Evolutionary Computation for Feature Selection and Feature Construction

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

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

http://www.sigevo.org/gecco-2016/

GECCO'16 Companion, July 20–24, 2016, Denver, CO, USA.
ACM 978-1-4503-4323-7/16/07.
http://dx.doi.org/10.1145/2908961.2927002

Instructors

- **Mengjie Zhang** is a Professor of Computer Science at the School of Engineering and Computer Science, Victoria University of Wellington (VUW), New Zealand. His research is mainly focused on evolutionary computation, particularly genetic programming, particle swarm optimization and learning classifier systems with application areas of image analysis, multi-objective optimization, classification with unbalanced data, feature selection and reduction, and job shop scheduling. He has published over 400 academic papers in refereed international journals and conferences. He has been serving as an associated editor or editorial board member for five international journals (including IEEE Transactions on Evolutionary Computation and the Evolutionary Computation Journal) and as a reviewer of over fifteen international journals. He has been serving as a steering committee member and a program committee member for over eighty international conferences.

- **Bing Xue** is a Lecturer at Victoria University of Wellington. She is with the Evolutionary Computation Research Group at VUW, and her research focuses mainly on evolutionary computation, machine learning and data mining, particularly, evolutionary computation for feature selection, feature construction, dimension reduction, symbolic regression, multi-objective optimisation, bioinformatics and big data. Bing is has been organising special sessions and issues on evolutionary computation for feature selection and construction. She is also the Chair of IEEE CIS Task Force on Evolutionary Computation for Feature Selection and Construction. Bing is a committee member of Evolutionary Computation Technical Committee, and Emergent Technologies Technical Committee, IEEE CIS. She has been serving as a guest editor, associated editor or editorial board member for international journals, and program chair, special session chair, symposium/special session organiser for a number of international conferences, and as reviewer for top international journals and conferences in the field.

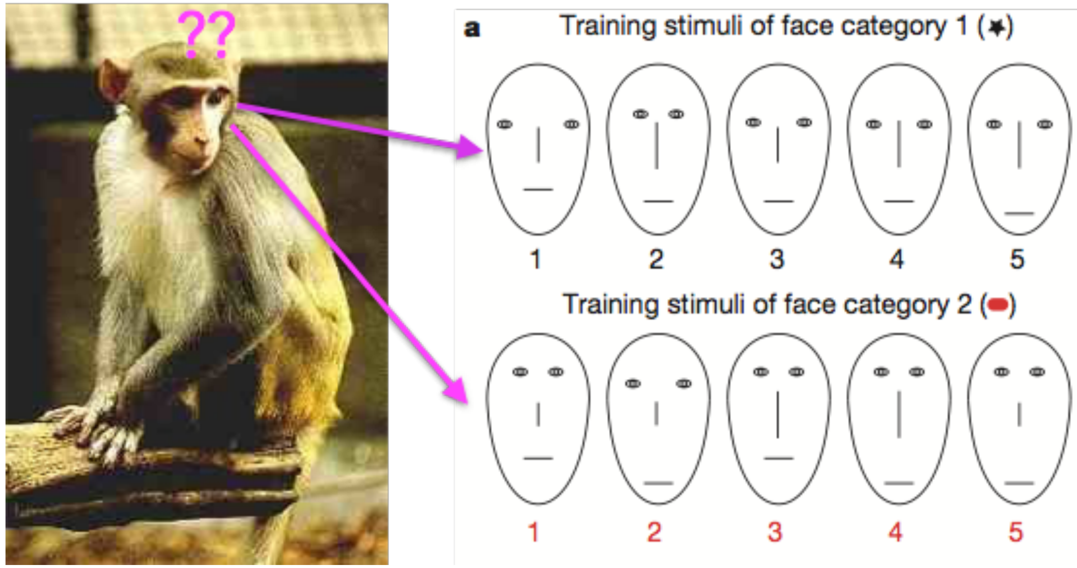
VICTORIA UNIVERSITY OF WELLINGTON
ECRG
Evolutionary Computation Research Group

Outline

- Feature Selection and Feature Construction
- Evolutionary Computation (EC) for Feature Selection
- Feature Selection Methods
- Feature Construction Methods
- Application on Images
- Application on Biology
- Issues and Challenges

Feature Selection: Example from Biology

- Monkeys performing classification task:
  - Diagnostic features:
    - Eye separation
    - Eye height
  - Non-Diagnostic features:
    - Mouth height
    - Nose length



[Acknowledgement: Nathasha Sigala, Nikos Logothetis: Visual categorization shapes feature selectivity in the primate visual cortex. *Nature* Vol. 415(2002)]

Feature Selection: Example from Biology

- Monkeys performing classification task
  - Diagnostic features:

"The data from the present study indicate that neuronal selectivity was shaped by **the most relevant subset of features** during the categorisation training."

—Nathasha Sigala, Nikos Logothetis

to one or both of the diagnostic features (and not for the non-diagnostic features)

[Acknowledgement: Nathasha Sigala, Nikos Logothetis: Visual categorization shapes feature selectivity in the primate visual cortex. *Nature* Vol. 415(2002)]

Data set (Classification) — Example 1

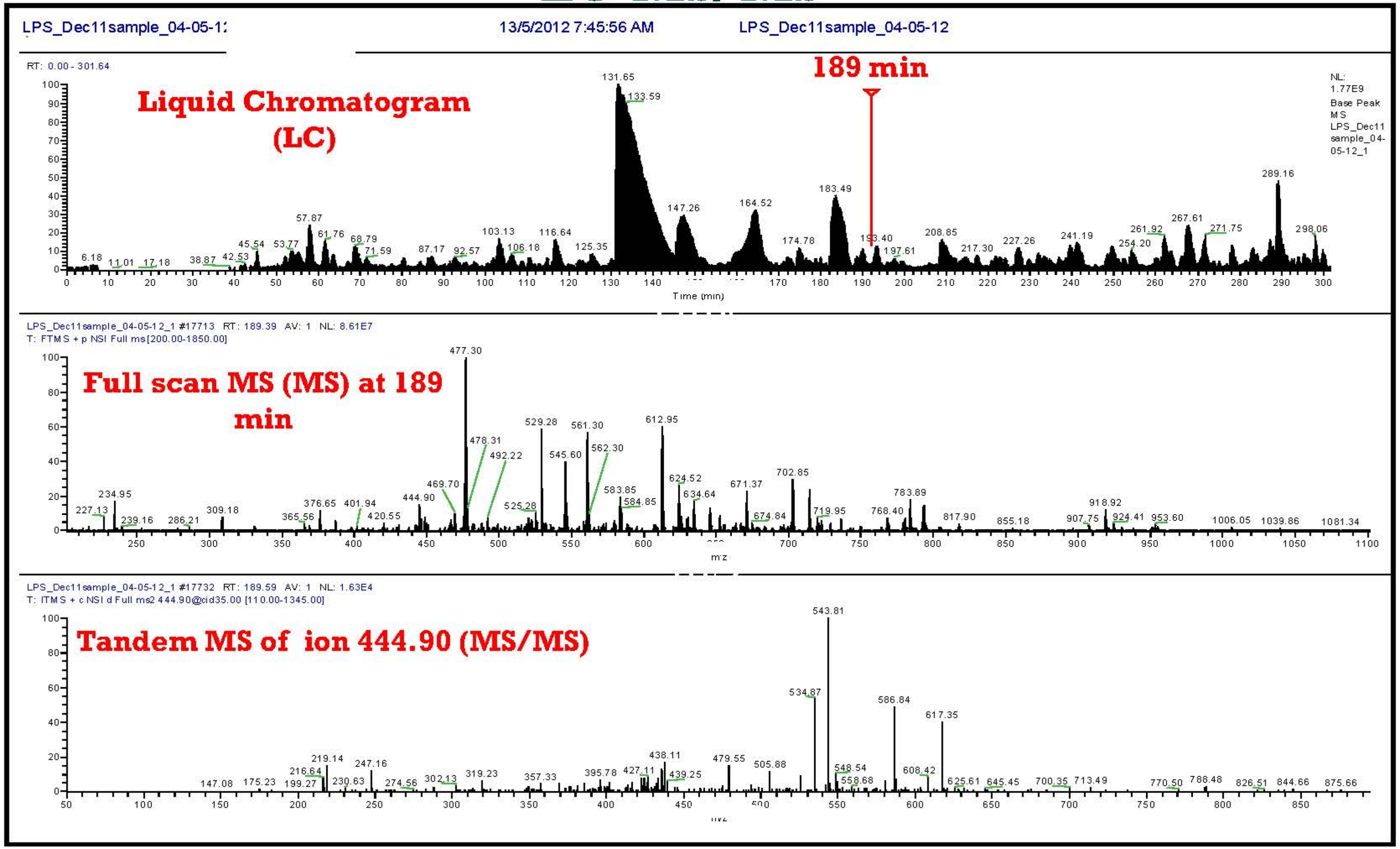
Credit card application:

- 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**

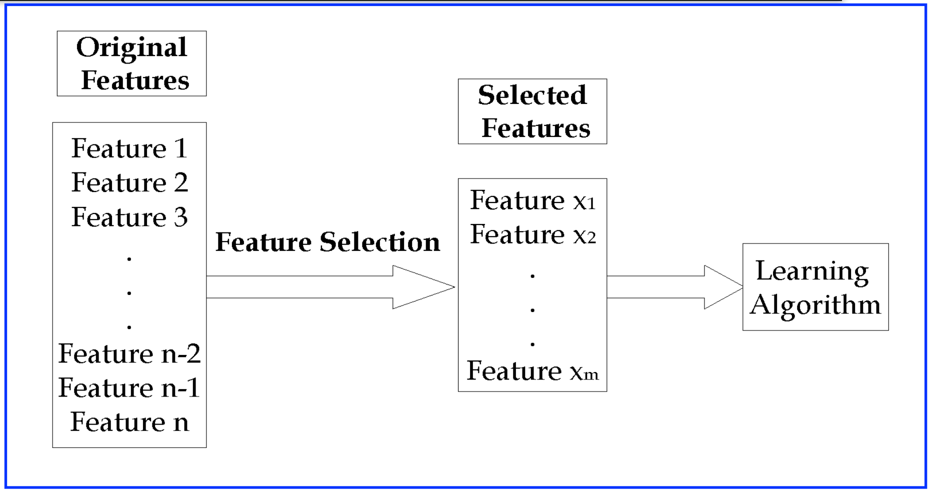
Cancer Diagnosis— Example 2

LC-MS/MS

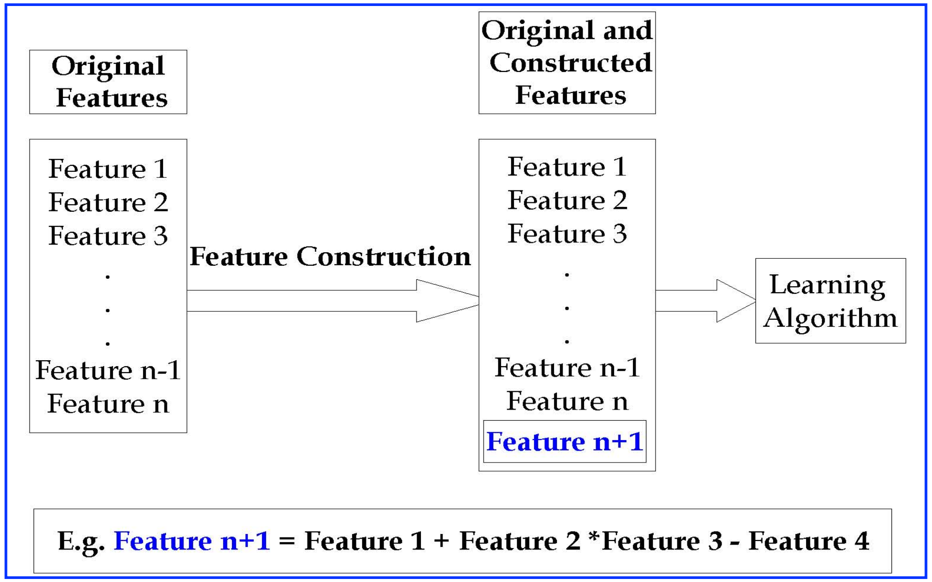


Feature Selection and Feature Construction

- 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.



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?

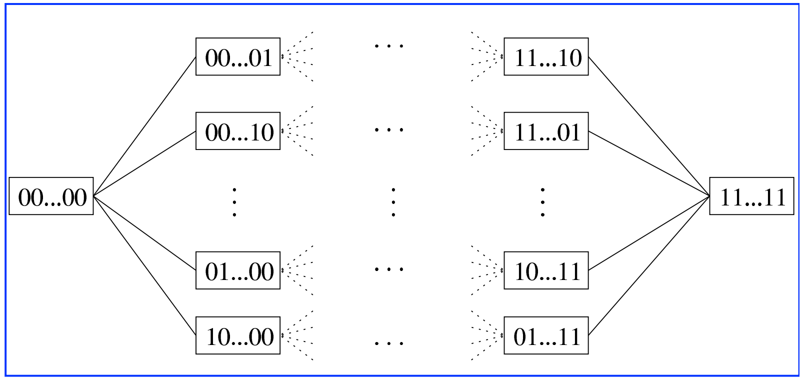
- The quality of input features can drastically affect the learning performance.
- Even if the quality of the original features is good, transformations might be required to make them usable for certain types of classifiers.
- Feature construction does not add to the cost of extracting (measuring) original 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 and interpretation
- Reduce the cost, e.g. save memory
- and ?

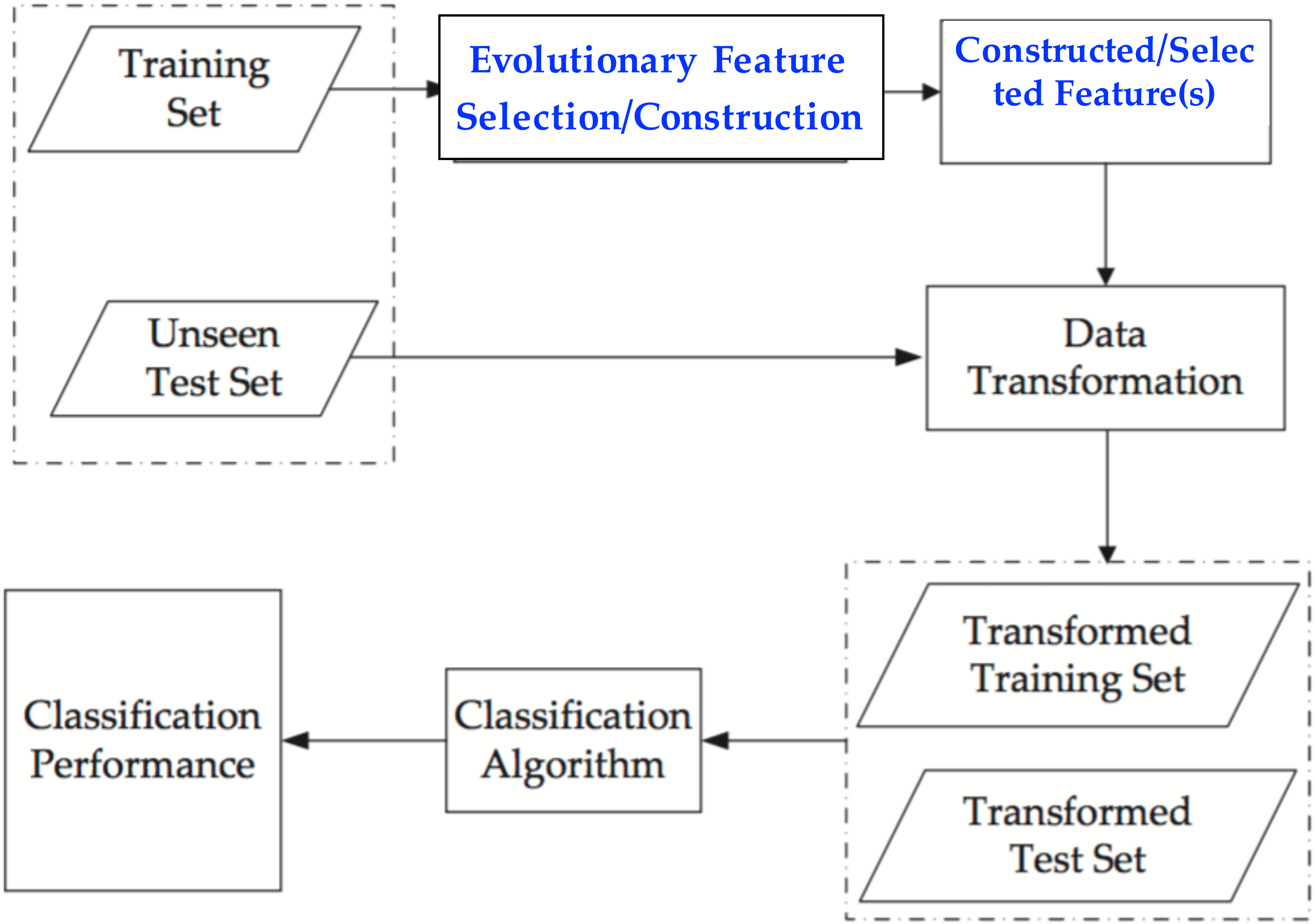
Challenges in FS and FC

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



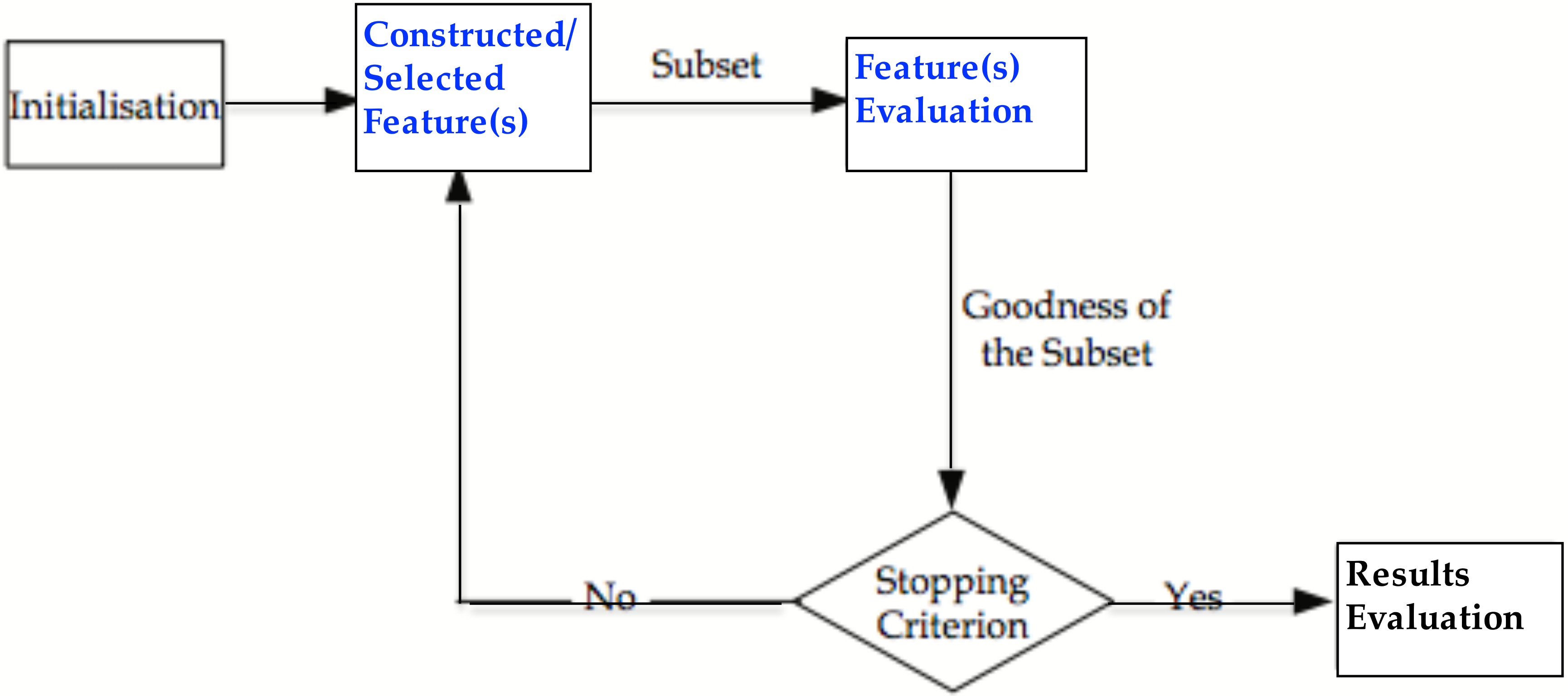
- Feature interaction
  - Relevant features may become redundant
  - Weakly relevant or irrelevant features may become highly useful
- Slow processing time, or even not possible
- **Multi-objective Problems**

