Feature weighting for antimicrobial peptides classification: a multi-objective evolutionary approach

Jesus A. Beltrán, Longendri Aguilera-Mendoza, Carlos A. Brizuela

Computer Sciences Deparment

CICESE Research Center

IEEE BIBM 2017- Kansas City, MO, USA.

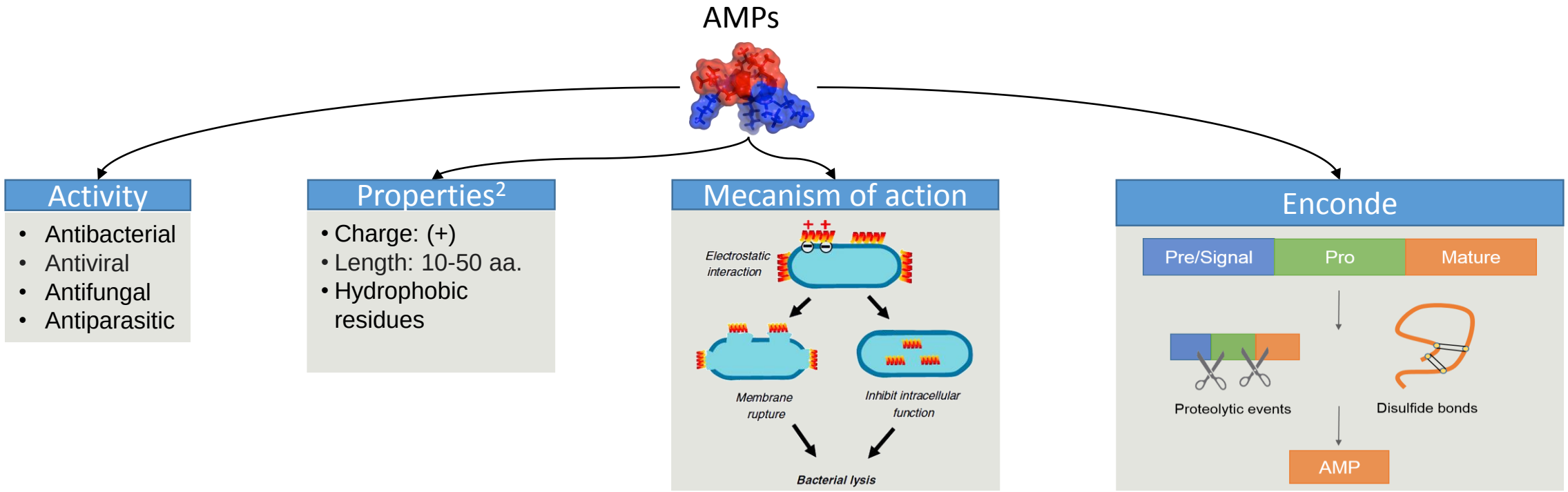
November 14th, 2017

Outline

1. Introduction
2. Problem Statement
3. Material and Methods
4. Experiments and Results
5. Conclusions

Introduction: Antimicrobial Peptides (AMPs)

- AMPs are a diverse class of natural occurring molecules that are produced as the first line of defense by multicellular organism [1].



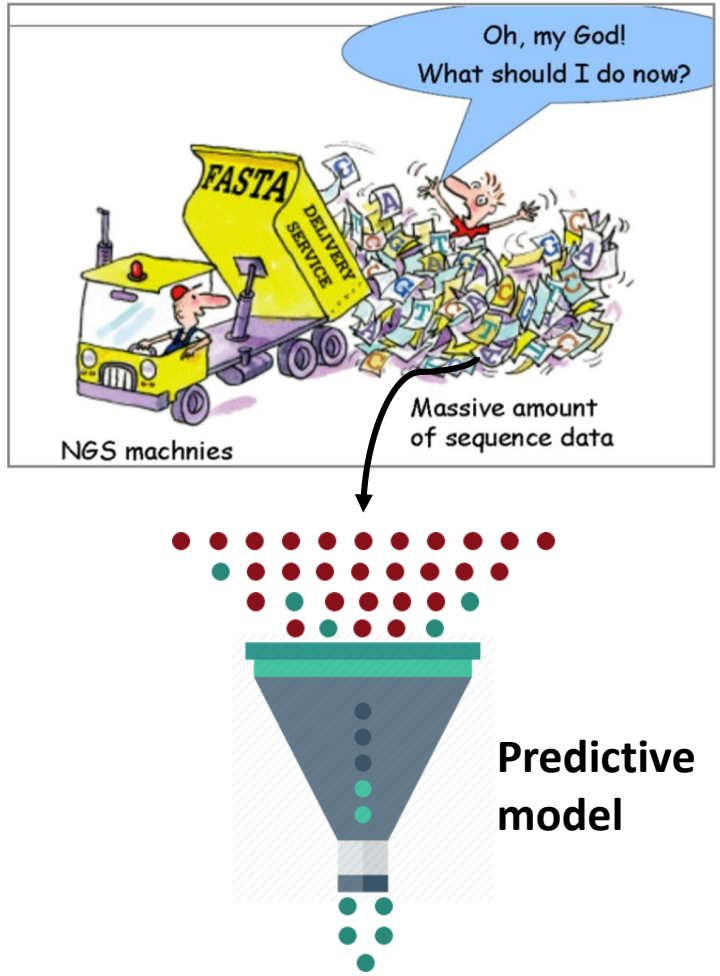
- AMPs might become crucial in fighting antibiotic-resistant bacteria and other infections.

[1] Zhang, L. J., & Gallo, R. L. (2016). Antimicrobial peptides. *Current Biology*, *26*(1), R14-R19.

[2] Wang, G., Li, X., & Zasloff, M. (2010). A database view of naturally occurring antimicrobial peptides: nomenclature, classification and amino acid sequence analysis. *Antimicrobial peptides: discovery, design and novel therapeutic strategies*, 1-21.

Introduction: Next-Generation Sequences (NGS)

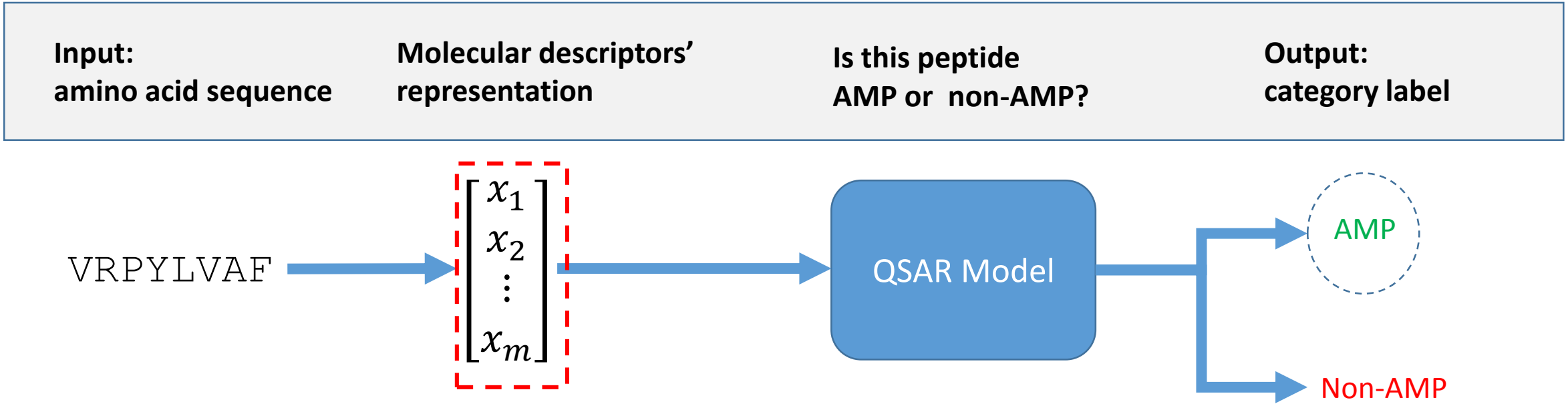
- NGS technologies are generating a large amount of data where peptide with antimicrobial activity could be found.
- The most important aspect of virtual screening (VS) is the derivation of predictive models for the identification of AMPs through peptide libraries.
- Select peptides with the potential to be antimicrobial before their synthesis in the wet lab.



