardo Rocha, Marcos Gonçalves, Fernando Mourão, Giovanni Sá, Thiago Salles, Guilherme Andrade, and Isac Sandin. "A genetic programming approach for feature selection in highly dimensional skewed data." *Neurocomputing* 273 (2018): 554-569.

GP for Embeded Feature Selection

General information about the datasets.

Dataset	Size	# Features	Density	Class distribution						
				# Classes	Minor class	1°quartile	Median	Mean	3°quartile	Major class
4UNI	8277	40,195	139.275	7	137	343	930	1182	1382	3759
REUT	8184	24,985	42.230	8	113	254.75	442	1023	946.5	3930
20NG	18,805	61,050	129.511	20	628	955	979	94,025	990	999
ACL-BIN	27,677	1,110,351	181.509	2	13,795	13816.75	13838.5	13838.5	13,860	13,882

Table 8
A comparison between the feature selection metrics considering the Top-4466 features of collection REUT.

	Mac.F_1(%)	Std. dev.	Mic.F_1(%)	Std. dev.
GP	**85.25**	**0.89**	**93.10**	**0.35**
GI	76.18	0.68	88.20	0.34
OR	71.12	2.50	83.37	2.57
χ^2	67.58	1.17	74.58	1.29
CC	76.26	0.98	88.27	0.47

Table 9
A comparison between the feature selection metrics considering the Top-42,096 features of collection ACL-BIN.

	Mac.F_1(%)	Std. dev.	Mic.F_1(%)	Std. dev.
GP	**86.22**	**0.21**	**86.22**	**0.21**
GI	85.23	0.49	85.32	0.47
OR	85.29	0.89	85.37	0.84
χ^2	78.56	6.87	79.51	5.95
CC	85.65	0.64	85.71	0.61

Table 10
A comparison between the feature selection metrics considering the Top-16,280 features of collection 20NG.

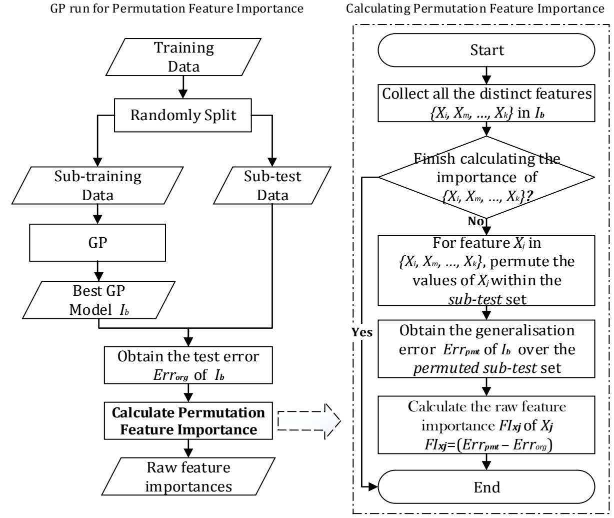
	Mac.F_1(%)	Std. dev.	Mic.F_1(%)	Std. dev.
GP	**82.39**	**0.41**	**83.06**	**0.40**
GI	80.19	0.40	80.84	0.40
OR	80.24	0.47	80.98	0.47
χ^2	79.92	0.38	80.70	0.34
CC	80.28	0.44	80.94	0.43

Table 11
A comparison between the feature selection metrics considering the Top-2824 features of collection 4UNI.

	Mac.F_1(%)	Std. dev.	Mic.F_1(%)	Std. dev.
GP	**55.46**	**0.68**	**62.85**	**0.94**
GI	48.27	1.75	55.00	1.36
OR	48.35	1.28	55.64	1.69
χ^2	50.95	0.86	59.20	1.13
CC	47.68	1.44	54.40	1.53

Viegas, Felipe, Leonardo Rocha, Marcos Gonçalves, Fernando Mourão, Giovanni Sá, Thiago Salles, Guilherme Andrade, and Isac Sandin. "A genetic programming approach for feature selection in highly dimensional skewed data." *Neurocomputing* 273 (2018): 554-569.

Feature Selection for Symbolic Regression



Qi. Chen, Mengjie. Zhang, and Bing. Xue. "Feature Selection to Improve Generalisation of Genetic Programming for High-Dimensional Symbolic Regression", IEEE Transaction on Evolutionary Computation, vol. 21, no. 5, pp. 792-806, 2017.

Feature Selection for Symbolic Regression

- Permutation feature importance:
  1. Randomly split the training data into a sub-training set and a sub-test set.
  2. Carry out a standard GP run and get the best-of-run individual I_b, which has the lowest training error over the sub-training set.
  3. Compute the generalisation error of I_b over the sub-test set, which is referred to $Err_{org}(I_b)$.
  4. For each distinct feature X_j in I_b, permute its values within the sub-test set, and get the test error of I_b on the permuted sub-test set, shown as $Err_{pmt}(I_b)$.
  5. Calculate the distance between $Err_{org}(I_b)$ and $Err_{pmt}(I_b)$, and use it to measure the raw feature importance of the feature $FI_{raw}(X_j)$, i.e.

$$FI_{raw}(X_j) = Err_{pmt}(I_b) - Err_{org}(I_b) \tag{3.3}$$

Dick, G., Rimoni, A. P., And Whigham, P. A re-examination of the use of genetic programming on the oral bioavailability prob- lem. In *Proceedings of the 17th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO)* (2015), ACM, pp. 1015– 1022.

Feature Selection for Symbolic Regression

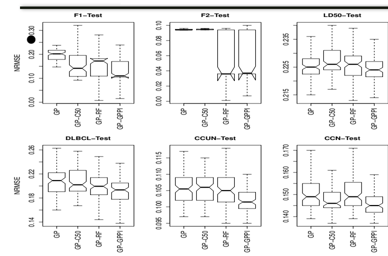


Figure 3.5: Distribution of the Corresponding **Test** NRMSEs.

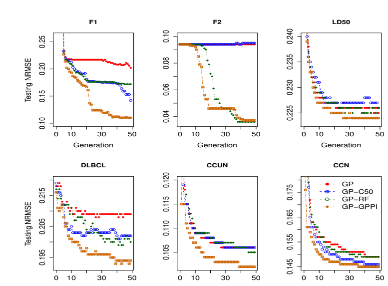


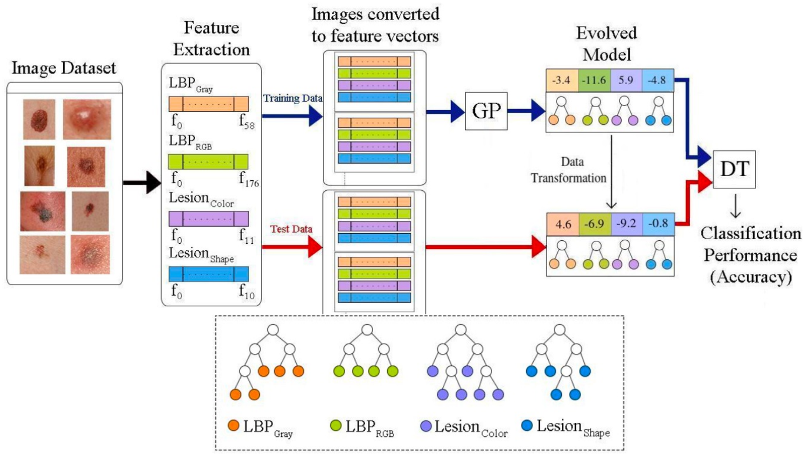
Figure 3.6: The **Testing** Error Evolution Plots.

