Multi-objective Ensemble Learning and Its Applications

Xin Yao
Department of Computer Science and Engineering, and Research Institute of Trustworthy Autonomous Systems
Southern University of Science and Technology (SUSTech)
Shenzhen, China
and
School of Computer Science, University of Birmingham, UK

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

What Is an Ensemble?

- In this talk, an ensemble indicates a collection of individual learning machines, also called individual/base learners or, simply, individuals.
- For example, we could have an ensemble of neural networks (NNs), decision trees, Bayesian classifiers, genetic programming classifiers, etc.
- We could even have heterogeneous base learners in an ensemble.
- There are many related topics to ensembles, e.g., committee machines, multiple classifiers, mixture of experts, stacked generalisation, etc.

What Is an Ensemble?

- In this talk, an ensemble indicates a collection of individual learning machines, also called individual/base learners or, simply, individuals.
- For example, we could have an ensemble of neural networks (NNs), decision trees, Bayesian classifiers, genetic programming classifiers, etc.
- We could even have heterogeneous base learners in an ensemble.
- There are many related topics to ensembles, e.g., committee machines, multiple classifiers, mixture of experts, stacked generalisation, etc.

What Is an Ensemble?

- In this talk, an ensemble indicates a collection of individual learning machines, also called individual/base learners or, simply, individuals.
- For example, we could have an ensemble of neural networks (NNs), decision trees, Bayesian classifiers, genetic programming classifiers, etc.
- We could even have heterogeneous base learners in an ensemble.
- There are many related topics to ensembles, e.g., committee machines, multiple classifiers, mixture of experts, stacked generalisation, etc.

What Is an Ensemble?

- **In this talk, an ensemble indicates a collection of individual learning machines, also called individual/base learners or, simply, individuals.**
- **For example, we could have an ensemble of neural networks (NNs), decision trees, Bayesian classifiers, genetic programming classifiers, etc.**
- **We could even have heterogeneous base learners in an ensemble.**
- **There are many related topics to ensembles, e.g., committee machines, multiple classifiers, mixture of experts, stacked generalisation, etc.**

What Is an Ensemble?

- In this talk, an ensemble indicates a collection of individual learning machines, also called individual/base learners or, simply, individuals.
- For example, we could have an ensemble of neural networks (NNs), decision trees, Bayesian classifiers, genetic programming classifiers, etc.
- We could even have heterogeneous base learners in an ensemble.
- There are many related topics to ensembles, e.g., committee machines, multiple classifiers, mixture of experts, stacked generalisation, etc.

Why Ensembles?

- For a large and complex problem, designing a monolithic system to solve it is often very difficult.
- Divide-and-conquer is a common strategy in solving such problems.
- Ensemble approaches could be viewed as an automatic approach toward divide-and-conquer.

Why Ensembles?

- **For a large and complex problem, designing a monolithic system to solve it is often very difficult.**
- **Divide-and-conquer is a common strategy in solving such problems.**
- **Ensemble approaches could be viewed as an automatic approach toward divide-and-conquer.**

Why Ensembles?

- For a large and complex problem, designing a monolithic system to solve it is often very difficult.
- Divide-and-conquer is a common strategy in solving such problems.
- Ensemble approaches could be viewed as an automatic approach toward divide-and-conquer.

Why Ensembles?

- For a large and complex problem, designing a monolithic system to solve it is often very difficult.
- Divide-and-conquer is a common strategy in solving such problems.
- Ensemble approaches could be viewed as an automatic approach toward divide-and-conquer.

X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 28(3):417-425, June 1998.

Potential Benefits of Ensembles

- For *regression problems,* it has been shown that an ensemble performs no worse than any of its individual learners under some mild assumptions/conditions.
- For *classification problems,* there are ample empirical evidences to show ensemble's advantages over single individuals.

Potential Benefits of Ensembles

- **For *regression problems*, it has been shown that an ensemble performs no worse than any of its individual learners under some mild assumptions/conditions.**
- **For *classification problems*, there are ample empirical evidences to show ensemble's advantages over single individuals.**

Potential Benefits of Ensembles

- **For *regression problems*, it has been shown that an ensemble performs no worse than any of its individual learners under some mild assumptions/conditions.**
- **For *classification problems*, there are ample empirical evidences to show ensemble's advantages over single individuals.**

What is it in an ensemble that makes it better?

Key to Ensembles: Diversity

- There have been many studies demonstrating that a diverse ensemble provide better generalisation.
- What do you mean by “diverse”? How do you define diversity? How do you generate diversity? …
  – Still ongoing research as to how diversity can be best defined and used in practice.

- G. Brown, J. L. Wyatt, R. Harris and X. Yao, ``Diversity Creation Methods: A Survey and Categorisation,'' *Information Fusion*, 6(1):5-20, January 2005.
- E. K. Tang, P. N. Suganthan and X. Yao, ``An Analysis of Diversity Measures,'' *Machine Learning*, 65:247-271, 2006.

Key to Ensembles: Diversity

- **There have been many studies demonstrating that a diverse ensemble provide better generalisation.**
- **What do you mean by "diversity"? How do you define diversity? How do you generate diversity? …**
  - **Still ongoing research as to how diversity can be best defined and used in practice.**

- **G. Brown, J. L. Wyatt, R. Harris and X. Yao, ``Diversity Creation Methods: A Survey and Categorisation,'' *Information Fusion*, 6(1):5-20, January 2005.**
- **E. K. Tang, P. N. Suganthan and X. Yao, ``An Analysis of Diversity Measures,'' *Machine Learning*, 65:247-271, 2006.**

Key to Ensembles: Diversity

- **There have been many studies demonstrating that a diverse ensemble provide better generalisation.**
- **What do you mean by "diversity"? How do you define diversity? How do you generate diversity? …**
  - **Still ongoing research as to how diversity can be best defined and used in practice.**

- G. Brown, J. L. Wyatt, R. Harris and X. Yao, ``Diversity Creation Methods: A Survey and Categorisation,'' *Information Fusion*, 6(1):5-20, January 2005.
- E. K. Tang, P. N. Suganthan and X. Yao, ``An Analysis of Diversity Measures,'' *Machine Learning*, 65:247-271, 2006.

Negative Correlation Learning (NCL)

- **NCL defines a simple error function for each individual as follows, where *$p_i(n)$* is a diversity measure:**

$$E_i = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{2}(F_i(n) - d(n))^2 + \lambda p_i(n)\right)$$

where

$$p_i(n) = (F_i(n) - d(n))\sum_{j \neq i}(F_j(n) - d(n))$$

$F(n)$ is the ensemble output.

Y. Liu and X. Yao, ``Ensemble learning via negative correlation,'' *Neural Networks*, 12(10):1399-1404, December 1999.

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Where Are Multiple Objectives?

- The only objective of learning is to maximise generalization, isn't it?
- Where do multiple objectives come from?

Where Are Multiple Objectives?

- **The only objective of learning is to maximise generalization, isn't it? What else?**
- **Where do multiple objectives come from?**

Where Are Multiple Objectives?

- The only objective of learning is to maximise generalization, isn't it?
- Where do multiple objectives come from?

Well …

- **Let’s just recall all the loss functions we have seen in the literature.**

- **Don’t they all have two or more terms added together through some coefficients (beautifully called hyper-parameters nowadays)?**

Well …

- **Let's just recall all the loss functions we have seen in the literature.**

- **Don't they all have two or more terms added together through some coefficients (beautifully called hyper-parameters nowadays)?**

- **Can you imagine how many arts (or how much human time or how many papers) have been generated to tune these hyper-parameters?**

Well …

- **Let's just recall all the loss functions we have seen in the literature.**
- **Don't they all have two or more terms added together through some coefficients (beautifully called hyper-parameters nowadays)?**
- **Can you imagine how many arts, or how much PhD student time, or how many papers, have been generated to tune these hyper-parameters?**

Recall Negative Correlation Learning

- Accuracy and diversity are two conflicting objectives.

$$E_i = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{2}(F_i(n) - d(n))^2 + \lambda p_i(n)\right)$$

where

$$p_i(n) = (F_i(n) - d(n))\sum_{j \neq i}(F_j(n) - d(n))$$

$F(n)$ is the ensemble output.

Y. Liu and X. Yao, ``Ensemble learning via negative correlation,'' *Neural Networks*, 12(10):1399-1404, December 1999.

Learning Is Inherently Multi-objective

- 1/Error = Accuracy + λ Diversity
- These are in essence two separate objectives.

- In general, we have

  Loss = Accuracy + λ_1 Regularisation + λ_2 Diversity + λ_3 …

Multi-objective Learning

- Multi-objective learning treats accuracy and diversity as two separate objectives in learning.
- Multi-objective optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), can be used as learning algorithms.
- The result from such an MOEA is a non-dominated set of solutions (i.e., learners), which ideally form the ensemble we are interested in.

Multi-objective Learning

- **Multi-objective learning treats accuracy and diversity as two separate objectives in learning.**
- Multi-objective optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), can be used as learning algorithms.
- The result from such an MOEA is a non-dominated set of solutions (i.e., learners), which ideally form the ensemble we are interested in.

Multi-objective Learning

- **Multi-objective learning treats accuracy and diversity as two separate objectives in learning.**
- **Multi-objective optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), can be used as learning algorithms.**
- **The result from such an MOEA is a non-dominated set of solutions (i.e., learners), which ideally form the ensemble we are interested in.**

- **A Chandra and X. Yao, ``Ensemble learning using multi-objective evolutionary algorithms,'' *Journal of Mathematical Modelling and Algorithms*, 5(4):417-445, December 2006.**
- **A Chandra and X. Yao, ``Evolving hybrid ensembles of learning machines for better generalisation,'' *Neurocomputing*, 69(7-9):686-700, March 2006.**

Multi-objective Learning

- Multi-objective learning treats accuracy and diversity as two separate objectives in learning.
- Multi-objective optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), can be used as learning algorithms.
- The result from such an MOEA is a non-dominated set of solutions (i.e., learners), which could ideally form an ensemble we are interested in.

- A Chandra and X. Yao, ``Ensemble learning using multi-objective evolutionary algorithms,'' *Journal of Mathematical Modelling and Algorithms*, 5(4):417-445, December 2006.
- A Chandra and X. Yao, ``Evolving hybrid ensembles of learning machines for better generalisation,'' *Neurocomputing*, 69(7-9):686-700, March 2006.

Flexibility and Generality

- For example, we can include an additional regularisation term, as an additional objective, in learning.
  - H. Chen and X. Yao, “Multiobjective Neural Network Ensembles based on Regularized Negative Correlation Learning,” *IEEE Transactions on Knowledge and Data Engineering*, 22(12):1738-1751, December 2010.
- Multi-objective learning can be used in online mode as well:
  - Z. Gong, H. Chen, B. Yuan and X. Yao, "Multiobjective Learning in the Model Space for Time Series Classification," *IEEE Transactions on Cybernetics*, 49(3):918-932, March 2019.

Flexibility and Generality

- **For example, we can include an additional regularisation term, as an additional objective, in learning.**
  - **H. Chen and X. Yao, "Multiobjective Neural Network Ensembles based on Regularized Negative Correlation Learning,"** ***IEEE Transactions on Knowledge and Data Engineering*****, 22(12):1738-1751, December 2010.**
- **Multi-objective learning can be used in online mode as well:**
  - **Z. Gong, H. Chen, B. Yuan and X. Yao, "Multiobjective Learning in the Model Space for Time Series Classification,"** ***IEEE Transactions on Cybernetics*****, 49(3):918-932, March 2019.**

Flexibility and Generality

- **For example, we can include an additional regularisation term, as an additional objective, in learning.**
  - H. Chen and X. Yao, “Multiobjective Neural Network Ensembles based on Regularized Negative Correlation Learning,” *IEEE Transactions on Knowledge and Data Engineering*, 22(12):1738-1751, December 2010.
- **Multi-objective learning can be used in online mode as well:**
  - Z. Gong, H. Chen, B. Yuan and X. Yao, "Multiobjective Learning in the Model Space for Time Series Classification," *IEEE Transactions on Cybernetics*, 49(3):918-932, March 2019.

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Class Imbalance Learning

- Class imbalance learning refers to learning from imbalanced data sets, in which some classes of examples (minority) are highly under-represented comparing to other classes (majority).
- Learning difficulty:
  - poor generalization on the minority class.
  - Learning objective: obtaining a classifier that will provide high accuracy for the minority class without severely jeopardizing the accuracy of the majority class.

Class Imbalance Learning

- **Class imbalance learning refers to learning from imbalanced data sets, in which some classes of examples (minority) are highly under-represented compared to other classes (majority).**
- **Learning difficulty:**
  - **poor generalization on the minority class.**
  - **Learning objective: obtaining a classifier that will provide high accuracy for the minority class without severely jeopardizing the accuracy of the majority class.**

Class Imbalance Learning

- **Class imbalance learning refers to learning from imbalanced data sets, in which some classes of examples (minority) are highly under-represented compared to other classes (majority).**
- **Learning difficulty:**
  - **poor generalization on minority classes.**
  - **<u>Learning objective</u>: obtaining a classifier that will provide high accuracy for minority classed without severely jeopardizing the accuracy of majority classes.**

S. Wang and X. Yao, ``Multi-Class Imbalance Problems: Analysis and Potential Solutions,'' *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 42(4):1119-1130, August 2012.

Multi-objective Class Imbalance Learning

- We can treat single class performances as separate objectives.
  - S. Wang, L. Minku and X. Yao, "A multi-objective ensemble method for online class imbalance learning," *Proc. of the 2014 International Joint Conference on Neural Networks (IJCNN'14)*, pp.3311-3318, IEEE Press, July 2014.
- We can use multi-objective algorithms for data sampling in class imbalance learning.
  - E. Fernandes, A. {de Carvalho} and X. Yao, ``Ensemble of Classifiers based on MultiObjective Genetic Sampling for Imbalanced Data,'' *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1104-1115, June 2020.
- We can of course use different evolutionary algorithms.
  - U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Evolving Diverse Ensembles using Genetic Programming for Classification with Unbalanced Data,'' *IEEE Transactions on Evolutionary Computation*, 17(3):368-386, June 2013.

Multi-objective Class Imbalance Learning

- **We can treat single class performances as separate objectives.**
  - S. Wang, L. Minku and X. Yao, "A multi-objective ensemble method for online class imbalance learning," *Proc. of the 2014 International Joint Conference on Neural Networks (IJCNN'14)*, pp.3311-3318, IEEE Press, July 2014.
- **We can use multi-objective algorithms for data sampling in class imbalance learning.**
  - E. Fernandes, A. {de Carvalho} and X. Yao, ``Ensemble of Classifiers based on MultiObjective Genetic Sampling for Imbalanced Data,'' *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1104-1115, June 2020.
- **We can of course use different evolutionary algorithms.**
  - U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Evolving Diverse Ensembles using Genetic Programming for Classification with Unbalanced Data,'' *IEEE Transactions on Evolutionary Computation*, 17(3):368-386, June 2013.