- C. D. Fjell, J. A. Hiss, R. E. Hancock, and G. Schneider, "Designing antimicrobial peptides: form follows function," Nature reviews Drug discovery, vol. 11, no. 1, pp. 37–51, 2012.
- D. Raventos, et al., "Improving on nature's defenses: optimization &high throughput screening of antimicrobial peptides," Combinatorial chemistry & high throughput screening, vol. 8, no. 3, pp. 219–233, 2005.

Introduction: Binary classification of antimicrobial activity

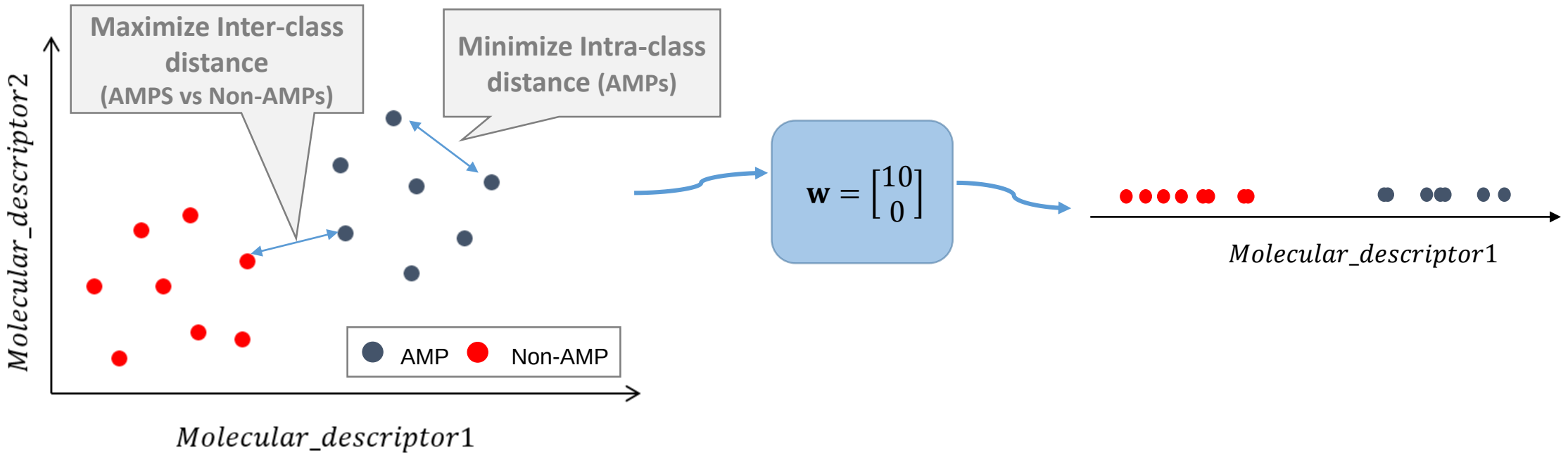
- Quantitative Structure-Activity Relationship (QSAR) modeling is widely practiced for predicting active (AMPs) and inactive (non-AMPs) peptides.
- Molecular descriptors define the chemical space whre each peptide is projected.
- Exploring the space of all possible subsets of descriptors is not feasible due to the exponential size of the search space, 2^(No. of molecular descriptors) .



H. Jenssen, "Descriptors for antimicrobial peptides," Expert opinion on drug discovery, vol. 6, no. 2, pp. 171–184, 2011.

Introudction: our approach

- Find a weight assignment for each molecular descriptor such that peptides with different biological activity are far away from each other, whereas peptides with antimicrobial activity are close together.



Dataset composed of n peptides with known biological activities.

Find a vector of weights

Peptide representation for the classification task.

S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

Problem Statement: notation and definition

- **Multi-Objective Feature Weigthing Problem (FWP).**

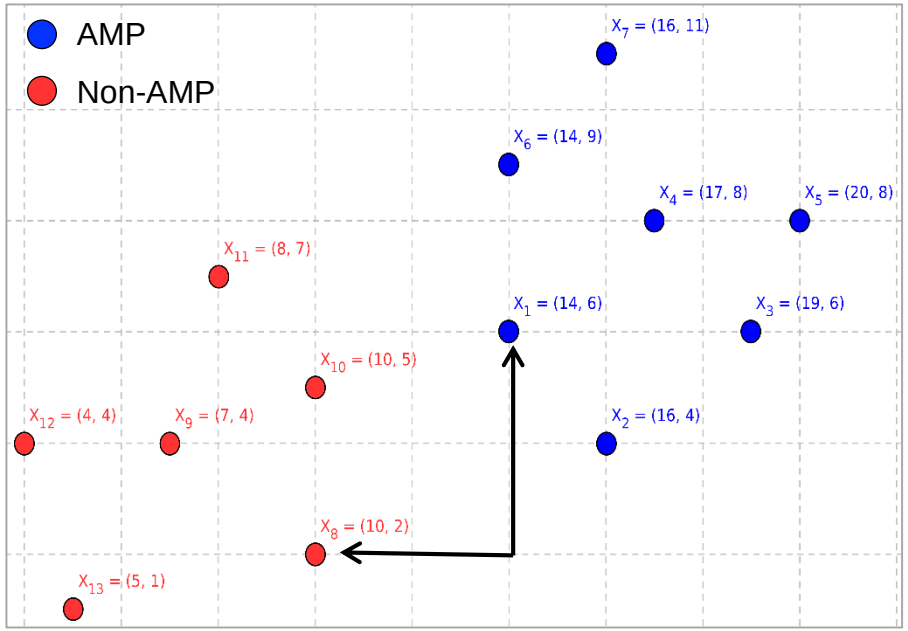
Given an input set of m candidate features and a labeled training dataset $\mathcal{D}$ with n instances, find a weight assignment for each feature such that intra-class and inter-class distances are optimized.

- **The weight vector**
  $\boldsymbol{w} = [w_1, \ldots, w_m]^T$ specifies the rescaling value of each feature.

$$\mathrm{w_i} = \begin{cases} [1, A], & if \ \ feature\ X_i\ is\ selected \\ 0, & if\ feature\ X_i\ is\ rejected \end{cases}$$

- **Weighted Manhatthan distance**
  Given two datapoints $\mathbf{x_p}, \mathbf{x_q}$ and a weight vector $\boldsymbol{w}$

$$d(\mathbf{w}, \mathbf{x_p}, \mathbf{x_q}) = \sum_{i=1}^{m} w_i |x_{pi} - x_{qi}| = \boldsymbol{w}^T |\boldsymbol{x_p} - \boldsymbol{x_q}|$$



S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

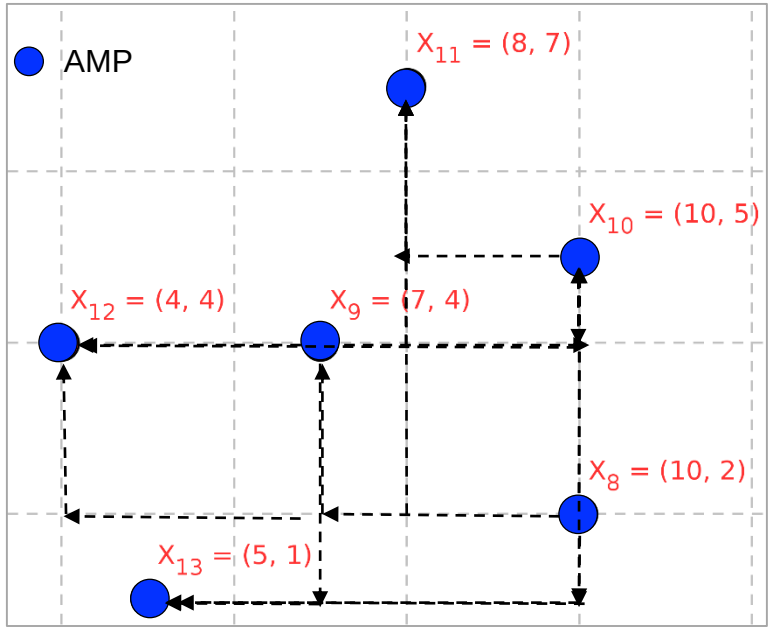
Problem Statement: notation and definition

- **Multi-Objective Feature Weigthing Problem (FWP).**

Given an input set of m candidate features and a labeled training dataset $\mathcal{D}$ with n instances, find a weight assignment for each feature such that intra-class and inter-class distance ares optimized.