Benchmark	Method	Training NRMSE (Medain±MAD)	Test NRMSE (Medain±MAD)	Significance Test (with GP-GPPI) (training, test)
F_1	LASSO	0.17	0.22	(−, −)
	RF	0.055±0.0013	0.16±0.0017	(−, −)
	GP	0.012±0.016	0.095±0.03	(+, −)
	GP-GPPI	0.037±0.043	0.049±0.064	
F_2	LASSO	0.11	0.09	(−, −)
	RF	0.040±4.20E-4	0.078±5.61E-4	(−, −)
	GP	0.002±2.97E-3	0.005±4.45E-3	(=, =)
	GP-GPPI	0.005±4.45E-3	0.004±2.97E-3	
LD50	LASSO	0.04	0.68	(+, −)
	RF	0.097±7.61E-4	0.23±0.0013	(+, −)
	GP	0.19±0.009	0.25±0.026	(+, −)
	GP-GPPI	0.21±4.45E-3	0.21±4.45E-3	
DLBCL	LASSO	0.18	0.22	(−, −)
	RF	0.058±7.77E-4	0.13±0.0014	(+, −)
	GP	0.088±0.012	0.182±0.032	(−, −)
	GP-GPPI	0.081±0.012	0.11±0.019	
CCUN	LASSO	0.13	0.15	(−, −)
	RF	0.030±1.18E-4	0.098±2.25E-4	(+, =)
	GP	0.073±1.48E-3	0.099±2.22E-3	(+, =)
	GP-GPPI	0.076±1.48E-3	0.097±2.97E-3	
CCN	LASSO	0.21	0.23	(−, −)
	RF	0.054±1.77E-4	0.141±3.44E-4	(+, −)
	GP	0.133±2.97E-3	0.143±2.97E-3	(+, −)
	GP-GPPI	0.139±2.22E-3	0.139±2.97E-3	

Qi. Chen, Mengjie. Zhang, and Bing. Xue. "Feature Selection to Improve Generalisation of Genetic Programming for High-Dimensional Symbolic Regression", IEEE Transaction on Evolutionary Computation, vol. 21, no. 5, pp. 792-806, 2017.

Multi-tree GP for FS: Wrapper and Embedded

- Multi-tree GP for FS: Wrapper and Embedded



Q. Ul Ain, H. Al-Sahaf, B. Xue, and M. Zhang, "A multi-tree genetic 1153 programming representation for melanoma detection using local and 1154 global features," in *Proc. 31st Australas. Joint Conf. Artif. Intell.*, 2018, 1155 pp. 111–123.
Qurrat Ul Ain, Harith Al-Sahaf, Bing Xue, Mengjie Zhang. "Generating Knowledge-guided Discriminative Features Using Genetic Programming for Melanoma Detection", IEEE Transactions on Emerging Topics in Computational Intelligence, vol. , Issue. , pp. , Online 2020 (DOI: 10.1109/TETCI.2020.2983426). 14pp

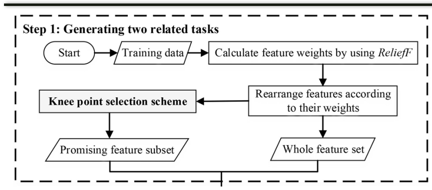
GP for FS: Wrapper and Embedded

- Multi-tree GP for FS: Wrapper and Embedded

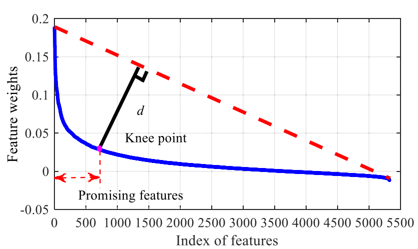
		PH^2		*Dermofit*	
		training	test	training	test
Multi-tree GP Wrapper	NB	92.21 ± 0.60	**85.70 ± 2.65**	89.40 ± 0.70	80.45 ± 2.18
	SVM	89.51 ± 0.79	81.52 ± 3.58	88.64 ± 0.74	80.33 ± 2.71
	$k-NN$	93.91 ± 0.55	61.26 ± 4.05	91.68 ± 0.60	69.27 ± 2.89
	$J48$	99.71 ± 0.15	85.18 ± 3.72	98.33 ± 0.25	**84.18 ± 4.11**
Multi-tree GP Embedded	–	79.69 ± 1.35	78.87 ± 2.92 +	75.63 ± 0.99	74.57 ± 1.86 +
Non-GP Methods	NB	93.85 ± 1.11	77.81 ± 08.44 +	86.42 ± 0.70	72.26 ± 11.62 +
	SVM	89.62 ± 1.37	70.00 ± 10.29 +	95.16 ± 0.84	70.02 ± 10.34 +
	$k-NN$	100.0 ± 0.00	75.63 ± 14.71 +	100.0 ± 0.00	72.08 ± 09.52 +
	$J48$	97.05 ± 2.71	71.25 ± 11.08 +	97.09 ± 1.31	73.98 ± 10.65 +
	RF	100.0 ± 0.00	76.56 ± 09.81 +	99.93 ± 0.22	71.30 ± 09.80 +
	MLP	78.92 ± 1.23	78.44 ± 10.96 +	79.83 ± 1.95	73.00 ± 08.51 +
Single tree GP Wrapper	LBP_{Gray}	88.32 ± 0.78	60.19 ± 4.73 +	74.44 ± 1.39	53.88 ± 3.44 +
	LBP_{rgb}	91.03 ± 0.94	65.70 ± 6.25 +	76.62 ± 1.56	53.80 ± 3.36 +
	$Lesion_{Color}$	87.54 ± 1.09	61.81 ± 4.56 +	87.55 ± 0.96	65.79 ± 5.90 +
	$Lesion_{Shape}$	84.27 ± 1.67	61.65 ± 4.28 +	85.89 ± 0.75	64.88 ± 3.69 +
Single tree GP Embedded	LBP_{Gray}	82.84 ± 1.35	65.96 ± 3.96 +	73.41 ± 1.87	59.91 ± 3.57 +
	LBP_{RGB}	84.42 ± 1.43	73.87 ± 2.34 +	75.52 ± 1.62	63.26 ± 3.19 +
	$Lesion_{Color}$	81.59 ± 2.31	65.70 ± 3.61 +	81.06 ± 1.31	74.13 ± 2.67 +
	$Lesion_{Shape}$	78.06 ± 1.97	49.89 ± 5.34 +	74.74 ± 2.67	61.74 ± 7.06 +

Q. Ul Ain, H. Al-Sahaf, B. Xue, and M. Zhang, "A multi-tree genetic 1153 programming representation for melanoma detection using local and 1154 global features," in *Proc. 31st Australas. Joint Conf. Artif. Intell.*, 2018, 1155 pp. 111–123.
Qurrat Ul Ain, Harith Al-Sahaf, Bing Xue, Mengjie Zhang. "Generating Knowledge-guided Discriminative Features Using Genetic Programming for Melanoma Detection", IEEE Transactions on Emerging Topics in Computational Intelligence, vol. , Issue. , pp. , Online 2020 (DOI: 10.1109/TETCI.2020.2983426). 14pp

Evolutionary Multitasking-Based Feature Selection Method

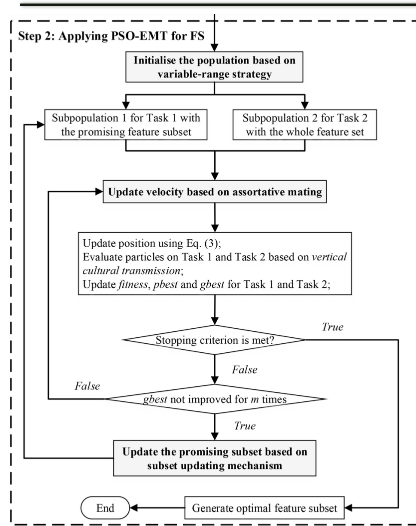


- ReliefF:
  - Feature ranking/weighting
  - Based on *h* nearest neighbors of the same category as the selected sample R , H_j.
  - Find *h* nearest neighbors different from category R_i



K. Chen, B. Xue, M. Zhang and F. Zhou, "An Evolutionary Multitasking-Based Feature Selection Method for High-Dimensional Classification," in IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2020.3042243.