General FS/FC System



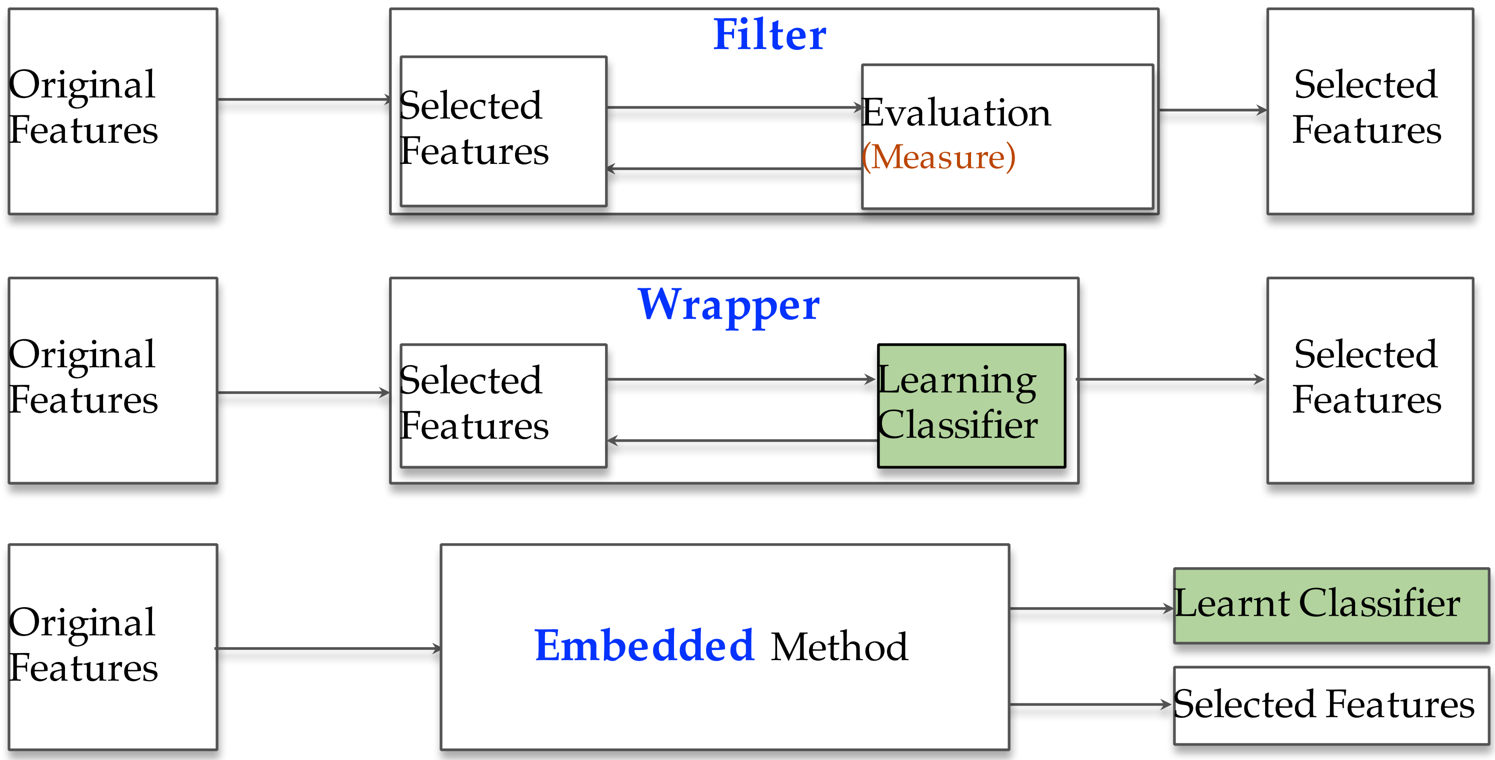
Feature FS/FC Process

- On training set:



Feature Selection Approaches

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



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
    - Feature ranking method
  - The FOCUS algorithm
  - Sequential forward/backward selection
  - Sequential forward/backward floating selection
  - Statistical feature selection methods
- Evolutionary Computation (EC) based approaches

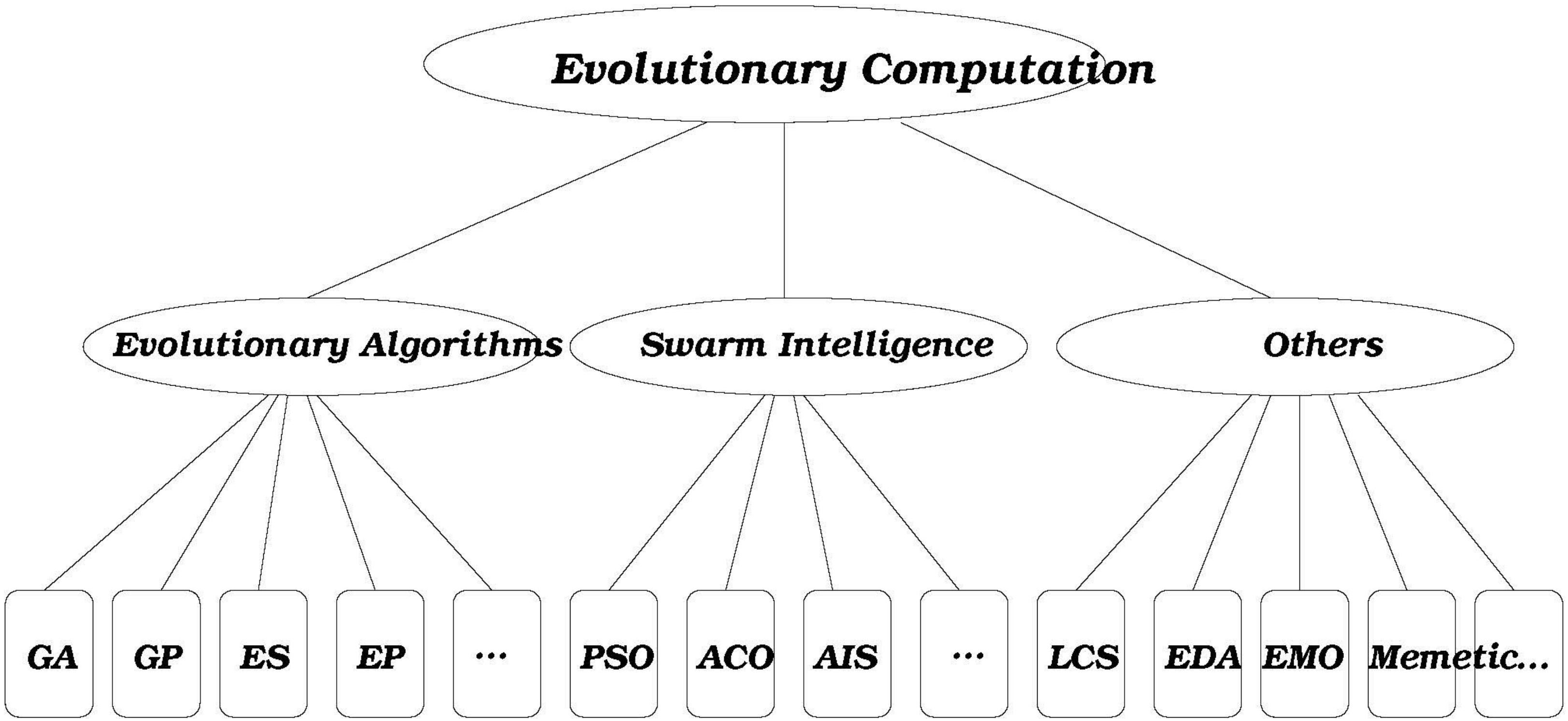
Tang, Jiliang, Salem Alelyani, and Huan Liu. "Feature selection for classification: A review." *Data Classification: Algorithms and Applications* (2014): 37.

Bolón-Canedo, Verónica, Noelia Sánchez-Maroño, and Amparo Alonso-Betanzos. "A review of feature selection methods on synthetic data." Knowledge and information systems 34.3 (2013): 483-519.

Zhai, Yiteng, Yew-Soon Ong, and Ivor W. Tsang. "The Emerging" Big Dimensionality"." IEEE Computational Intelligence Magazine 9.3 (2014): 14-26.

Evolutionary Computation (EC)

- A group of techniques inspired by the principles of biological evolution

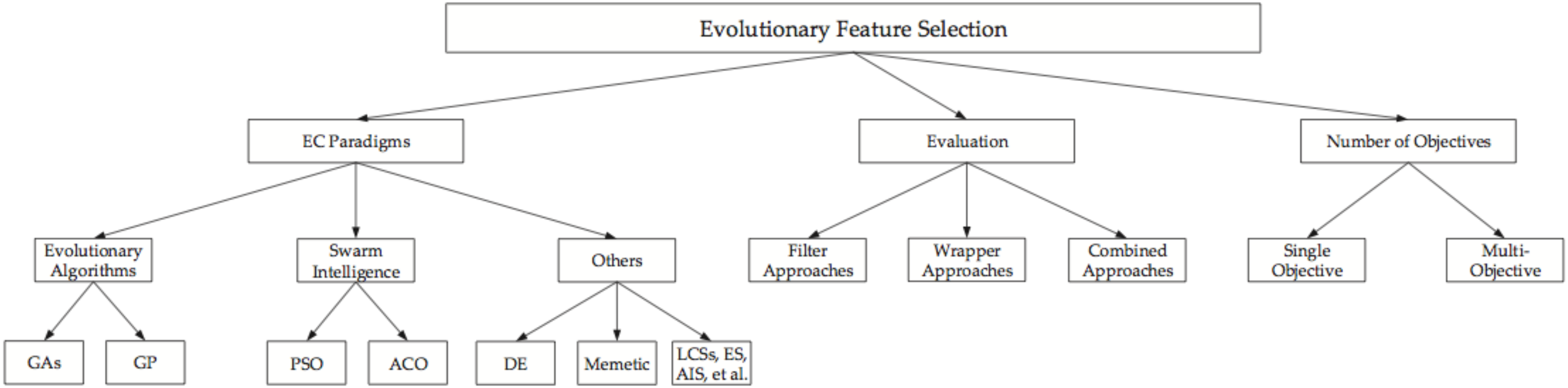


Why Evolutionary Computation ?

- 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

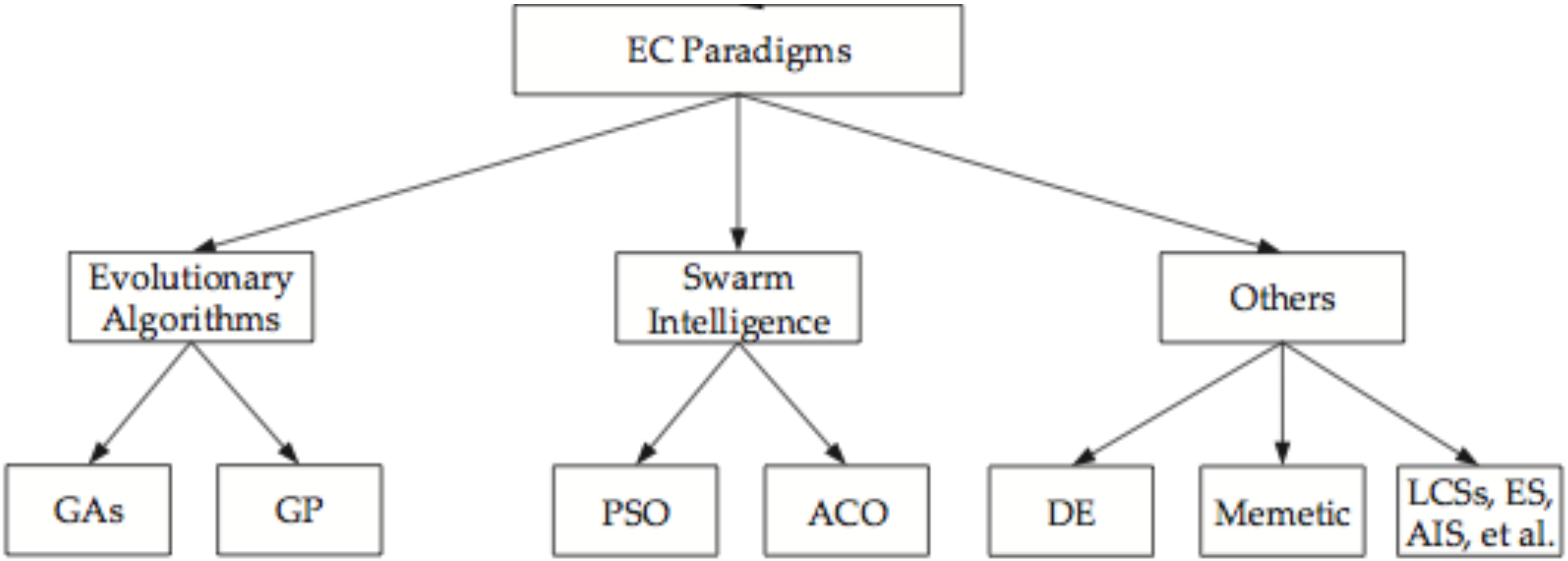
- EC Paradigms
- Evaluation
- Number of Objectives



Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

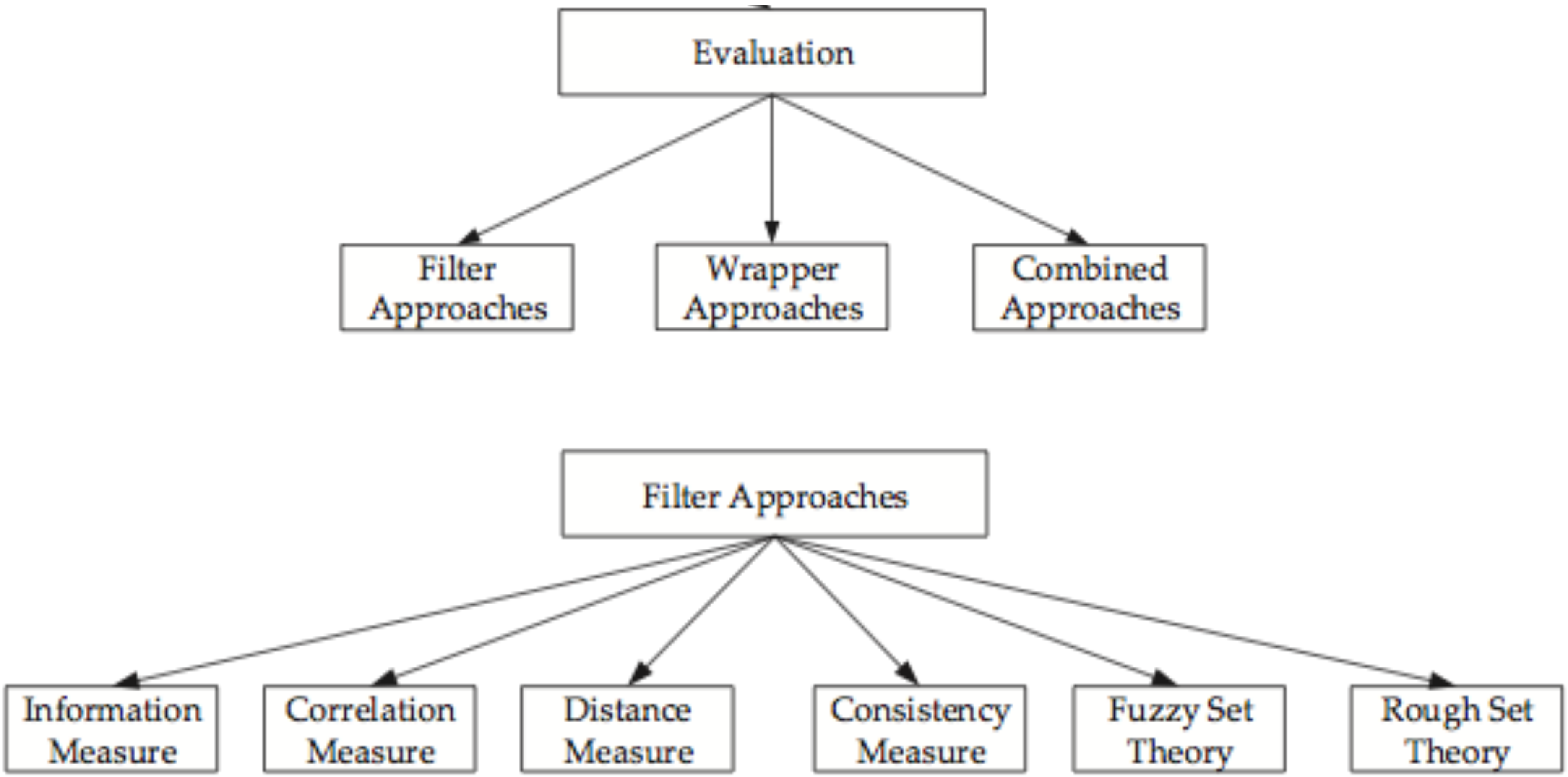
EC for Feature Selection

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



Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

EC for Feature Selection



Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

GAs for Feature Selection

- Over 25 years ago, first EC techniques
  - Filter, Wrapper, Single Objective, Multi-objective
- Representation
  - Binary string
- Search mechanisms
  - Genetic operators
- Multi-objective feature selection
- Scalability issue

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

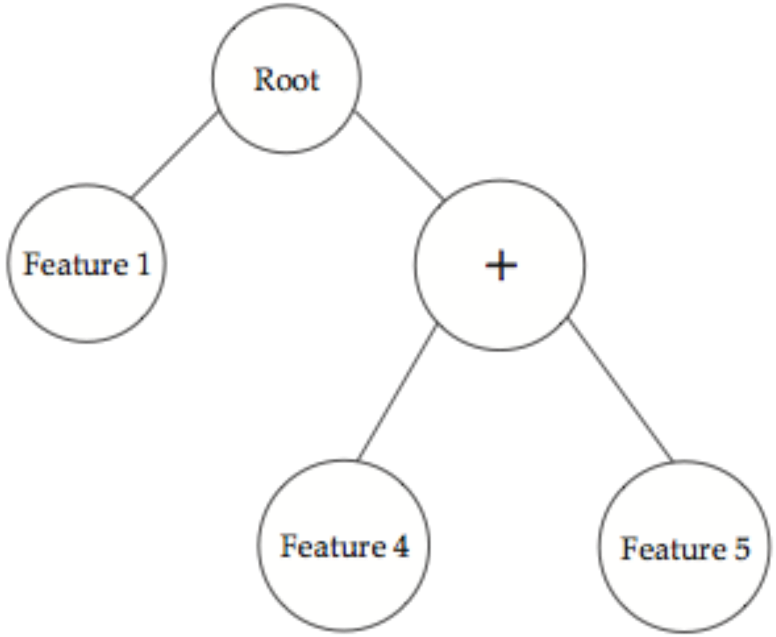
R. Leardi, R. Boggia, and M. Terrile, “Genetic algorithms as a strategy for feature selection,” Journal of Chemometrics, vol. 6, no. 5, pp. 267– 281, 1992.
Z. Zhu, Y.-S. Ong, and M. Dash, “Markov blanket-embedded genetic algorithm for gene selection,” Pattern Recognition, vol. 40, no. 11,pp. 3236–3248, 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.
Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

GP for Feature Selection

- Implicit feature selection
  - Filter, Wrapper, Single Objective, Multi-objective

- Embedded feature selection
- Feature construction

- <u>Computationally expensive</u>



L. Jung-Yi, K. Hao-Ren, C. Been-Chian, and Y. Wei-Pang, "Classifier design with feature selection and feature extraction using layered genetic programming," Expert Systems with Applications, vol. 34, no. 2, pp. 1384–1393, 2008.
Purohit, N. Chaudhari, and A. Tiwari, "Construction of classi- fier with feature selection based on genetic programming," in IEEE Congress on Evolutionary Computation (CEC), pp. 1–5, 2010.
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.
Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

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
- Scalability

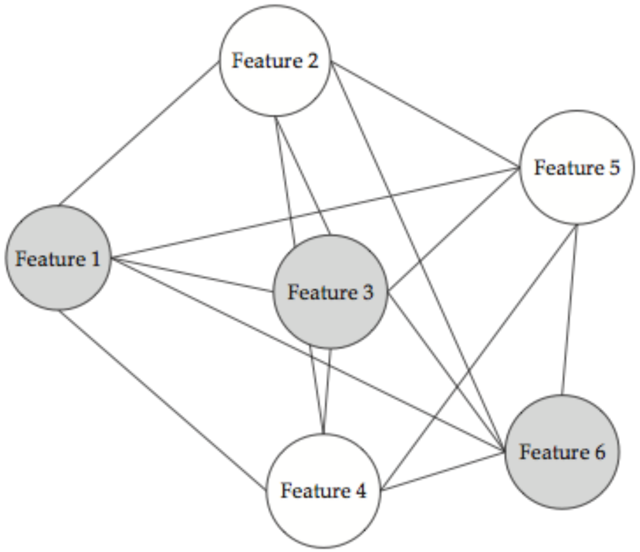
0.7	0.12	0.84	0.69	0.25	0.06	0.92	0.45	0.36	0.80	0.67	0.30	0.41

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.
L. Y. Chuang, H. W. Chang, C. J. Tu, and C. H. Yang, "Improved binary PSO for feature selection using gene expression data," Computational Biology and Chemistry, vol. 32, no. 29, pp. 29– 38, 2008.
C. L. Huang and J. F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," Application on Soft Computing, vol. 8, pp. 1381–1391, 2008.
B. Xue, M. Zhang, and W. N. Browne, "Multi-objective particle swarm optimisation (PSO) for feature selection," in Proceeding of the 14th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO), pp. 81–88, ACM, 2012.
Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

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.