- **Intra-class distance for the class of interest (AMPs)**

$$D_{intra}(\boldsymbol{w}, \mathcal{D}) = \sum_{p=1}^{n-1} \sum_{\substack{q=p+1 \\ y_p, y_q = AMP}}^{n} d(\boldsymbol{w}, \boldsymbol{x_p}, \boldsymbol{x_q})$$



S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

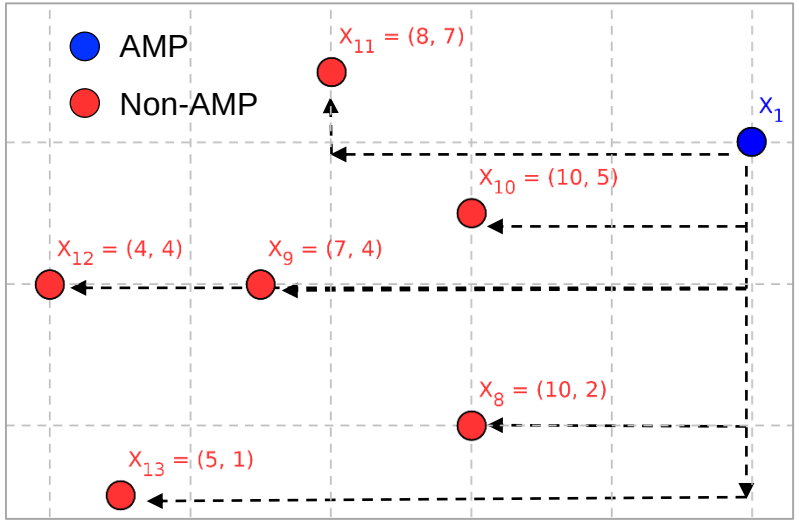
Problem Statement: notation and definition

- **Multi-Objective Feature Weigthing Problem (FWP).**

Given an input set of m candidate features and a labeled training dataset $\mathcal{D}$ with n instances, find a weight assignment for each feature such that intra-class and inter-class distances are optimized.

- **Inter-class distance is defined as:**

$$D_{inter}(\boldsymbol{w}, \mathcal{D}) = \sum_{p=1}^{n-1} \sum_{\substack{q=p+1 \\ yp \neq yq}}^{n} d(\boldsymbol{w}, \boldsymbol{x_p}, \boldsymbol{x_q})$$



S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

Problem Statement: a multi-objective approach

Feature weighting probem

- Let $\mathcal{D}$ be a training dataset with n instances and m candidate input feature, the multi-objective feature weigthing problem can be stated as:

$$\underset{\boldsymbol{w}}{\text{minimize}}\, F\,(\boldsymbol{w}) = [f_1\,(\boldsymbol{w}), f_2\,(\boldsymbol{w})]^{\boldsymbol{T}}$$

$$subject\ to\ \ w_i \in \{0\} \cup [1, \mathcal{A}]\ \ i = 1, \ldots, m,$$

Where $f_1(\boldsymbol{w}) = D_{intra}(\boldsymbol{w}, \mathcal{D}) + \dfrac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m},$

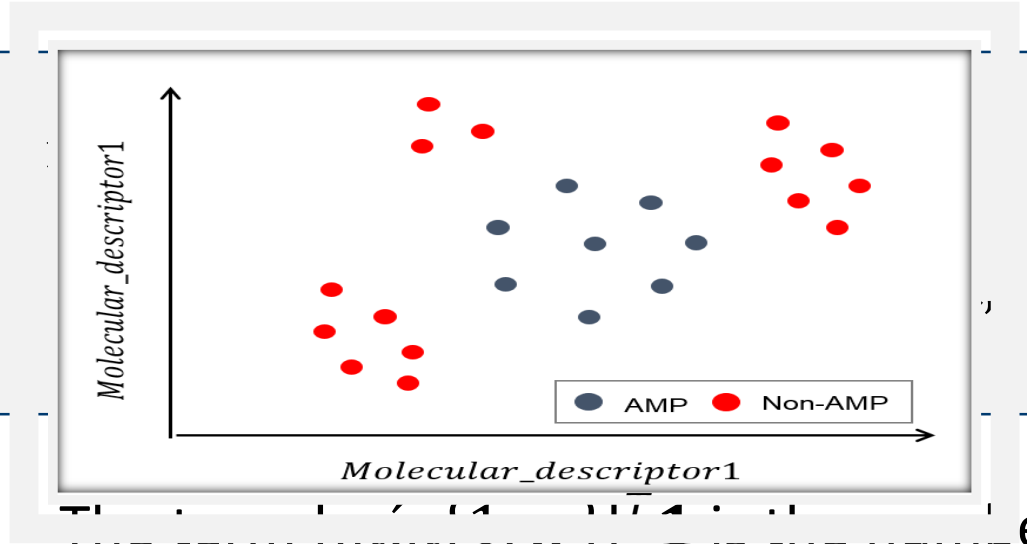
$f_2(\boldsymbol{w}) = -D_{inter}(\boldsymbol{w}, \mathcal{D}) + \dfrac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m}.$

- The term $[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}$ is the number of weights that are different from zero.

S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

Problem Statement: a multi-objective approach

We only minimize the intra-class distance of the AMP set, because the non-AMPs could contain peptide with different biological activity.



Where $f_1(\boldsymbol{w}) = D_{intra}(\boldsymbol{w}, \mathcal{D}) + \frac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m}$,

$f_2(\boldsymbol{w}) = -D_{inter}(\boldsymbol{w}, \mathcal{D}) + \frac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m}$.

er of weights that are different from zero.

To solve the MO optimization problem, we follow a similar approach to the one presented by Paul and Das (2015).

S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

Problem Statement: a multi-objective approach

The number of weights than are different from zero is used as a tiebreaker criterion for weight vectors with the same distances

$$\underset{\boldsymbol{w}}{\text{minimize}}\, F\,(\boldsymbol{w}) = [f_1\,(\boldsymbol{w}), f_2\,(\boldsymbol{w})]^{\boldsymbol{T}}$$

$$subject\ to\ \ w_i \in \{0\} \cup [1, \mathcal{A}]\ \ i = 1, \ldots, m,$$

Where $f_1(\boldsymbol{w}) = D_{intra}(\boldsymbol{w}, \mathcal{D}) + \dfrac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m},$

$f_2(\boldsymbol{w}) = -D_{inter}(\boldsymbol{w}, \mathcal{D}) + \dfrac{[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}}{m}.$