Multi-objective Class Imbalance Learning

- **We can treat single class performances as separate objectives.**
  - **S. Wang, L. Minku and X. Yao, "A multi-objective ensemble method for online class imbalance learning," *Proc. of the 2014 International Joint Conference on Neural Networks (IJCNN'14)*, pp.3311-3318, IEEE Press, July 2014.**
- **We can use multi-objective algorithms for data sampling in class imbalance learning.**
  - **E. Fernandes, A. {de Carvalho} and X. Yao, ``Ensemble of Classifiers based on MultiObjective Genetic Sampling for Imbalanced Data,'' *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1104-1115, June 2020.**
- **We can of course use different evolutionary algorithms.**
  - **U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Evolving Diverse Ensembles using Genetic Programming for Classification with Unbalanced Data,'' *IEEE Transactions on Evolutionary Computation*, 17(3):368-386, June 2013.**

Multi-objective Class Imbalance Learning

- **We can treat single class performances as separate objectives.**
  - **S. Wang, L. Minku and X. Yao, "A multi-objective ensemble method for online class imbalance learning,"** ***Proc. of the 2014 International Joint Conference on Neural Networks (IJCNN'14)*****, pp.3311-3318, IEEE Press, July 2014.**
- **We can use multi-objective algorithms for data sampling in class imbalance learning.**
  - **E. Fernandes, A. {de Carvalho} and X. Yao, ``Ensemble of Classifiers based on MultiObjective Genetic Sampling for Imbalanced Data,''** ***IEEE Transactions on Knowledge and Data Engineering*****, 32(6):1104-1115, June 2020.**
- **We can use different evolutionary algorithms.**
  - **U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Evolving Diverse Ensembles using Genetic Programming for Classification with Unbalanced Data,''** ***IEEE Transactions on Evolutionary Computation*****, 17(3):368-386, June 2013.**

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Software Effort Estimation (SEE)

Problem description:

- **Estimation of the effort required to develop a software project (e.g., in person-hours), based on features such as functional size (numerical), required reliability (ordinal), programming language (categorical), development type (categorical), team expertise (ordinal), etc.**

Existing Learning Approaches to SEE

- Uses historical projects as training examples for creating SEE models, e.g.,
  - Multilayer Perceptrons (MLPs)
  - Radial Basis Function networks (RBFs)
  - Regression Trees (RTs)
- Learned models can then be used as decision support tools.

Different Performance Measures Were Used

Mean Magnitude of the Relative Error:

$$MMRE = \frac{1}{T}\sum_{i=1}^{T} MRE_i,$$

where $MRE_i = |\hat{y}_i - y_i|/y_i$; $\hat{y}_i$ is the predicted effort; and y_i is the actual effort.

Percentage of estimations within 25% of the actual values:

$$PRED(25) = \frac{1}{T}\sum_{i=1}^{T} \begin{cases} 1, & \text{if } MRE_i \le \frac{25}{100} \\ 0, & \text{otherwise} \end{cases}.$$

Logarithmic Standard Deviation:

$$LSD = \sqrt{\frac{\sum_{i=1}^{T}\left(e_i + \frac{s^2}{2}\right)^2}{T-1}},$$

where s^2 is an estimator of the variance of the residual e_i and $e_i = \ln y_i - \ln \hat{y}_i$.

Challenges

- **There is no universally agreed single performance measure.**
- **The relationship among different measures in SEE is not well understood.**
- **Existing SEE approaches use at most one measure during the learning procedure.** ***It is unclear whether a model/learner trained using one measure would still perform well under a different measure.***
- **Many papers did not even report the measure they used in training!**

Challenges

- **There is no universally agreed single performance measure.**
- **The relationship among different measures in SEE is not well understood.**
- **Existing SEE approaches use at most one measure during the learning procedure.** ***It is unclear whether a model/learner trained using one measure would still perform well under a different measure.***
- **Many papers did not even report the measure they used in training!**

Challenges

- **There is no universally agreed single performance measure.**
- **The relationship among different measures in SEE is not well understood.**
- **Existing SEE approaches use at most one measure during the learning procedure. *It is unclear whether a model/learner trained using one measure would still perform well under a different measure.***
- **Many papers did not even report the measure they used in training!**

Challenges

- **There is no universally agreed single performance measure.**
- **The relationship among different measures in SEE is not well understood.**
- **Existing SEE approaches use at most one measure during the learning stage. *It is unclear whether a model/learner trained using one measure would still perform well under a different measure.***
- **Many papers did not even report the measure they used in training!**

Challenges

- **There is no universally agreed single performance measure.**
- **The relationship among different measures in SEE is not well understood.**
- **Existing SEE approaches use at most one measure during the learning stage. *It is unclear whether a model/learner trained using one measure would still perform well under a different measure.***
- **Some papers did not even report the measure they used in training!**

SEE by Multi-objective Learning

- **We proposed to formulate SEE as a multi-objective learning problem.**

- **Each performance measure is considered explicitly as a separate objective in learning.**

Research Questions

1. **What is the relationship among different performance measures for SEE?**

2. **Can we use different performance measures as a source of diversity to create better SEE models?**

L. L. Minku and X. Yao, "Software Effort Estimation as a Multi-objective Learning Problem," ***ACM Transactions on Software Engineering and Methodology*****, 22(4), Article No. 35, October 2013, 32 pages.**

Relationship among Different Performance Measures (RQ1)

Interesting insight:

1) **MMRE, PRED(25) and LSD behaved more differently than one might have initially thought.**

2) **MMRE and LSD can present even opposite behaviours.**

Multi-objective Ensemble Learning (RQ2)

- **Multi-objective ensemble learning does improve the performance of single objective learning.**
- **The use of different measures as separate objectives helped to increase the diversity in the ensemble and improve ensemble learning performance.**
- **The ensembles did well even on those performance measures that were *not* used in multi-objective learning.**

Multi-objective Ensemble Learning (RQ2)

- **Multi-objective ensemble learning does improve the performance of single objective learning.**
- **The use of different measures as separate objectives helped to increase the diversity in the ensemble and improve ensemble learning performance.**
- **The ensembles did well even on those performance measures that were *not* used in multi-objective learning.**

Multi-objective Ensemble Learning (RQ2)

- **Multi-objective ensemble learning does improve the performance of single objective learning.**
- **The use of different measures as separate objectives helped to increase the diversity in the ensemble and improve ensemble performance.**
- **The ensembles did well even on those performance measures that were *not* used in multi-objective learning.**

Multi-objective Ensemble Learning (RQ2)

- **Multi-objective ensemble learning does improve the performance of single objective learning.**
- **The use of different measures as separate objectives helped to increase the diversity in the ensemble and improve ensemble performance.**
- **The ensembles performed well even according to those measures that were *not* used in multi-objective learning.**

Pareto Ensemble vs. Other Approaches

Number of times ranked best:

Approach	LSD	MMRE	PRED(25)	MdMRE	MAE	MdAE
Pareto Ens	1	6	5	5	5	7
RT	4	3	2	3	2	2
Bag+RT	5	0	1	0	2	1
Bag+MLP	0	2	2	2	1	1
Log + EBA	0	2	2	2	1	1
Bag+RBF	3	0	1	0	2	0
Rand+MLP	0	0	0	1	0	0
RBF	0	0	0	0	0	1
Total	13	13	13	13	13	13

- **Pareto ensemble was most often ranked first, except for LSD.**
- **It is more robust, even according to measures that were not used in multi-objective learning.**

Overview

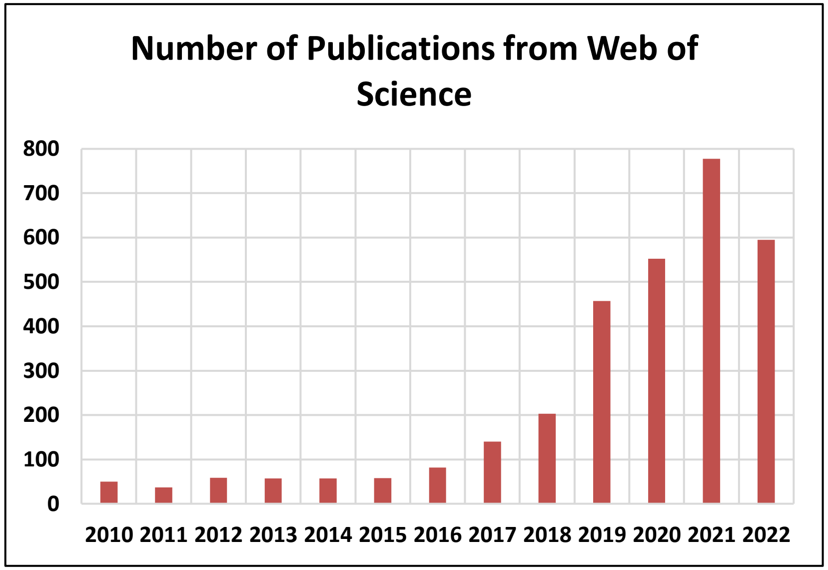
- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Different Dimensions of Trustworthiness

Some Principles	Selected Keywords
Transparency	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Fairness and Justice	Justice, fairness, consistency, inclusion, equality, equity, (non-) bias, (non-) discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Responsibility	Responsibility, accountability, liability, acting with integrity
Non-maleficence	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion, reliability, robustness
Privacy	Privacy, personal or private information, data protection,
Beneficence	Benefits, beneficence, well-being, peace, social good, common good, non-violence
Freedom and Autonomy	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment, human rights
Solidarity	Solidarity, social security, cohesion, inclusion, inclusiveness
Sustainability	Sustainability, environment, nature, energy, resources
Trust	Trust, trustworthiness, trustworthy
Dignity	Dignity

Fair Machine Learning

1. I. Zliobaite, “A survey on measuring indirect discrimination in machine learning,” arXiv preprint arXiv:1511.00148, 2015.
2. I. Zliobaitė, “Measuring discrimination in algorithmic decision making,” Data Mining and Knowledge Discovery, vol. 31, no. 4, pp. 1060–1089, 2017.
3. S. Corbett-Davies and S. Goel, “The measure and mismeasure of fairness: A critical review of fair machine learning,” arXiv preprint arXiv:1808.00023, 2018.
4. S. Verma and J. Rubin, “Fairness definitions explained,” in 2018 IEEE/ACM International Workshop on Software Fairness (FairWare), IEEE, pp. 1–7, 2018.
5. B. Hutchinson and M. Mitchell, “50 Years of Test (Un)fairness,” Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 49–58, 2019.
6. S. Caton and C. Haas, “Fairness in machine learning: A survey,” arXiv preprint arXiv:2010.04053, 2020.
7. N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A Survey on Bias and Fairness in Machine Learning,” ACM Computing Surveys, vol. 54, no. 6, pp. 1–35, 2021.
8. M. Du, F. Yang, N. Zou, and X. Hu, “Fairness in Deep Learning: A Computational Perspective,” IEEE Intelligent Systems, vol. 36, no. 4, pp. 25–34, 2021.
9. D. Pessach, E. Shmueli, “A Review on Fairness in Machine Learning,” ACM Computing Surveys, vol. 55, no. 3, pp. 1-44, 2022.
10. M. Wan, D. Zha, N. Liu, and N. Zou, “In-Processing Modeling Techniques for Machine Learning Fairness: A Survey,” ACM Transactions on Knowledge Discovery from Data, 2022, doi: 10.1145/3551390.



Search on Web of Science by using keys of “fairness and bias in artificial intelligence” or “algorithmic bias” or “algorithmic fairness” or “fairness-aware machine learning” or “fairness in machine learning”. (Accessed by Nov. 3, 2022)

How to Measure Fairness?

There are <u>many</u> （20+） fairness metrics proposed so far, which can be divided into 5 main categories [1]：

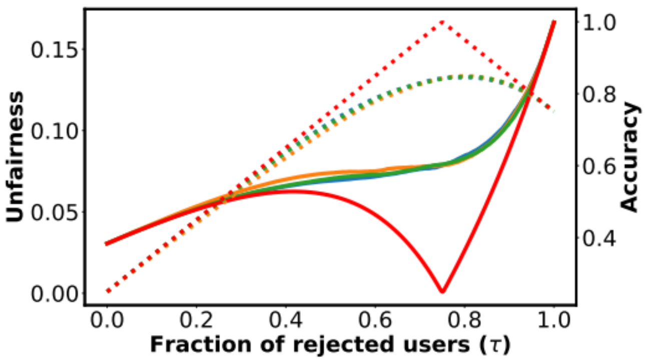
1. **Metrics based on predicted outcome**
2. **Metrics based on predicted and actual outcomes**
3. **Metrics based on predicted probabilities and actual outcome**
4. **Metrics based on similarity**
5. **Metrics based on causal reasoning**

[1] S. Verma and J. Rubin, "Fairness definitions explained," in *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, IEEE, 2018, pp. 1–7.

Two Challenges

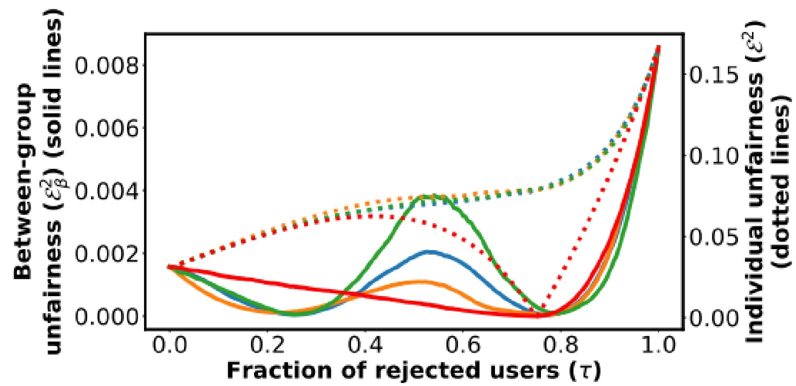
There are <u>inherent conflicts</u> (1) between model accuracy and fairness, and (2) among different fairness metrics.