S. Vieira, J. Sousa, and T. Runkler, “Multi-criteria ant feature selection using fuzzy classifiers,” in Swarm Intelligence for Multi-objective Problems in Data Mining, vol. 242 of Studies in Computational Intelligence, pp. 19–36, Heidelberg, 2009.

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.

R. Jensen, “Performing feature selection with aco,” in Swarm Intelli- gence in Data Mining, vol. 34 of Studies in Computational Intelligence, pp. 45–73, / Heidelberg, 2006.

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.

Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

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

A. Al-Ani, A. Alsukker, and R. N. Khushaba, "Feature subset selection using differential evolution and a wheel based search strategy," Swarm and Evolutionary Computation, vol. 9, pp. 15–26, 2013.
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.
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.
S. Palanisamy and S. Kanmani, "Artificial bee colony approach for optimizing feature selection," International Journal of Computer Science Issues (IJCSI), vol. 9, no. 3, pp. 432–438, 2012.
Z. Zhu, S. Jia, and Z. Ji, "Towards a memetic feature selection paradigm [application notes]," IEEE Computational Intelligence Mag- azine, vol. 5, no. 2, pp. 41–53, 2010.
Y. Wen and H. Xu, "A cooperative coevolution-based pittsburgh learn- ing classifier system embedded with memetic feature selection," in IEEE Congress on Evolutionary Computation, pp. 2415–2422, 2011.
Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, online on 30 Nov 2015

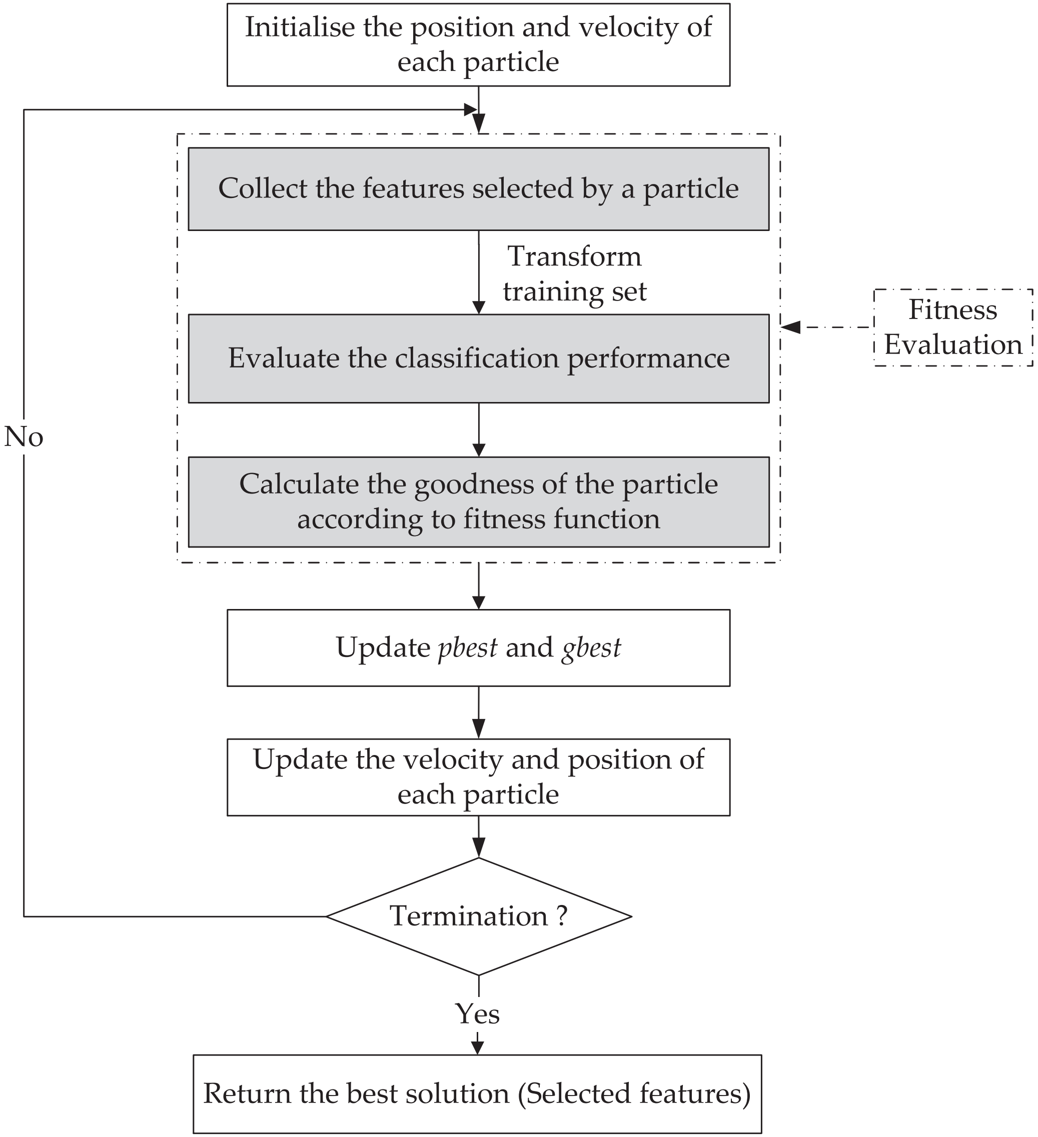
Related Areas (Applications)

- 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.

Bing Xue, Mengjie Zhang, Will Browne, Xin Yao. "A Survey on Evolutionary Computation Approaches to Feature Selection", IEEE Transaction on Evolutionary Computation, doi: 10.1109/TEVC.2015.2504420, published online on 30 Nov 2015

Feature Selection

PSO for FS: initialisation and updating

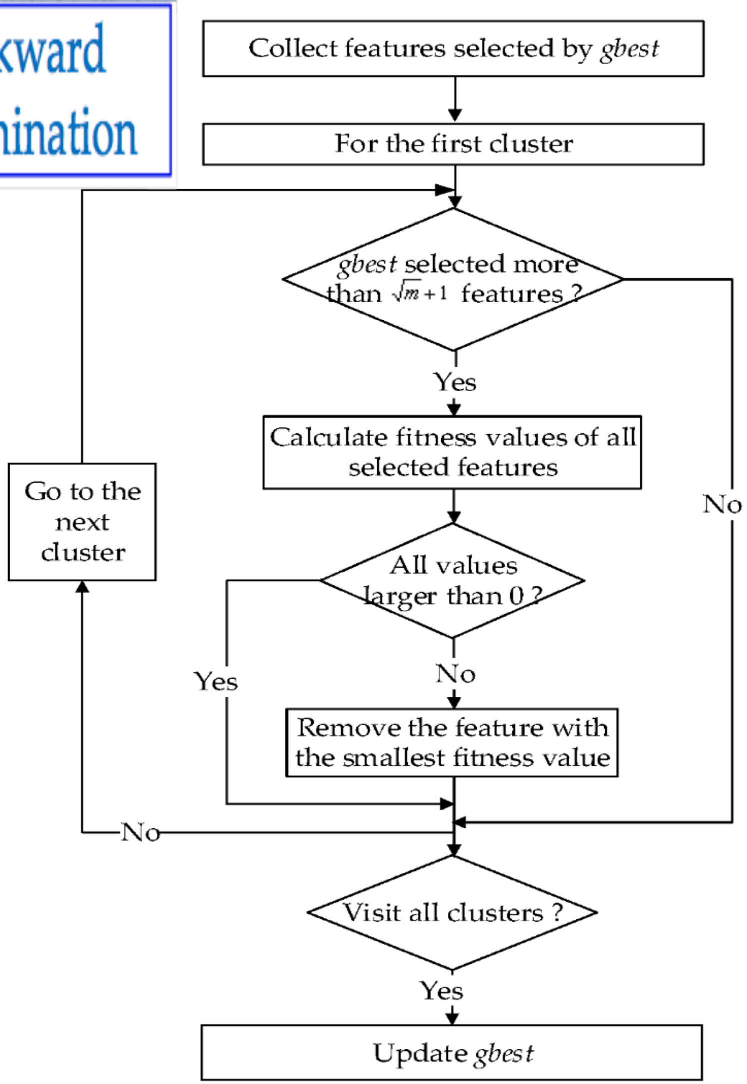


- Initialisation:
  - Forward selection
  - Backward selection
  - Mixture of both
- Updating:
  - Consider the number of features in the pest and gbest updating

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

PSO FS: with backward elimination

Backward Elimination



- **Filter Measure**

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

- 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

$$Rel(s_i) = I(s_i; c)$$

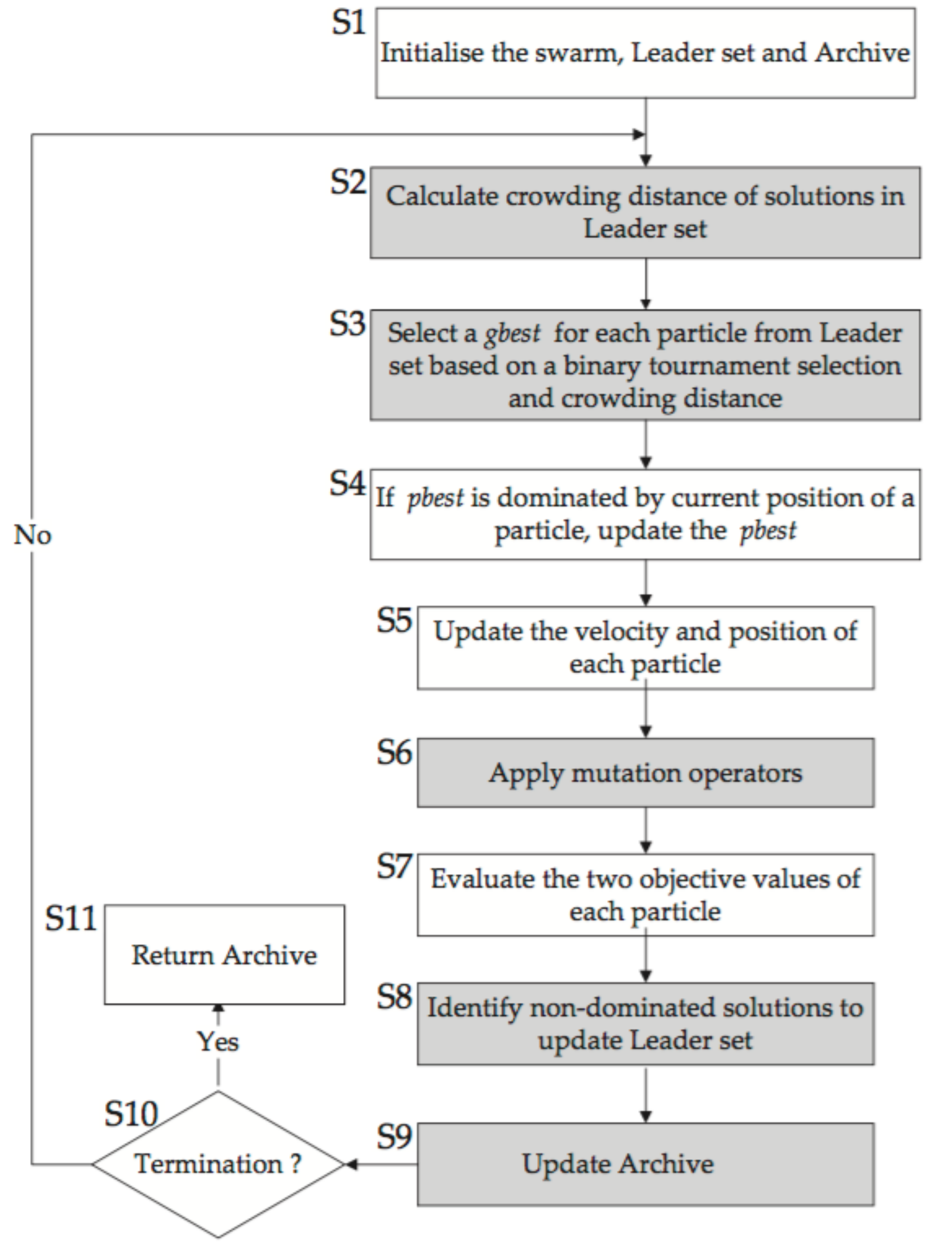
$$Red(s_i) = \frac{1}{|S|-1} \sum_{s_i, s_j \epsilon S, s_i, \neq s_j} I(s_i; s_j)$$

Rel(s_i) **:** relevance contribution of s_i in S
Red(s_i): redundancy in S caused by s_i

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.

Multi-objective PSO for FS

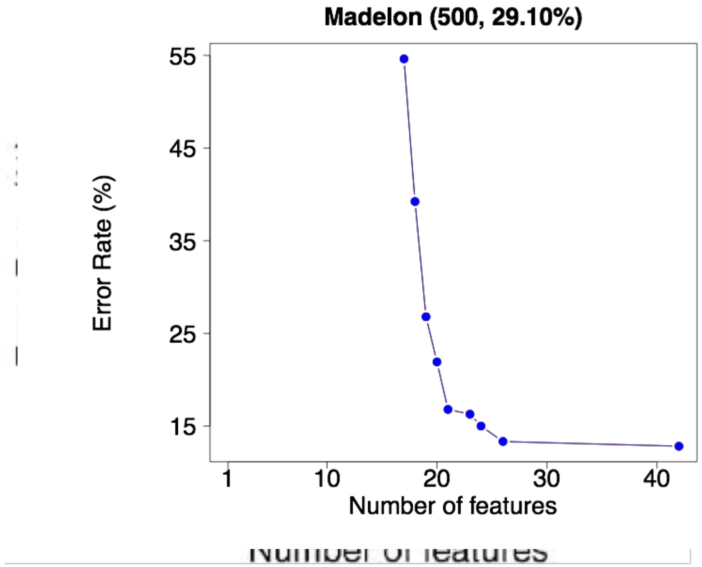
- Introduce and develop the first multi-objective PSO approach to feature selection
  - Simultaneously minimise the number of features and the error rate
  - ~121 citations since June 2013



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



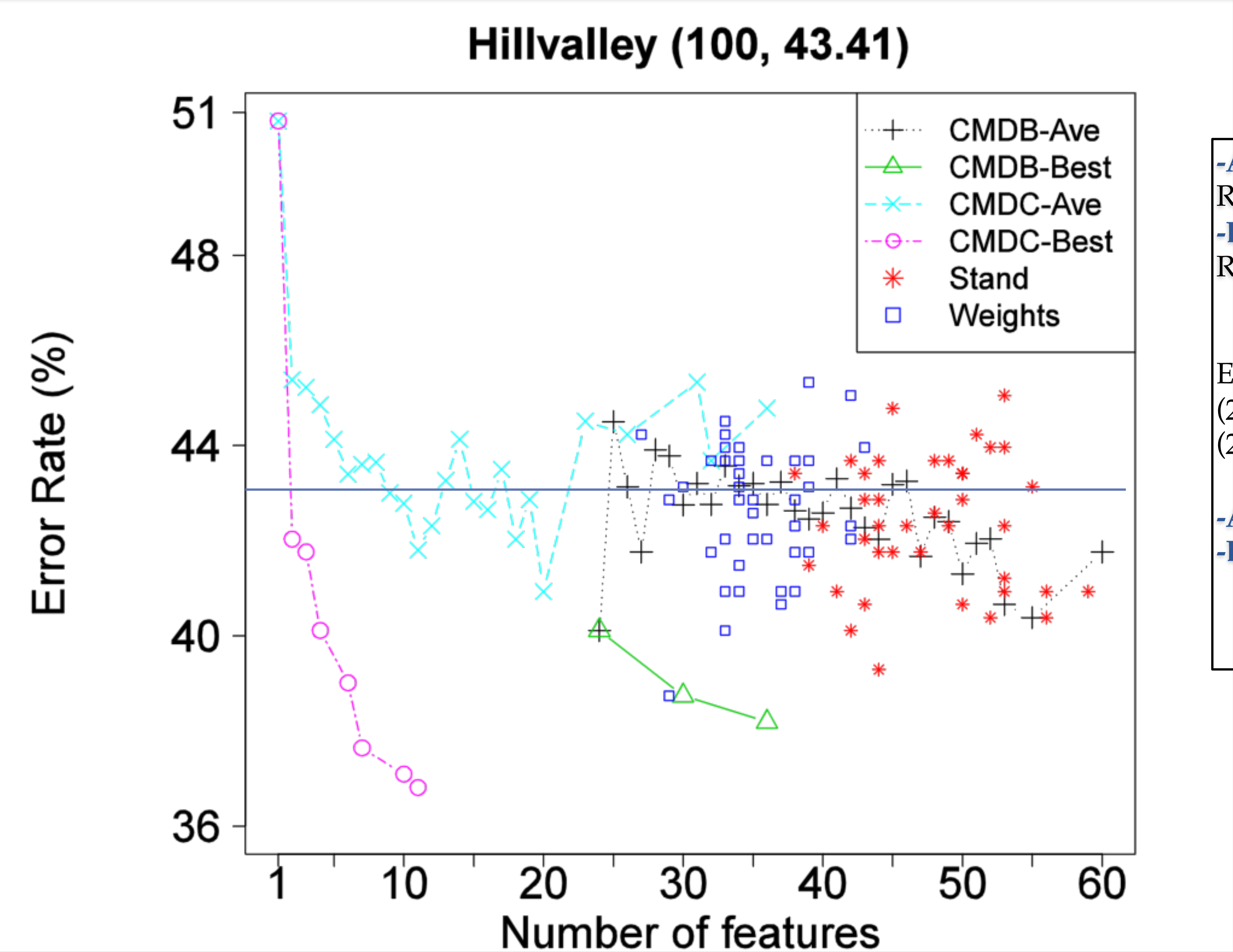
T-TEST ON HYPERVOLUME RATIOS ON TRAINING ACCURACY

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.