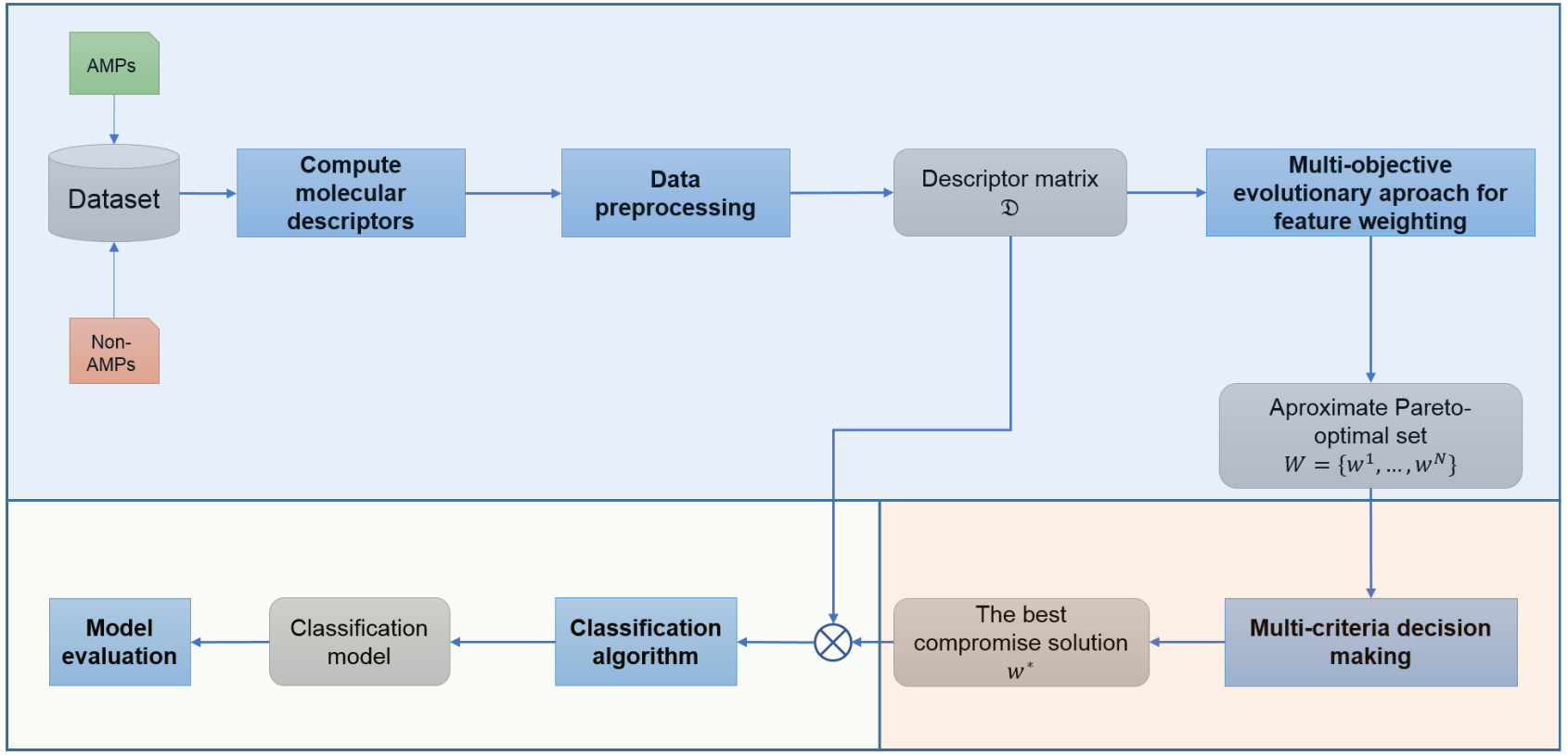
- The term $[mín\{1, \boldsymbol{w}\}]^T \mathbf{1}$ is the number of weights that are different from zero.

To solve the MO optimization problem, we follow a similar approach to the one presented by Paul and Das (2015).

S. Paul and S. Das, "Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach," Pattern Recognition Letters, vol. 65, pp. 51–59, 2015.

Materials and Methods

1) Solve the multi-objective problem



3) Generate a classifcation model

2) Choose a few weight vectors.

Figure 1. The overall scheme of the feature weighting framework. The rectangles with bold texts represent processes, and the rounded rectangles represent the inputs and outputs of processes.

Materials and Methods: training dataset

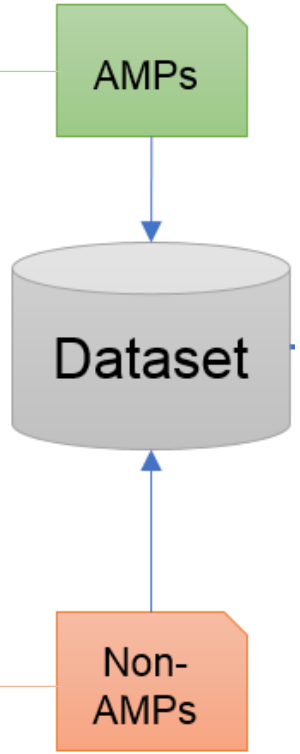
- We used a training dataset of 115 AMP and 116 non-AMP sequences, the methodology to construct this dataset was introduced by Fernandes et al. 2012

AMPs

- Retrieved from a Antimicrobial Peptide database.
- Experimentally validated sequences.
- All sequences have lengths between 10 and 100 residues.

Non-AMPs

- Retrieved from the Protein Data Bank (PDB).
- All sequences have a length between 10 and 100 residues.
- Sequences that are not predicted as membranes or extracellular proteins.



F. C. Fernandes, D. J. Rigden, and O. L. Franco, “Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application,” Peptide Science, vol. 98, no. 4, pp. 280–287, 2012.

Materials and Methods: computing molecular descriptors

- To codify the peptide sequences into numerical value, we used two different packages:
  - **Tango software**: α-helix propensity, β-sheet propensity, turn structure propensity, and in vitro aggregation.
  - **In-house Java Peptide Descriptor from Sequences (JPEDES):** 0D and 1D molecular descriptor.
- Each peptide sequence is converted into a 271-dimensional vector.

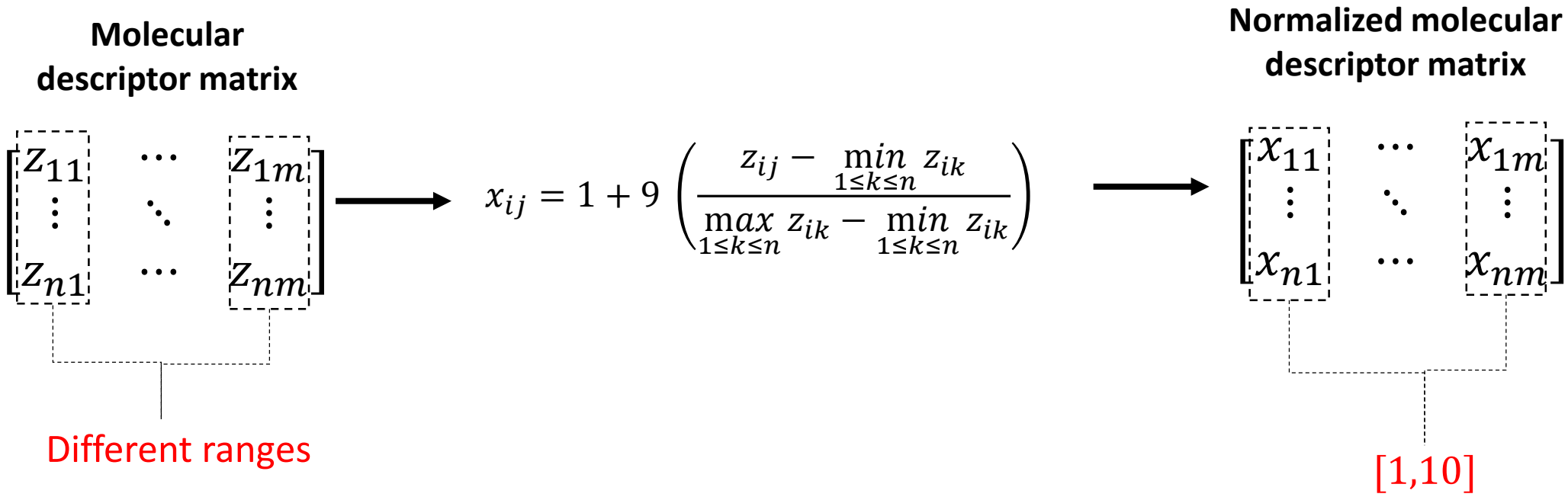
$$S = (AA_1, AA_2, \dots AA_l) \Rightarrow Z = [z_1, z_2, \dots, z_{271}]^{\mathrm{T}}$$

- F. Rousseau, J. Schymkowitz, and L. Serrano, "Protein aggregation and amyloidosis: confusion of the kinds?" Current opinion in structural biology, vol. 16, no. 1, pp. 118–126, 2006.
- A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins," Nature biotechnology, vol. 22, no. 10, pp. 1302–1306, 2004.
- R. Linding, J. Schymkowitz, F. Rousseau, F. Diella, and L. Serrano, "A comparative study of the relationship between protein structure and-aggregation in globular and intrinsically disordered proteins," Journal of molecular biology, vol. 342, no. 1, pp. 345–353, 2004.