(a) Conflict between accuracy and fairness



dotted line: unfairness
solid line: accuracy

(b) Conflict among multiple fairness metrics



dotted line: individual unfairness
solid line: group unfairness

Image source: Figure 4 from [17] T. Speicher, H. Heidari, N. Grgic-Hlaca, K. P. Gummadi, A. Singla, A. Weller, and M. B. Zafarl, "A unified approach to quantifying algorithmic unfairness: Measuring individual & group unfairness via inequality indices," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2239–2248

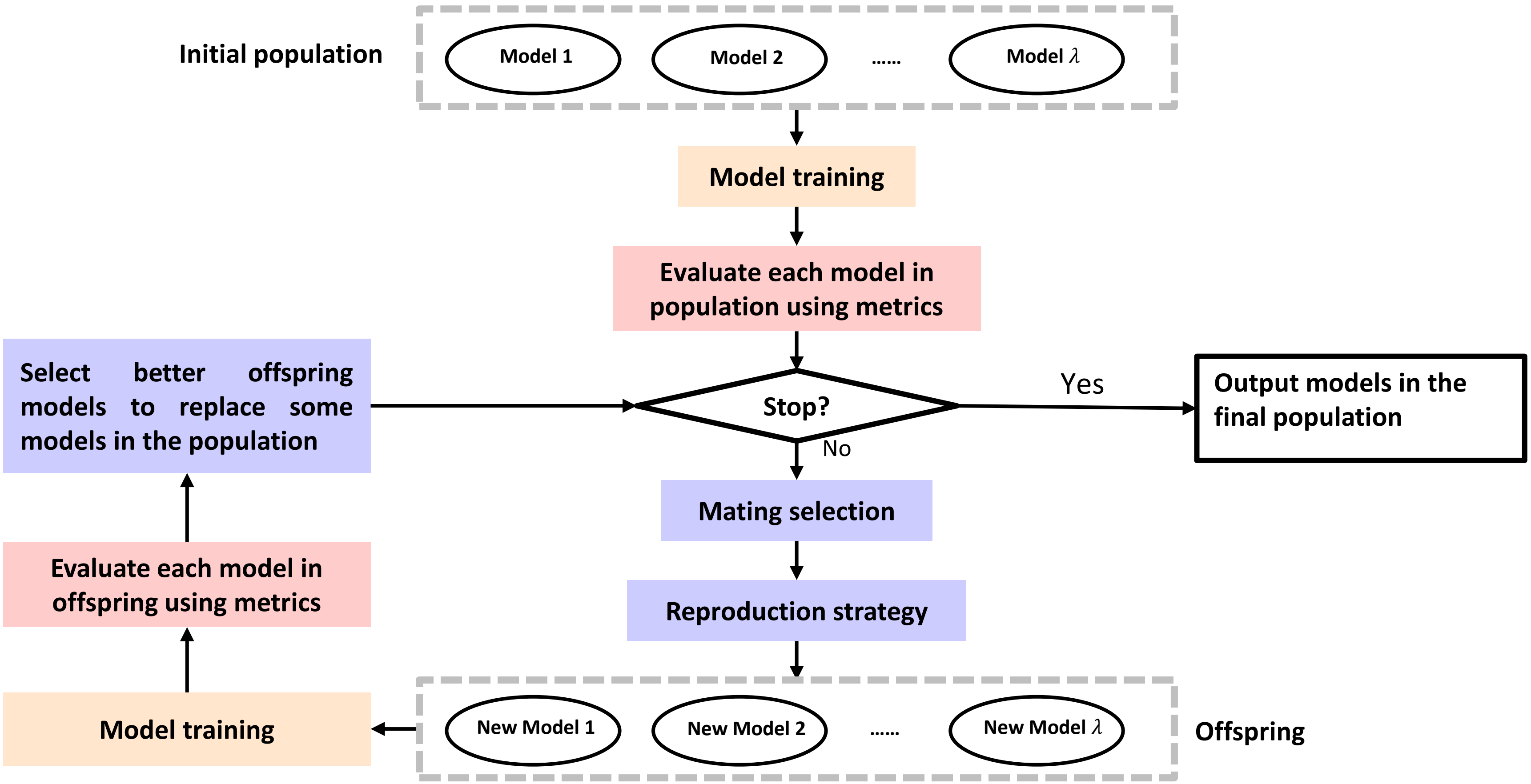
Multi-objective Learning Approach to Fairer Machine Learning

Can a machine learning model be made fairer by simultaneously considering accuracy and *multiple* fairness metrics?

Yes, we can use a multi-objective learning approach to improve fairness.

Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

Multi-objective Evolutionary Learning



Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

Findings from Our Experimental Studies

- Our algorithm can simultaneously optimize several fairness measures without sacrificing accuracy.
- It can even improve fairness measures that are <u>not</u> used in model training.
- It can generate an ensemble model combined from evolved models to balance accuracy and multiple fairness measures.

Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

Findings from Our Experimental Studies

- **Our algorithm can simultaneously optimize several fairness measures without sacrificing accuracy.**
- **It can even improve fairness measures that are <u>not</u> used in model training.**
- **It can generate an ensemble model combined from evolved models to balance accuracy and multiple fairness measures.**

Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

Findings from Our Experimental Studies

- **Our algorithm can simultaneously optimize several fairness measures without sacrificing accuracy.**
- **It can even improve fairness measures that are <u>not</u> used in model training.**
- **It can generate an ensemble model combined from evolved models to balance accuracy and multiple fairness measures.**

Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

Findings from Our Experimental Studies

- **Our algorithm can simultaneously optimize several fairness measures without sacrificing accuracy.**
- **It can even improve fairness measures that are <u>not</u> used in model training.**
- **It can generate an ensemble model combined from evolved individuals to better balance accuracy and multiple fairness measures.**

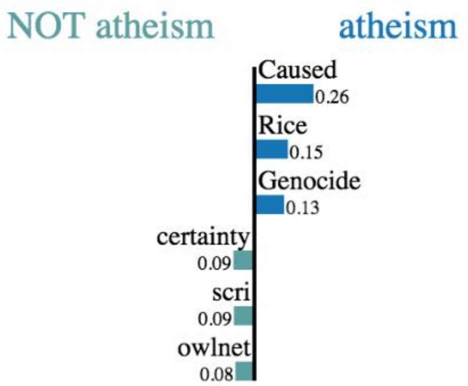
Q. Zhang, J. Liu, Z. Zhang, J. Wen, B. Mao and X. Yao, "Mitigating Unfairness via Evolutionary Multi-objective Ensemble Learning," in *IEEE Transactions on Evolutionary Computation*, 2022, doi: 10.1109/TEVC.2022.3209544.

From Fairness to Trustworthiness

- **In fact, multi-objective learning can be used to tackle other trustworthy AI issues, such as model explainability.**

What Is Feature Attribution Explanation (FAE)

Local **feature attribution explanation (FAE) describes how much each input feature contributes to the output of the model for a given data point.**



FAE explains a tabular data point [1]

FAE explains an image data point [2]

[1] Ribeiro M T , Singh S , Guestrin C . "Why Should I Trust You?": Explaining the Predictions of Any Classifier[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 2016.
[2]Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]//Proceedings of the 31st international conference on neural information processing systems. 2017: 4768-4777.

Evaluating FAE Methods

Various evaluation metrics have been proposed to assess the explanation quality of FAE methods, such as [1]:

- **Faithfulness:** $u_F(f, g; \boldsymbol{x}) = \underset{S \in \binom{[d]}{|S|}}{corr} \left(\sum_{i \in S} g(f, \boldsymbol{x})_i, f(\boldsymbol{x}) - f(\boldsymbol{x}_{[\boldsymbol{x}_s = \bar{\boldsymbol{x}}_s]})\right)$
- **Sensitivity:** $u_A(f, g, \boldsymbol{x}) = \frac{1}{|N_r|} \sum_{\boldsymbol{z} \in N_r} \frac{D\left(g(f,\boldsymbol{x}), g(f,\boldsymbol{z})\right)}{\rho(\boldsymbol{x},\boldsymbol{z})}$
- **Complexity:** $u_C(f, g; \boldsymbol{x}) = -\sum_{i=1}^{d} \frac{|g(f,\boldsymbol{x})_i|}{\sum_{j \in [d]} |g(f,\boldsymbol{x})_j|} \ln\left(\frac{|g(f,\boldsymbol{x})_i|}{\sum_{j \in [d]} |g(f,\boldsymbol{x})_j|}\right)$

However, none of the existing FAE methods actually used multiple metrics at the same time.

[1] Bhatt U , Weller A , Moura J . Evaluating and Aggregating Feature-based Model Explanations[C]// Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20. 2020.

Our Proposal: Multi-Objective FAE

■ Problem Formulation

The multi-objective explanation problem can be defined as:

$$\underset{\boldsymbol{g}(\boldsymbol{f},\boldsymbol{x})}{\textbf{Maximize}}\, \boldsymbol{u}(\boldsymbol{g};\boldsymbol{f},\boldsymbol{x}) = [\boldsymbol{u_1}(\boldsymbol{g};\boldsymbol{f},\boldsymbol{x}),\dots,\boldsymbol{u_k}(\boldsymbol{g};\boldsymbol{f},\boldsymbol{x})]$$

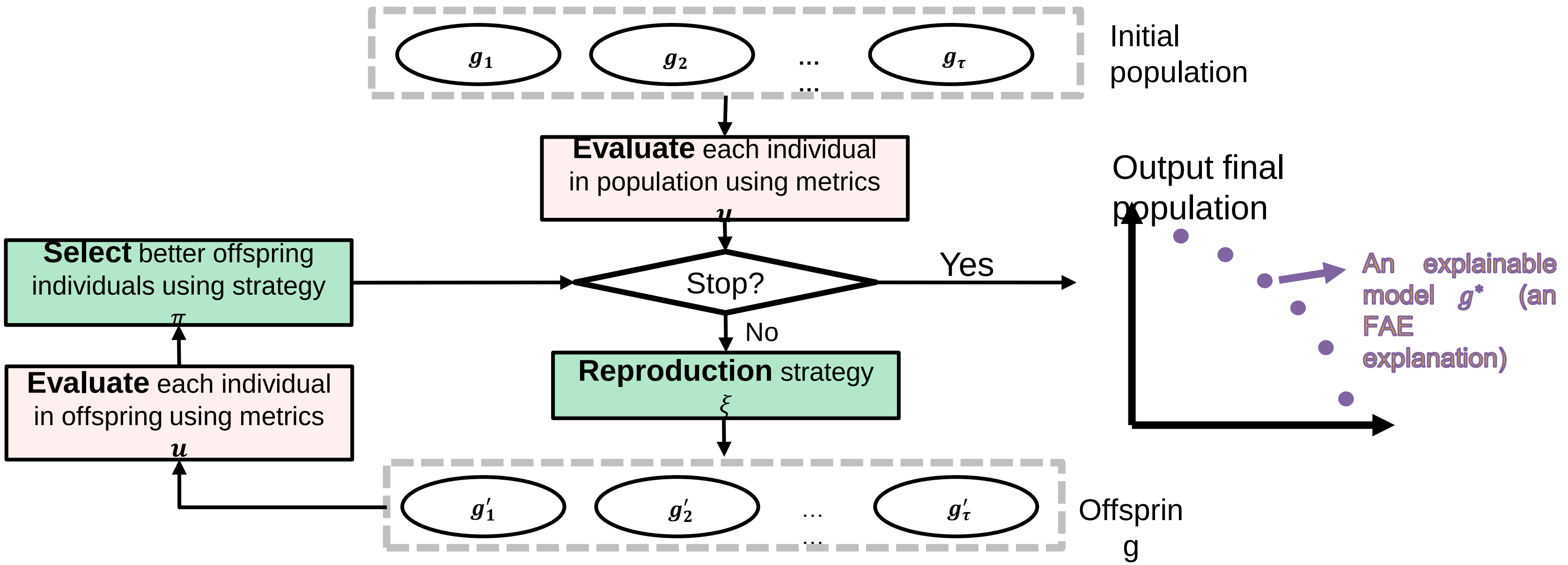
- $\boldsymbol{f}$ **is the black-box model**
- $\boldsymbol{x}$ **is the explained data point**
- $\boldsymbol{g}$ **is a local FAE function**
- $\boldsymbol{u_i}(\cdot)$ $\textbf{for}$ $\boldsymbol{i} = \boldsymbol{1},\dots,\boldsymbol{k}$ **is one of the explanation quality metrics**
- $\boldsymbol{g}(\boldsymbol{f},\boldsymbol{x}) \in \mathbb{R}^{\boldsymbol{d}}$ **is the explanation result obtained after optimizing metrics** $\boldsymbol{u}(\cdot)$

Z. Wang, C. Huang, Y. Li and X. Yao, "Multi-objective Feature Attribution Explanation For Explainable Machine Learning," ACM Transactions on Evolutionary Learning and Optimization, Accepted, August 2023.

Multi-Objective FAE Framework

Input: A trained model f, an explained data point x, an FAE g, a set of explanation quality evaluation metrics $u = \{u_1, \dots, u_k\}$, selection and reproduction strategies π and ξ, respectively.

Output: Set of explainable models $G^* = \{g_1^*, \dots, g_\tau^*\}$.



Multi-Objective FAE Algorithm

■ Framework Instantiation: A Specific Algorithm

Optimization objectives [1]: faithfulness u_F, sensitivity u_A, complexity u_C

Black-box model: artificial neural network (ANN)

Multi-objective optimizer: non-dominated sorting genetic algorithm-III (NSGA-III)

[1] Bhatt U , Weller A , Moura J . Evaluating and Aggregating Feature-based Model Explanations[C]// Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20. 2020.

Experimental Studies: Experimental Setup

<u>Dataset</u>: Adult, Iris, German, Wine, Mushroom, Car, Bank, and Glass

<u>Compared FAE methods</u>: LIME, SHAP, Grad, IG, GI, SG

<u>Crossover and mutation probabilities</u>: 1.0 and $\frac{1}{n}$, respectively

Z. Wang, C. Huang, Y. Li and X. Yao, "Multi-objective Feature Attribution Explanation For Explainable Machine Learning," ACM Transactions on Evolutionary Learning and Optimization, Accepted, August 2023.

Experimental Studies: Research Questions

Q1. Do faithfulness, sensitivity, and complexity conflict with each other?

Q2. Can MOFAE simultaneously optimize these conflicting metrics and be competitive against existing state-of-the-art FAE methods?

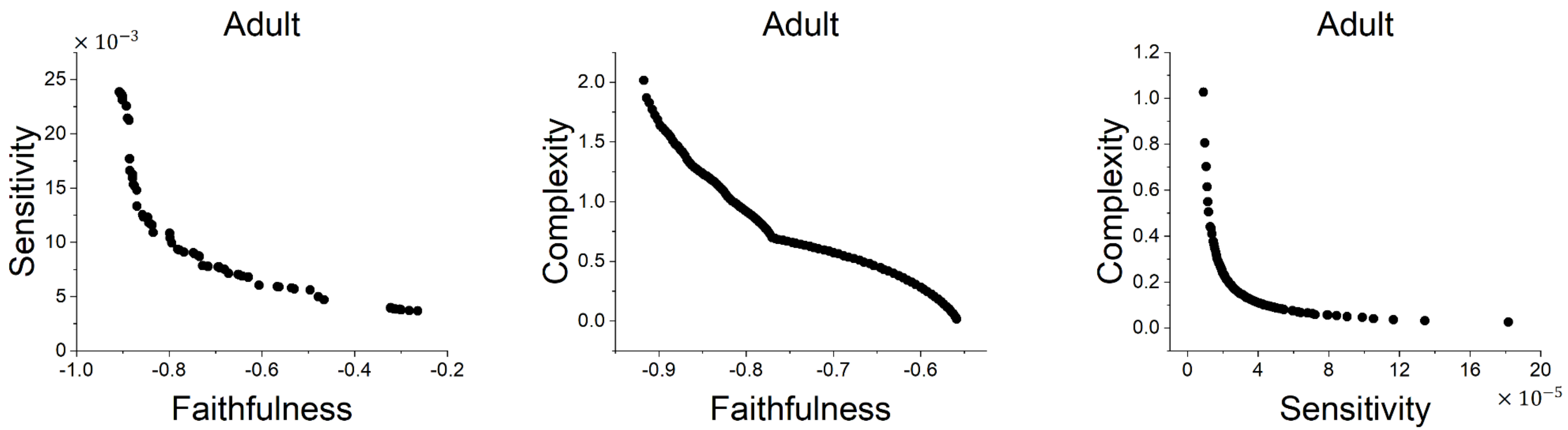
Q3. Can our method find a set of explainable models (i.e., explanations) with different trade-offs among the objectives (i.e., metrics)?

Z. Wang, C. Huang, Y. Li and X. Yao, "Multi-objective Feature Attribution Explanation For Explainable Machine Learning," ACM Transactions on Evolutionary Learning and Optimization, Accepted, August 2023.

Experimental Results (Q1)

Q1. Do faithfulness, sensitivity, and complexity conflict with each other?

We optimized every two objectives



Using the Adult dataset as an example, the three objectives indeed conflict with each other.