Multi-objective PSO for FS : Binary VS continuous



-Ave: Average Results
-Best: Best Results

Example:
(20, 40), (20,42)
(20, 41), (20,43)

-Ave: (20, 41.5)
-Best: (20, 40)

Bing Xue, Mengjie Zhang, Will N. Browne. "Multi-Objective Particle Swarm Optimisation (PSO) for Feature Selection". Proceedings of 2012 Genetic and Evolutionary Computation Conference (GECCO 2012). ACM Press. Philadelphia, USA. 7-11 July 2012. pp. 81-88

Probability based BPSO (PBPSO)

- Updating equations:

$$x_{id}(t+1) = \begin{cases} 1 - x_{id}(t), & if\ rand() < p_{id} \\ x_{id}(t), & otherwise \end{cases}$$

$$p_{id} = p_0 + p_{pd} + p_{gd}$$

$$p_{pd} = \begin{cases} p_1, & if\ \ x_{id}(t) \neq y_{id}(t) \\ 0, & otherwise \end{cases}$$

$$p_{gd} = \begin{cases} p_2, & if\ \ x_{id}(t) \neq \hat{y}_{id}(t) \\ 0, & otherwise \end{cases}$$

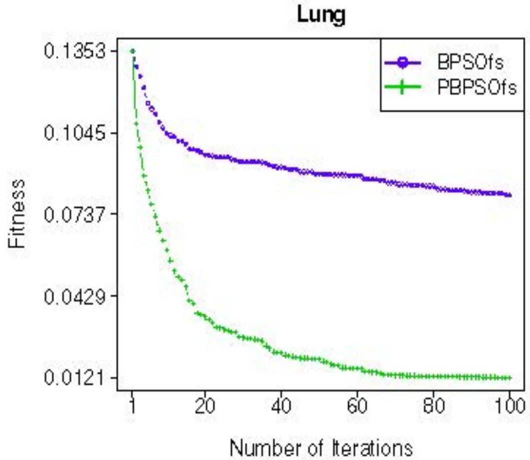
where $y_{id}(t)$ represents ***pbest***, and $\hat{y}_{id}$ represents ***gbest***

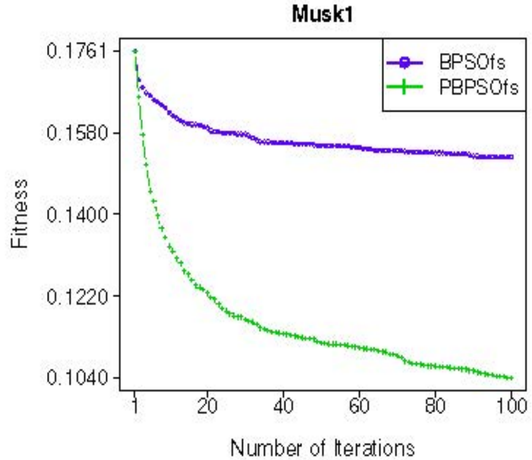
$$p_0 + p_1 + p_2 = 1$$

Bing Xue, Su Nguyen, Mengjie Zhang. "A New Binary Particle Swarm Optimisation Algorithm for Feature Selection". Proceedings of the 17th European Conference on Applications of Evolutionary Computation (EvoApplications 2014). Lecture Notes in Computer Science. Vol. 8602. Granada, Spain 23rd - 25th April 2014. pp. 501-513

Probability based BPSO (PBPSO)

Dataset	Method	AveSize	BestAcc	AveAcc ± StdAcc	Test 1	Test 2	Time
Hillvalley	All	100	56.59				
	BPSOfs	39.35	60.16	56.88 ± 1.6322	=		50.28
	PBPSOfs	31.08	61.26	58.25 ± 1.6952	+	+	46.1
Musk1	All	166	83.92				
	BPSOfs	75.52	90.91	84.21 ± 2.8401	=		13.96
	PBPSOfs	69.3	88.81	85.38 ± 1.8087	+	+	12.7
Arrhythmia	All	279	94.46				
	BPSOfs	99.7	95.14	94.21 ± 0.3937	-		16.97
	PBPSOfs	63.42	95.48	94.71 ± 0.3405	+	+	12.44
Madelon	All	500	70.9				
	BPSOfs	243.85	78.59	75.81 ± 1.4905	+		991.6
	PBPSOfs	212.42	81.15	78.91 ± 1.2565	+	+	953.63
Isolet5	All	617	98.45				
	BPSOfs	225.15	98.59	98.25 ± 0.1354	-		384.16
	PBPSOfs	169.35	98.87	98.61 ± 0.1248	+	+	328.34
Multiple Features	All	649	98.63				
	BPSOfs	237.05	99.1	98.89 ± 0.0923	+		692.54
	PBPSOfs	176.15	99.27	99.01 ± 0.1043	+	+	551.44





Bing Xue, Su Nguyen, Mengjie Zhang. "A New Binary Particle Swarm Optimisation Algorithm for Feature Selection". Proceedings of the 17th European Conference on Applications of Evolutionary Computation (EvoApplications 2014). Lecture Notes in Computer Science. Vol. 8602. Granada, Spain 23rd - 25th April 2014. pp. 501-513

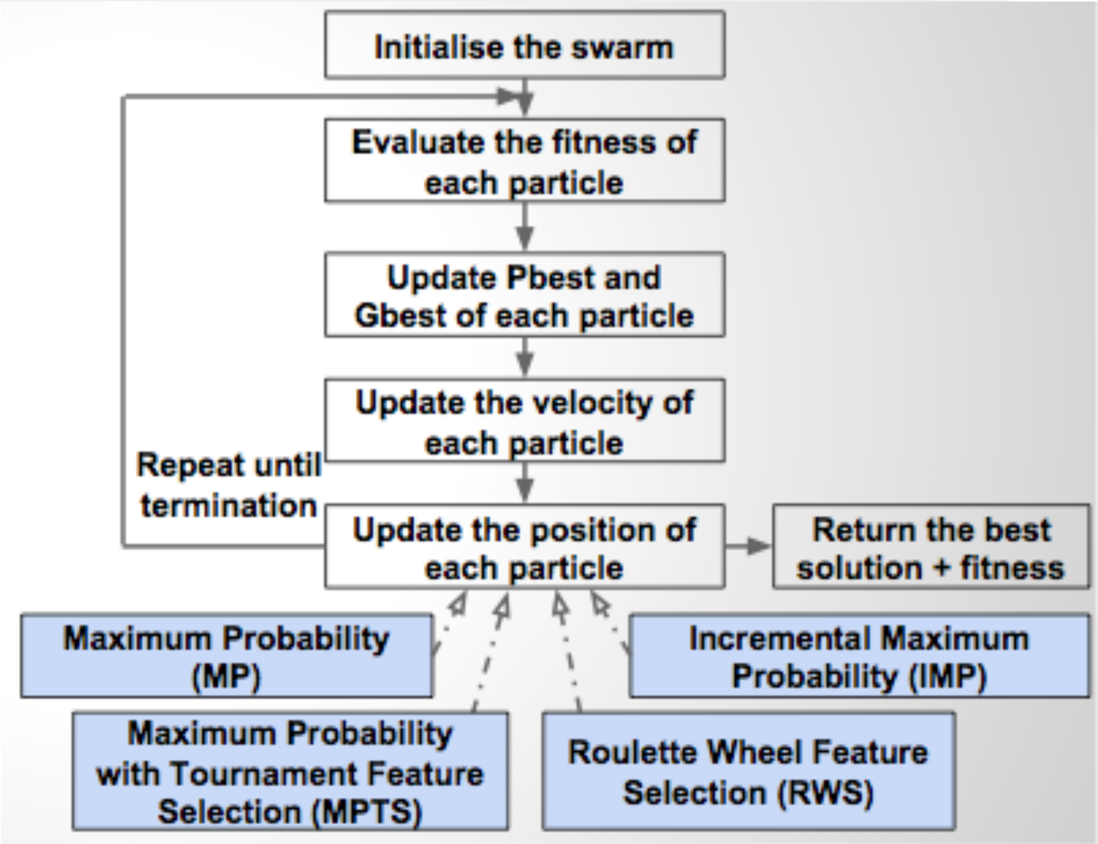
EC and Statistical Grouping for FS

- Introduce statistical feature clustering to feature selection and develop the first approach
  - reduce the size of the search space
  - #features: from 600 to ~12
  - implicitly consider feature interaction
  - Example:
    - our method achieved accuracy 100%: {10, 7, 3}
    - Single feature ranking: 7, 10, 12, 1, 9, 11, 6, 2, 13, 5, 4, 3

Bing Xue, Micthell C. Lane, Ivy Liu, Mengjie Zhang, "Particle Swarm Optimisation for Feature Selection Based on Statistical Clustering", Evolutionary Computation (Journal, MIT Press), *Passed first round review with positive comments*

EC and Statistical Grouping for FS

- Development of four new particle position update algorithms that automatically select a single feature from each feature cluster
- As features are grouped by similarity, a single feature is expected to provide ***enough*** information about its feature cluster



Mitchell C. Lane, Bing Xue, Ivy Liu, Mengjie Zhang. "Gaussian Based Particle Swarm Optimisation and Statistical Clustering for Feature Selection". Proceedings of the 14th European Conference on Evolutionary Computation in Combinatorial Optimisation (EvoCOP 2014). Lecture Notes in Computer Science. Volume 8600, Granada, Spain 23rd - 25th April 2014. pp. 133--144

Mitchell C. Lane, Bing Xue, Ivy Liu and Mengjie Zhang. "Particle Swarm Optimisation and Statistical Clustering for Feature Selection". Proceedings of the 26th Australasian Joint Conference on Artificial Intelligence (AI2013) Lecture Notes in Computer Science. Vol. 8272. Springer. Dunedin, New Zealand, December 2013. pp. 214-220

Information Theory Feature Selection

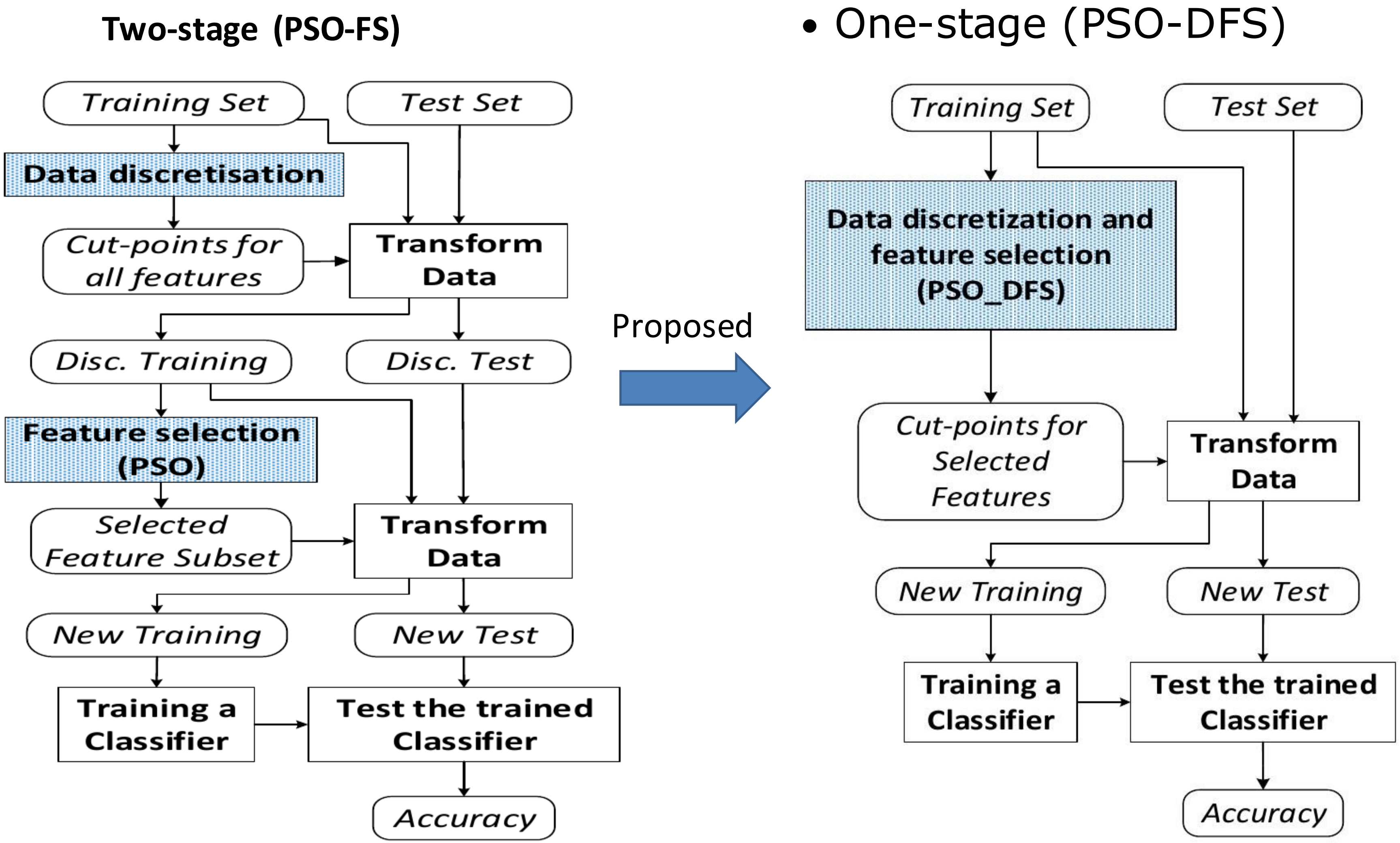
- Information theory in evolutionary feature selection
  - Fast algorithm — mutual information
  - New measures, evaluate multiple features
  - Evolutionary multi-objective filter feature selection

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.

Bare-Bone Particle Swarm Optimisation



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

Feature Selection Though Data Discretisation

	F_1	F_2	F_3	F_4	F_5	…	F_D
Min	5.5	-9.2	7.6	-8.5	6.9	…	-9.3
Max	10.2	20.5	12.7	4.5	50.1	…	-1.6

Particle's Position	7.2	10.4	12.7	-3.8	6.9	…	-4.5

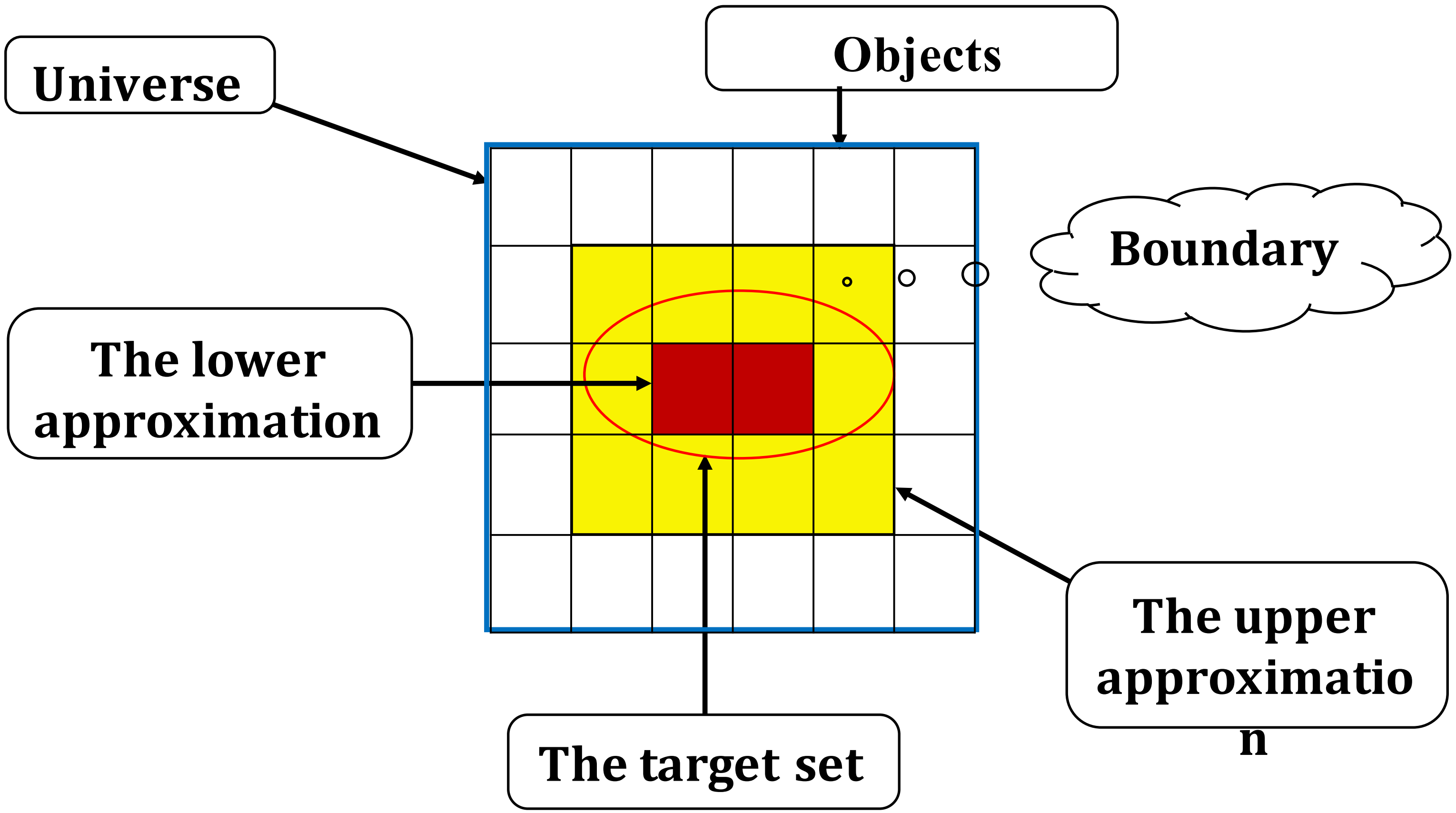
Not selected

Candidate solution

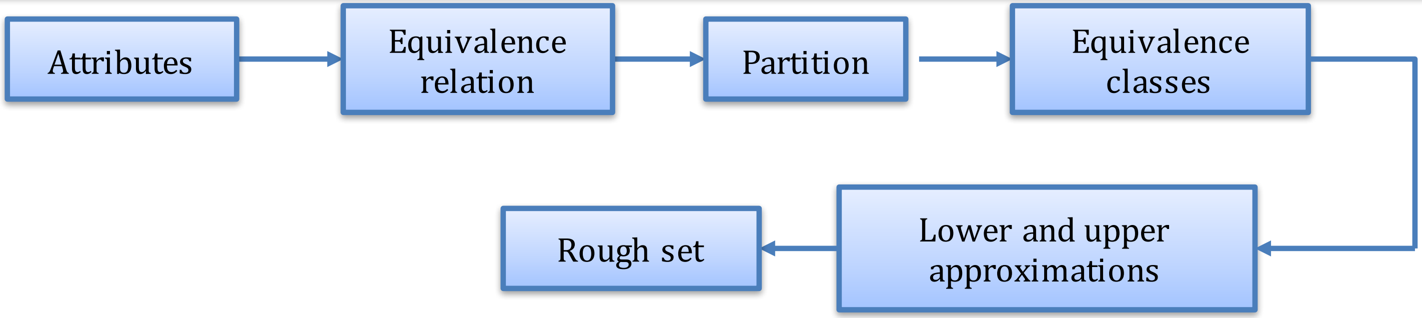
Binh Tran Ngan, Mengjie Zhang, Bing Xue. "Bare-Bone Particle Swarm Optimisation for Simultaneously Discretising and Selecting Features For High-Dimensional Classification". Proceedings of the 19th European Conference on the Applications of Evolutionary Computation (EvoApplications 2016, EvoIASP 2016). Lecture Notes in Computer Science. Vol. 9597. Porto, Portugal, March 30 - April 1, 2016. pp. 701-718

Filter FS based on Rough Set

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



FS based on Rough Set



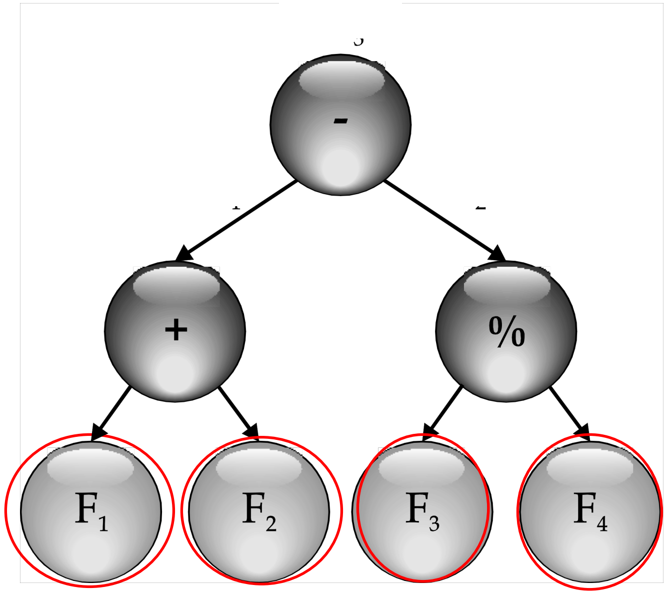
$$Fitness = \frac{\sum_{i=1}^{n} |\underline{apr}P\ U_i|}{|U|} + \frac{\sum_{x\in\{EqC\}} \frac{|x|}{|U|}}{|EqC|}$$

- 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

GP for Embedded Feature Selection

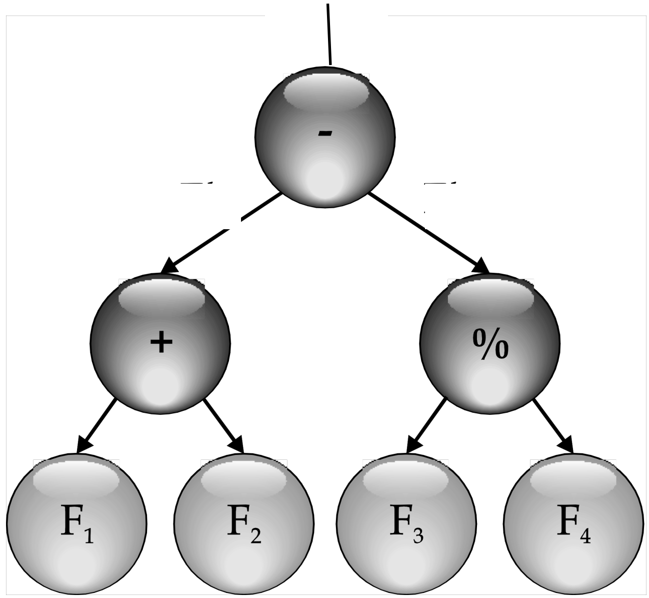
Feature Selection



Classification

< 0
Class +

>=
Class -



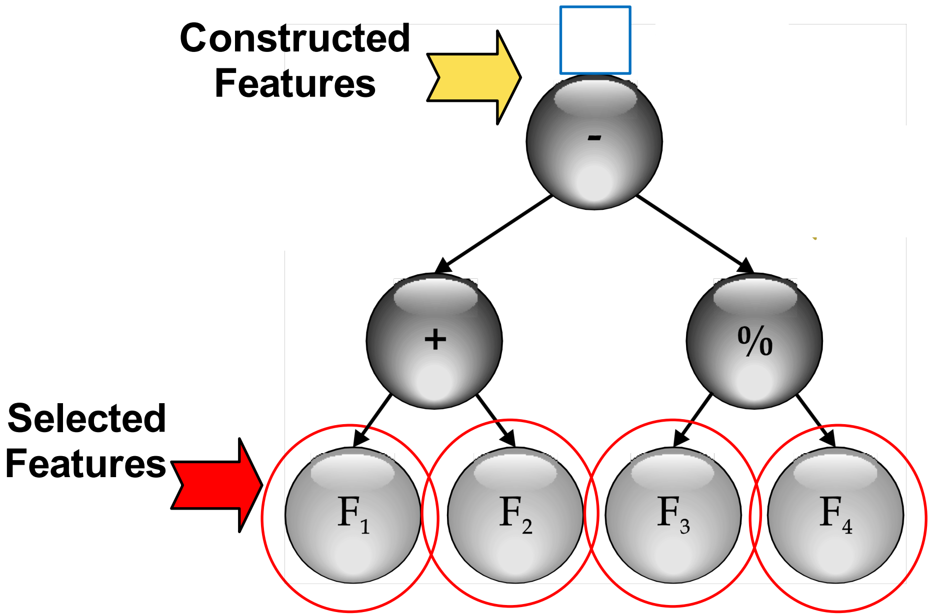
Soha Ahmed, Mengjie Zhang, Lifeng Peng and Bing Xue. "Genetic Programming for Measuring Peptide Detectability". Proceedings of the 10th International Conference on Simulated Evolution and Learning (SEAL 2014). Lecture Notes in Computer Science. Vol. 8886. Dunedin, New Zealand. December 15-18, 2014. pp.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

Feature Construction

Why Use GP for Feature Construction?