Materials and Methods: data preprocessing

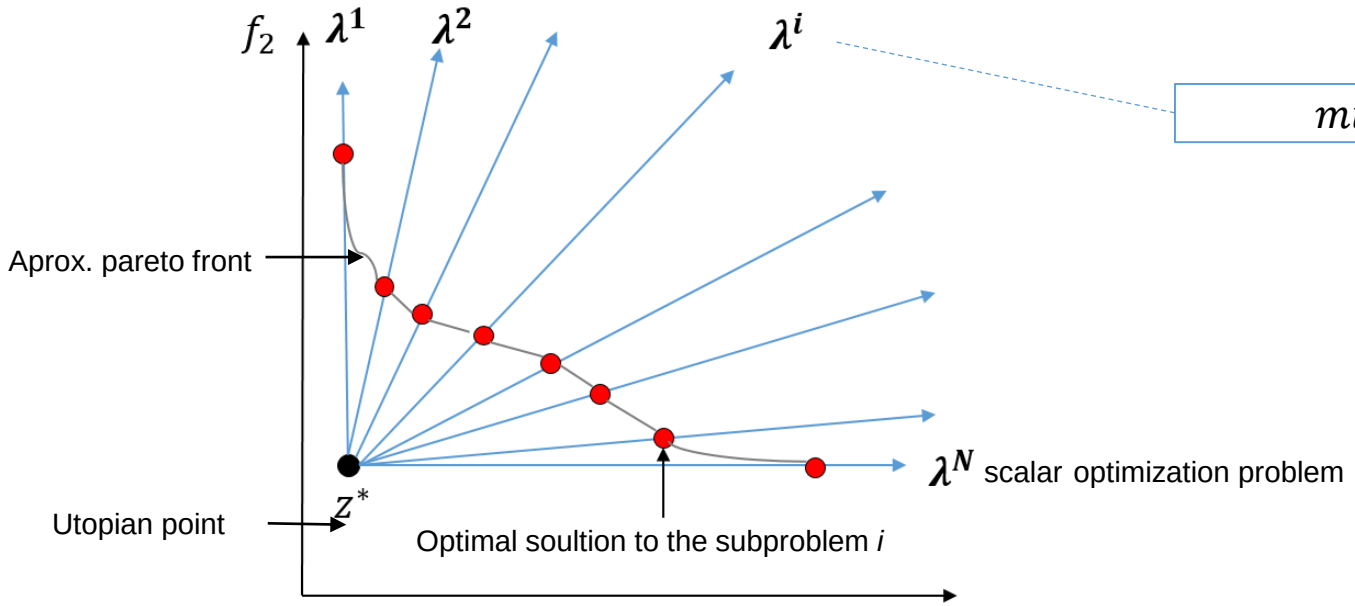
Normalization

- Molecular descriptors measured over the training data might have different ranges.



Materials and Methods: Multi-Objective Evolutionary Algorithm

- The multi-objetive evolutionary algorithm based on decomposition (MOEA/D-DE) proposed by Zhang Q. and Li. H. (2007).
- MOEA/DE outperforms NSGA-II on continuous MO optimization problems.
- MOEA/D-DE **decomposes** the **MO optimization problem** into N **scalar optimization problem** by using the Tchebycheff approach.



- Tchebycheff approach

$$min\ g^{te}(\boldsymbol{x}\,\boldsymbol{\lambda^i}, \boldsymbol{z}^*) = ma\,x_{1\leq i\leq m}\{\lambda_i|\ f_i(x) - z_i^*|\}$$

The **N subproblems** are solved in parallel.

Zhang, Q., & Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, *11*(6), 712-731.

Multi-Objective Evolutionary approach for feature weigthing (MOEA-FW)

- The MOEA/D algorithm offers a set of approximate N optimal solution.

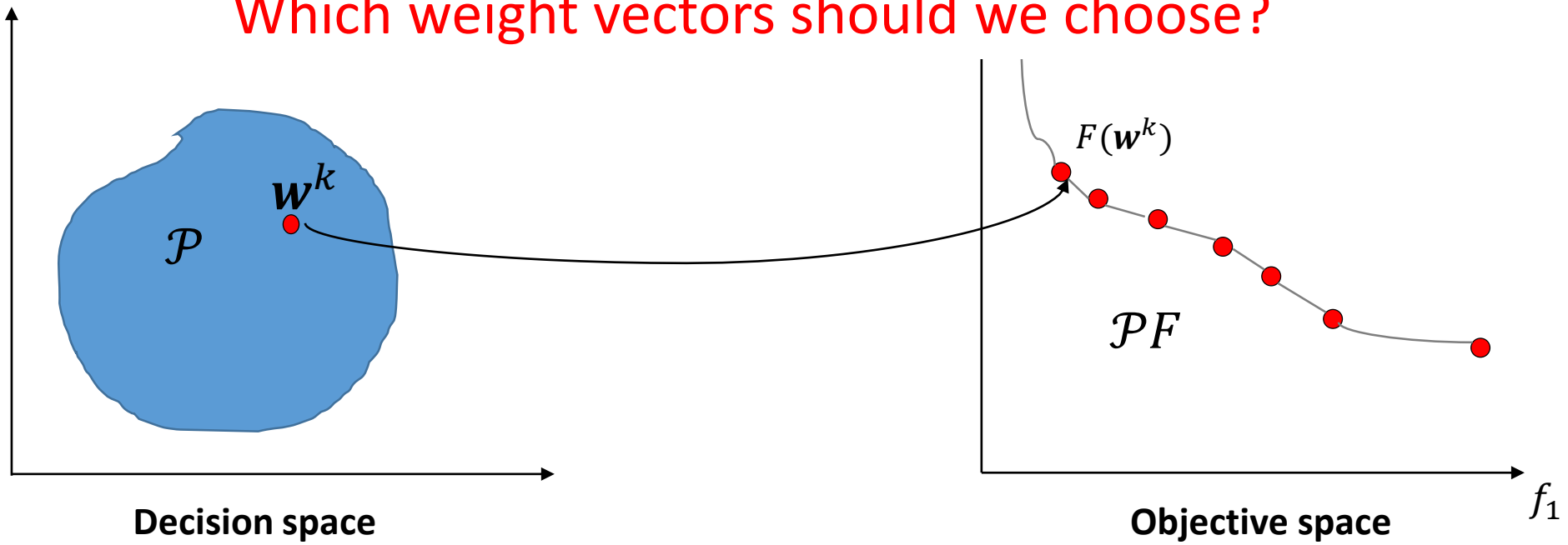
Approximated Pareto set

$\mathcal{P}^* = \{\boldsymbol{w}^1, \dots, \boldsymbol{w}^N\}$

Approximated Pareto front

$PF = \{F(\boldsymbol{w}^1), \dots, F(\boldsymbol{w}^N)\}$

Which weight vectors should we choose?



MOEA-FW: multi-criteria decision making approach

- **Step 1:** measure the degree of satisfaction
  $\boldsymbol{\mu}^k = [\mu_1^k, \mu_2^k]^T$

$$\mu_i^k = \begin{cases} 1 & \text{if } f_i(\mathbf{w^k}) = f_i^{\min}, \\ \frac{f_i^{\max} - f_i(\mathbf{w^k})}{f_i^{\max} - f_i^{\min}} & \text{if } f_i^{\min} < f_i^k < f_i^{\max}, \\ 0 & \text{if } f_i(\mathbf{w^k}) = f_i^{\max}, \end{cases}$$

- **Step 2:** let a weight vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]^T$ used the weighted sum approach to combine μ_1 and μ_2 in a single number.

$$g^{bcs}(\mu|\lambda_1) = \lambda_1\mu_1 + (1-\lambda_1)\mu_2$$

- **Step 3:** find the highest weighted sum g^{bcs}

$$k^* = \underset{k\in[1,N]}{\arg\max} \quad g^{bcs}(\mu^{\mathbf{k}}|\lambda_1)$$