Z. Wang, C. Huang, Y. Li and X. Yao, "Multi-objective Feature Attribution Explanation For Explainable Machine Learning," ACM Transactions on Evolutionary Learning and Optimization, Accepted, August 2023.

Experimental Results (Q2)

Q2. Can MOFAE simultaneously optimize these conflicting metrics and be competitive with other state-of-the-art FAE methods?

Using the Adult dataset as an example.

From Fig.1, we can conclude that better overall explanations can be generated as the optimization progress.

From Fig.2 and Table 1, we can see that:

1. **No MOFAE solutions were dominated by others.**
2. **Many solutions of MOFAE can dominate others.**
3. **Some solutions of MOFAE and other methods are non-dominated.**

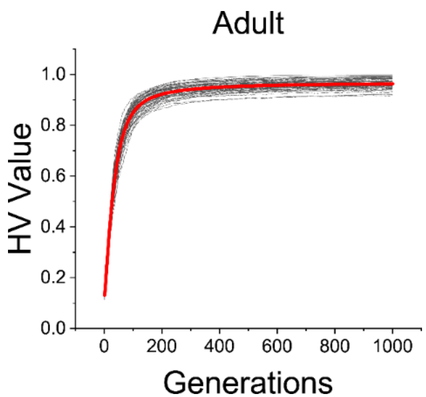


Fig. 1. HV Convergence curve

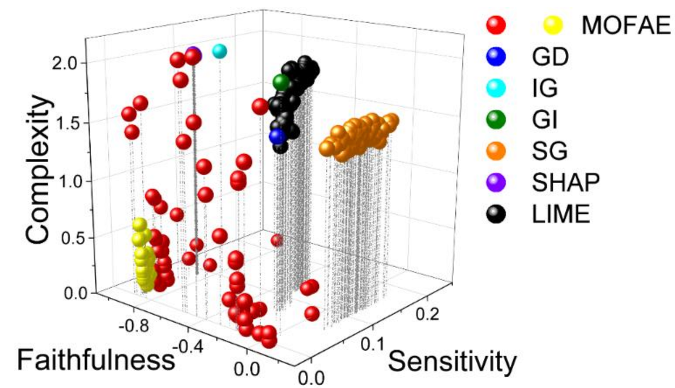


Fig. 2. Comparing methods through 3D plot

Table 1. Dominate relationship between MOFAE and other FAE methods

DataSet	Relationship	GD	IG	GI	SG	SHAP	LIME
Adult	Dominate	73.38%	9.43 %	61.39%	81.52%	52.50%	75.80%
	Non-dominate	26.62%	90.57%	38.61%	18.48%	47.50%	24.20%
	Be dominated	0 %	0 %	0 %	0 %	0 %	0 %

Experimental Results (Q3)

Q3. Can our method find a set of explanations with different trade-offs among the objectives (i.e., metrics)?

As can be seen from Table 2, the diversity of MOFAE is significantly better than other FAE methods.

* Not compared with Grad, IG, GI because these three methods do not have any diversity.

Table 2. Comparison of the diversity of MOFAE and other FAE methods

Indicator	DataSet	MOFAE	SG	SHAP	LIME
MS	Adult	**1.9345**	0.2717	0.0705	0.6006
	Iris	**1.6477**	0.5328	0	1.1175
	German	**2.4647**	0.0510	0.0765	0.2547
	Mushroom	**1.7467**	0.2353	0.2387	0.3792
	Wine	**2.1523**	0.6306	0.0610	0.4456
	Car	**1.6205**	0.6830	0	0.1996
	Bank	**1.9964**	0.5406	0.1312	0.2547
	Glass	**2.0692**	0.4555	0	0.3281
PD	Adult	**46472**	18000	1688.1	22175
	Iris	**34849**	10914	0	33037
	German	**95416**	3401.9	985.51	14679
	Mushroom	12924	11530	5726	**21260**
	Wine	**55594**	17551	2700.2	22889
	Car	40554	**47117**	0	10777
	Bank	**44373**	32286	2400.4	12602
	Glass	**61630**	28956	0	19327

Z. Wang, C. Huang, Y. Li and X. Yao, "Multi-objective Feature Attribution Explanation For Explainable Machine Learning," ACM Transactions on Evolutionary Learning and Optimization, Accepted, August 2023.

Overview

- Introduction to ensemble learning
- Multi-objective learning and ensembles
- Multi-objective class imbalance learning
- Multi-objective software effort estimation
- Multi-objective approach to trustworthy AI
- Concluding remarks

Concluding Remarks

- **Ensemble learning can be effectively combined with multi-objective learning.**
- **Many real-world learning problems are inherently multi-objective.**
  - **Accuracy**
  - **Regularisation**
  - **Trustworthiness**
  - **…**
- **More research into multi-objective ensemble learning is needed.**

Other Multi-objective Learning Problems

- L. Li, X. Yao, R. Stolkin, M. Gong and S. He ``An Evolutionary Multi-objective Approach to Sparse Reconstruction,'' *IEEE Transactions on Evolutionary Computation*, 18(6):827-845, December 2014.
- P. Wang, M. Emmerich, R. Li, K. Tang, T. Baeck and X. Yao, ``Convex Hull-Based Multi-objective Genetic Programming for Maximizing Receiver Operating Characteristic Performance,'' *IEEE Transactions on Evolutionary Computation*, 19(2):188-200, April 2015.
- P. Wang, K. Tang, T. Weise, E. P. K. Tsang and X. Yao, ``Multiobjective Genetic Programming for Maximizing ROC Performance,'' *Neurocomputing*, 125:102-118, 11 February 2014.

Thank you!

Ensemble Member Selection

- **Sometimes it is unnecessary to include the entire set of classifiers found by MOEAs in an ensemble. A subset would be sufficient, or even better.**
  - **X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(3):417-425, June 1998.**
- **There are various methods in the literature for selecting a diverse subset of classifiers from a large set, e.g.,**
  - **U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Reusing Genetic Programming for Ensemble Selection in Classification of Unbalanced Data,'' *IEEE Transactions on Evolutionary Computation*, 18(6):893-908, December 2014.**

Tackling Many Objectives

Xin Yao
Department of Computer Science and Engineering, and
Research Institute of Trustworthy Autonomous Systems
Southern University of Science and Technology (SUSTech)
Shenzhen, China
and
School of Computer Science, University of Birmingham, UK

Outline

- Introduction
- Objective Reduction
- Alternative Dominance Relationship
- Improved Two-Archive Algorithm (Two_Arch2)
- Conclusions and Future Work

What Is Multi-objective Optimisation?

More than one objective to be optimized simultaneously.

$$
\begin{aligned}
&\text{min/max} \;\; f_m(\boldsymbol{x}), && m = 1, 2, \cdots, M \\
&\text{subject to } g_j(\boldsymbol{x}) \geq 0, && j = 1, 2, \cdots, J \\
&\qquad\qquad\quad h_k(\boldsymbol{x}) = 0, && k = 1, 2, \cdots, K \\
&\qquad\qquad\quad \underset{\substack{\text{lower}\\\text{bound}}}{x_i^{(L)}} \leq x_i \leq \underset{\substack{\text{upper}\\\text{bound}}}{x_i^{(U)}}, && i = 1, 2, \cdots, n
\end{aligned}
$$

Multi-objective Evolutionary Algorithms

- Evolutionary algorithms (EAs) have been widely used in the last 20 years for multi-objective optimisation.
- They can provide a set of non-dominated solutions simultaneously in a single run.
- They do not require the objective functions and constraints to be convex, smooth, or even continuous.
- They can deal with uncertainty and dynamics better than other alternatives.

Widely Used?

- **Deb, K., Pratap. A, Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II."** ***IEEE Transaction on Evolutionary Computation*****, 6(2), 181-197. (29,491 Google Scholar citations)**

Unfortunately

- **NSGA-II and other early MOEAs work well only with two or three objectives.**
- **They do not work well when the number of objectives goes beyond three.**
- **There is a scalability issue in terms of the number of objectives.**
  - V. Khare, X. Yao and K. Deb, "Performance Scaling of Multi-objective Evolutionary Algorithms," In *Proc. of the 2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO'03)*, Lecture Notes in Computer Science, Vol. 2632, Carlos M. Fonseca, Peter J. Fleming, Eckart Zitzler, Kalyanmoy Deb and Lothar Thiele (Eds.), Springer-Verlag, April 2003, pp.376-390.
- **Many Objective Optimisation (MaOP) indicates that the number of objectives is greater than three.**

Problem-solving Strategies

1. Develop more sophisticated solution approaches to more complex problems.
2. Simplify a complex problem so that an existing solution approach can be applied.

Can we simplify MaOPs?

Can we reduce the number of objectives?

Outline

- Introduction
- Objective Reduction
- Alternative Dominance Relationship
- Improved Two-Archive Algorithm (Two_Arch2)
- Conclusions and Future Work

Objective Reduction

- If two objectives are positively correlated, we only need to optimise one of them.
- There are many methods that could be used to reduce the number of objectives.
- We present one example here.

<u>Nonlinear</u> Correlation Information Entropy (NCIE)

- **NCIE is an entropy measure.**
- **NCIE first divides variables *X* and *Y* into *b*b* uniform rank grids. Then, the probabilities p_{ij} can be approximated by counting the samples in those grids. In other words, p_{ij} in the *ij*-th grid can be calculated by the number of solutions in the *ij*-th grid (n_{ij}/N).**
- **Parameter *b* can be set as *N^0.5*.**

$$H^r(X) = -\sum_{i=1}^{b} \frac{n_i}{N} log_b(\frac{n_i}{N})$$

$$H^r(X,Y) = -\sum_{i=1}^{b}\sum_{j=1}^{b} \frac{n_{ij}}{N} log_b(\frac{n_{ij}}{N})$$

$$NCIE(X,Y) = H^r(X) + H^r(Y) - H^r(X,Y)$$

Objective Reduction Based on NCIE

- **Correlation analysis is based on the matrix of modified NCIE R^N of a non-dominated population.**

$$R^N = \{Sgn(cov_{ij})NCIE_{ij}\}, (1 \le i, j \le m)$$

- **Objective selection aims to choose the most conflicting objectives.**
- **Our approach is applied in every generation of MOEAs to update the correlation information among objectives.**

H. Wang and X. Yao, "Objective Reduction Based on Nonlinear Correlation Information Entropy," ***Soft Computing*****, June 2016, Volume 20, Issue 6, pp 2393–2407.**

Objective Selection: An Example

- **Select the most conflicting objective.**
- **Remove the objectives that are positively correlated to the selected objective.**

	f_1	f_2	f_3	f_4	f_5
f_1	1.0000	0.4959	0.4244	0.5348	-0.3552
f_2	0.4959	1.0000	0.3972	0.4686	-0.3381
f_3	0.4244	0.3972	1.0000	0.4765	-0.4352
f_4	0.5348	0.4686	0.4765	1.0000	-0.4488
f_5	-0.3552	-0.3381	-0.4352	-0.4488	1.0000
$\sum NCIE < 0$	-0.3552	-0.3381	-0.4352	-0.4488	-1.5773

- **f_5 is selected because it has the most conflicting degree with other objectives.**
- **There is no objective positively correlated to f_5, thus, there is no redundant objective with f_5 in the remaining objectives.**
- **f_4 is then selected. f_1 , f_2, and f_3 are removed because they are all positively correlated to f_4 .**
- **Output $\{f_5, f_4\}$.**

Objective reduction can remove redundant objectives, but what if there is no redundancy among objectives?

Outline

- Introduction
- Objective Reduction
- Alternative Dominance Relationship
- Improved Two-Archive Algorithm (Two_Arch2)
- Conclusions and Future Work

Why Are Many Objectives Hard to Handle?

- The number of non-dominated solutions increases exponentially as the number of objectives grows.
- As a result, there is no selection pressure in MaOEAs to drive the evolutionary search.
- Can we use alternative dominance relationship other than Pareto dominance in order to distinguish currently non-dominated solutions?

Θ-dominance --- Intuition

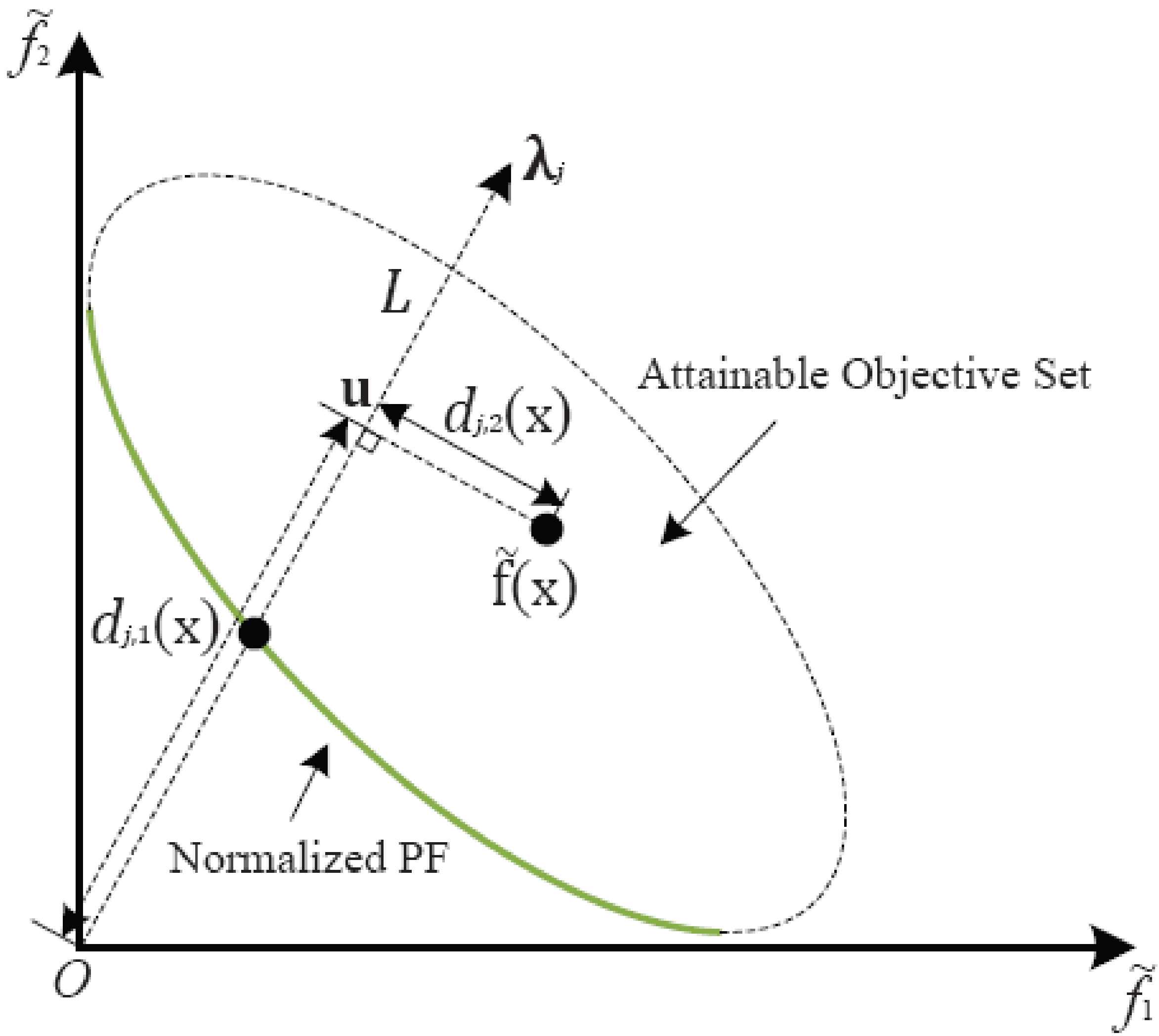


Fig. 3. Illustration of distances $d_{j,1}(\mathbf{x})$ and $d_{j,2}(\mathbf{x})$.