Evolutionary Multitasking-Based Feature Selection Method



- Fitness function:

$$\text{fitness} = \alpha * \gamma_R(D) + (1-\alpha) * \frac{|S|}{|N|}$$

- Knowledge transfer
  - Multipopulation framework
  - Assortative mating - crossover
  - Vertical cultural transmission
    - assign the skill factor for each generated individual
- Variable-Range Strategy
  - linearly reduced from [0, 1] to [0, δ]
- Subset Updating Mechanism:
  - Update candidate features in Task 1, size unchanged

K. Chen, B. Xue, M. Zhang and F. Zhou, "An Evolutionary Multitasking-Based Feature Selection Method for High-Dimensional Classification," in IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2020.3042243.
L. Feng et al., "An empirical study of multifactorial PSO and muli-tifactorial DE," in Proc. IEEE Congr. Evol. Comput., San Sebastian, Spain, 2017, pp. 921–928.

Evolutionary Multitasking-Based Feature Selection Method

Dataset	Method	Time (m)	Size	Best	Mean ± Std	W
11Tumor	FULL		12533.00	71.37		+
	PSO	418.54	6205.00	75.59	71.81 ± 1.75	+
	CSO	6278.54	589.36	87.45	83.50 ± 1.70	+
	AMSO	91.22	319.00	85.06	83.10 ± 1.31	+
	VLPSO	67.41	249.30	85.21	80.92 ± 2.39	+
	PSO-EMT	106.53	541.45	**89.09**	**86.15 ± 1.45**	
Lung	FULL		12600.00	78.12		+
	PSO	574.17	6234.70	82.72	78.77 ± 1.53	+
	CSO	5419.71	230.41	93.47	88.94 ± 1.75	+
	AMSO	255.32	193.47	92.64	89.97 ± 1.80	+
	VLPSO	78.00	176.00	92.86	89.55 ± 1.68	+
	PSO-EMT	134.59	617.61	**93.55**	**91.09 ± 0.94**	

Traditional methods

11Tumor	CFS	4681.40	379.00	83.91		+
	FCBF	25.06	394.00	82.94		+
	ReliefF	15.54	1114.00	84.91		+
	SBMLR	13.87	15.00	70.13		+
	SPEC	2.28	6158.00	83.30		+
	PSO-EMT	6391.56	541.45	**89.09**	**86.15 ± 1.45**	
Lung	CFS	10029.00	550.00	93.31		−
	FCBF	37.95	453.00	92.06		−
	ReliefF	16.69	1440.00	90.17		+
	SBMLR	28.13	30.00	92.62		−
	SPEC	3.11	2378.00	81.29		+
	PSO-EMT	8075.17	617.61	**93.55**	91.09 ± 0.94	

With VS without knowledge transfer

Dataset	Method		Time (m)	Size	Best	Mean ± Std	W
11Tumor	PSO-EMT	Task 1	106.53*	**541.45**	86.15 ± 1.45	
		Task 2		**4999.10**	**84.15 ± 1.48**	
	PSO-EMT^{m-}	Task 1	25.14	599.71	**86.51 ± 1.74**	≈
		Task 2	213.05	5673.12	84.01 ± 1.76	≈
Lung	PSO-EMT	Task 1	134.59*	**617.61**	**91.09 ± 0.94**	
		Task 2		**2472.30**	**89.23 ± 1.34**	
	PSO-EMT^{m-}	Task 1	36.91	725.31	90.67 ± 1.60	≈
		Task 2	265.34	3561.24	86.74 ± 1.35	+

With VS without subset updating

11Tumor	PSO-EMT^{v-}	121.09	**535.25**	89.06	**86.20 ± 1.53**	≈
	PSO-EMT	**106.53**	541.45	**89.09**	86.15 ± 1.45	
Lung	PSO-EMT^{v-}	169.36	641.28	93.03	90.02 ± 1.37	+
	PSO-EMT	**134.59**	**617.61**	**93.55**	**91.09 ± 0.94**	

With VS without the variable-range strategy

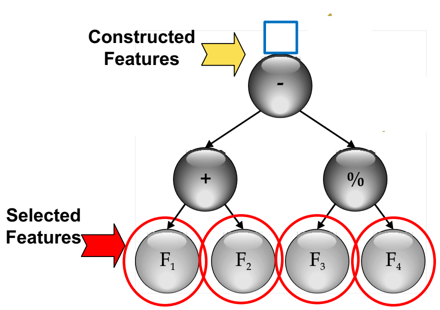
Leuk2	PSO-EMT^{s-}	12.78	272.36	94.34	88.91 ± 2.49	+
	PSO-EMT	**12.19**	**224.44**	**94.46**	**90.07 ± 2.47**	
Brain2	PSO-EMT^{s-}	12.01	526.26	78.00	72.21 ± 3.63	≈
	PSO-EMT	**11.51**	**499.69**	**80.00**	**72.27 ± 4.09**	
Leuk3	PSO-EMT^{s-}	18.91	414.50	96.07	93.49 ± 1.98	+
	PSO-EMT	**14.72**	**268.08**	**97.32**	**94.51 ± 1.50**	
11Tumor	PSO-EMT^{s-}	**105.74**	**539.22**	**90.85**	**86.33 ± 1.60**	≈
	PSO-EMT	106.53	541.45	89.09	86.15 ± 1.45	
Lung	PSO-EMT^{s-}	145.45	676.74	**94.04**	**91.16 ± 1.19**	≈
	PSO-EMT	**134.59**	**617.61**	93.55	91.09 ± 0.94	

K. Chen, B. Xue, M. Zhang and F. Zhou, "An Evolutionary Multitasking-Based Feature Selection Method for High-Dimensional Classification," in IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2020.3042243.

Feature Construction

Why use GP for Feature Construction?

- GP is flexible in making mathematical and logical functions
- There isn't much structural (topological) information in the search space of possible functions, so using a meta-heuristic approach (such as evolutionary computation) seems reasonable.