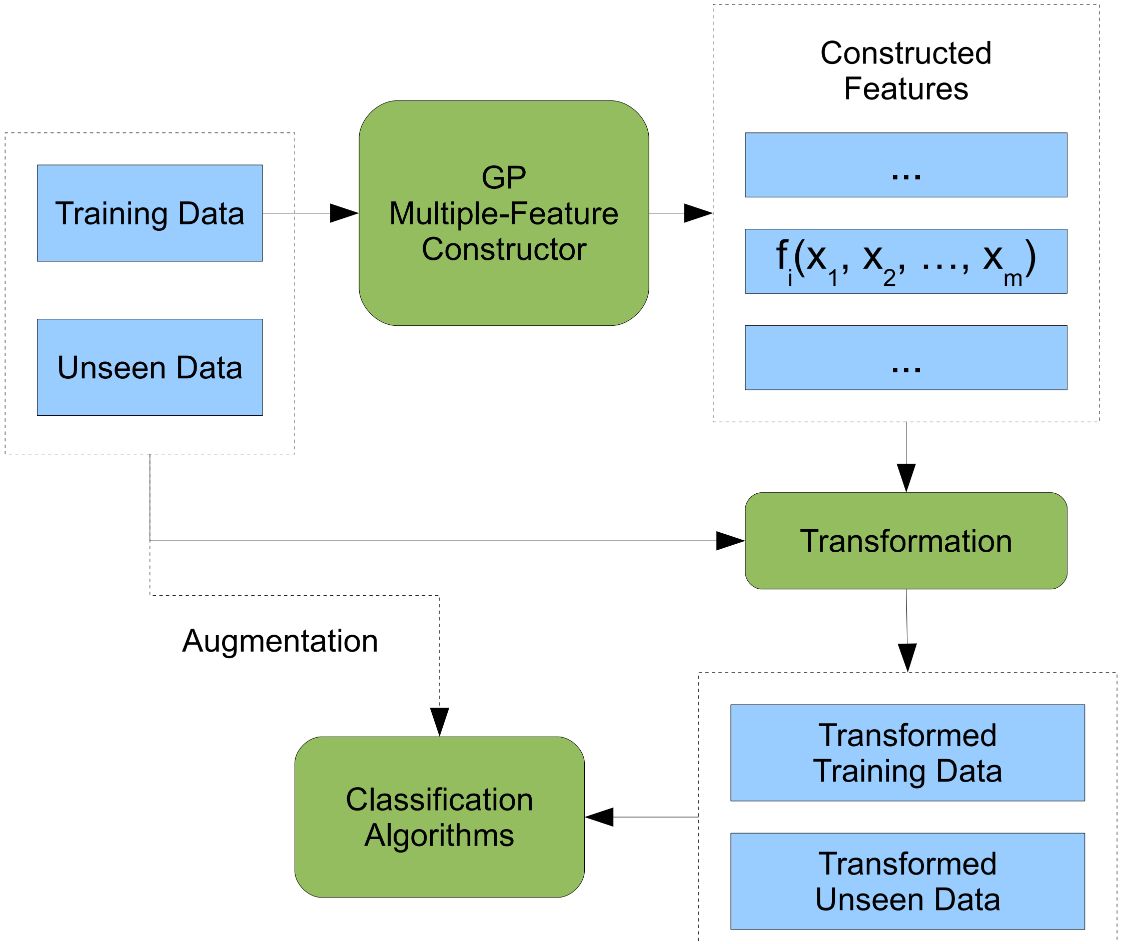
- GP is flexible in making mathematical and logical functions

- 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 FC: A System Diagram

- One constructed feature for one class



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

GP for FC Measure: Entropy of Class Intervals

Defining a measure of goodness for a single feature:

- The interval of *a class* along a feature is determined by the dispersion of the instances of that class along the feature axis. The dispersion of instances themselves is related to the distribution of data points in that class.
- An interval ***I*** is represented with a pair (*lower, upper*) which shows the lower and upper boundaries of the interval. ***Ic*** is used to indicate an interval for class ***c***.
- The interval of class ***c*** could be formulated as follows if the class distributions were normal.

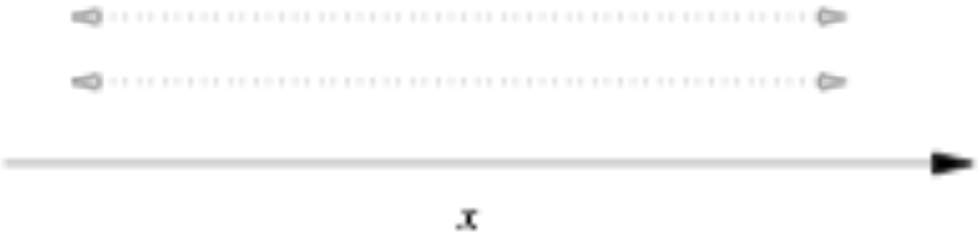
$$I_c = [\mu_c - 3\sigma_c, \mu_c + 3\sigma_c]$$

- However, the normality assumption is not always satisfied.

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

GP for FC Measure: Examples of good and bad class intervals

- Overlapping intervals

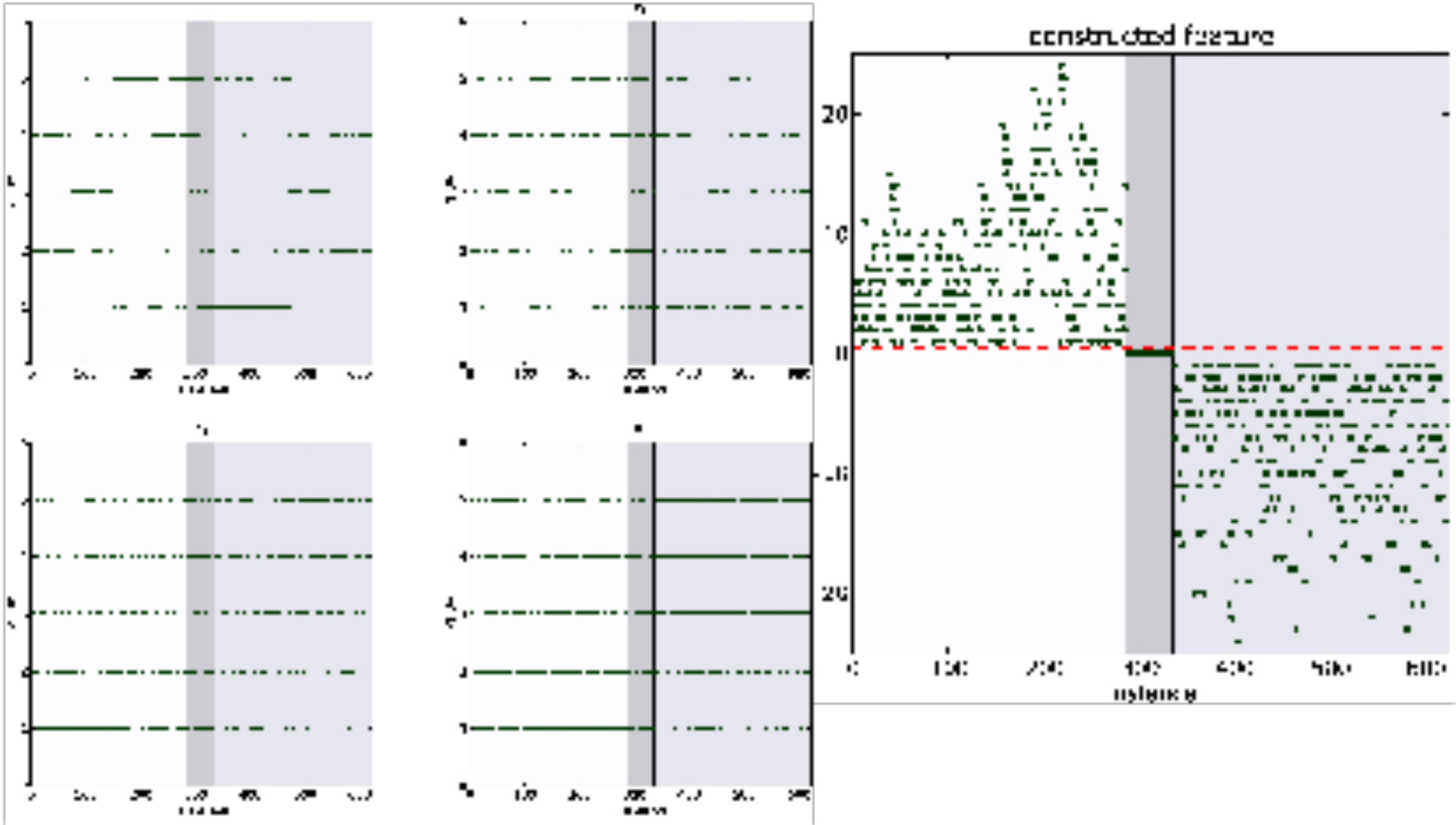


- Non-overlapping intervals

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

GP for FC Measure: Original VS Constructed

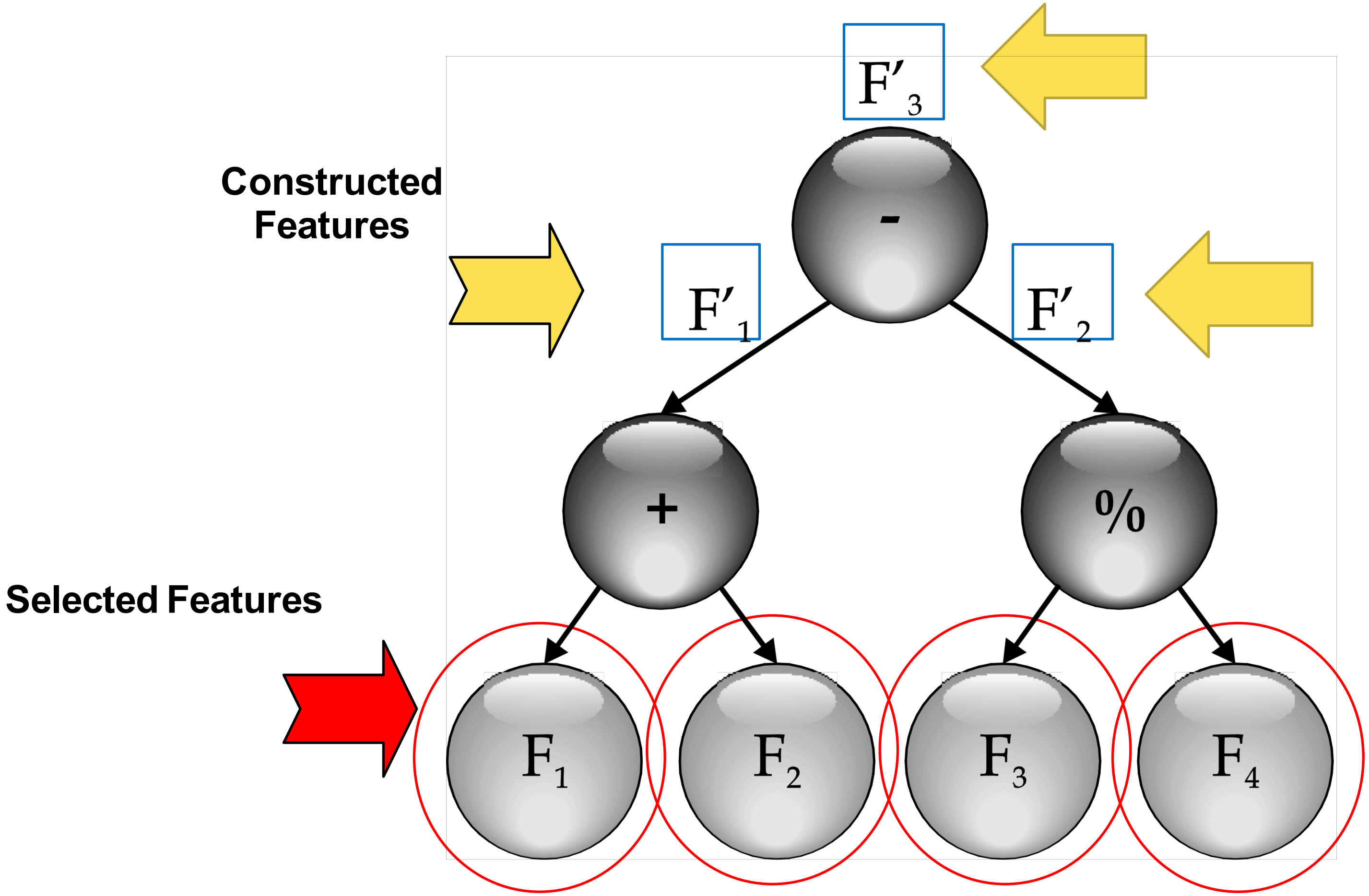
- 4 features, 3 classes



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

GP for FC Measure: Multiple feature construction

- Construct multiple features from a single tree

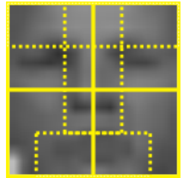
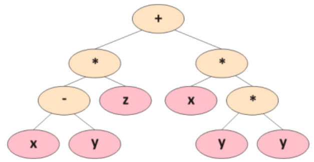


Soha Ahmed, Mengjie Zhang, Lifeng Peng and Bing Xue. "Multiple Feature Construction for Effective Biomarker Identification and Classification using Genetic Programming". Proceedings of 2014 Genetic and Evolutionary Computation Conference (GECCO 2014). ACM Press. Vancouver, BC, Canada. 12-16 July 2014.pp.249--256

Image Recognition/Classification

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 a GP-based classifier	

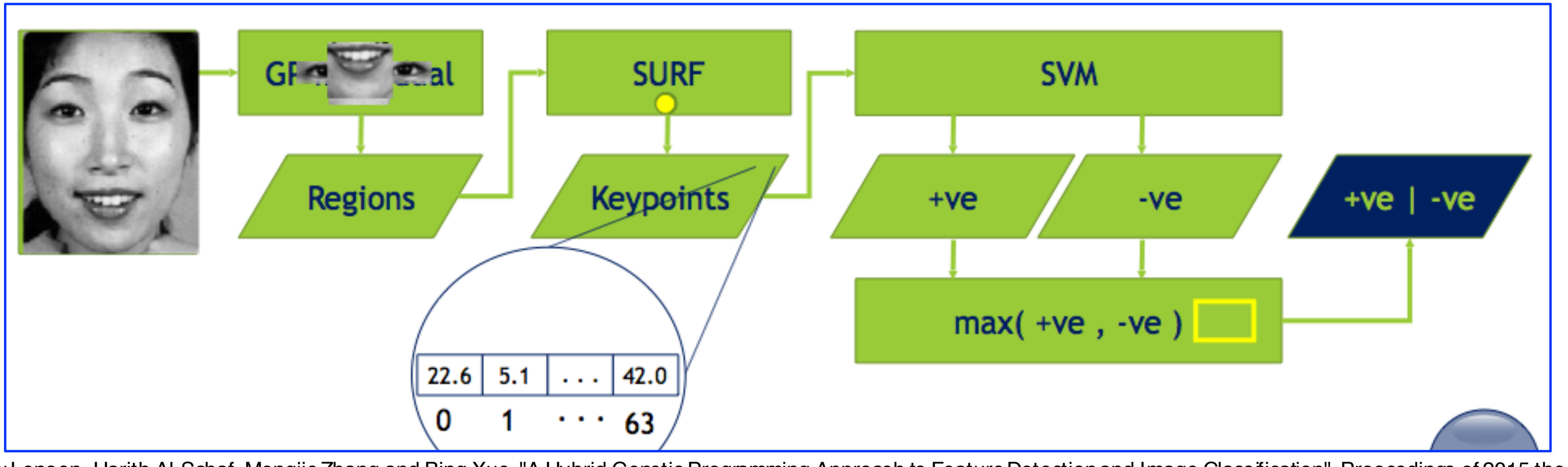
Images: GP-Surff

- Improve domain-independent object classification in images by using GP techniques.

Designing a program representation that is capable of detecting sub-regions of the image that are rich in features;

Constructing a classification system to extract features from the selected regions and then use a SVM classifier and voting scheme to predict the class label; and

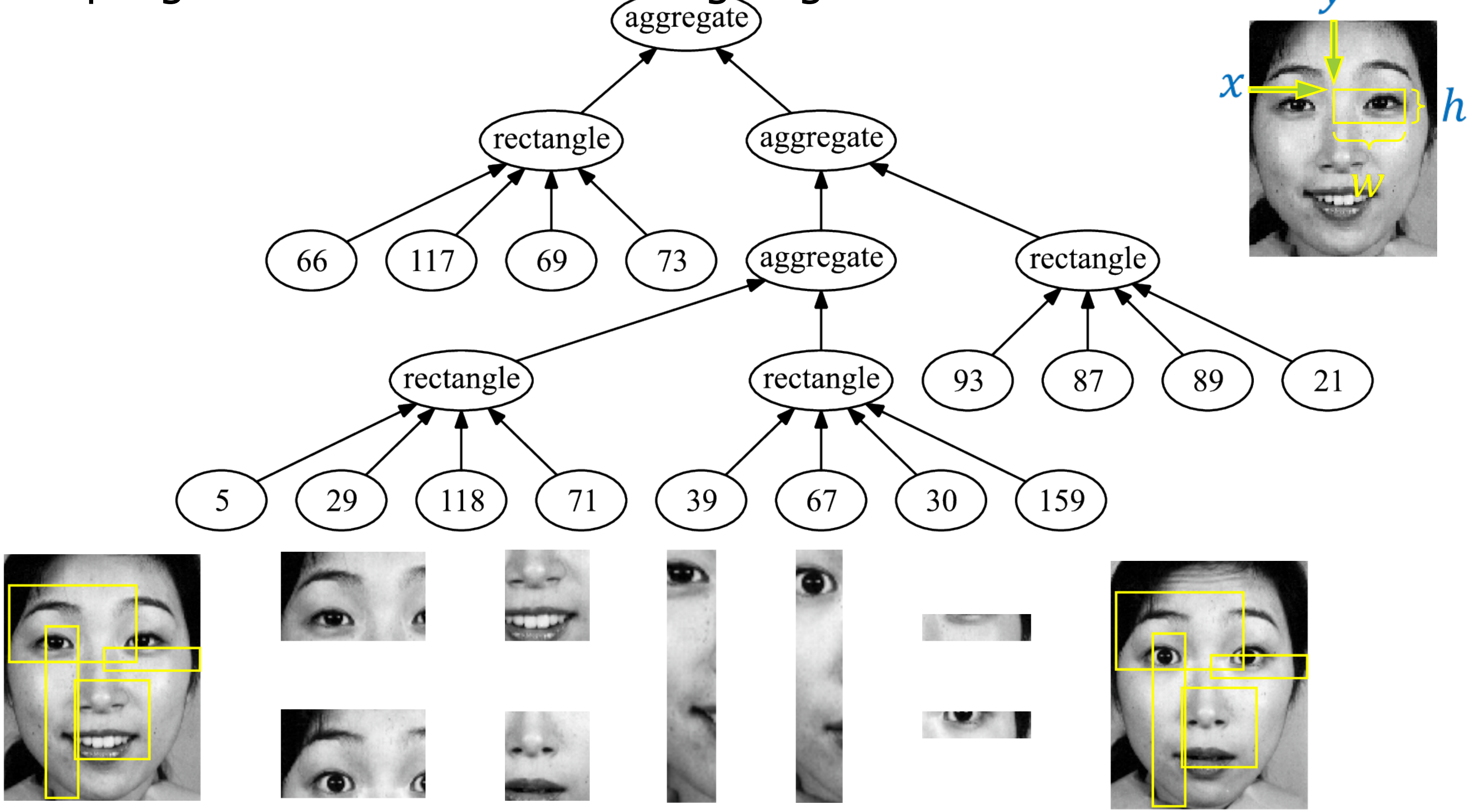
Investigating whether the regions detected by the new method are similar to those designed by domain experts.