- ***f*'s are normalised objective functions.**
- **λ is the reference direction (point).**

Y. Yuan, H. Xu, B. Wang and X. Yao, "A New Dominance Relation Based Evolutionary Algorithm for Many-Objective Optimization," *IEEE Transactions on Evolutionary Computation*, 20(1):16-37, February 2016.

Θ-dominance --- Definition

Definition 7: Given two solutions $\mathbf{x}, \mathbf{y} \in S_t$, $\mathbf{x}$ is said to θ-dominate $\mathbf{y}$, denoted by $\mathbf{x} \prec_\theta \mathbf{y}$, iff $\mathbf{x} \in C_j$, $\mathbf{y} \in C_j$, and $\mathcal{F}_j(\mathbf{x}) < \mathcal{F}_j(\mathbf{y})$, where $j \in \{1, 2, \ldots, N\}$.

$$\mathcal{F}_j(\mathbf{x}) = \bar{d}_{j,1}(\mathbf{x}) + \theta d_{j,2}(\mathbf{x})$$

Y. Yuan, H. Xu, B. Wang and X. Yao, "A New Dominance Relation Based Evolutionary Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 20(1):16-37, February 2016.**

Balancing Convergence and Diversity

- **The form of $F_j(x)$ indicates that the balance between convergence and diversity is essential in MaOEAs.**
- **Why not manipulating the balance explicitly?**
  - **Y. Yuan, H. Xu, B. Wang, B. Zhang and X. Yao, "Balancing Convergence and Diversity in Decomposition-Based Many-Objective Optimizers,"** ***IEEE Transactions on Evolutionary Computation*****, 20(2):180-198, April 2016.**
- **Strike the balance stochastically.**
  - **B. Li, K. Tang, J. Li and X. Yao, ``Stochastic Ranking Algorithm for Many-Objective Optimization Based on Multiple Indicators,'‘** ***IEEE Transactions on Evolutionary Computation*****, 20(6):924-938, December 2016.**

What if alternative dominance relationships still do not provide a satisfactory solution to a MaOP?

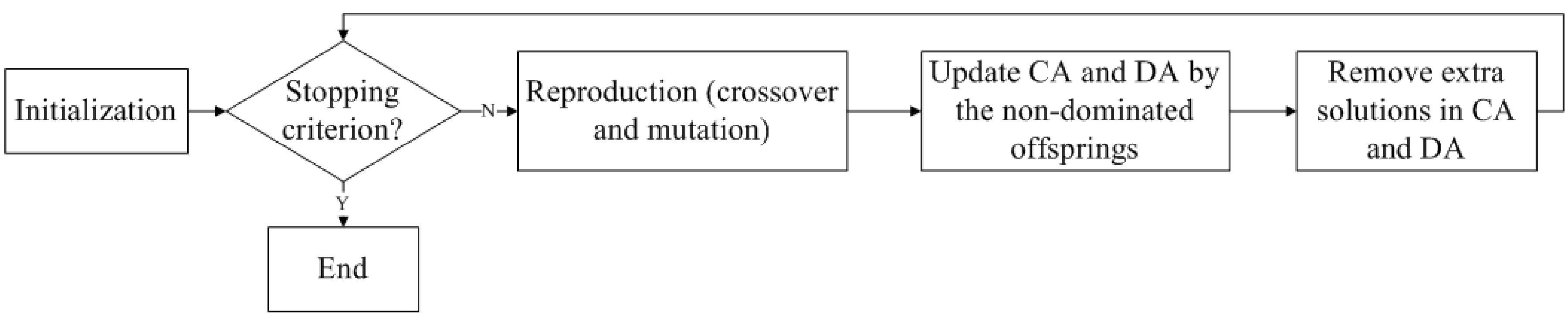
Outline

- Introduction
- Objective Reduction
- Alternative Dominance Relationship
- Improved Two-Archive Algorithm (Two_Arch2)
- Conclusions and Future Work

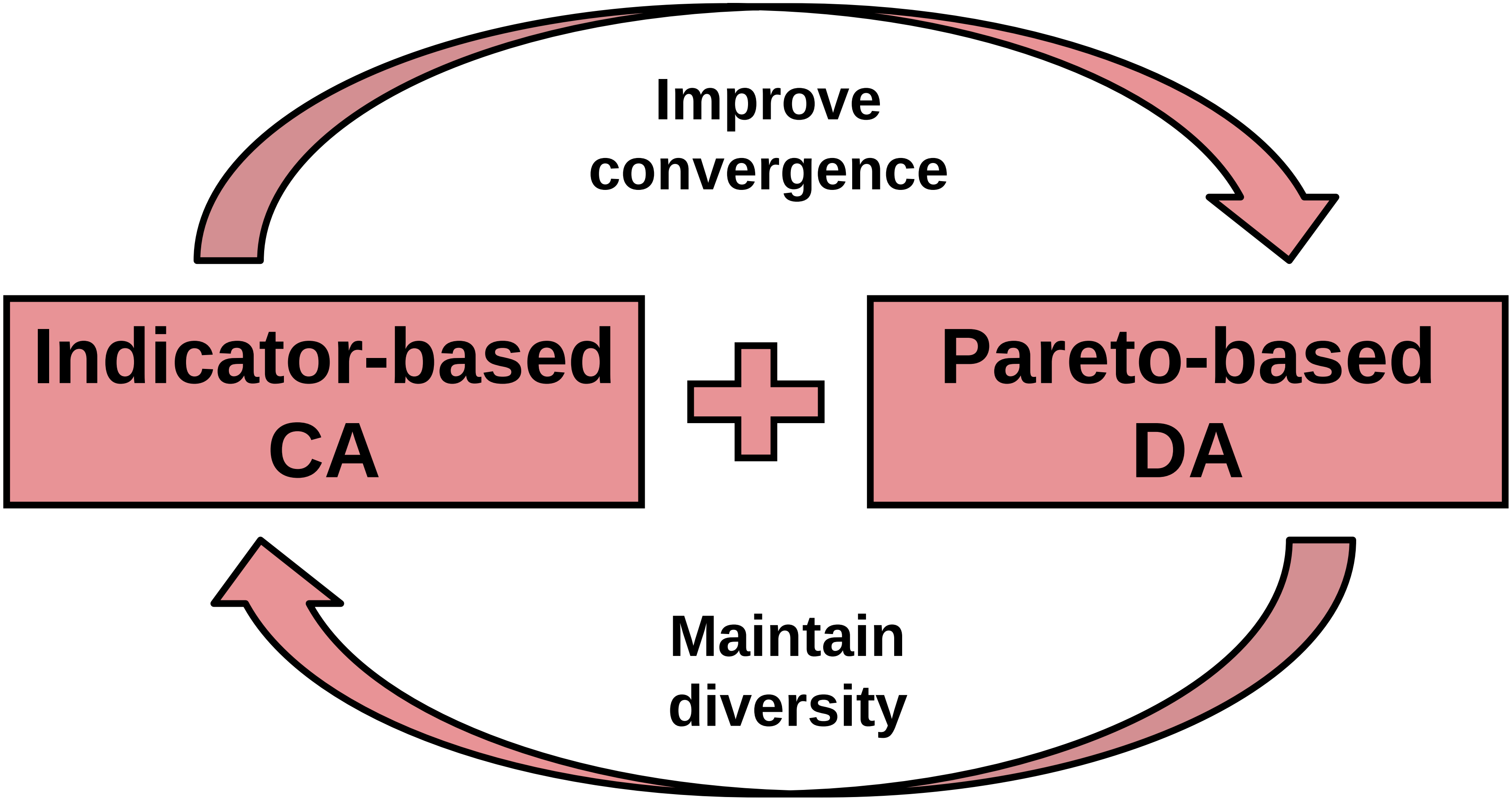
Original Two-Archive Algorithm

Two-Archive algorithm (Two_Arch) maintains two archives (CA and DA) to promote convergence and diversity explicitly.

- **K. Praditwong and X. Yao, "A New Multi-objective Evolutionary Optimisation Algorithm: The Two-Archive Algorithm," *Proc. of the 2006 International Conference on Computational Intelligence and Security (CIS'2006)*, 3-6/11/2006, Ramada Pearl Hotel, Guangzhou, China. IEEE Press, Volume 1, pp.286-291.**



<u>Improved</u> Two-Archive Algorithm: Main Idea



Two_Arch2: Main Steps

Step 1: Initialization.
Step 2: Output <u>DA</u> if the stopping criterion is met, otherwise continue.
Step 3: Generate new solutions from CA and DA by crossover and mutation.
Step 4: Update CA and DA separately, go to Step 2.

H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

Convergence Archive (CA)

- **The quality indicator $I_{\varepsilon+}$ in IBEA is used in selection of CA. $I_{\varepsilon+}$ is an indicator that describes the minimum distance that one solution needs to dominate another solution in the objective space.**

$$I_{\varepsilon+}(x_1, x_2) = min_{\varepsilon}(f_i(x_1) - \varepsilon \leq f_i(x_2), 1 \leq i \leq m)$$

- **The fitness is assigned as below, the solution with the smallest fitness is removed from CA first.**

$$F(x_1) = \sum_{x_2 \in P \setminus \{x_1\}} -e^{-I_{\varepsilon+}(x_2, x_1)/0.05}$$

Diversity Archive (DA)

- **Update DA**
  - **When DA overflows, boundary solutions (solutions with maximal or minimal objective values) are firstly selected.**
  - **In the iterative process, the most different solution from the current DA is added until reaching the size.**
- **L_p-norm distance is adopted as the similarity measure in DA.**
- **DA is used as the final output of Two_Arch2.**

H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

Degraded Euclidean Distance (Distance Concentration) in High-Dimensional Space

- **The Euclidean distance (L_2-norm) degrades its similarity indexing performance in a high-dimensional space.**
- **Most of existing diversity maintenance methods use the Euclidean distance to measure similarity among solutions for MaOPs.**

C. C. Aggarwal, A. Hinneburg and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space." Springer, 2001.

Similarity in High-Dimensional Space

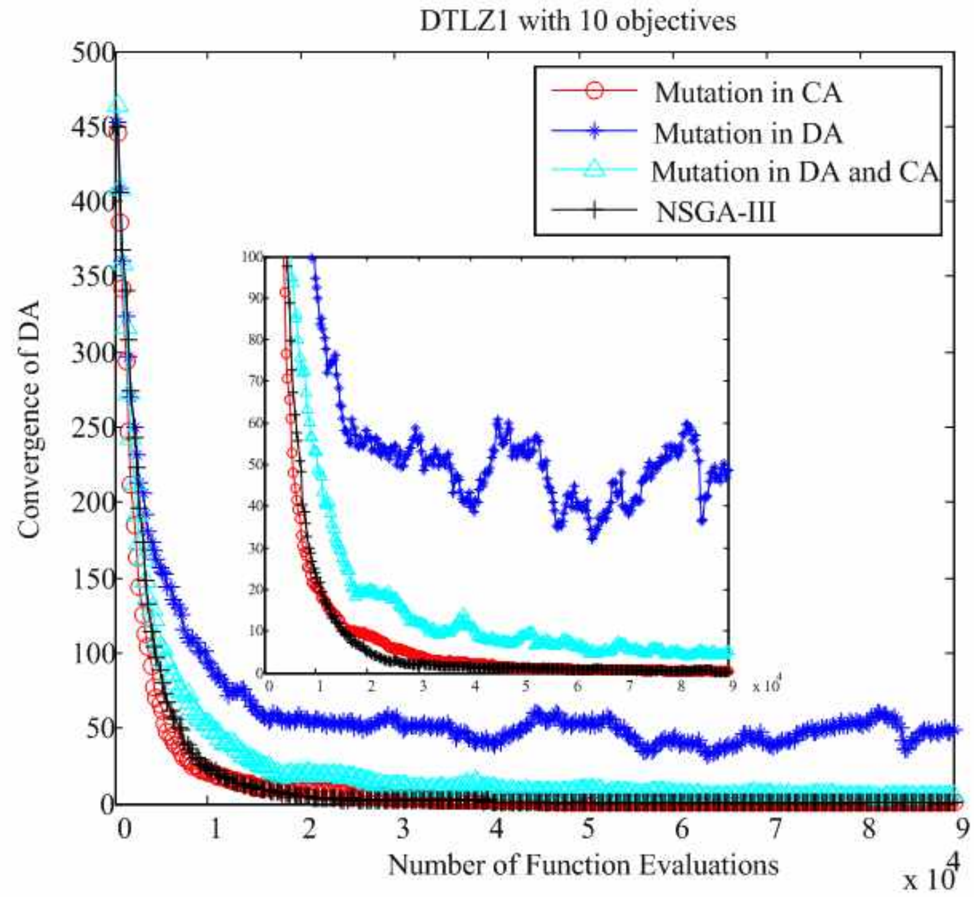
- **The fractional distances (L_p-norm, $p<1$) perform better in a high-dimensional space.**

- **$L_{1/m}$-norm is employed in Two_Arch2, where *m* is the number of objectives.**

H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

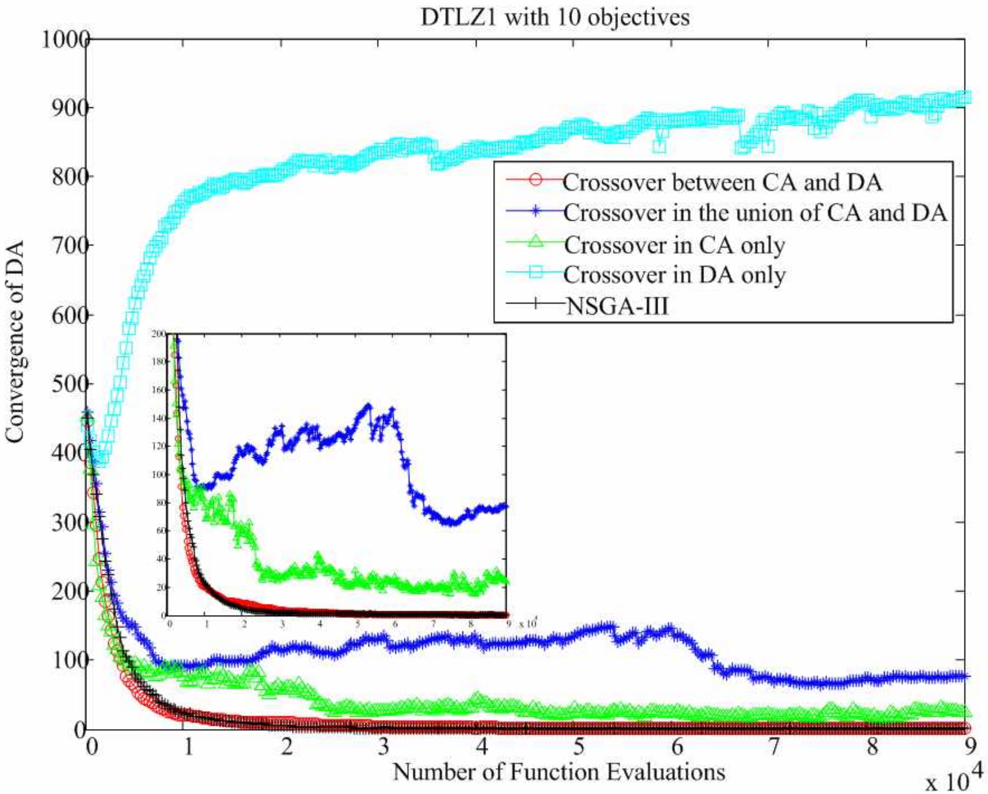
Interaction between CA and DA: Mutation

- **Mutation to DA does not speed up convergence, and disturbs the guidance of CA to DA.**
- **Mutation is applied to CA only in Two_Arch2.**