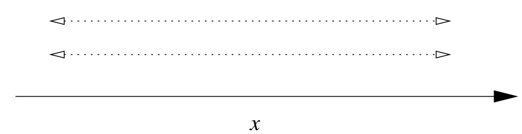


GP for Feature Construction

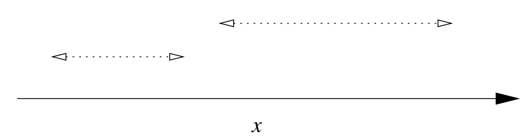
- The interval of class ***c*** could be formulated as follows if the class distributions were normal, measuring the purity, purity of each class

$$I_c = [\mu_c - 3\sigma_c, \mu_c + 3\sigma_c]$$

- Overlapping intervals, bad



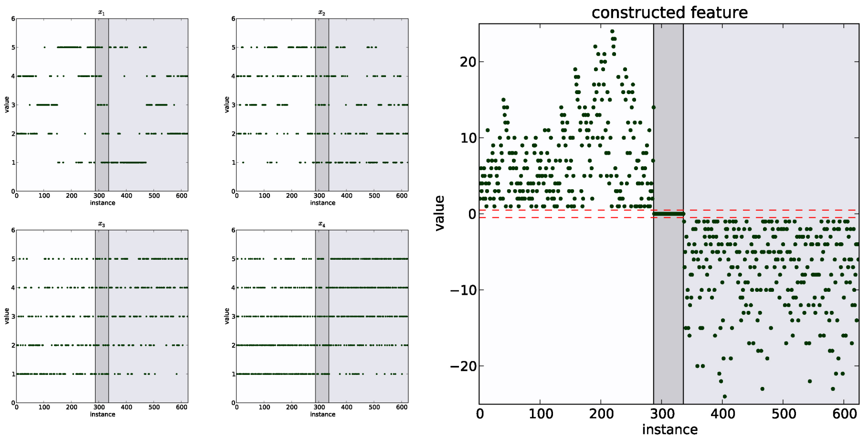
- Non-overlapping intervals, good



Neshatian, K.; Mengjie Zhang; Andreae, P., "A Filter Approach to Multiple Feature Construction for Symbolic Learning Classifiers Using Genetic Programming," *IEEE Transactions on Evolutionary Computation* , vol.16, no.5, pp.645-661, Oct. 2012

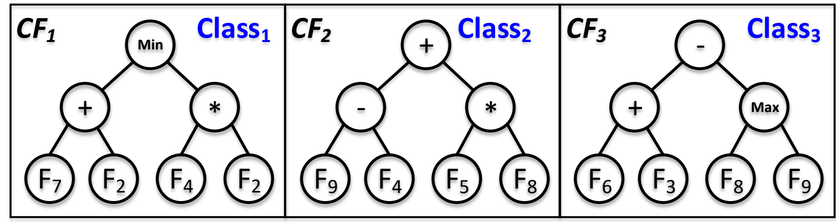
GP for FC Measure: Original VS Constructed

- 4 features, 3 classes



Neshatian, K.; Mengjie Zhang; Andreae, P., "A Filter Approach to Multiple Feature Construction for Symbolic Learning Classifiers Using Genetic Programming," in *Evolutionary Computation, IEEE Transactions on* , vol.16, no.5, pp.645-661, Oct. 2012

Multi-tree GP for FC



1. **Constructed feature set (CF): CF_1, CF_2, CF_3**
   - CF_1 = Min ((F_7 + F_2), (F_4 * F_2))
   - CF_2 = (F_9 - F_4) + (F_5 * F_8)
   - CF_3 = (F_6 + F_3) - Max (F_8 , F_9))
2. **Selected feature set (Ter): F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9**
3. **Combination set (CFTer): CF_1, CF_2, CF_3, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9**

Binh Ngan Tran, Bing Xue, Mengjie Zhang. "Class Dependent Multiple Feature Construction Using Genetic Programming for High-Dimensional Data". Proceedings of the 30th Australasian Joint Conference on Artificial Intelligence (AI2017) Lecture Notes in Computer Science. Vol. 10400. Springer. Melbourne, Australia, August 19-20th, 2017. pp. 182-194.
Binh Tran and Bing Xue and Mengjie Zhang. "Genetic programming for multiple-feature construction on high-dimensional classification", Pattern Recognition, vol. 93, pp. 404-417, 2019

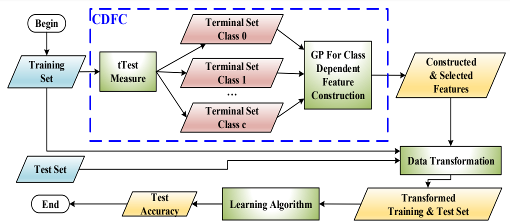
Class-Dependent Terminal Sets

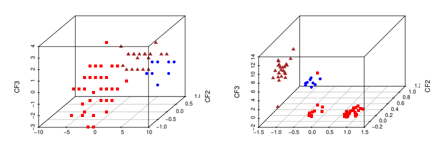
- A constructed feature aims at discriminating instances of a class (c) from other classes => It is constructed based on features that are relevant to class c.
- Class-relevant measure:

$$Rel_{f,c} = \begin{cases} 0, & \text{if p-value} \geq 0.05 \\ \frac{|\textbf{t-value}(f_{class=c}, f_{class\neq c})|}{\textbf{p-value}}, & \text{otherwise} \end{cases}$$

- Half of the top-ranked features will be used to form the terminal set of class c.
  - => eliminate irrelevant features
  - => narrow the search space
- $Fitness = \alpha \cdot Accuracy + (1 - \alpha) \cdot Distance$

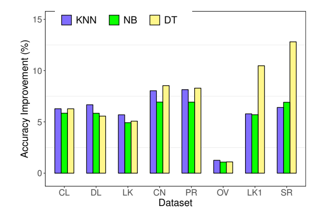
Class-Dependent Multi-tree GP for FC





(a) 1TGPFC on Leukemia1 (b) CDFC on Leukemia1

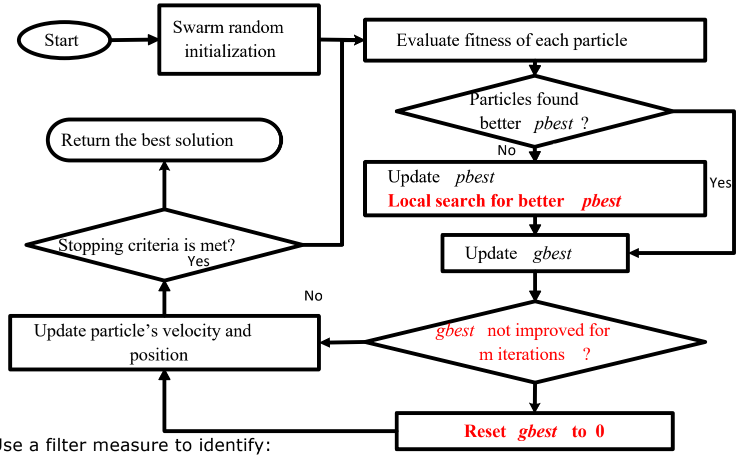
Figure 6.10: Constructed features on Leukemia1.



(d) Accuracy Improvement

Binh Ngan Tran, Bing Xue, Mengjie Zhang. "Class Dependent Multiple Feature Construction Using Genetic Programming for High-Dimensional Data". Proceedings of the 30th Australasian Joint Conference on Artificial Intelligence (AI2017) Lecture Notes in Computer Science. Vol. 10400. Springer. Melbourne, Australia, August 19-20th, 2017. pp. 182-194.
Binh Ngan Tran, Evolutionary Computation for Feature Manipulation in Classification on High-dimensional Data, PhD thesis, Victoria University of Wellington, New Zealand, http://researcharchive.vuw.ac.nz/xmlui/handle/10063/7078

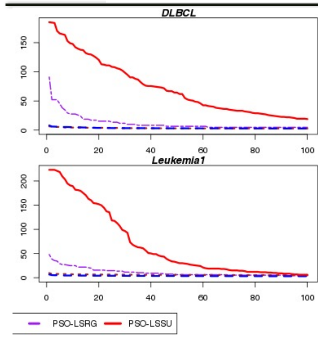
Biology: PSO with local search on *pbest* and resetting *gbest*



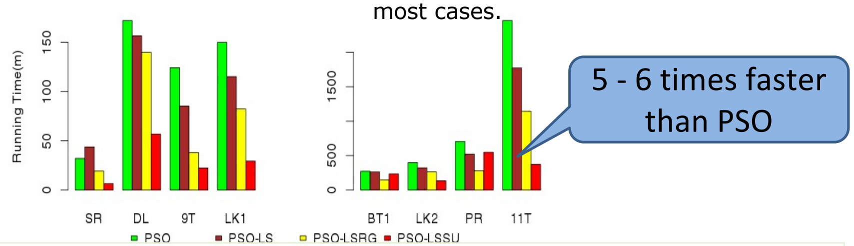
- Use a filter measure to identify:
  - Relevant features: correlated to the class label.
  - Redundant features: correlated with each other.
- Symmetric uncertainty (SU) is a normalised version of information gain (IG).