Andrew Lensen, Harith Al-Sahaf, Mengjie Zhang and Bing Xue. "A Hybrid Genetic Programming Approach to Feature Detection and Image Classification". Proceedings of 2015 the 30th International Conference on Image and Vision Computing New Zealand (IVCNZ 2015). IEEE Press. Auckland. 23 - 24 Nov 2015. pp. (to appear)

Images: GP-Surff

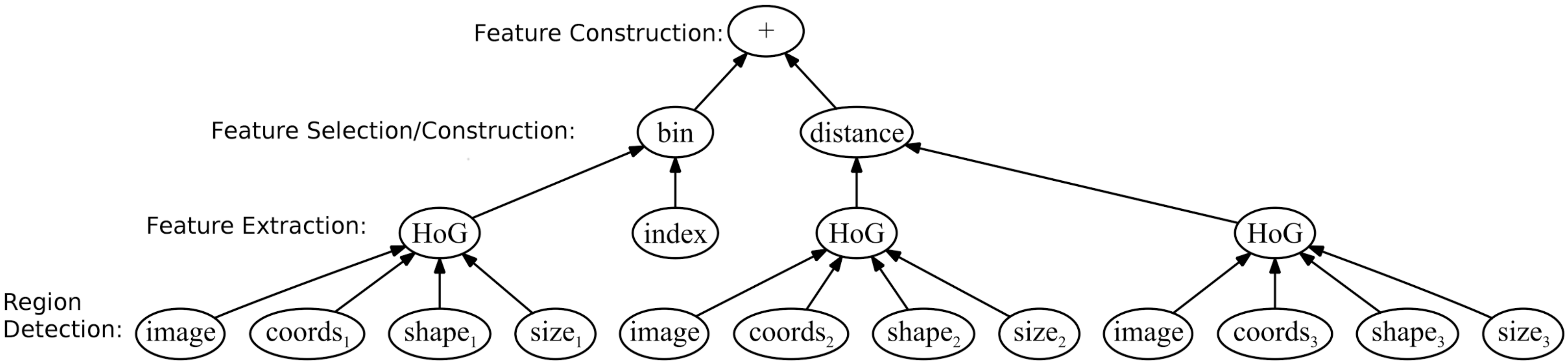
- A program evolved on JAFFE, average over 95% test accuracy
- The program detect 4 interesting regions



Andrew Lensen, Harith Al-Sahaf, Mengjie Zhang and Bing Xue. "A Hybrid Genetic Programming Approach to Feature Detection and Image Classification". Proceedings of 2015 the 30th International Conference on Image and Vision Computing New Zealand (IVCNZ 2015). IEEE Press. Auckland. 23 - 24 Nov 2015. pp. (to appear)

Images: GP-HoG Method

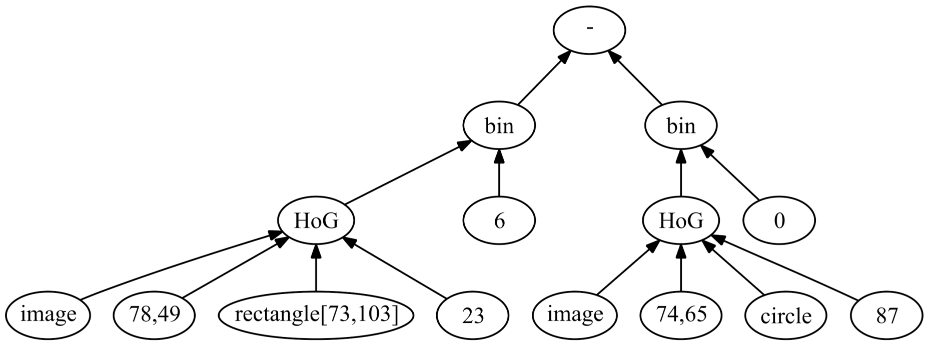
- GP-HoG uses strongly typed GP to perform three tasks in the same tree structure.
- All layers are trained simultaneously and coherently.
- Output of the tree is thresholded.

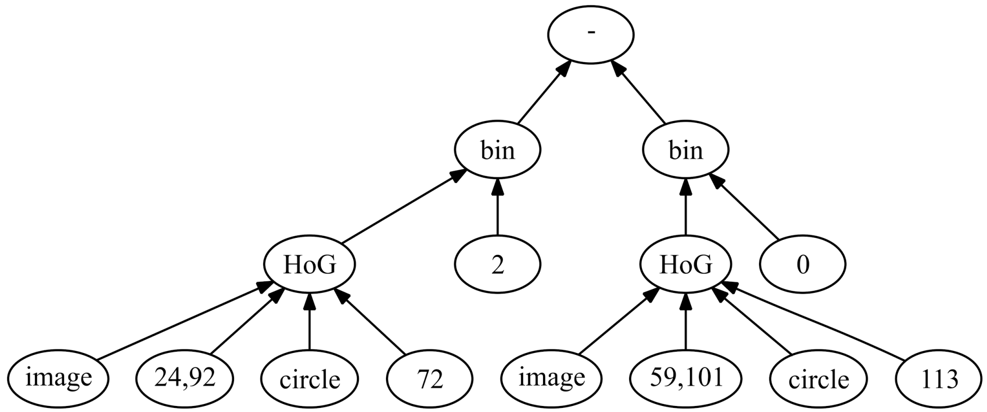


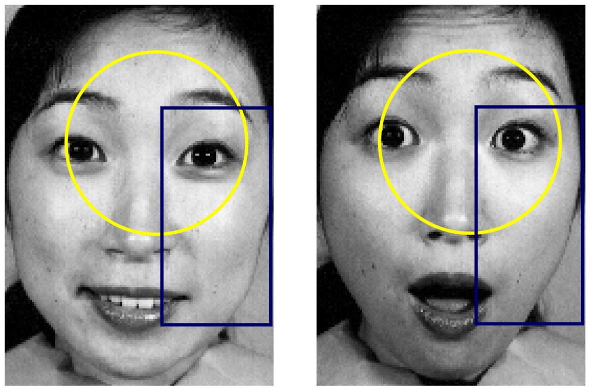
Andrew Lensen, Harith Al-Sahaf, Mengjie Zhang, Bing Xue. "Genetic Programming for Region Detection, Feature Extraction, Feature Construction and Classification in Image Data". Proceedings of the 19th European Conference on Genetic Programming (EuroGP 2016). Lecture Notes in Computer Science. Vol. 9594. Porto, Portugal, March 30 - April 1, 2016. pp. 51-67

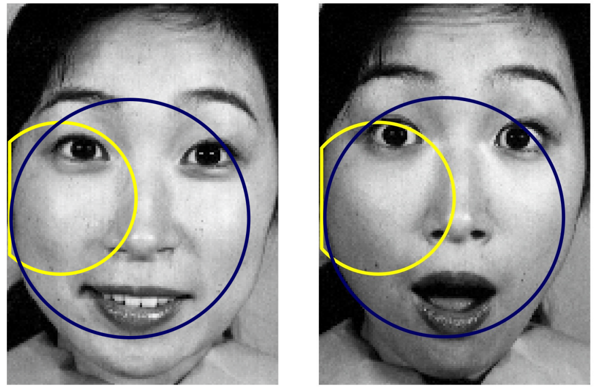
Images: GP-HoG Method

- The below tree has 98% training and 95% test performance on the Jaffe dataset despite being very simple.
- The below tree has 95% training and 100% test performance on the Jaffe dataset.









Andrew Lensen, Harith Al-Sahaf, Mengjie Zhang, Bing Xue. "Genetic Programming for Region Detection, Feature Extraction, Feature Construction and Classification in Image Data". Proceedings of the 19th European Conference on Genetic Programming (EuroGP 2016). Lecture Notes in Computer Science. Vol. 9594. Porto, Portugal, March 30 - April 1, 2016. pp. 51-67

Biology

Biology

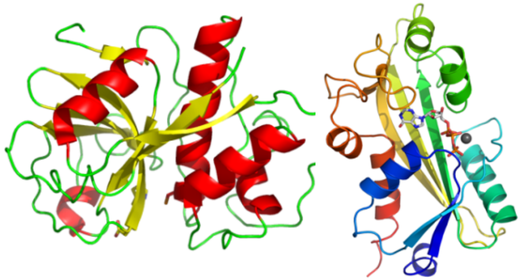
- Due to the nature, the MS data production process is very expensive (costs around 2,000 NZD daily) and time consuming (around two weeks to produce a single sample).
- The number of samples available is very small and the number of features in each sample is extremely large.
- Moreover, the features of interest are too small.
- The classification of MS data is challenging.

Biological Datasets

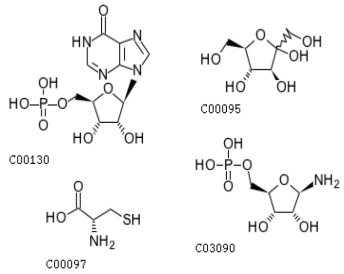
Data set	# Features	# Samples	# Classes
Pancreatic Cancer	6771	181	2
Ovarian Cancer1	15154	253	2
Ovarian Cancer 2	15000	216	2
Prostate Cancer	15000	322	4
Toxpath	7105	115	4
Arcene	10,000	200	2
Apple-plus	**773**	**40**	**4**
Apple-minus	365	40	4

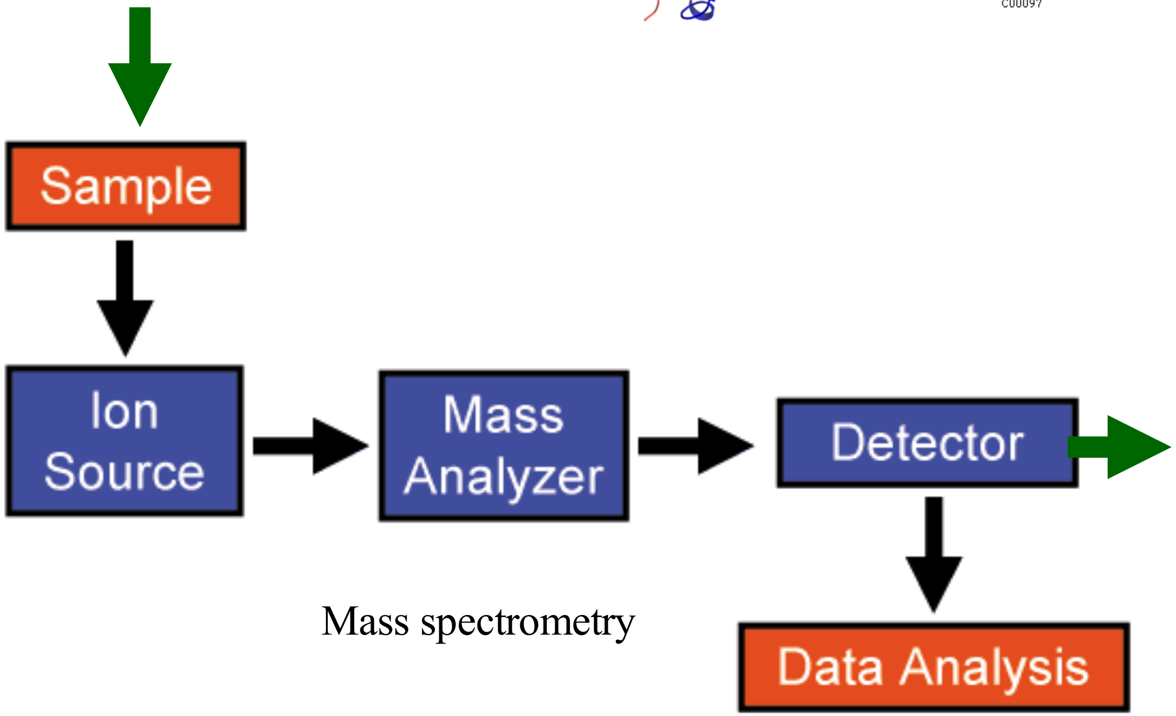
Biology

Proteins

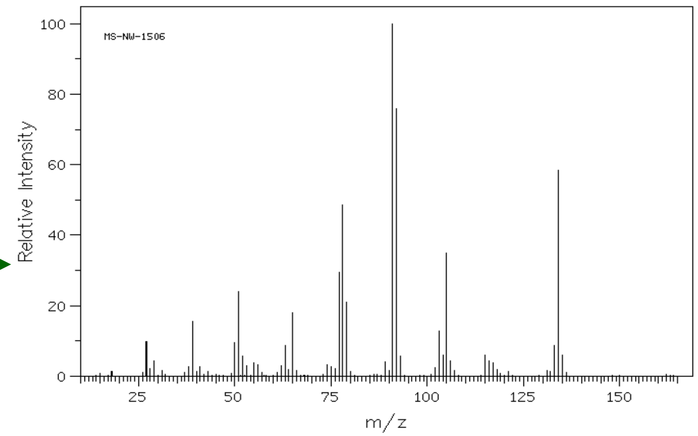


Metabolites





Mass spectrometry



Spectrum

Biology

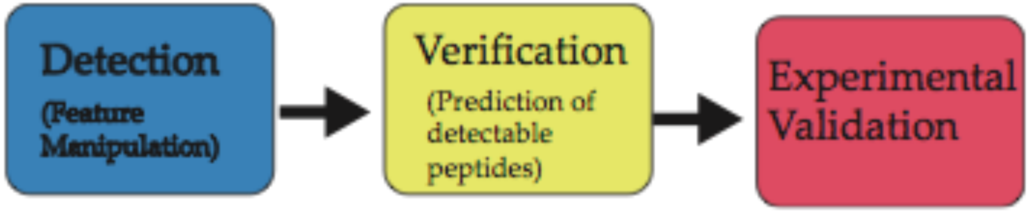
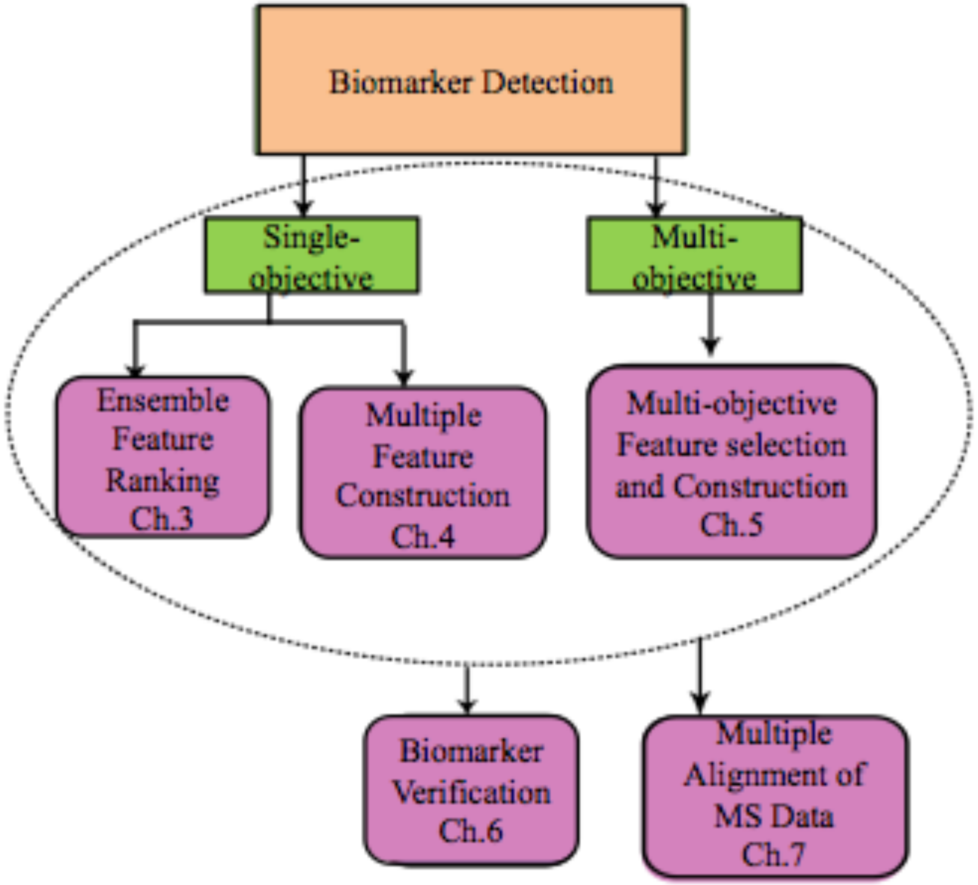


Figure 1.1: Stages of the biomarker identification process



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

Biology: Feature ranking and GP FS

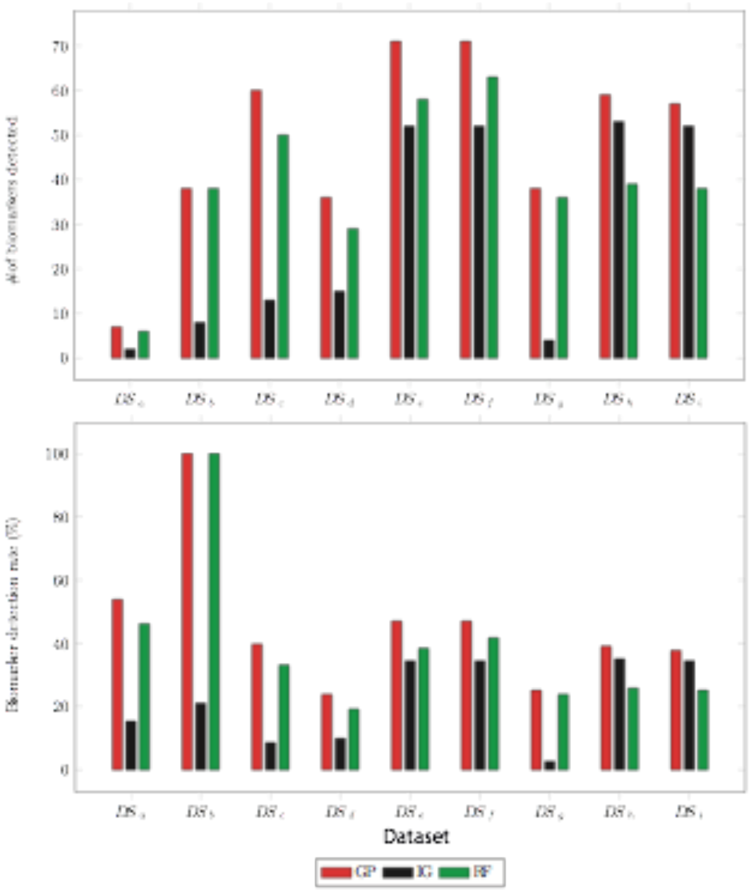
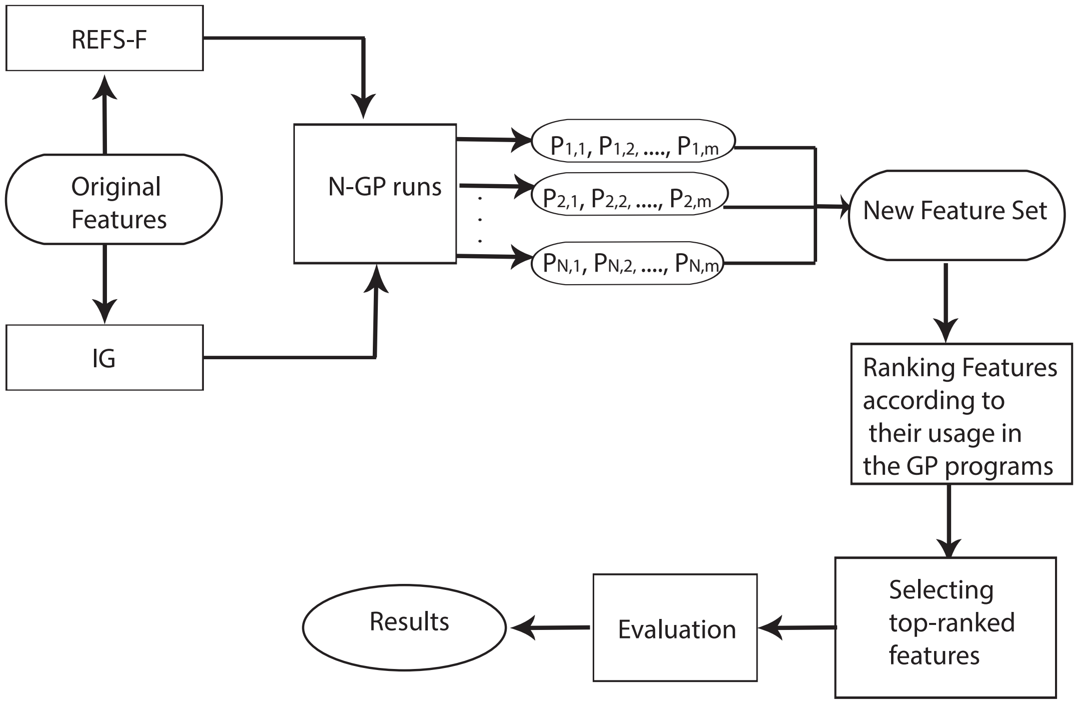


Figure 3.6: Biomarker detection of the proposed method in comparison with IG and RF.