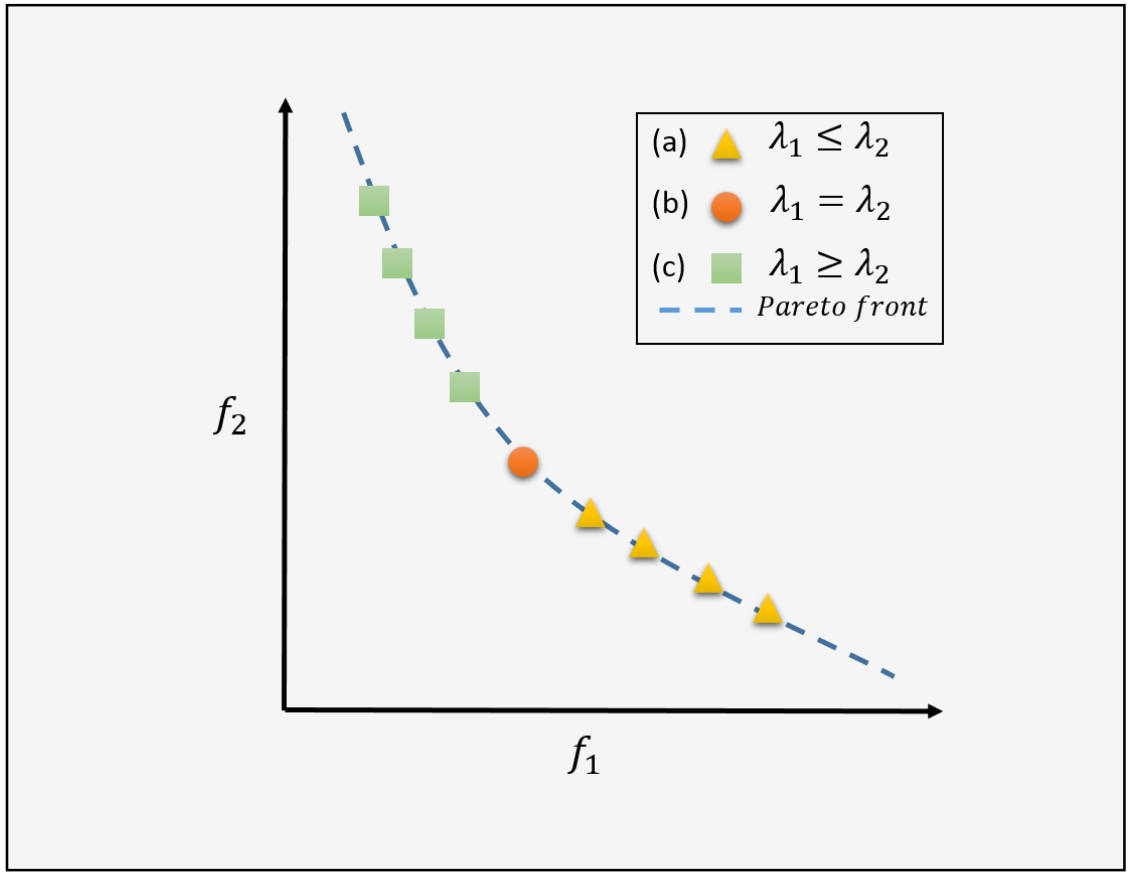


Illustration of the weighted sum approach. (a) f_1 is less important than f_2. (b) f_1 is equally important as f_2. (C) f_2 is less important than f_1.

MOEA-FW: Multi-criteria decision making approach

- We selected five of the best compromise (g^{bcs}) using λ_1 equal to 0.4, 0.45, 0.5, 0.55 and 0.60.
- Each best compromise solution was applied to dataset $\mathcal{D}$ as follows:

Normalized molecular descriptor matrix $\boldsymbol{w}^{k*}$ **Weighted molecular descriptor matrix**

$$\mathcal{D} = \left[\begin{array}{cccc|c} x_{11} & x_{12} & \cdots & x_{1m} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2m} & y_2 \\ \vdots & & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} & y_n \end{array}\right] \quad \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \quad \Rightarrow \quad \hat{\mathcal{D}} = \left[\begin{array}{cccc|c} w_1x_{11} & w_2x_{12} & \cdots & w_mx_{1m} & y_1 \\ w_1x_{21} & w_2x_{22} & \cdots & w_mx_{2m} & y_2 \\ \vdots & & \ddots & \vdots & \vdots \\ w_1x_{n1} & w_2x_{n2} & \cdots & w_mx_{nm} & y_n \end{array}\right]$$

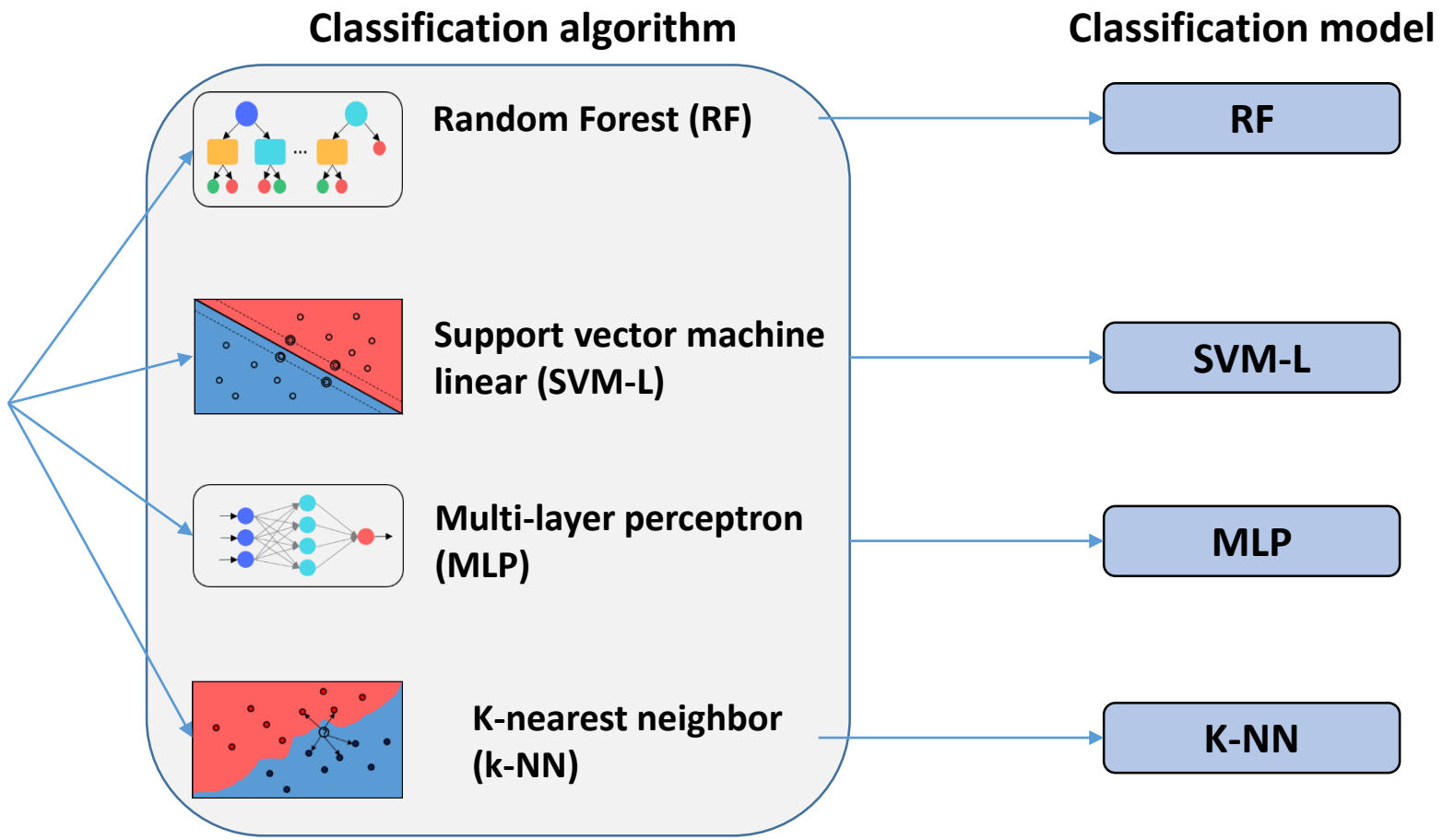
The rejected descriptors corresponds to columns whose values are zero and these columns were deleted.

Materials and Methods : classification algorithms

- For each weighted molecular descriptor matrix $\widehat{\mathcal{D}}$, we build four classification models.

Weighted molecular descriptor matrix

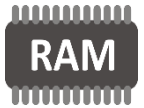
$$\hat{\mathcal{D}} = \left[\begin{array}{cccc|c} w_1x_{11} & w_2x_{12} & \cdots & w_mx_{1m} & y_1 \\ w_1x_{21} & w_2x_{22} & \cdots & w_mx_{2m} & y_2 \\ \vdots & & \ddots & \vdots & \vdots \\ w_1x_{n1} & w_2x_{n2} & \cdots & w_mx_{nm} & y_n \end{array}\right]$$



Experiments and Results: experimental setup

Set-up enviroment

2.40 GHz

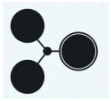


6 GB

8.1

8.0.2.