Binh Tran, Mengjie Zhang and Bing Xue, "A PSO Based Hybrid Feature Selection Algorithm For High-Dimensional Classification". Proceedings of 2016 IEEE World Congress on Computational Intelligence/ IEEE Congress on Evolutionary Computation (WCCI/CEC 2016). Vancouver, Canada. 24-29 July, 2016. pp.3801-3808.

Biology: PSO with local search on *pbest* and resetting *gbest*



- A PSO based hybrid FS algorithm for high-dimensional classification.
- PSO-LSSU combines wrapper and filter measures:
  - The fitness function.
  - The local search.
- PSO-LSSU achieved much smaller feature subsets with significantly better classification performance than the compared methods in most cases.



Binh Tran, Mengjie Zhang and Bing Xue, "A PSO Based Hybrid Feature Selection Algorithm For High-Dimensional Classification". Proceedings of 2016 IEEE World Congress on Computational Intelligence/ IEEE Congress on Evolutionary Computation (WCCI/CEC 2016). Vancouver, Canada. 24-29 July, 2016. pp.3801-3808.

GP for FS and FC in Biology

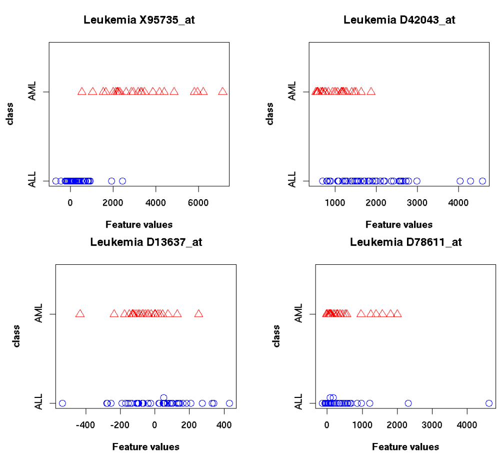


Figure 1: Leukemia constructed feature

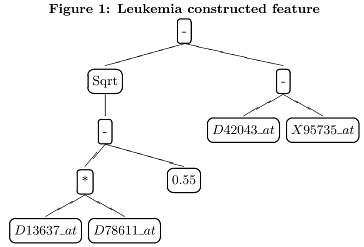
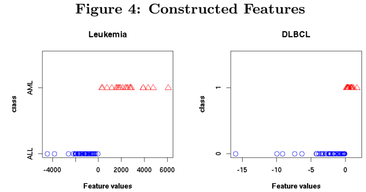


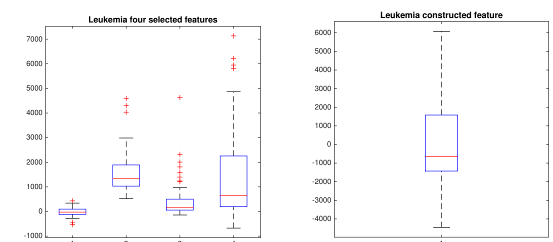
Figure 4: Constructed Features

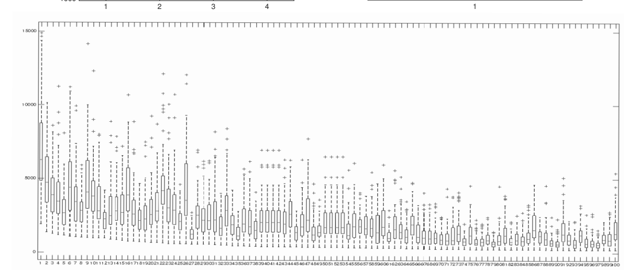


Binh Tran, Bing Xue and Mengjie Zhang. "Genetic Programming for Feature Construction and Selection in Classification on High-dimensional Data", Memetic Computing, vol 8, Issue 1, pp3-15. 2016

GP for FS and FC

- Feature distribution

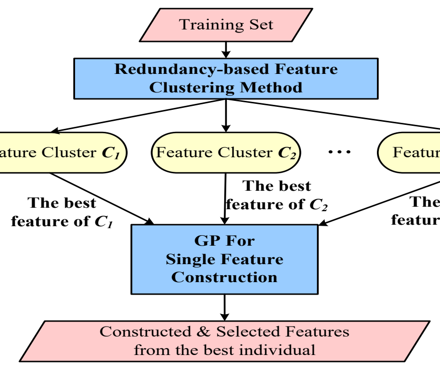




Binh Tran, Bing Xue and Mengjie Zhang. "Genetic Programming for Feature Construction and Selection in Classification on High-dimensional Data", Memetic Computing, vol 8, Issue 1, pp3-15. 2016

Feature Clustering for GP-Based FC on High-Dimensional Data

Eliminating redundant features may improve GP performance in FC



Binh Ngan Tran, Bing Xue, Mengjie Zhang. "Using Feature Clustering for GP-Based Feature Construction on High-Dimensional Data". Proceedings of the 20th European Conference on Genetic Programming (EuroGP 2017). Lecture Notes in Computer Science. Vol. 10196. Amsterdam. 18-21 April 2017. pp.210--226.

Feature Clustering for GP-Based FC on High-Dimensional Data

- Results of cluster analysis

Dataset	#Features	#Clusters	%Dimensionality reduction	ASC
Colon	2000	104.10	0.95	0.80
DLBCL	5469	819.20	0.85	0.96
Leukemia	7129	901.30	0.87	0.98
CNS	7129	79.30	0.99	1.00
Prostate	10509	1634.80	0.84	0.85
Ovarian	15154	601.20	0.96	0.31
Alizadeh	1095	93.60	0.91	0.94
Yeoh	2526	97.60	0.96	1.00

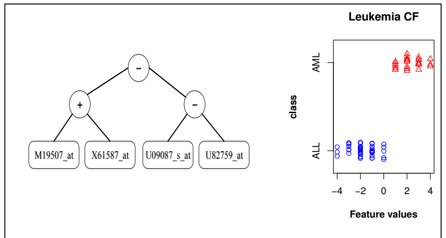


Figure 5.5: Leukemia constructed feature.

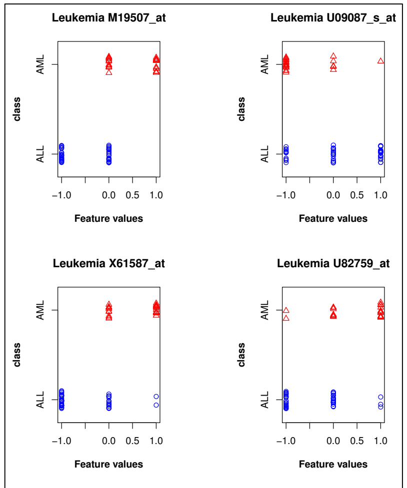
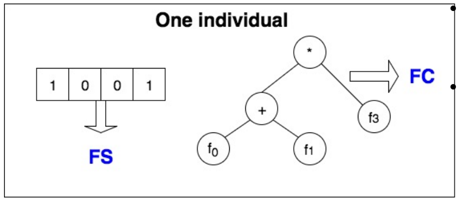
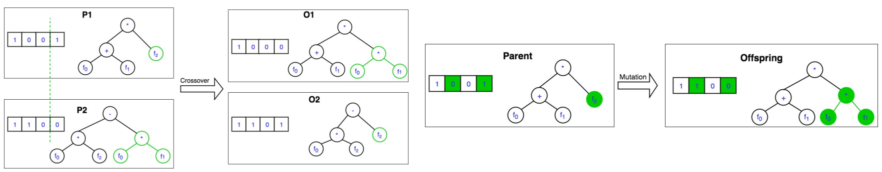


Figure 5.6: Leukemia original features.