Soha Ahmed, Mengjie Zhang, Lifeng Peng. "Improving Feature Ranking for Biomarker Discovery in Proteomics Mass Spectrometry Data using Genetic Programming". Connection Science. **Vol. 26**, Issue 3, 2014. pp. 215-243

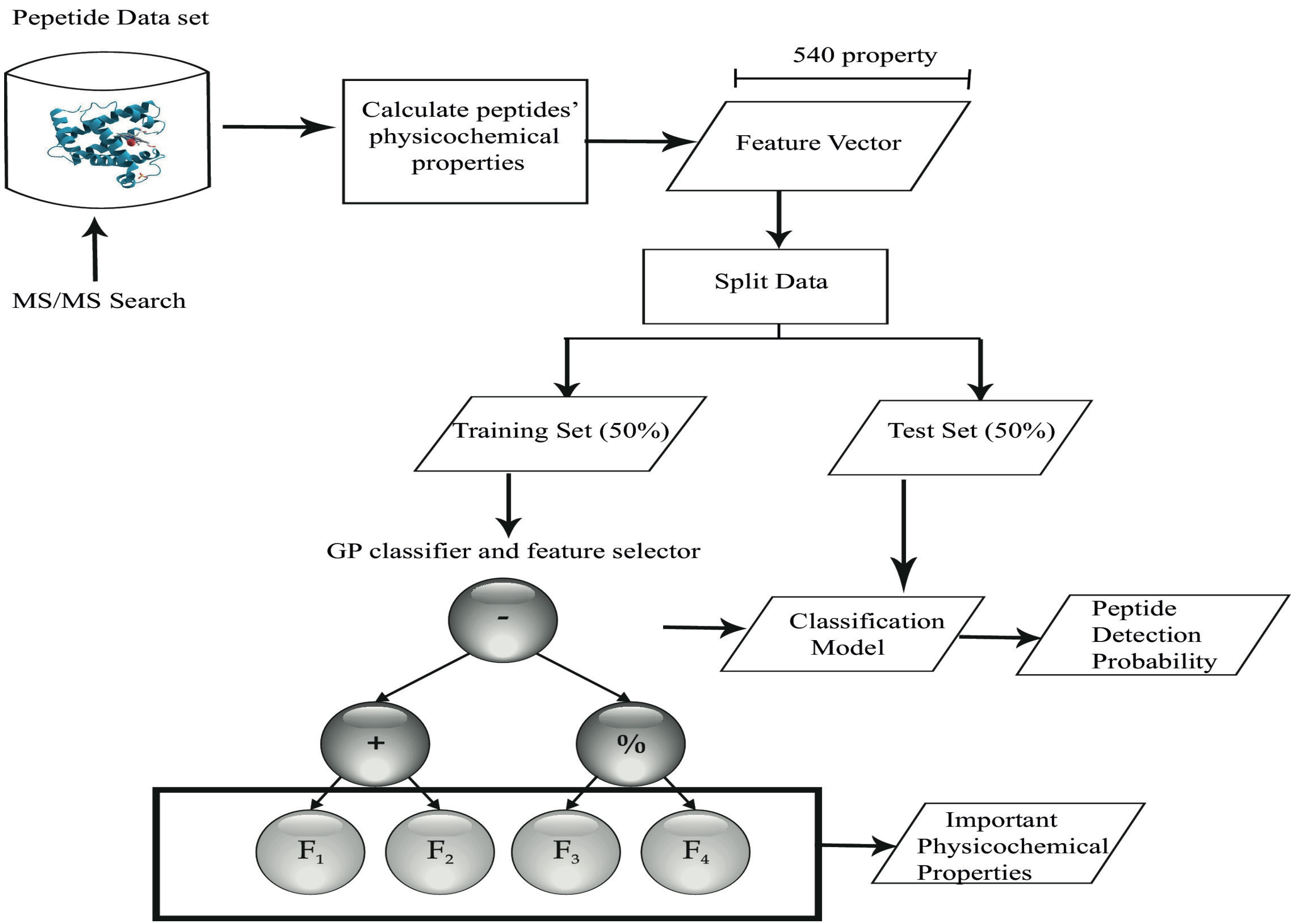
Biomarker Identification

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	✓	✗

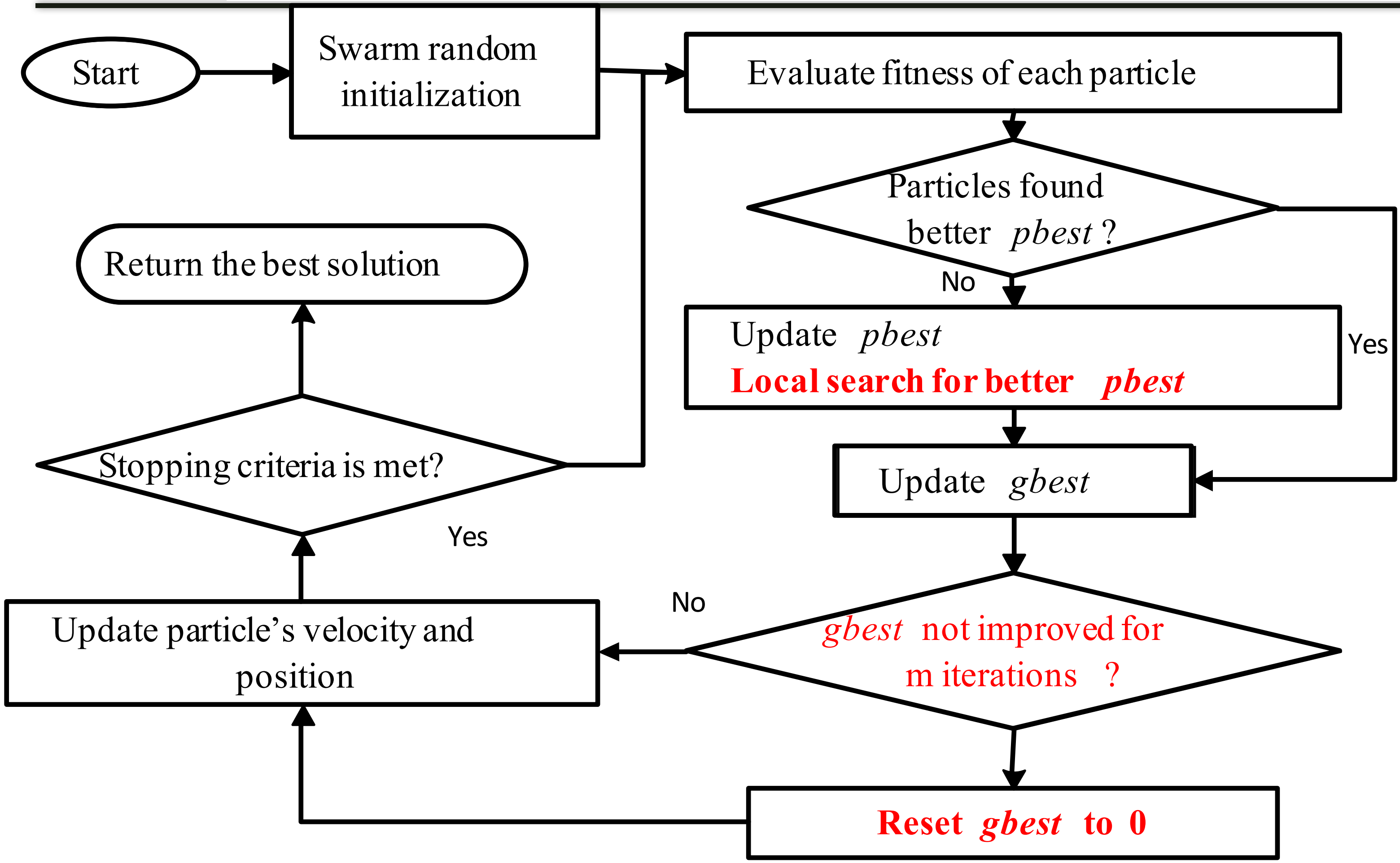
Soha Ahmed, Mengjie Zhang, Lifeng Peng and Bing Xue. "Multiple **Feature Construction** for Effective Biomarker Identification and Classification using Genetic Programming". Proceedings of 2014 Genetic and Evolutionary Computation Conference (GECCO 2014). ACM Press. Vancouver, BC, Canada. 12-16 July 2014.pp.249--256

Biology:GP for Measuring Peptide Detectability



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

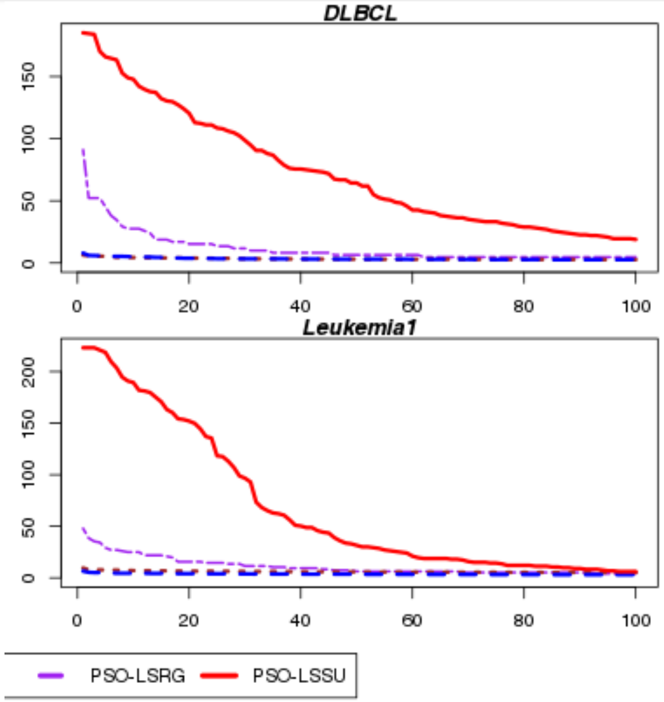
Biology: PSO with local search on *pbest* and resetting *gbest* (PSO-LSRG)



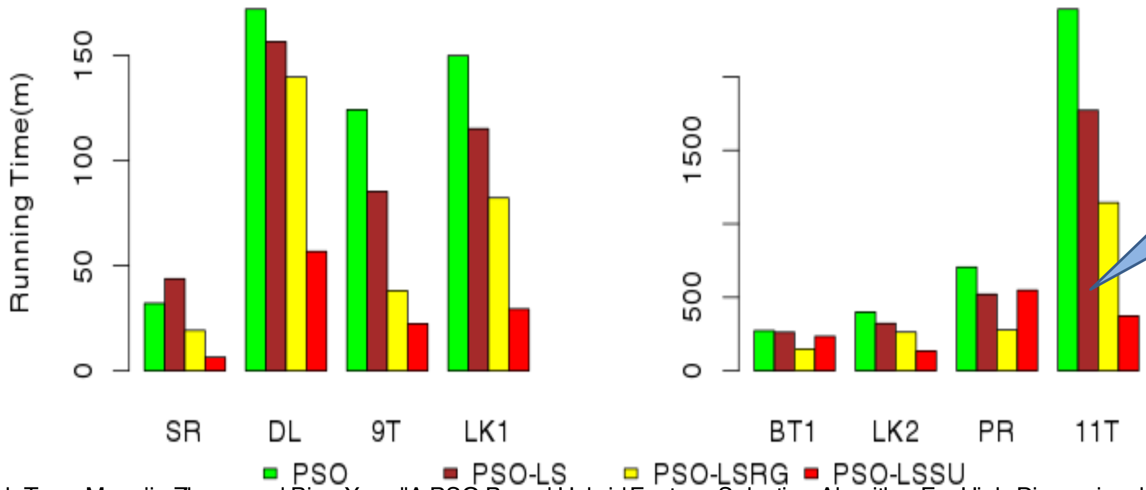
- Use a filter measure to identify:
  - Relevant features: correlated to the class label.
  - Redundant features: correlated with each other.
- Symmetric uncertainty (SU) is a normalised version of information gain (IG).

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 2016 /CEC2016). Vancouver, Canada. 24-29 July, 2016. pp(to appear)

Biology: PSO with local search on *pbest* and resetting *gbest* (PSO-LSRG)



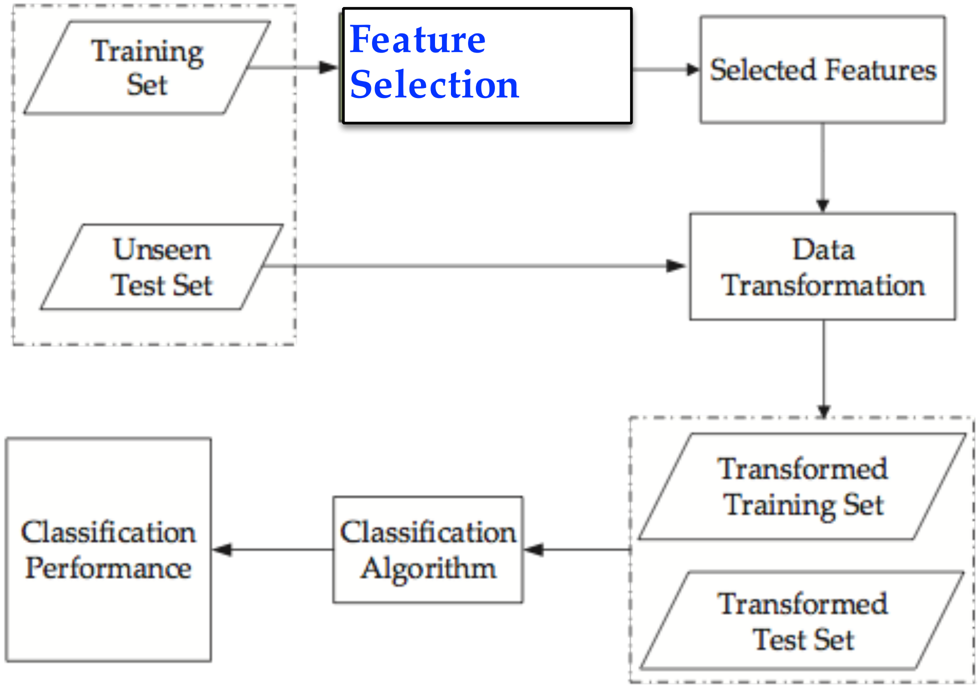
- A PSO based hybrid FS algorithm for high-dimensional classification.
- PSO-LSSU combines wrapper and filter measures:
  - The fitness function.
  - The local search.
- PSO-LSSU achieved much smaller feature subsets with significantly better classification performance than the compared methods in most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



5 - 6 times faster than PSO

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 2016 /CEC2016). Vancouver, Canada. 24-29 July, 2016. pp(to appear)

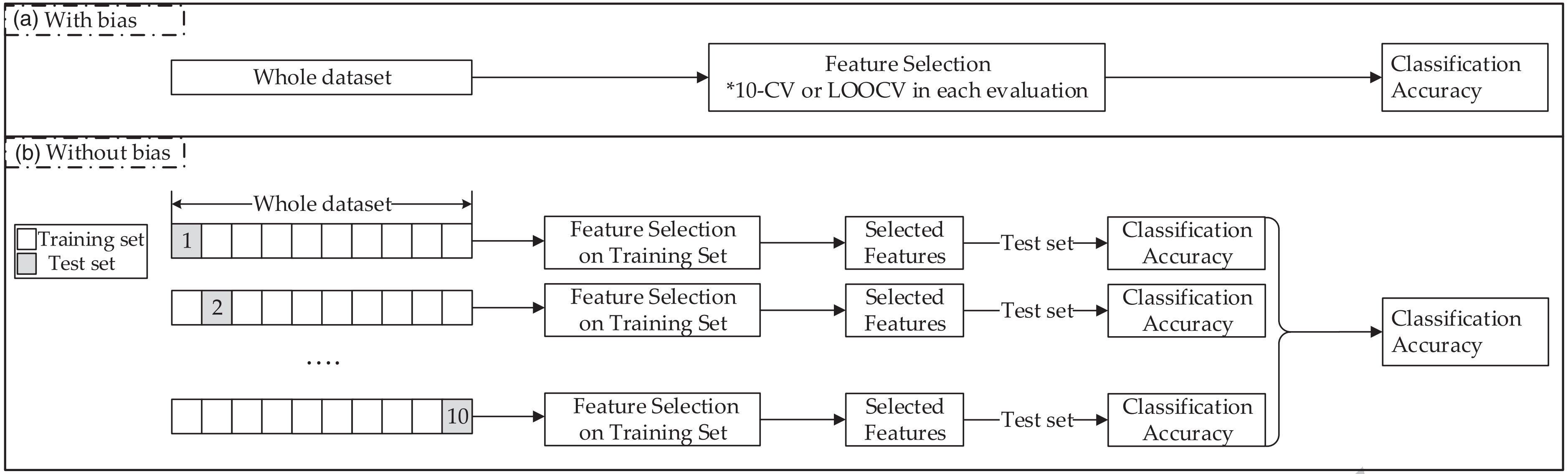
Feature Selection/Construction Bias



- 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. "Investigation on Particle Swarm Optimisation for Feature Selection on High-dimensional Data: Local Search and Selection Bias", Connection Science, Accepted April 2016