H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

Interaction between CA and DA: Crossover



- **The crossover between CA and DA has the fastest convergence speed.**
- **The crossover between CA and DA is employed in Two_Arch2.**

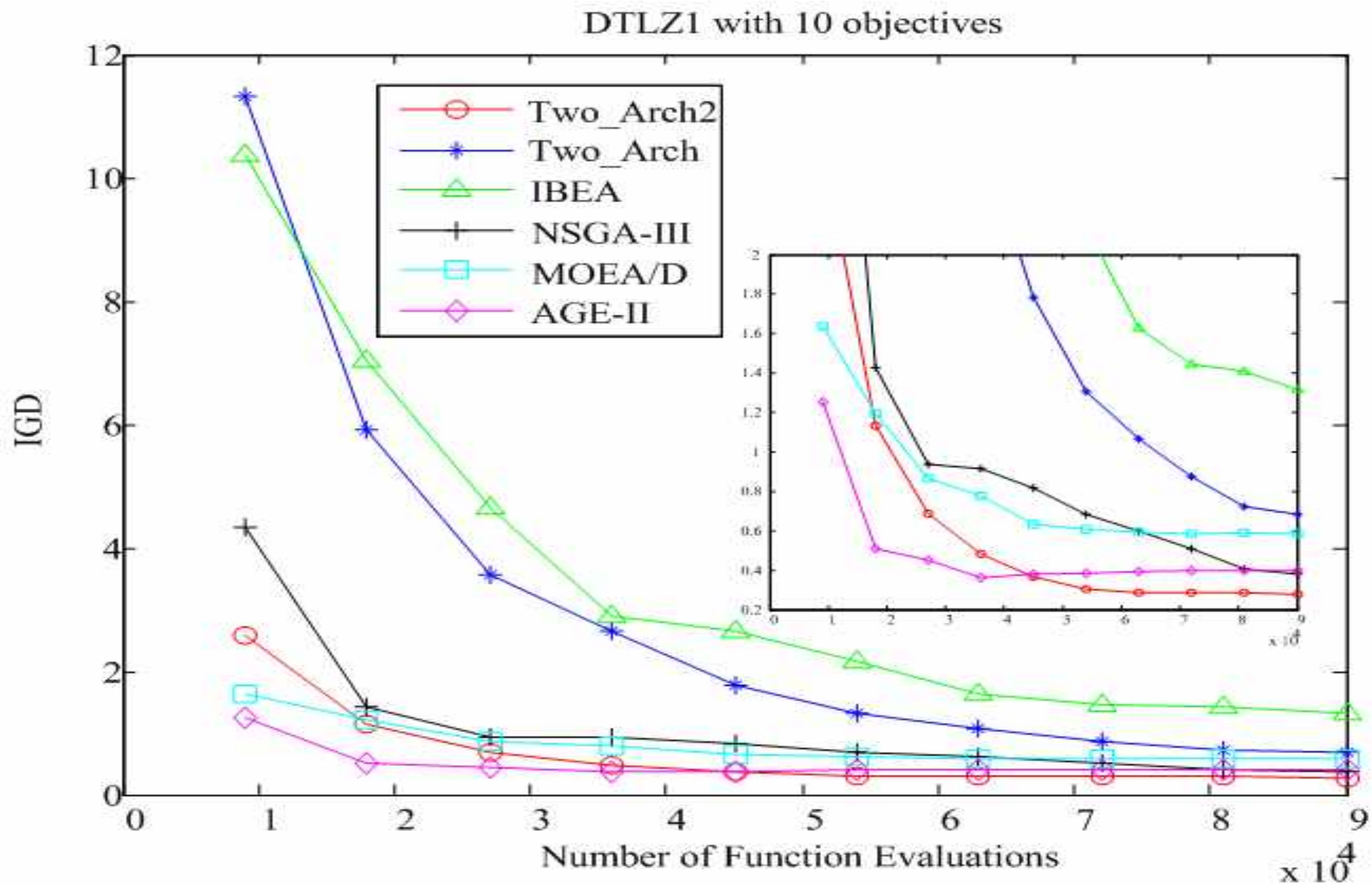
H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

Experimental Comparisons

- **Two_Arch2: Our algorithm.**
- **Two_Arch: The original version of Two_Arch, to show the improvement of Two_Arch2 over Two_Arch.**
- **IBEA: An indicator-based ($I_{\varepsilon+}$) MOEA with good convergence but poor diversity.**
- **NSGA-III: A newly-proposed MOEA, which is widely used.**
- **MOEA/D: An aggregation function-based MOEA.**
- **AEG-II: A Pareto-based MOEA with the ε-grid approximation in the objective space.**

H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," *IEEE Transactions on Evolutionary Computation*, 19(4):524-541, August 2015.

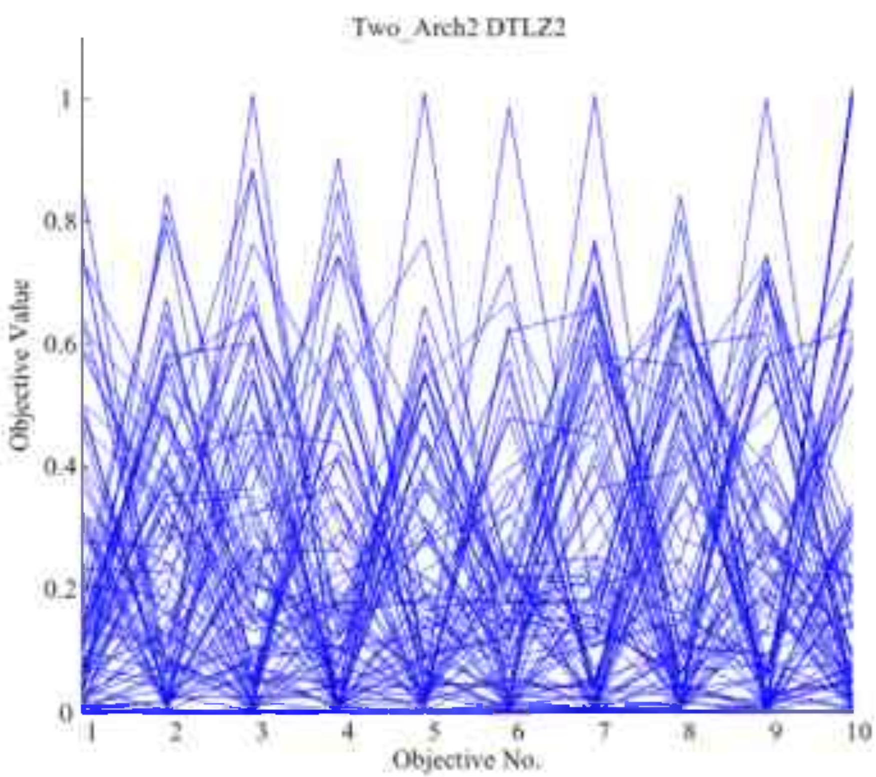
DTLZ1 with 10 Objectives

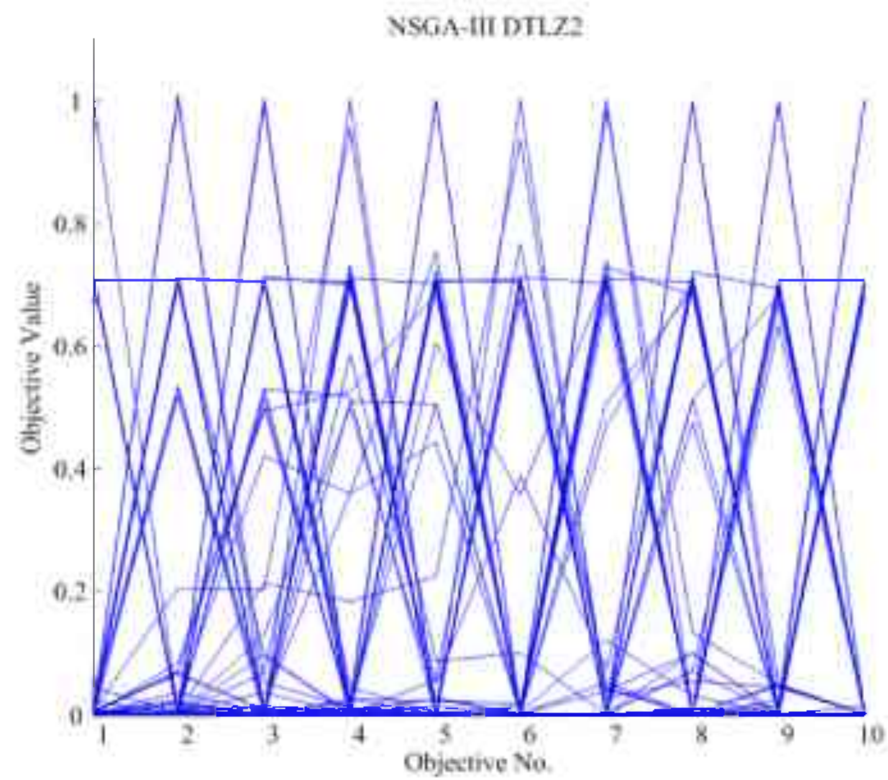


H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

Two_Arch2 vs. NSGA-III on DTLZ2 with 10 Objectives

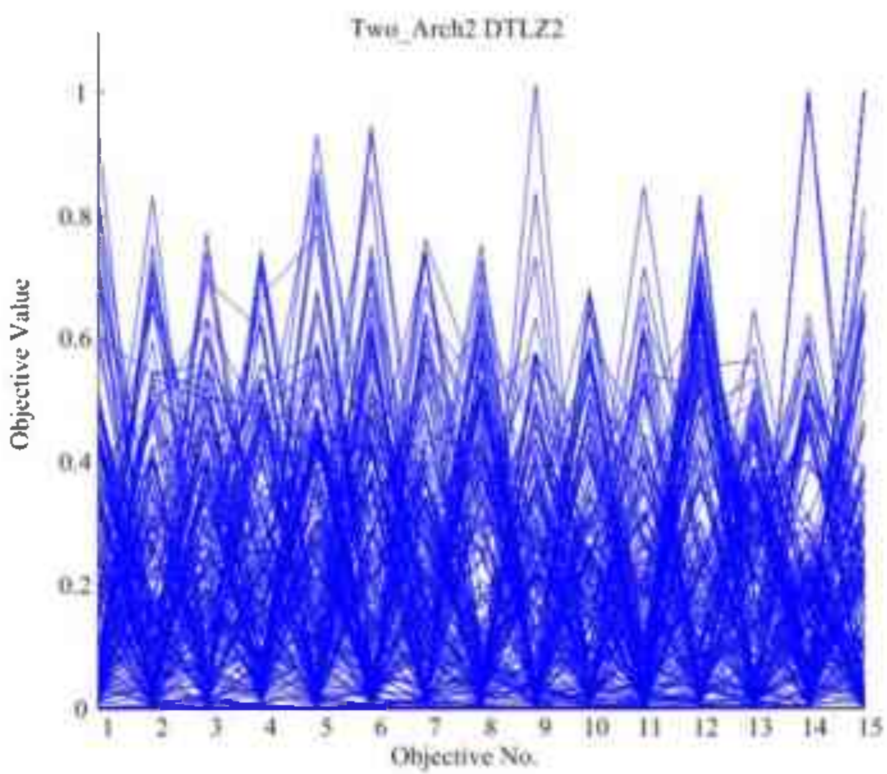
	Convergence	Diversity	Extreme point
Two_Arch2	Good	Good	Fair
NSGA-III	Good	Fair	Good

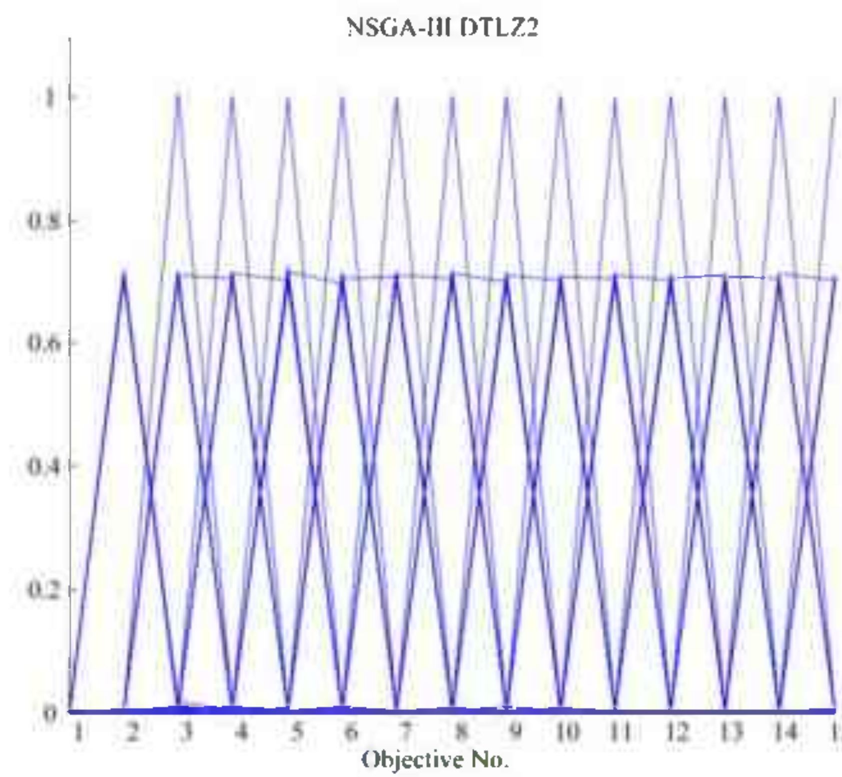




Two_Arch2 vs. NSGA-III on DTLZ2 with 15 Objectives

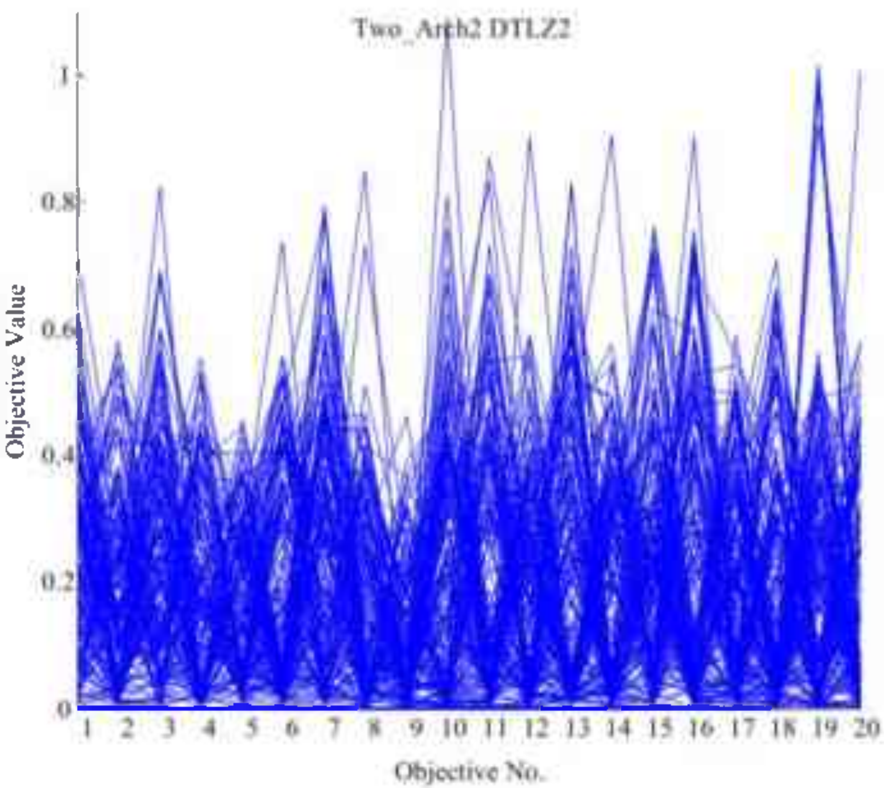
	Convergence	Diversity	Extreme point
Two_Arch2	Good	Good	Poor
NSGA-III	Good	Fair	Good