Hybrid GAs-GP Representation for FS and FC



- A combination of the vector and tree-based representations.
- The high-level feature and selected original features are evaluated as one feature set, which ensures to consider the interactions between two kinds of features.



Hoai Bach Nguyen, Bing Xue, and Peter Andreae. "A Hybrid GA-GP Method for Feature Reduction in Classification". Proceedings of the 11th International Conference on Simulated Evolution and Learning (SEAL 2017). Lecture Notes in Computer Science. Vol. 10593. Shenzhen, China. November 10-13, 2017. pp. 591-604.

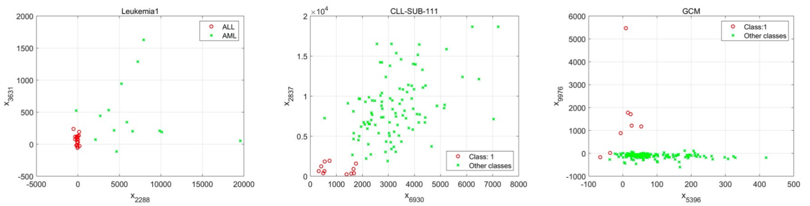
Multi-objective GP for FC and FS

- c-class problem into c binary classification problems
- evolve c sets of binary classifiers employing a steady-state multi-objective GP with three minimizing objectives.
  - (i) false positives (FPs),
  - (ii) false negatives (FNs),
  - (iii) the number of leaf nodes of the corresponding encoding tree.
- Each binary classifier is composed of a binary tree and a linear support vector machine (SVM)
- During crossover and mutation, the SVM-weights are used to determine the usefulness of the corresponding nodes

Nag, Kaustuv, and Nikhil R. Pal. "Feature Extraction and Selection for Parsimonious Classifiers with Multiobjective Genetic Programming." IEEE Transactions on Evolutionary Computation (2019).

Multi-objective GP for FC and FS

Dataset	%TA[a]	FS[b]	TS[c]	$(F/T)^d$	$\%F^e$	$(\%F/T)^f$
Colon	81.10 (3.01)	188.4	5.24	3.09	9.42	0.15
TOX-171	79.06 (1.34)	494.2	6.50	3.71	8.60	0.06
Leukemia1	92.33 (1.71)	211.2	3.00	1.98	2.96	0.03
Leukemia2	93.17 (1.37)	354.8	3.00	1.99	4.98	0.03
CLL-SUB-111	77.30 (3.26)	586.8	5.45	3.18	5.17	0.03
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



(a) Feature x_{3631}-vs.-x_{2288} of Leukemia1. (b) Feature x_{2837}-vs.-x_{6930} of CLL-SUB-111. (c) Feature x_{9976}-vs.-x_{5396} of GCM.

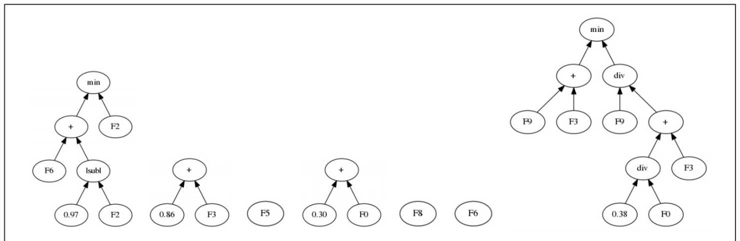
Fig. 4: Linearly separable feature pairs for one class vs. other classes of (a) Leukemia1, (b) CLL-SUB-111, and (c) GCM.

[a]Test Accuracy (standard deviation is provided within parenthesis), [b]Number of Features Selected per Classifier, [c]Tree Size, [d]Number of Features per Tree, [e]Percentage of Features Selected, [f]Percentage of Features Selected per Tree.

Nag, Kaustuv, and Nikhil R. Pal. "Feature Extraction and Selection for Parsimonious Classifiers with Multiobjective Genetic Programming." IEEE Transactions on Evolutionary Computation (2019).

GP for FC in Clustering: Multi-Tree

- Each tree creates a single constructed feature.
- Each individual contains *t* trees, to give *t* constructed features.



(a) The evolved multi-tree GP individual with 7 trees.

$$\left[min(F_6+|0.97-F_2|, F_2),\ (0.86+F_3),\ F_5,\ (0.3+F_0),\ F_8,\ F_6,\ min\left((F_9+F_3), \frac{F9}{(\frac{0.38}{F_0}+F_3)}\right)\right]$$

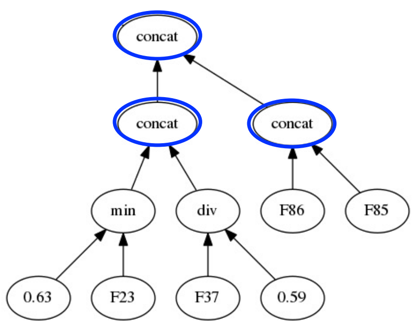
(b) Constructed feature set generated by the programs.

Andrew Lensen, Bing Xue, and Mengjie Zhang. "New Representations in Genetic Programming for Feature Construction in k-means Clustering". Proceedings of the 11th International Conference on Simulated Evolution and Learning (SEAL 2017). Lecture Notes in Computer Science. Vol. 10593. Shenzhen, China. November 10-13, 2017. pp. 543--555.

GP Representation – Vector

- Having to set *t* is annoying. Can we use a single tree?
- Introduce a new concat operator which can create vectors of CFs.
  - Automatically build up a suitable length vector.
  - Extend the function set to work on vectors.

Example output gives 4 features:
[min(0.63,F23), F37/0.59, F86, F85]

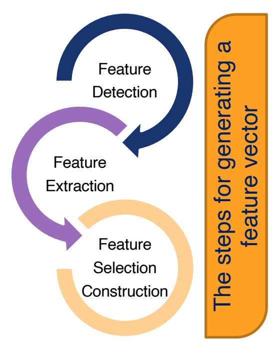


- However, each tree must be larger.

Andrew Lensen, Bing Xue, and Mengjie Zhang. "New Representations in Genetic Programming for Feature Construction in k-means Clustering". Proceedings of the 11th International Conference on Simulated Evolution and Learning (SEAL 2017). Lecture Notes in Computer Science. Vol. 10593. Shenzhen, China. November 10-13, 2017. pp. 543--555.