- **MOEA Framework 2.1:** to solve the multiobjective optimization problems:



- **WEKA library 3.8.0:** classification algortihms (RF, KNN, MLP, SVM-L)

TABLE I
PARAMETER SETTINGS FOR THE MOEA-FW

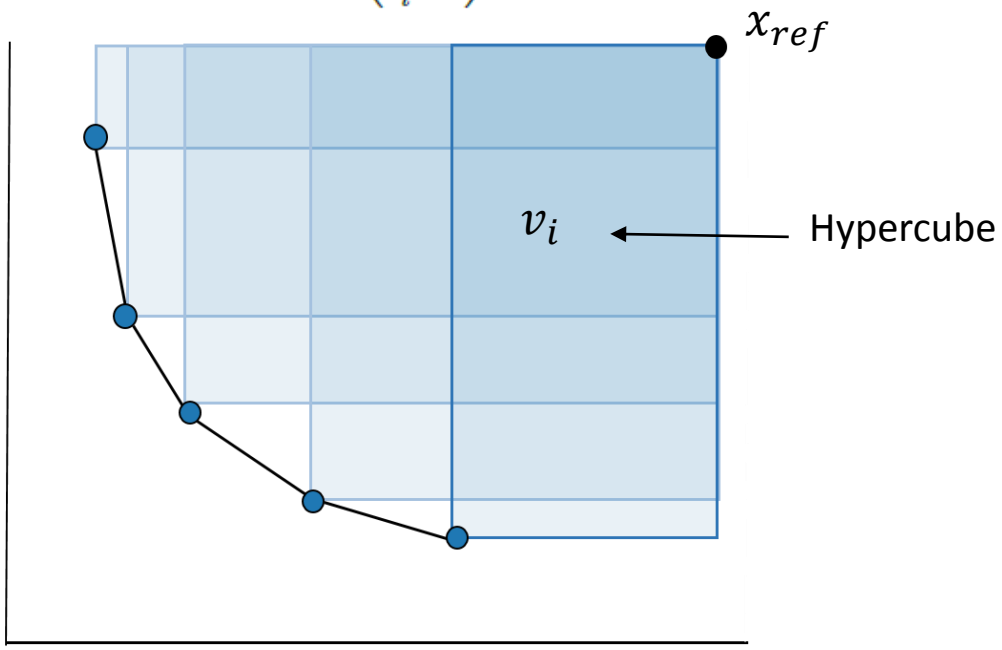
Symbol	Value	Description
Control parameters in DE crossover and polynomial mutation		
CR	1.0	The crossover rate
F	0.5	The Scaling factor
η	20	The distribution index for polynomial mutation
p_m	$\frac{1}{n}$	The mutation rate
Run time and stop condition		
N_{pop}	500	The population size
N_{gen}	1000	The maximum number of generations
N_r	30	The number of trials
Control parameters in MOEA/D-DE		
T	20	The size of neighborhood
δ	0.9	The probability for parents selection from the neighborhood
n_r	2	The maximum number of solutions replaced by each offspring

- Q. Zhang and H. Li, "Moea/d: A multiobjective evolutionary algorithm based on decomposition," IEEE Transactions on evolutionary computation, vol. 11, no. 6, pp. 712–731, 2007.

Experiments and Results: performance evaluation

Hypervolume I_H

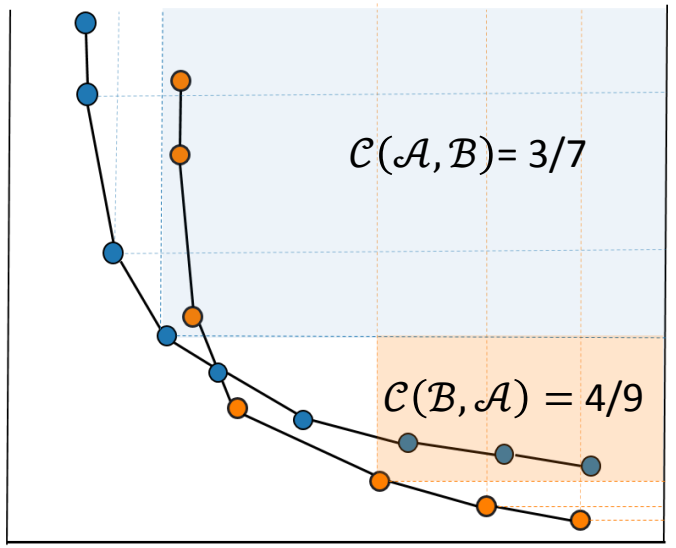
$$I_H = Vol\left(\bigcup_i^N v_i\right)$$



- Higher values of I_H indicates better results.

Coverage indicator $\mathcal{C}(\mathcal{A},\mathcal{B})$

$$C(\mathcal{A},\mathcal{B}) = \frac{|\{\mathbf{w} \in \mathcal{B} | \exists \mathbf{v} \in \mathcal{A} : \mathbf{v} \succeq \mathbf{w}\}|}{|\mathcal{B}|}$$

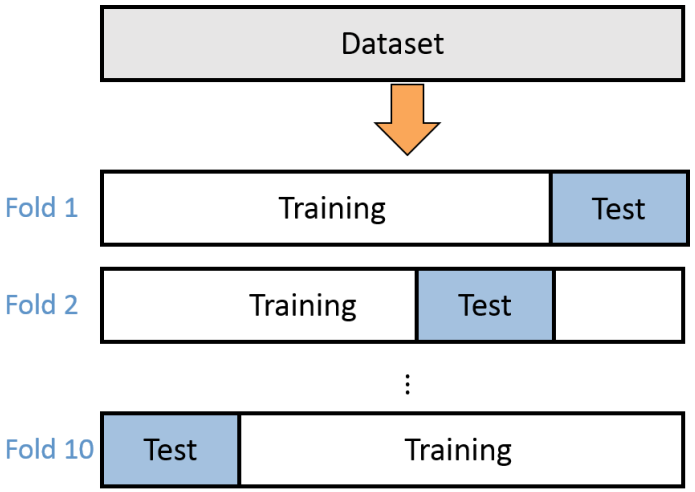


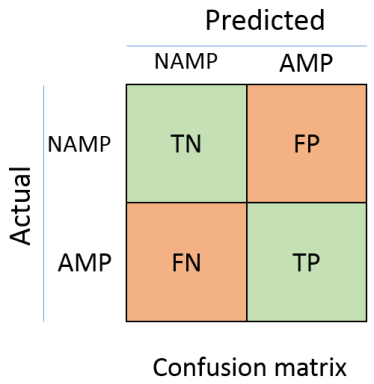
- $\mathcal{C}(\mathcal{A},\mathcal{B}) = \mathbf{1}$ means that all **solution in B are dominated by** at least one solution in **A**.

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," IEEE Transactions on evolutionary computation, vol. 7, no. 2, pp. 117–132, 2003.

Experiments and Results: performance evaluation

10-Fold Cross-Validation





Confusion matrix

- Accuracy

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

- Mattews correlation coefficent

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TN + FP)(TP + FP)(TN + FN)}}$$

- Precision

$$Precision = \frac{TP}{TP + FP}$$

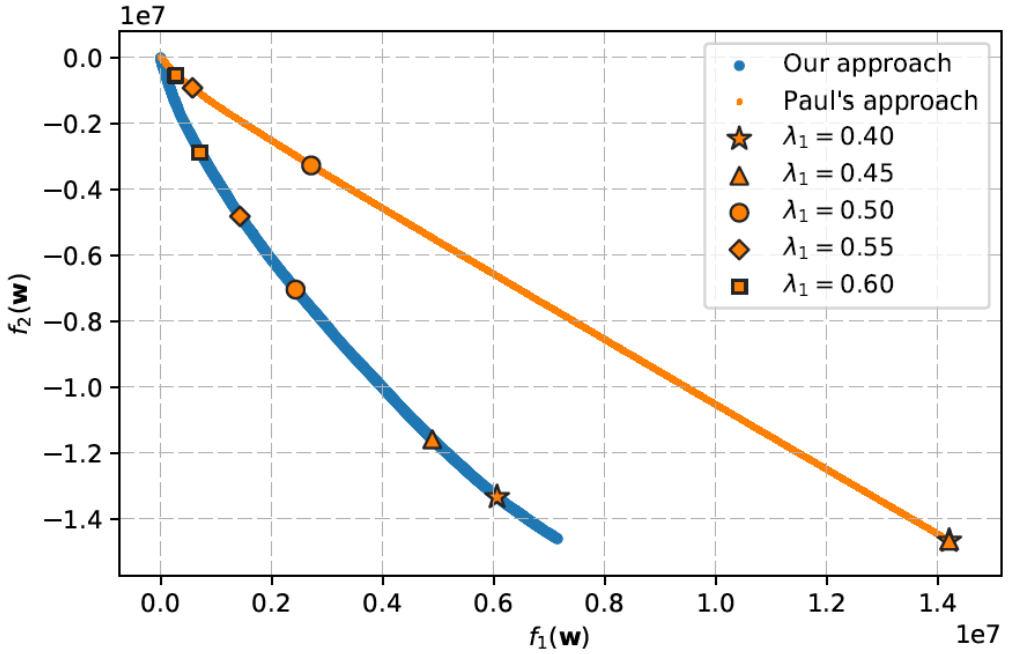
- Recall

$$Recall = \frac{TP}{TP + FN}$$

A Higher score denote a more predictive model.