GP for FS and FC

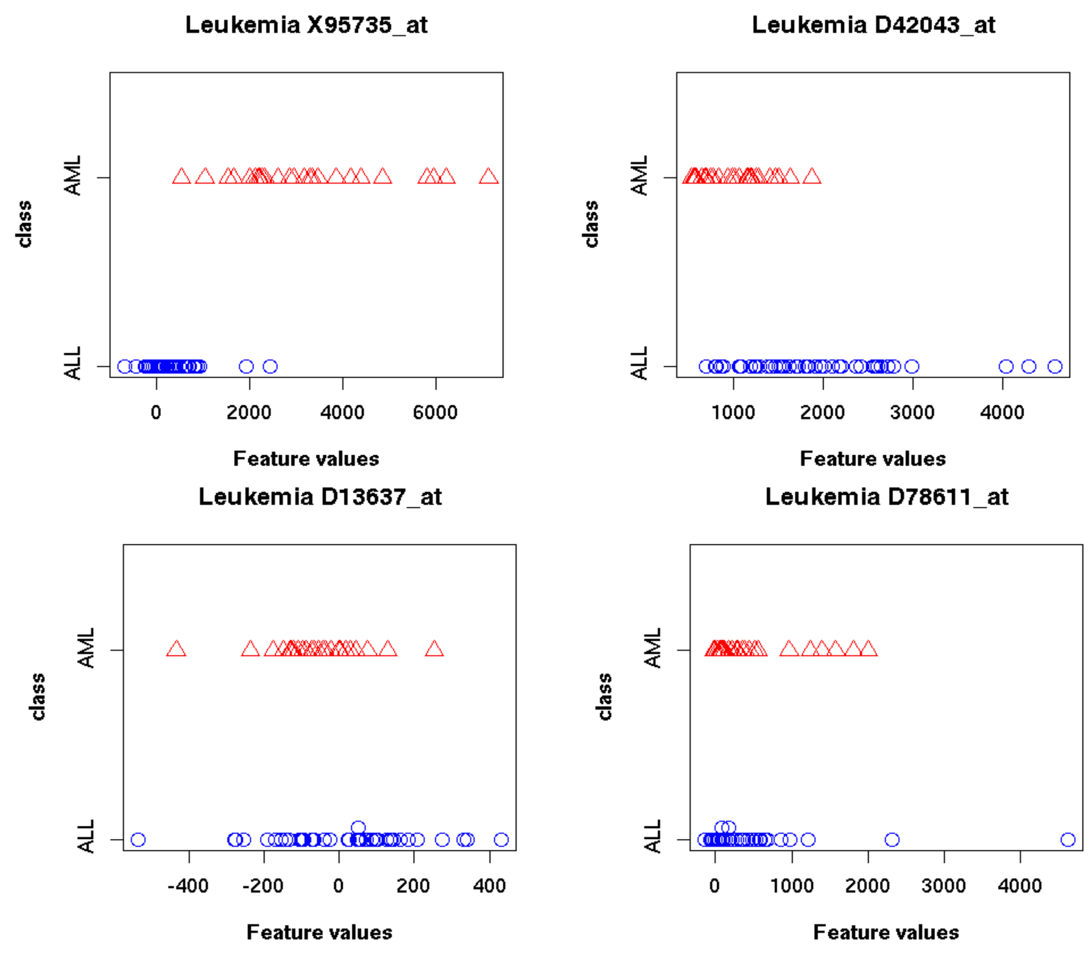


Figure 1: Leukemia constructed feature

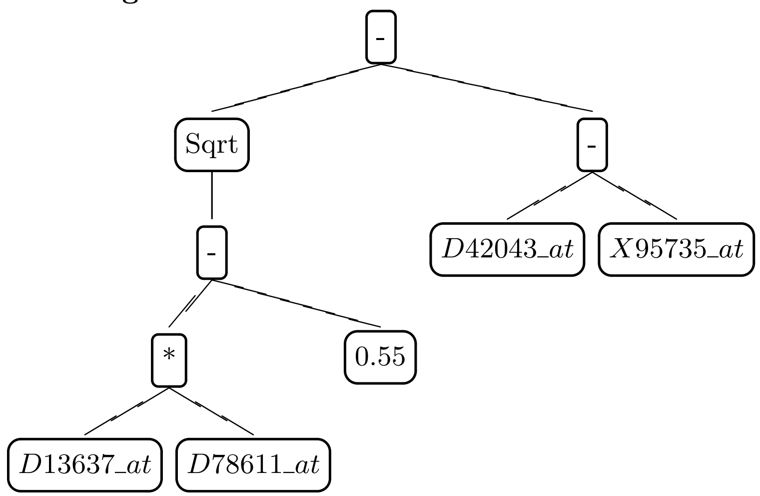
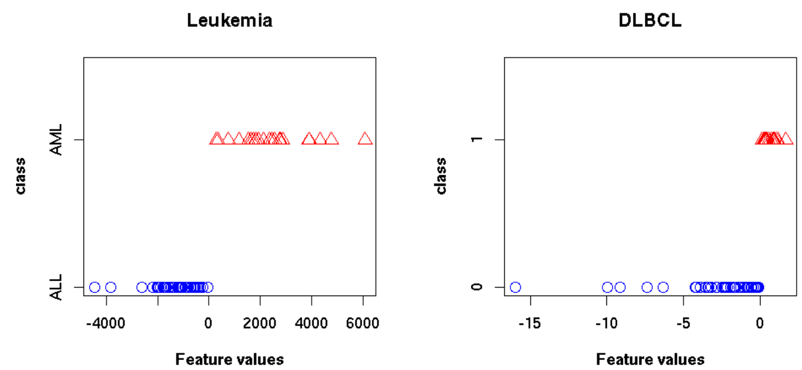


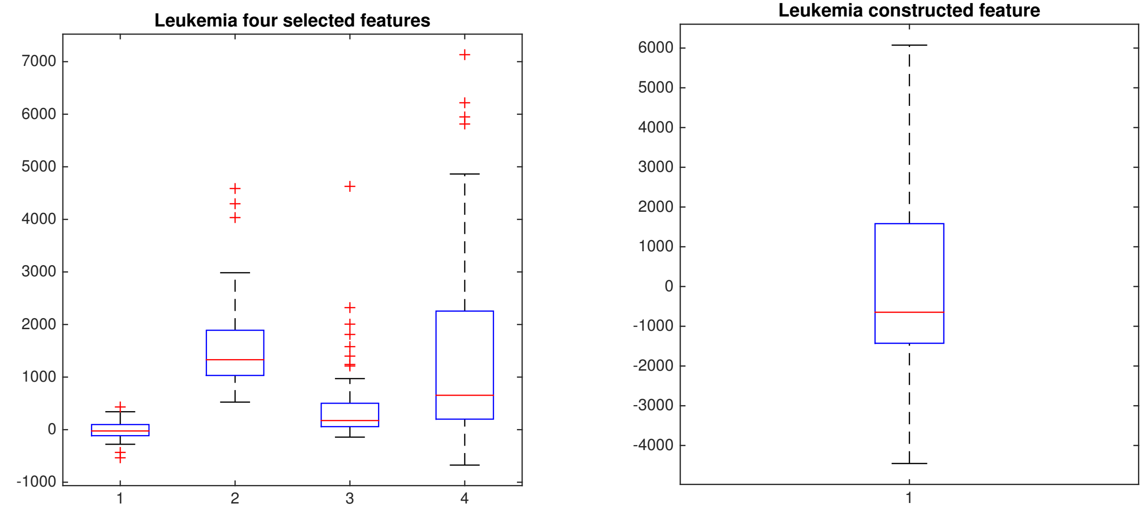
Figure 4: Constructed Features

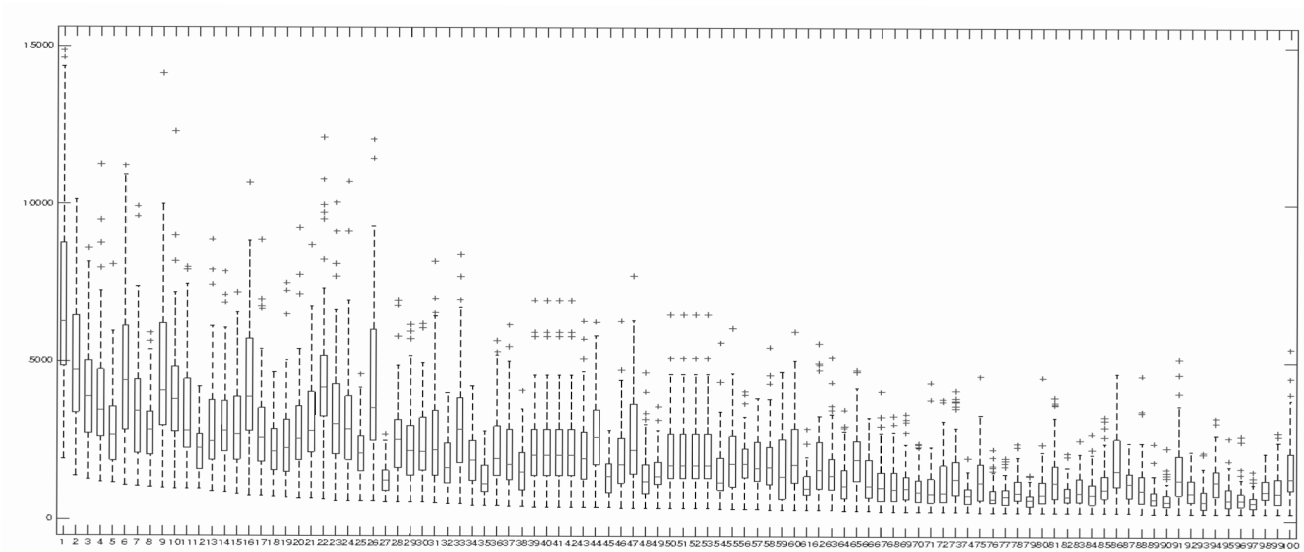


Binh Tran, Bing Xue and Mengjie Zhang. "Genetic Programming for Feature Construction and Selection in Classification on High-dimensional Data", Memetic Computing, Accepted 4 December 2015. (DOI:10.1007/s12293-015-0173-y)

GP for FS and FC

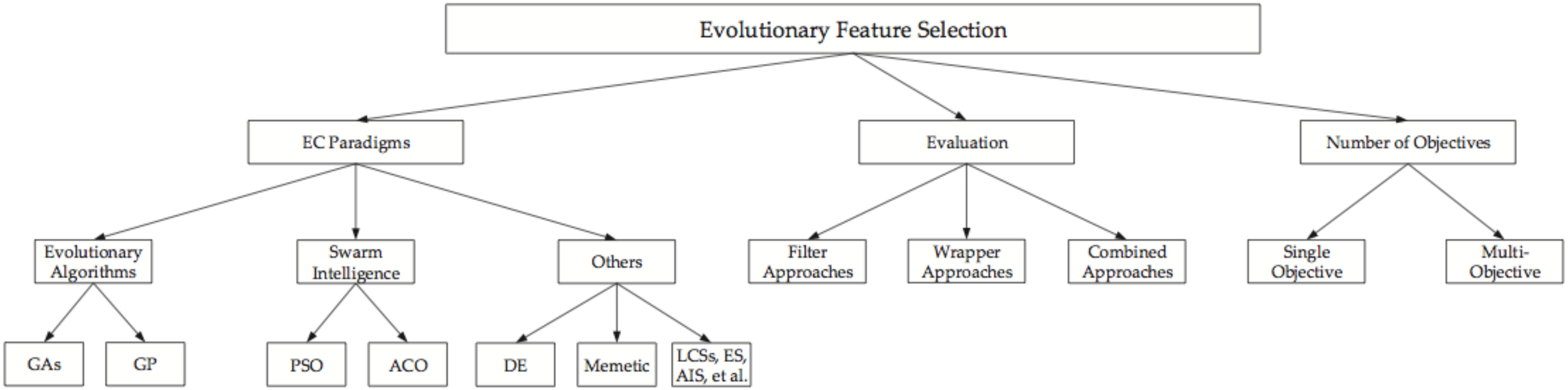
Figure 9: Feature distributions

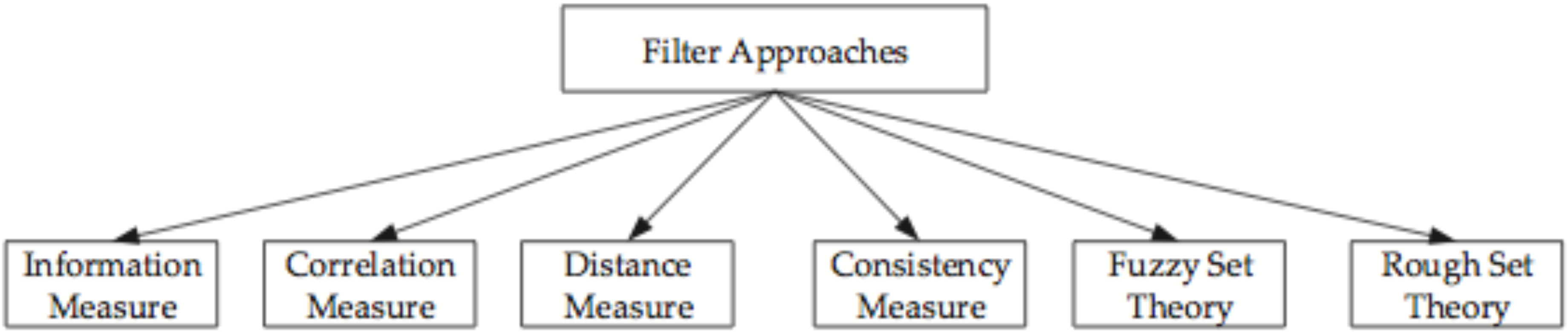




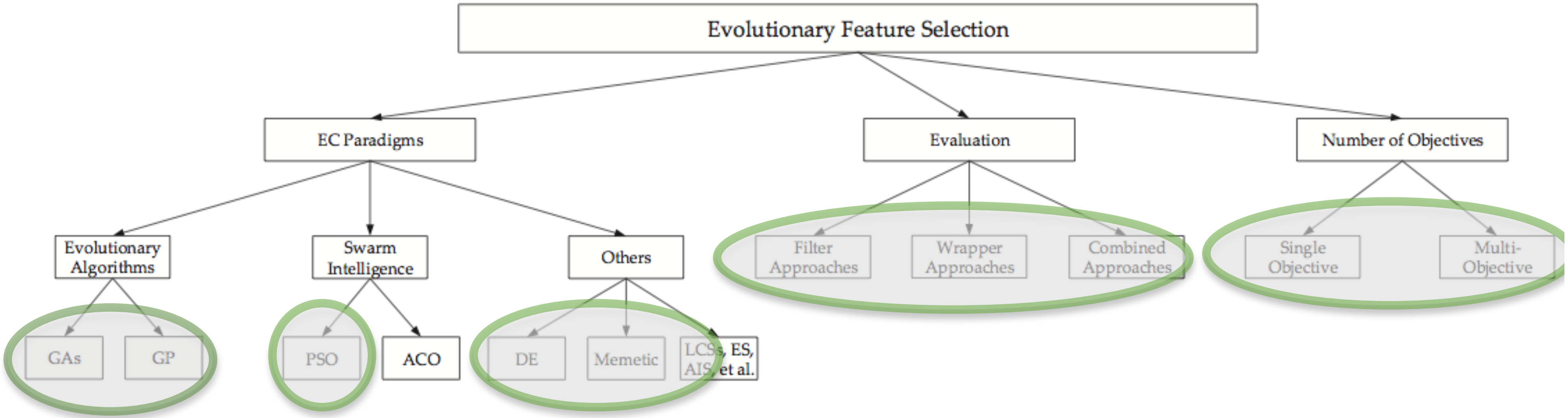
Binh Tran, Bing Xue and Mengjie Zhang. "Genetic Programming for Feature Construction and Selection in Classification on High-dimensional Data", Memetic Computing, Accepted 4 December 2015. (DOI:10.1007/s12293-015-0173-y)

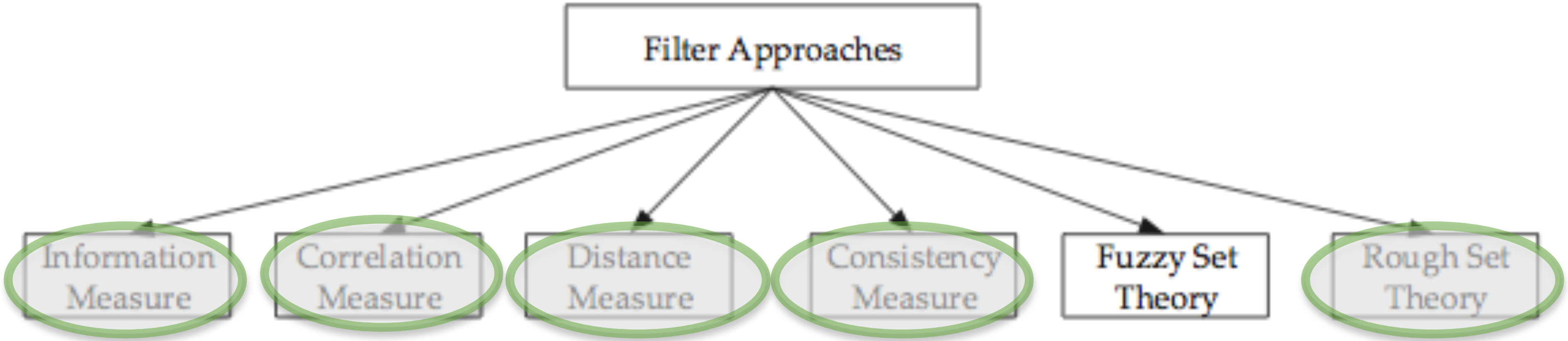
Work from ECRG





Work from ECRG





Weakness and Issues

- 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

- Efficient and effective filter measure for the fitness function:
  - reduce the computational cost,
  - smooth the landscape of the search space,
  - improve the learning and generalisation performance, and
  - increase the interpretability/understandability of the obtained feature set

- 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
  - Evolutionary multi-objective optimisation
  - Combinatorial optimisation
  - Memetic computing
  - Large-scale optimisation
  - Surrogate models
  - Adaptive parameter control techniques
- Feature construction
  - both feature selection and feature construction
- Instance selection and construction
- Combining EC with *machine learning approaches*
- Feature selection and feature construction for other machine learning tasks: clustering and symbolic regression

Acknowledgement

- Thanks A/Prof Will Browne, A/Prof Peter Andreae, A/Prof Ivy (I-Ming) Liu, A/Prof Lin Shang, Dr Lifeng Peng, Dr Kourosh Neshatian, Dr Yi Mei, Dr Su Nguyen, Dr Soha Ahmed, Harith Al-Sahaf, Liam Cervante, Andrew Lensen, Hoai Bach Nguyen, Binh Tran, Qi Chen, Emrah Hancer, and others

- Funding Agents:
  1. Marsden Fund of New Zealand award number(s): VUW1209
  2. University Research Fund at Victoria University of Wellington award number(s): 210375/3557, 209861/3580

Acknowledgement

- Thanks everyone in the Evolutionary Computation Research Group at Victoria University of Wellington, New Zealand

Activities in 2016

- Task Force on Evolutionary Computation for Feature Selection and Construction, IEEE CIS
- IEEE Symposium on Computational Intelligence in Feature Analysis, Selection, and Learning in Image and Pattern Recognition (FASLIP) in IEEE SSCI 2016
- Australian Conference on Artificial Lift and Computational Intelligence (ACALCI 2017)
- Special session on Evolutionary Machine Learning in Image Analysis and Pattern Recognition in The 20th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES2016)
- Special session on Evolutionary Feature Selection and Construction in IEEE WCCI 2016 /CEC2016
- Special session on Transfer Learning in Evolutionary Computation in IEEE WCCI 2016 /CEC2016
- Special Issue on Evolutionary Optimisation, Feature Reduction and Learning, Soft Computing (Journal) , Springer

Proposed Activities

- Task Force on Evolutionary Computation for Feature Selection and Construction, IEEE CIS
- Special session on Evolutionary Feature Selection and Construction in CEC2017
- Special session on Evolutionary Machine Learning in Image Analysis and Pattern Recognition in IES2017
- IEEE Symposium on Computational Intelligence in Feature Analysis, Selection, and Learning in Image and Pattern Recognition (FASLIP) in IEEE SSCI 2017
- Special session on Transfer Learning in Evolutionary Computation in IEEE WCCI 2016 /CEC2016
- The tutorial on EC for Feature Selection and Feature Construction, GECCO 2017
- The tutorial on EC for Feature Selection and Feature Construction, CEC2017
- Special Issue

Call for Papers

IEEE Symposium on Computational Intelligence in Feature Analysis, Selection, and Learning in Image and Pattern Recognition (FASLIP)

2016 IEEE Symposium Series on Computational Intelligence (SSCI 2016)

http://ssci2016.cs.surrey.ac.uk/

December 6-9, 2016, Athens, Greece

Deadline: 15 August 2016

PhD Scholarships Available in Evolutionary Computation

Join our internationally renowned and friendly research team:

- Up to eight funded PhDs (fees + stipend) available 3 times a year,
- 3yr duration in English, with expert supervision.

Five major EC strategic research directions:

- Feature selection/construction for classification, regression, clustering
- Combinatorial optimisation: scheduling, routing, web services
- Computer vision and image analysis
- Multi- and many- criteria optimisation
- Transfer learning

Techniques include: Genetic Programming, Learning Classifier Systems, Particle Swarm Optimisation, Differential Evolution, and many others.

Wellington voted as 'Coolest little capital in the World!'

VUW is the top-rated research university in New Zealand.

Requirements: MSc/ME; GPA >= 3.5/4; research experience/publications

Come and find us after one of our many talks or apply at:
http://www.victoria.ac.nz/fgr/prospective-phds/scholarships

Find us: http://ecs.victoria.ac.nz/Groups/ECRG/WebHome or [Search: 'ECRG VUW']

Call for Postdoctoral Fellows

- Evolutionary Computation Research Group, Victoria University of Wellington, NZ
- Postdoc in Evolutionary Computations
- Salary: $70,000 – 85,000
- Areas:
  - Evolutionary Feature Selection and High Dimensionality Reduction
  - Evolutionary Image Analysis
  - Classification and Clustering
  - Transfer Learning
- Huawei NZ Funded Project
- Contact: Mengjie.Zhang@ecs.vuw.ac.nz or Bing.Xue@ecs.vuw.ac.nz

Thank you