Image Recognition/Classification

- GP for image ananlysis: evolve image descriptors
- Keypoints identification, feature extraction, feature construction/selection



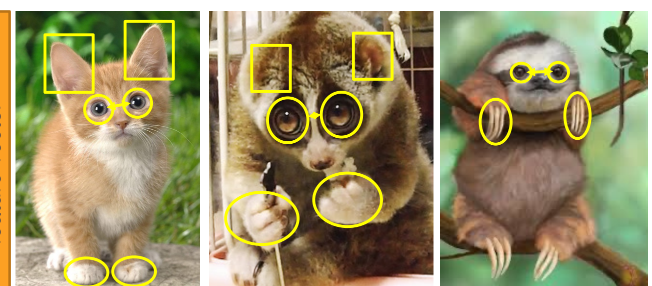
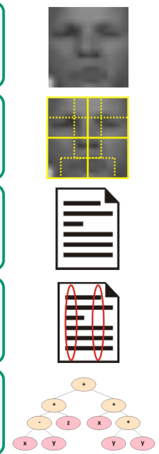


Image Recognition/Classification

- The traditional way
- Domain-specific pre-extracted features approach (DS-GP)

Input	The input is raw image pixel values
Design	The feature areas need to be designed by domain-experts
Feature Extraction	Transform the pixel values of the selected areas to a different domain
Feature Selection	Select a subset out of the extracted features (optional)
Classification	Feed the extracted features (with or without selection) to GP-based classifier



GP for Image Classification

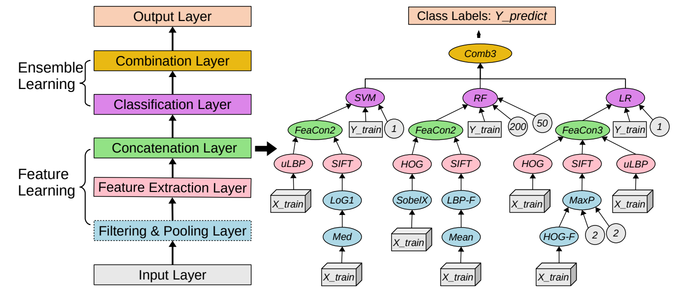


Fig. 8.2 The new program structure of FELGP and an example solution/program that can be evolved by FELGP

Y. Bi, B. Xue and M. Zhang, "Genetic Programming With a New Representation to Automatically Learn Features and Evolve Ensembles for Image Classification," in *IEEE Transactions on Cybernetics*, vol. 51, no. 4, pp. 1769-1783, April 2021, doi: 10.1109/TCYB.2020.2964566.

Biological Datasets

Data set	# Features	# Samples	# Classes
Pancreatic Cancer	6771	181	2
Ovarian Cancer1	15154	253	2
Ovarian Cancer 2	15000	216	2
Prostate Cancer	15000	322	4
Toxpath	7105	115	4
Arcene	10,000	200	2
Apple-plus	773	40	4
Apple-minus	365	40	4

Biomarker Identification

m/z values in Apple-plus data set (12 biomarkers)	New Method (9 ✓)	Method 2 (3✓)
331.21	✗	✓
471.09	✓	✓
107.05, 169.05, 238.05, 275.09, 456.11, 459.13	✓	✗
456.62, 475.10	✗	✗
449.11	✓	✓
229.09	✓	✗

Apple minus m/z (5 biomarkers)	New Method (5 ✓)	Method 2 (2✓)
463.0	✓	✗
447.09	✓	✓
273.03	✓	✓
435.13	✓	✗
227.07	✓	✗

Soha Ahmed, Mengjie Zhang, Lifeng Peng and Bing Xue."Multiple Feature Construction for Effective Biomarker Identification and Classification using Genetic Programming". Proceedings of 2014 Genetic and Evolutionary Computation Conference (GECCO 2014). ACM Press. 2014.pp.249—256
Soha Ahmed, Genetic Programming for Biomarker Detection in Classification of Mass Spectrometry Data, PhD thesis, 2015, School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

GP for FS/FC for Visualisations

- Multi-objective:
  - visualizations quality;
  - visualizations complexity;
- Based on t-distributed Stochasti Neighbor Embedding (t-SNE)

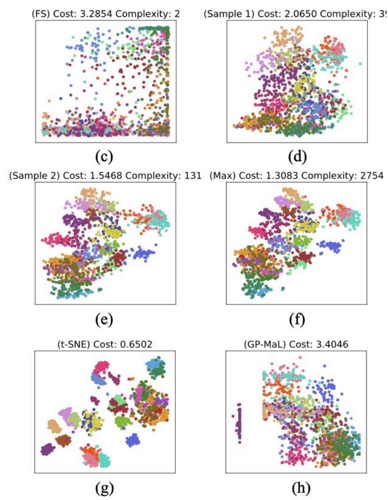


Fig. 6. Isolet.

Andrew Lensen, Mengjie Zhang, and Bing Xue. "Genetic Programming for Evolving a Front of Interpretable Models for Data Visualization", IEEE Transactions on Cybernetics, vol. , Issue. , pp. ,. Online 21 Feb 2020

GP for FS/FC for Visualisations

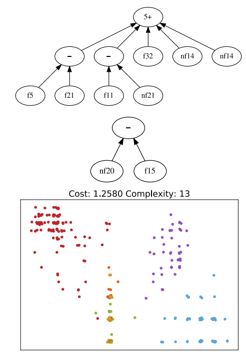


Fig. 11. Complexity of 13.

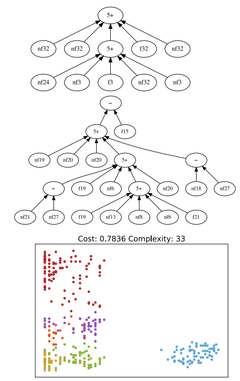


Fig. 14. Complexity of 33.

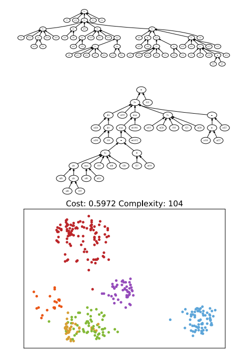
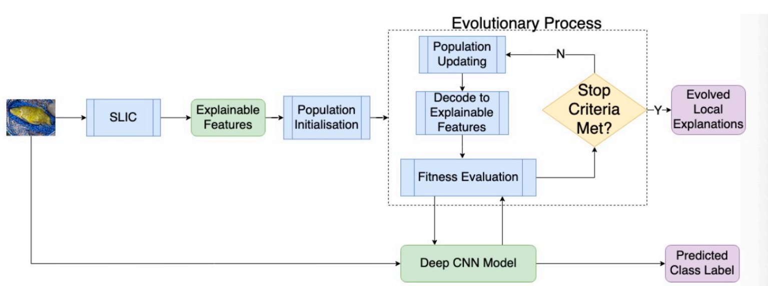


Fig. 16. Complexity of 104.

Andrew Lensen, Mengjie Zhang, and Bing Xue. "Genetic Programming for Evolving a Front of Interpretable Models for Data Visualization", IEEE Transactions on Cybernetics, vol. , Issue. , pp. ,. Online 21 Feb 2020

Feature Selection for Explainable ML



Bin Wang, Wenbin Pei, Bing Xue, and Mengjie Zhang. "Evolving Local Interpretable Model-agnostic Explanations for Deep Neural Networks in Image Classification". Proceedings of 2021 Genetic and Evolutionary Computation Conference (GECCO 2021 Companion)

Feature Selection for Explainable ML

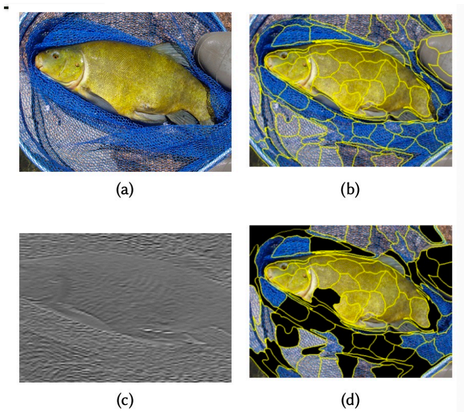


Figure 4: Explainable features vs unexplainable feature maps. (a). The original image. (b). Superpixels, i.e. segments obtained by SLIC. (c). A sample unexplainable feature map of ResNet. (d). Evolved explainable features.