Experiments and Results

The consolidated non-dominated front after 30 runs of the MOEA-FW and the Paul et al. [8] approach for Fernandes' dataset. Each orange point represents the best compromise solution given λ_1 .



Measure	MOEA-FW	Paul's approach
I_H	**0.60**	0.52
$\mathcal{C}$	**0.99**	0.00

The consolidated non-dominated front obtained by our approach are better than the ones generated by Paul's approach.

Experiments and Results

- Performance comparison of the best compromise solution given λ_1, generated by MOEA-FW and Paul's approach for four different classification algorithm.

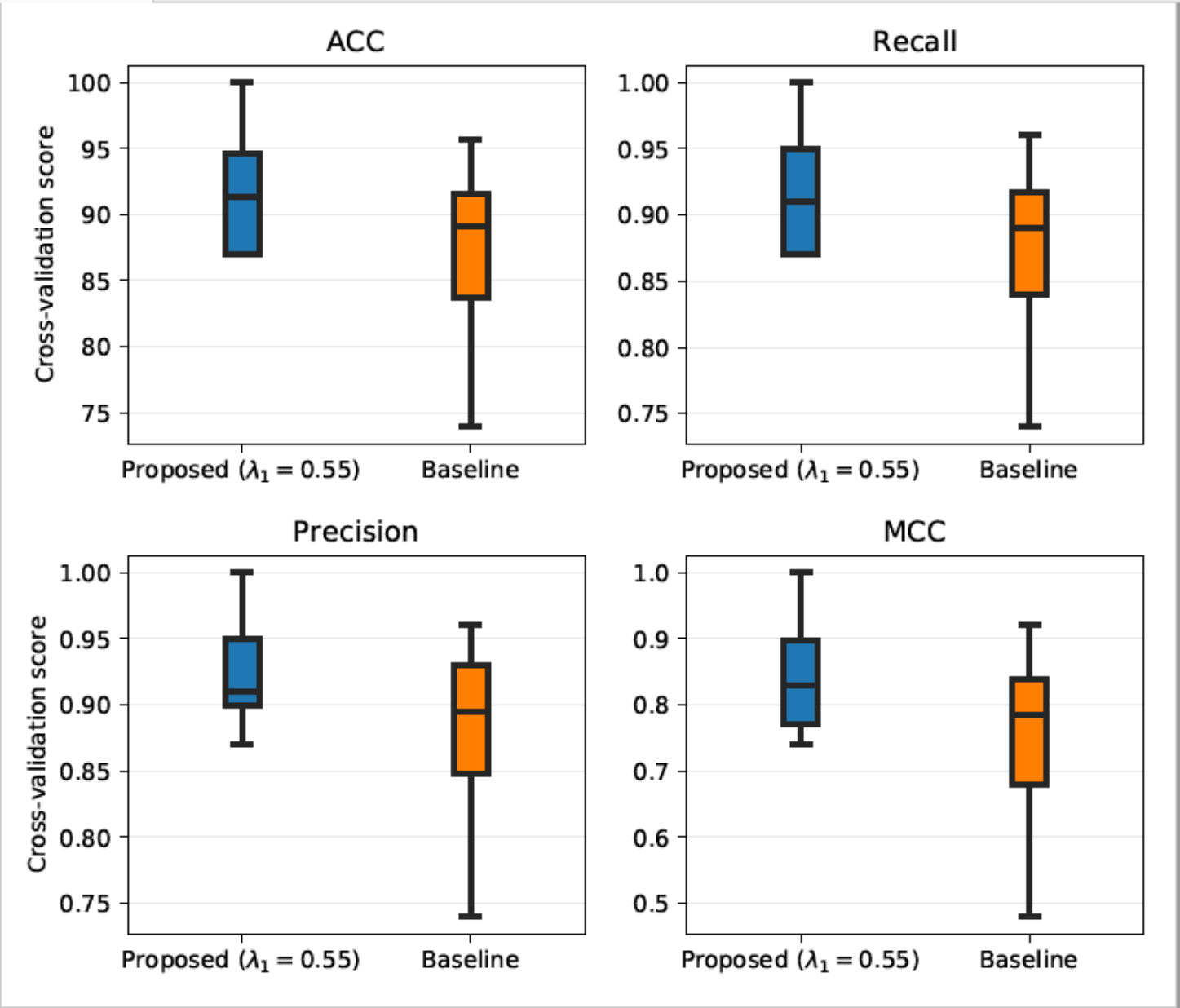
Method	Best compromise solution	Candidate input features	Num. of selected features	Dim. reduction (%)	Average classification accuracy (%)				
					RF[a]	K-NN[a]	MLP[a]	SVM-L[a]	Average
MOEA-FW	$\lambda_1 = 0.40$	271	222	18.08	89.18	86.58	87.01	87.45	87.56
	$\lambda_1 = 0.45$		187	31.00	87.45	86.58	87.45	87.01	87.12
	$\lambda_1 = 0.50$		116	57.20	88.75	**89.18**	**89.18**	88.75	88.97
	$\lambda_1 = 0.55$		87	67.90	**91.34**[b]	88.75	87.88	89.18	**89.29**
	$\lambda_1 = 0.60$		54	80.07	88.75	88.31	85.28	88.31	87.66
Paul's approach	$\lambda_1 = 0.40$	271	268	1.21	88.28	85.71	87.01	72.25	83.31
	$\lambda_1 = 0.45$		268	1.21	88.28	85.71	87.01	72.25	83.31
	$\lambda_1 = 0.50$		13	95.21	87.41	88.28	89.17	**89.61**	88.62
	$\lambda_1 = 0.55$		2	99.26	83.08	87.86	57.10	86.97	78.75
	$\lambda_1 = 0.60$		1	99.63	76.21	77.05	67.45	82.25	75.74

[a] Classification algorithm: RF=Random Forest; K-NN=k-Nearest Neighbor (k=11); MLP=Multi-layer Perceptron; SVM-L=Support Vector Machine-Linear.

[b] The bold values are the highest accuracy for a given classification algorithm.

Experiments and Results: our approach vs control

- On average, the MOEA-FW shows a significant improvement of the classifier over baseline.
- With method proposed, we obtained a precision and recall of 0.922, MCC of 0.83, and an ACC of 91.34%



Conclusions and Future work

- This work modeled the molecular descriptors weighting problem as a multi-objective (MO) optimization problem to obtain a good peptide representation for the classification task.
- To solve this problem, a variant of a general methodology based on a multiobjective evolutionary algorithm (MOEA/D-DE) was introduced.
- The results show that the performance of a baseline classifier (all features) increases when using the descriptors selected by the MOEA-FW algorithm.
- To assesst the performance of MOEA-FW algorithm over high dimensional spaces.
- The obtained classifier is aimed at searching for AMPs in various transcriptomes.

Contact information

Jesus_Beltran2

abeltran@gmail.com

Thank you!

Feature weighting for antimicrobial peptides classification: a multi-objective evolutionary approach

Jesus A. Beltrán, Longendri Aguilera-Mendoza, Carlos A. Brizuela

Computer Sciences Deparment

CICESE Research Center

IEEE BIBM 2017- Kansas City, MO, USA.
November 14th, 2017