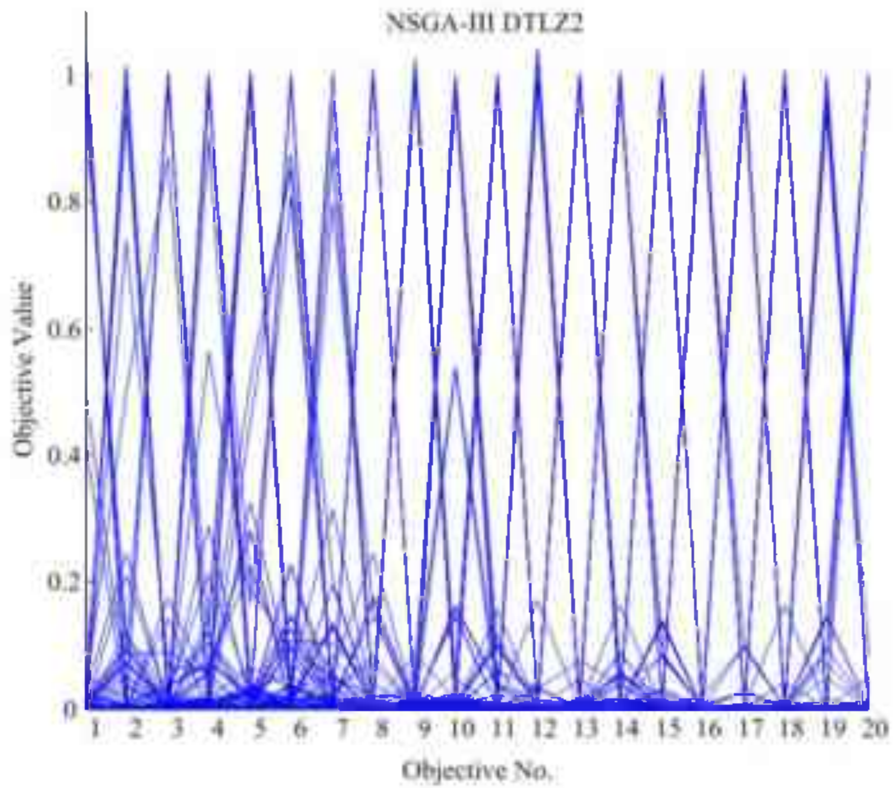




Two_Arch2 vs. NSGA-III on DTLZ2 with 20 Objectives

	Convergence	Diversity	Extreme point
Two_Arch2	Good	Good	Poor
NSGA-III	Good	Fair	Good





Two_Arch2 vs. NSGA-III

	Two_Arch2	NSGA-III
Convergence methodology	$I_{\varepsilon+}$	Pareto dominance
Convergence degeneration	No	No
Diversity maintenance	$L_{1/m}$-norm-based distance	Minimal perpendicular distances to reference points
Diversity degeneration	No	Increase with the dimension of objective space
Manual settings	None	Reference points

More Experimental Results

More experimental results are in

- **H. Wang, L. Jiao and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization,"** ***IEEE Transactions on Evolutionary Computation*****, 19(4):524-541, August 2015.**

including Matlab code.

Outline

- Introduction
- Objective Reduction
- Alternative Dominance Relationship
- Improved Two-Archive Algorithm (Two_Arch2)
- Conclusions and Future Work

Conclusions

- **There are three major approaches to dealing with a large number of objectives:**
  1. **Objective reduction**
  2. **Alternative dominance relationship**
  3. **New algorithms**

- **This talk touches on only a tiny proportion of all the work. For more comprehensive review:**
  - **B. Li, J. Li, K. Tang and X. Yao, "Many-Objective Evolutionary Algorithms: A Survey," *ACM Computing Surveys*, 48(1), Article 13, 35 pages, September 2015.**

Future Work

1. **Dynamic number of objectives, e.g.,**
   - **R. Chen, K. Li and X. Yao, "Dynamic Multiobjectives Optimization With a Changing Number of Objectives,"** ***IEEE Transactions on Evolutionary Computation*****, vol. 22, no. 1, pp. 157-171, Feb. 2018.**

2. **Constraint handling, e.g.,**
   - **K. Li, R. Chen, G. Fu and X. Yao, "Two-Archive Evolutionary Algorithm for Constrained Multi-Objective Optimization,"** ***IEEE Transactions on Evolutionary Computation*****, online on 19/7/2018. DOI: 10.1109/TEVC.2018.2855411**

Ensemble Approaches to Class Imbalance Learning

Xin Yao
Department of Computer Science and Engineering, and
Research Institute of Trustworthy Autonomous Systems
Southern University of Science and Technology (SUSTech)
Shenzhen, China
and
School of Computer Science, University of Birmingham, UK

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Motivation

- **Given a classification task, most machine learning methods assume that:**
  - **Every class has the same misclassification cost.**
  - **Aim to maximize the classification accuracy.**
- **However, many real-world applications have very unbalanced distributions among classes:**
  - **E.g. fault diagnosis, software defect prediction, etc.**
  - **Minority class: rare cases, high misclassification cost.**

Class Imbalance Learning

- **Class imbalance learning refers to learning from imbalanced data sets, in which some classes of examples (minority) are highly under-represented compared to other classes (majority).**
- **Learning difficulty: poor generalization on the minority class.**
- **Learning objective: obtaining a classifier that will provide high accuracy for the minority class without severely jeopardizing the accuracy of the majority class.**

Existing Work

- **Re-sampling techniques** try to rebalance the data distribution.
  - Over-sampling minority classes
  - Under-sampling majority classes
- **Cost-sensitive methods** increase the misclassification cost of minority classes.
  - Hard to quantify costs in practice
- **Classification ensembles** combine multiple learners to improve performance.
  - Advantages: can improve minority and overall performance.

Existing Work

- **Re-sampling techniques** try to rebalance the data distribution.
  - Over-sampling minority classes
  - Under-sampling majority classes
- Cost-sensitive methods increase the misclassification cost of minority classes.
  - Hard to quantify costs in practice
- **Classification ensembles** combine multiple learners to improve performance.
  - Advantages: can improve minority and overall performance.

Why Ensembles?

- For a large and complex problem, designing a monolithic system to solve it is often very difficult.
- Divide-and-conquer is a common strategy in solving such problems.
- Ensemble approaches could be viewed as an automatic approach toward divide-and-conquer.
- Ensemble learning has some nice theoretical properties that explain why and when it works.
- It is often straightforward to implement.

Y. Liu and X. Yao, ``Ensemble learning via negative correlation,'' *Neural Networks*, 12(10):1399-1404, December 1999.

What is an Ensemble?

$$O = \sum_{j=1}^{N} w_j o_j$$

- **There are many different ways to determine or learn w_j.**
  - X. Yao and Y. Liu, ``Making use of population information in evolutionary artificial neural networks,'' *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(3):417-425, June 1998.
- **There are many different training algorithms, e.g., bagging, boosting, negative correlation learning, etc.**
  - Y. Liu and X. Yao, ``Ensemble learning via negative correlation,'' *Neural Networks*, 12(10):1399-1404, December 1999.

Diversity is Essential in Ensembles

- **An ensemble of positively correlated individuals provide few advantages over single individuals.**
- **There have been many studies demonstrating that a diverse ensemble provide better generalisation.**
- **What do we mean by "diverse"? How do we define diversity? How do we generate diversity? … [1,2]**
  - **Still ongoing research as to how diversity can be best defined and used in practice.**

[1] G. Brown, J. L. Wyatt, R. Harris and X. Yao, "Diversity Creation Methods: A Survey and Categorisation," *Information Fusion*, 6(1):5-20, January 2005.

[2] E. K. Tang, P. N. Suganthan and X. Yao, "An Analysis of Diversity Measures," *Machine Learning*, 65:247-271, 2006.

- **If diversity is so important in ensembles, what role(s) does it play in dealing with class imbalance classification problems?**

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Diversity in Class Imbalance Ensembles

Strong correlations have been found between diversity and generalisation performance measures [1]:

- **Diversity showed a positive impact on the minority class, by making the ensemble produce smoother and less over-fitting classification boundaries for the minority class;**
- **Diversity was shown to be beneficial to both AUC and G-mean (overall performance).**

[1] S. Wang and X. Yao, "Relationships Between Diversity of Classification Ensembles and Single-Class Performance Measures," ***IEEE Transactions on Knowledge and Data Engineering*****, 25(1):206-219, January 2013.**

Making Use of Diversity: AdaBoost.NC

Building on the existing AdaBoost algorithm:

- **Apply random oversampling to rectify the imbalanced distribution first.**
- **Encourage diversity: introduce diversity information (amb) into the weights of training examples in the sequential training procedure of AdaBoost.**
  - **The weight-updating rule of AdaBoost is modified such that both high classification errors *and* high diversity will lead to larger weights.**

S. Wang, H. Chen and X. Yao, "Negative correlation learning for classification ensembles". *Proc. of IJCNN'10*, pp.2893-2900. IEEE Press, 2010.

AdaBoost.NC: Result Summary

AdaBoost.NC with random over-sampling is effective in classifying minority class examples correctly *without* sacrificing the overall performance (in terms of AUC), in comparison to other methods.

S. Wang, H. Chen and X. Yao, "Negative correlation learning for classification ensembles". *Proc. of IJCNN'10*, pp.2893-2900. IEEE Press, 2010.

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Multi-class Imbalance Learning

- **Multi-class imbalance: there are more than two classes with uneven class distributions.**
  - **E.g. In software defect prediction: there are different types of defects.**
- **Most existing imbalance learning techniques are only designed for and evaluated in two-class scenarios.**
- **Existing methods are not effective or even cause a negative effect when there is more than one minority/majority class [1].**

[1] S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 42(4):1119-1130, August 2012.

Research Questions

- **Two types of multi-class imbalance : multi-minority and multi-majority.**
  1. **Are there any differences between multiple minority and multiple majority classes?**
  2. **Would these two types of problem pose the same or different challenges to a learning algorithm?**
- **More effective learning algorithms:**
  3. **Can AdaBoost.NC be extended to tackle multi-class imbalance directly?**
  4. **Is class decomposition necessary for multi-class problems?**

S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 42(4):1119-1130, August 2012.

Experimental Analysis: Multi-minority

- **Multi-minority reduces the performance of ensemble learning, and**
- **Data re-sampling does not help.**

S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," ***IEEE Transactions on Systems, Man and Cybernetics, Part B*****, 42(4):1119-1130, August 2012.**

Experimental Analysis: Multi-majority

- **The ensembles suffered from significant performance reduction because of multi-majority.**

S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 42(4):1119-1130, August 2012.

Key Messages

- **Both multi-minority and multi-majority negatively affect the overall and minority-class performance.**
- **In particular, the multi-majority case tends to be more challenging, in terms of F-measure and recall.**
- **Neither oversampling nor undersampling is satisfactory:**
  - **Random oversampling suffers from overfitting as no new information is introduced into minority classes to facilitate classification;**
  - **The effect of random undersampling is weakened when there are more minority classes.**

S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 42(4):1119-1130, August 2012.

Potential Solution: AdaBoost.NC

- **AdaBoost.NC can better balance the performance across multiple classes with a high G-mean.**
- **Using class decomposition is unnecessary in tackling multi-class imbalance problems.**

S. Wang and X. Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions," ***IEEE Transactions on Systems, Man and Cybernetics, Part B*****, 42(4):1119-1130, August 2012.**

Why do we always separate sampling from learning?

Embedding Sampling into Learning

- **Sampling/re-sampling does not have to be done separately before learning.**
- **Consider a *single* classifier, we can use the following simple strategy called *DyS* [1]:**
  - **For every training example that is fed to the current classifier, the probability of it being actually used for training the classifier is first estimated.**
  - **Then the classifier is trained on this example according to that probability.**

[1] M. Lin, K. Tang and X. Yao, "A Dynamic Sampling Approach to Training Neural Networks for Multi-class Imbalance Classification," ***IEEE Transactions on Neural Networks and Learning Systems*****, 24(4):647-660, April 2013.**

Estimating the Probability

- **Very simple: considering both the current status of training and imbalanceness [1]:**
  - **The examples that were misclassified will be selected to update the classifier.**
  - **For the examples that were correctly classified, those from minority classes are emphasized more than those from majority classes.**

[1] M. Lin, K. Tang and X. Yao, "A Dynamic Sampling Approach to Training Neural Networks for Multi-class Imbalance Classification," ***IEEE Transactions on Neural Networks and Learning Systems*****, 24(4):647-660, April 2013.**

Advantages of DyS

- Handle multiple classes easily.
- No need to worry about whether it’s over- or under-sampling.
- Applicable to different classifiers.
- Very simple.

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Online Class Imbalance Learning

- **A relatively new area of research, combining online learning with class imbalance learning.**
- **Online learning here means learning from data examples "one-by-one" without storing and reprocessing observed examples.**
- **Online class imbalance learning deals with data streams where data arrive continuously and the class distribution is imbalanced.**
- **It introduces new research challenges.**

Defining Class Imbalance Is Difficult

To handle class imbalance online, we first need to define it by answering the following three questions:

1. Is the data stream currently imbalanced?
2. Which classes belong to the minority/majority?
3. What is the imbalance rate currently?

S. Wang, L. L. Minku and X. Yao, ``Resampling-Based Ensemble Methods for Online Class Imbalance Learning,'' IEEE Transactions on Knowledge and Data Engineering, 27(5):1356-1368, May 2015.

Online Bagging + Re-sampling

- Sampling is added to online bagging to handle imbalanced data streams [1]:
  - Over-sampling based online bagging (OOB)
  - Under-sampling based online bagging (UOB)
- We will look at two-class problems first.
- Use a simple parameter to adjust the re-sampling rate.
- No need to use any class imbalance detection method.