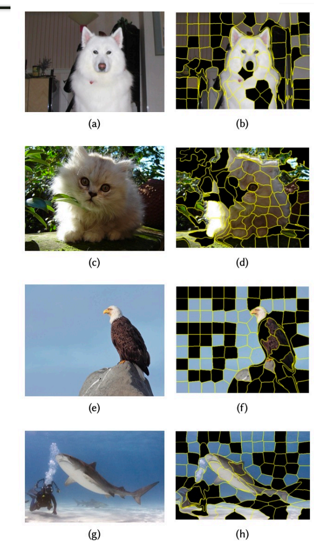
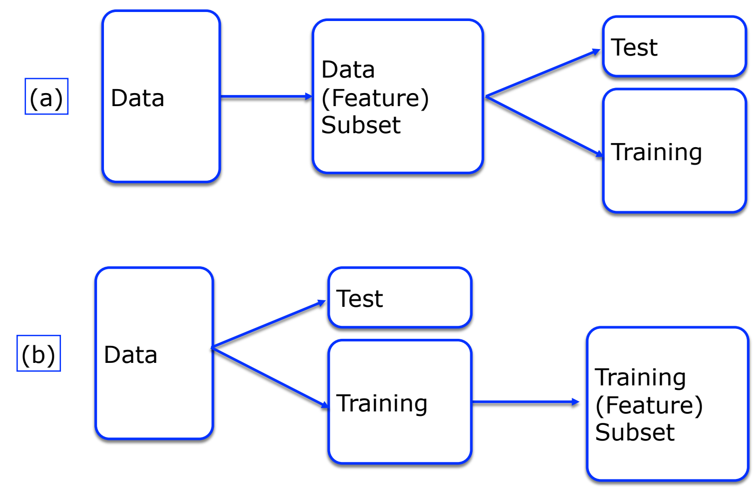


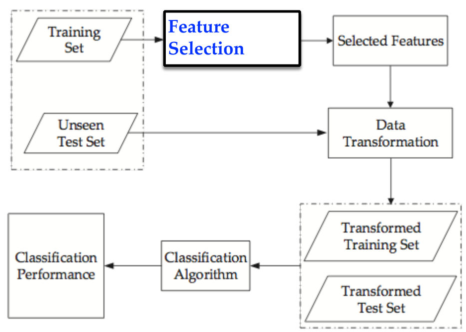
Figure 6: Evolved local interpretable explanations based on ResNet. From the top to the bottom, a samoyed, a Persian cat, a bald eagle and a tiger shark are presented.

Issues and Challenges

Any problem ?



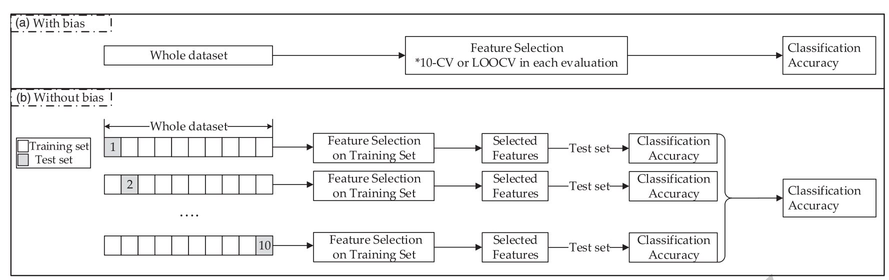
Feature Selection/Construction Bias



- If the whole dataset is used during FS/FC process, the experiments(or evaluation) have FS/FC Bias
- What if only a small number of instances available ?
  - In classification, use k-fold cross validation
  - How to use k-fold cross validation in FS/FC to evaluate a FS/FC system ?

Feature Selection Bias

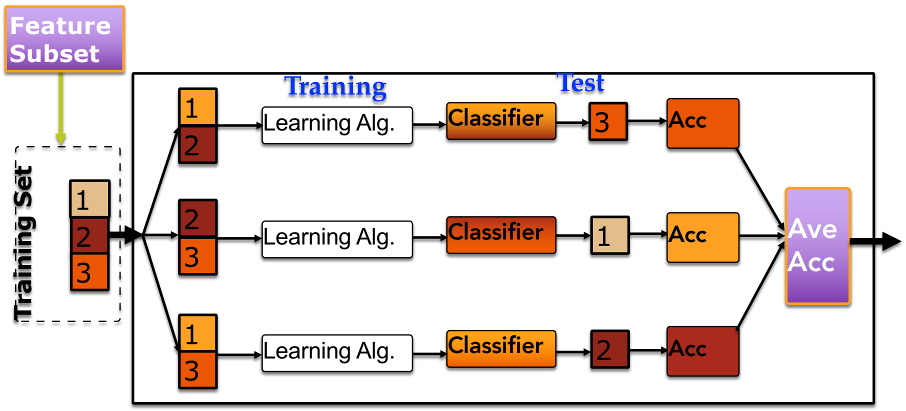
- Many works on bio-data containing feature selection
  - which leads to biased results
  - conclusion might change



Binh Tran, Bing Xue and Mengjie Zhang. "Genetic Programming for Feature Construction and Selection in Classification on High-dimensional Data", Memetic Computing, vol 8, Issue 1, pp3-15. 2016

K-CV in Each Evaluation — Inner Loop

- 3-CV as an *inner loop* to evaluate each feature subset
- In **each** evaluation to get Acc



Kohavi, Ron, and George H. John. "Wrappers for feature subset selection." Artificial intelligence 97.1-2 (1997): 273-324.

Weakness and Issues

- Large Search space:
  - Large search space: bit-string/vector with a length equal to the total number of features
  - Classification accuracy or existing filter measures in the fitness function, which often cannot lead to a smooth fitness landscape or with low locality
- Long computational time
  - A large number of evaluations
  - Wrapper: each evaluation involves a learning process of a machine learning or data mining algorithm
  - Filters are computationally cheaper than wrappers
- Poor scalability
  - the dimensionality of the search space often equals to the total number of features, thousands, or even millions
  - the number of instances is large

Weakness and Issues

- Feature selection or construction bias issue
- Generalisation issue
  - especially wrappers: selected or constructed features can easily overfit the wrapped learning algorithm and the training data, leading to poor performance on unseen test data
  - Feature construction

Future Directions

- Fitness function:
  - reduce the computational cost,
  - smooth the landscape of the search space,
  - improve the learning and generalisation performance, and
  - efficient and effective filter measure
  - hybridise wrapper and filter measures
  - Surrogate models
- Representation
  - Reduce the search space
  - Incorporate more information of about the features, e.g. relative importance of features, feature interactions or feature similarity
  - Embedded feature selection or construction

Future Directions

- Search mechanism
  - Combinatorial optimisation
  - Memetic computing
  - Large-scale optimisation
  - Adaptive parameter control techniques
- Multi-objective feature selection or feature construction
  - How to keep non-dominated solutions
  - Objective space, solution space and search space
  - Distance measure, e.g. crowding distance
  - How to maintain archive set

Future Directions

- Explainable machine learning:
  - increase the interpretability/understandability of the obtained feature set
  - Simple models via feature selection/construction
- Feature construction
  - Construct multiple features
  - both feature selection and feature construction
- Transfer learning/Multi-tasking via or for feature selection and/or construction
- Instance selection and construction
- Combining EC with *machine learning approaches*
- Feature selection and feature construction for other machine learning tasks: clustering, regression, text mining, etc.

Thank you