[1] S. Wang, L. L. Minku and X. Yao, ``Resampling-Based Ensemble Methods for Online Class Imbalance Learning,'' IEEE Transactions on Knowledge and Data Engineering, 27(5):1356-1368, May 2015.

Main Findings

- **Both UOB and OOB performed significantly better than just OB, showing the importance of sampling.**
- **In most cases, UOB has the best performance in terms of the minority recall and G-mean.**
- **However, OOB is more robust against the imbalance rate changes.**
- **It is possible to combine the advantages of OOB and UOB.**

S. Wang, L. L. Minku and X. Yao, "Resampling-Based Ensemble Methods for Online Class Imbalance Learning," IEEE Transactions on Knowledge and Data Engineering, 27(5):1356-1368, May 2015.

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Multi-objective Class Imbalance Learning

- **Multi-objective learning treats single class performances as separate objectives.**
- **Multi-objective optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), are used as learning algorithms.**
- **The result from such an MOEA is a non-dominated set of solutions (i.e., learners), which ideally form an ensemble we are interested in.**
  - **A Chandra and X. Yao, ``Ensemble learning using multi-objective evolutionary algorithms,'' *Journal of Mathematical Modelling and Algorithms*, 5(4):417-445, December 2006.**

Multi-objective Genetic Programming

- **Still treat accuracy and diversity as separate objectives.**
- **Diversity is introduced explicitly to MOGP for classifying imbalanced data. Two alternatives to diversity definition:**
  - **Negative correlation learning (NCL)**
    - **Y. Liu and X. Yao, "Negatively correlated neural networks can produce best ensembles,"** ***Australian Journal of Intelligent Information Processing Systems,*** **vol. 4, pp. 176–185, 1997.**
  - **Pairwise Failure Crediting (PFC)**
    - **A. Chandra and X. Yao, "Ensemble learning using multi-objective evolutionary algorithms,"** ***Journal of Mathematical Modelling and Algorithms,*** **vol. 5, pp. 417–445, 2006.**
- **No need for sampling/re-sampling in MOGP.**
  - **U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Evolving Diverse Ensembles using Genetic Programming for Classification with Unbalanced Data,'' IEEE Transactions on Evolutionary Computation, 17(3):368-386, June 2013.**

Ensemble Member Selection

- **Sometimes it is unnecessary to include the entire set of classifiers found by MOEAs in an ensemble. A subset would be sufficient or even better.**
- **There are various methods for selecting a diverse subset of classifiers from a large set, e.g.,**
  - **X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(3):417-425, June 1998.**
  - **U. Bhowan, M. Johnston, M. Zhang and X. Yao, ``Reusing Genetic Programming for Ensemble Selection in Classification of Unbalanced Data,'' IEEE Transactions on Evolutionary Computation, 18(6):893-908, December 2014.**

Overview

- Introduction
- Diversity in ensembles
- Multi-class imbalance learning
- Online class imbalance learning
- Multi-objective class imbalance learning
- Concluding remarks

Concluding Remarks

- Ensembles are competitive learning methods to tackle class imbalance problems.
- Diversity is the key issue in ensemble learning.
- Insight into diversity's roles enables us to design better ensemble algorithms.
- Online class imbalance learning is a promising future research direction.
- We need more theoretical analysis of the algorithms.

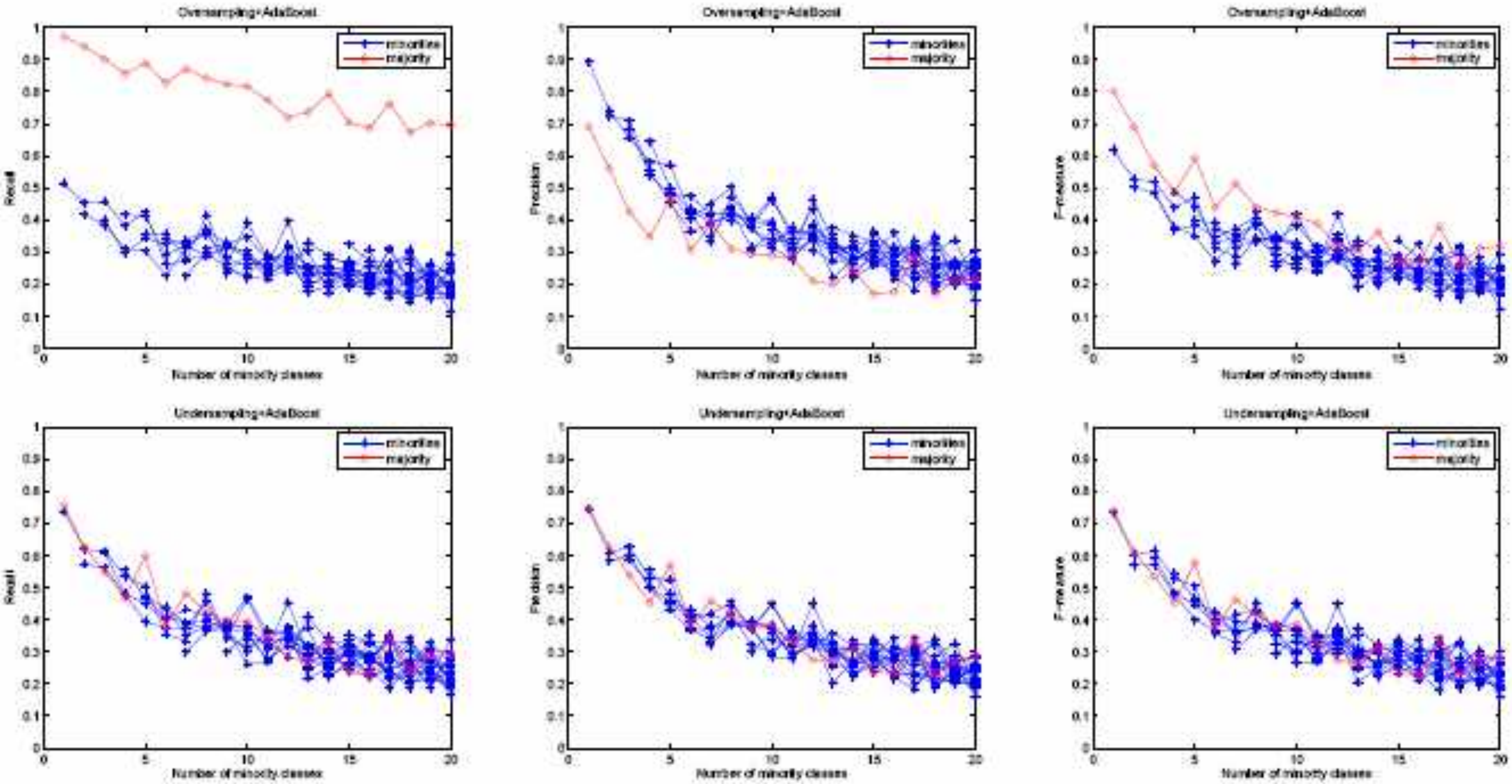
Existing Work

- **Use class decomposition**: Converting a multi-class problem into a set of two-class sub-problems; then use two-class imbalance techniques to handle each obtained binary sub-task.
- Class decomposition schemes include (given a c-class task, c > 2):
  - **one-against-all (OAA)**: Each of the c classes is trained against all other classes. It results in c binary classifiers, making data more imbalanced.
  - **one-against-one (OAO)**: Each of the c classes is trained against every one of the other classes. It results in c(c-1)/2 binary classifiers. When c is large, the training time can be very long.
  - **P-against-Q (PAQ)**: Using P of the c classes against the other Q of the c classes, the training process is repeated several times. Different P classes are chosen at each time.
- ***No work treated multi-class imbalance problems as multi-class.***

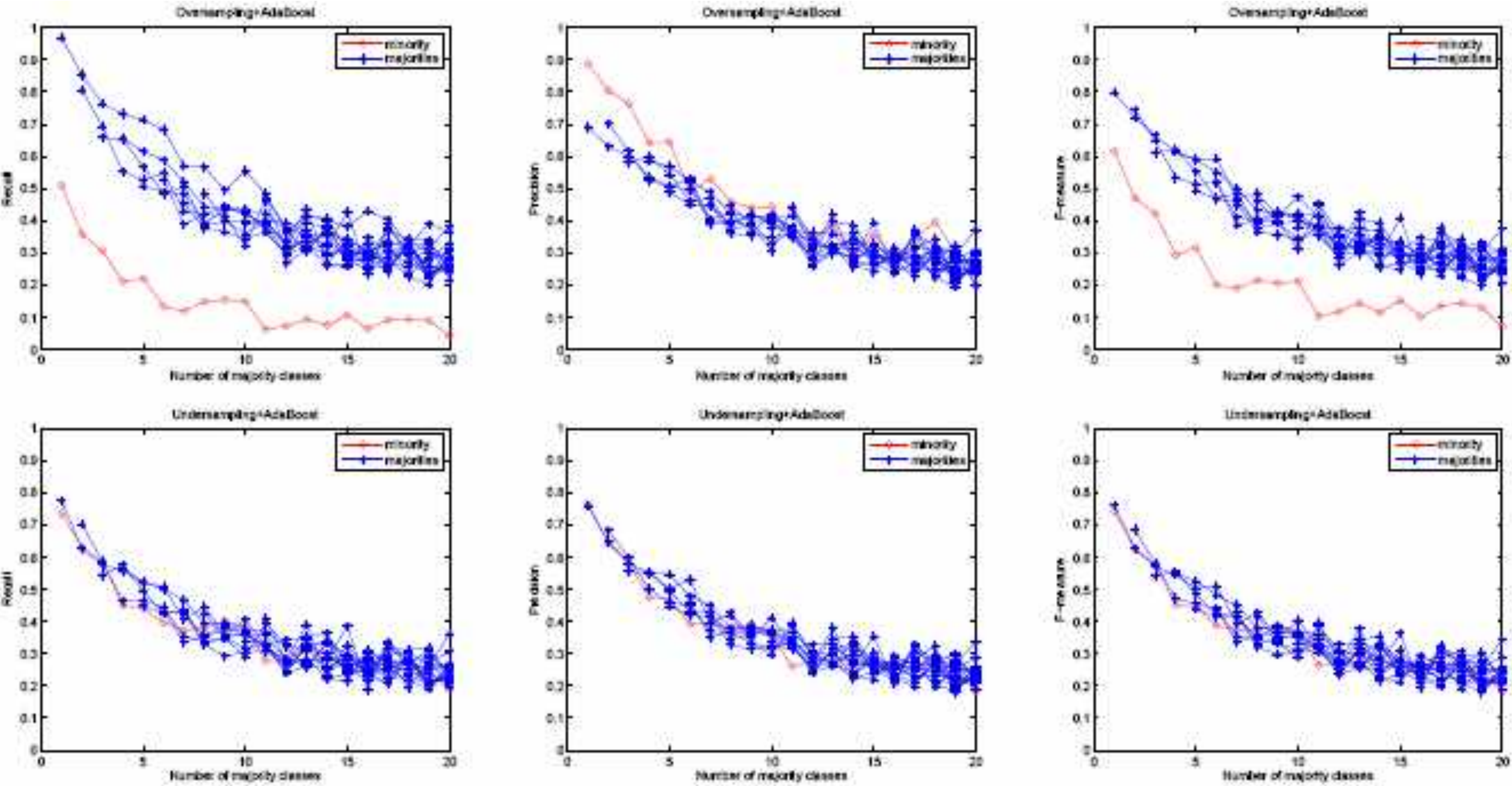
Experimental Analysis: Setup

- **Data generation:**
  - Data points in each class are generated randomly from Gaussian distributions, where the mean and standard deviation of each attribute are random real values in [0,10].
  - Small imbalanced data: each example has 2 attributes; each minority class has 10 examples and each majority class has 100 examples.
  - Large imbalanced data: each example has 20 attributes; each minority class has 100 examples and each majority class has 1000 examples.
- **Experimental settings:**
  - Multi-minority: the number of minority classes is varied from 1 to 20, and only one majority class exists.
  - Multi-majority: the number of majority classes is varied from 1 to 20, and only one class is generated as the minority.
  - 3 ensemble methods, each containing 51 C4.5 decision tree as base learners:
    - the conventional AdaBoost that is trained from the original imbalanced data (baseline model);
    - random oversampling + AdaBoost;
    - random undersampling + AdaBoost.

Experimental Results: Multi-minority



Experimental Results: Multi-majority



Detecting Imbalance: Time-decayed Class Size

- **Given a data stream (x_t, y_t), where y_t can be c_1, c_2, … c_N, define the class size as**

$$w_k^{(t)} = \theta w_k^{(t-1)} + (1-\theta)[(x_t, c_k)], \qquad (k = 1, \ldots, N),$$

Detecting Imbalance: Recall

If x_t's real label y_t is c_i, the recall of class c_i, denoted by $R_i^{(t)}$, is updated by [6]:

$$R_i^{(t)} = \theta' R_i^{(t-1)} + (1 - \theta')[x_t \leftarrow c_i]. \tag{2}$$

For the recall of the other classes, denoted by $R_j^{(t)}$ $(j \neq i)$, it is updated by:

$$R_j^{(t)} = R_j^{(t-1)}, \qquad (j = 1, \ldots, N, j \neq i). \tag{3}$$

Minority vs. Majority

- $w_i - w_j > \delta_1 \ (0 < \delta_1 < 1)$
- $R_i - R_j > \delta_2 \ (0 < \delta_2 < 1)$.

If both conditions are satisfied, then class c_j is sent to the minority class label set Y_{min} and class c_i is sent to the majority class label set Y_{maj}. If Y_{min} and Y_{maj} are not empty, it means that the data stream is imbalanced. This can then be used to invoke the class imbalance techniques running in the online model to tackle the imbalanced distribution.

OOB and UOB

Input: an ensemble with M base learners, current training example (x_t, y_t), and current class size $w^{(t)} = (w_+^{(t)}, w_-^{(t)})$.

for each base learner f_m $(m = 1, 2, \ldots, M)$ **do**

  if $y_t = +1$ and $\begin{cases} w_+^{(t)} < w_-^{(t)} & \text{for OOB} \\ w_+^{(t)} > w_-^{(t)} & \text{for UOB} \end{cases}$

    set $K \sim Poisson(w_-^{(t)}/w_+^{(t)})$

  else if $y_t = -1$ and $\begin{cases} w_-^{(t)} < w_+^{(t)} & \text{for OOB} \\ w_-^{(t)} > w_+^{(t)} & \text{for UOB} \end{cases}$

    set $K \sim Poisson(w_+^{(t)}/w_-^{(t)})$

  else

    set $K \sim Poisson(1)$

  end if

  update f_m K times

end for

Combining OOB and UOB (I): WEOB1

- **Suppose OOB has G-mean value g_o and UOB has G-mean value g_u at the current moment. Let α_o and α_u denote the weights of OOB and UOB respectively.**

$$\alpha_o = \frac{g_o}{g_o + g_u}, \qquad \alpha_u = \frac{g_u}{g_o + g_u}.$$

Combining OOB and UOB (II): WEOB2

$$\begin{cases} \alpha_o = 1, \alpha_u = 0 & \text{if } g_o \geq g_u, \\ \alpha_o = 0, \alpha_u = 1 & \text{if } g_o < g_u. \end{cases}$$