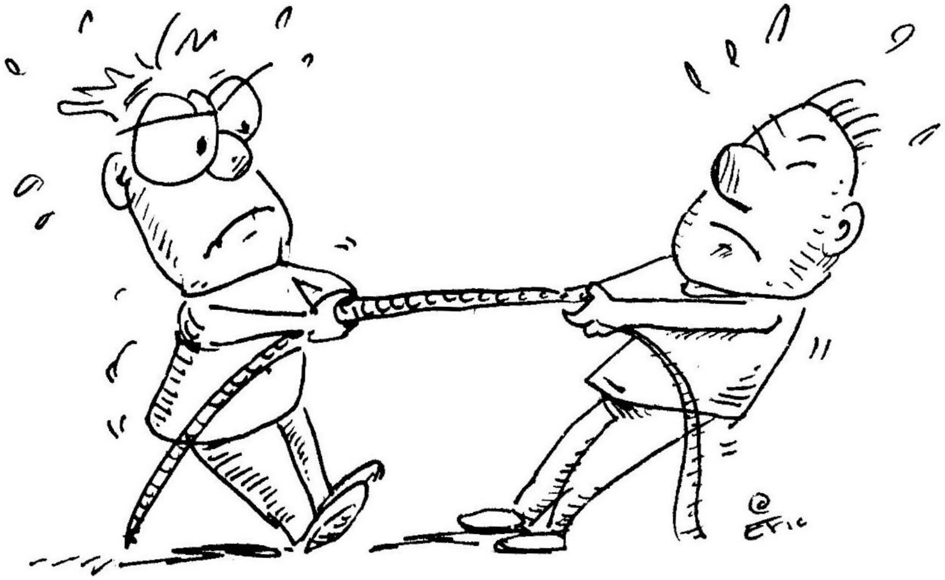


Hugo Jair Escalante

# MULTI-OBJECTIVE EVOLUTIONARY OPTIMIZATION: FEW APPLICATIONS

# Contents

- Single and multi objective optimization
- Multi-objective evolutionary algorithms (NSGA-II)
- Maximizing diversification of search results
- Prototype generation for classification
- Discussion

Multi-objective Evolutionary Algorithms: two applications

# SINGLE/MULTI OBJECTIVE OPTIMIZATION

# Mathematical optimization

From Wikipedia, the free encyclopedia

*For other uses, see Optimization (disambiguation).*

In mathematics, computer science, or management science, **mathematical optimization** (alternatively, **optimization** or **mathematical programming**) is the selection of a best element (with regard to some criteria) from some set of available alternatives.[1]

In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. The generalization of optimization theory and techniques to other formulations comprises a large area of applied mathematics. More generally, optimization includes finding "best available" values of some objective function given a defined domain, including a variety of different types of objective functions and different types of domains.

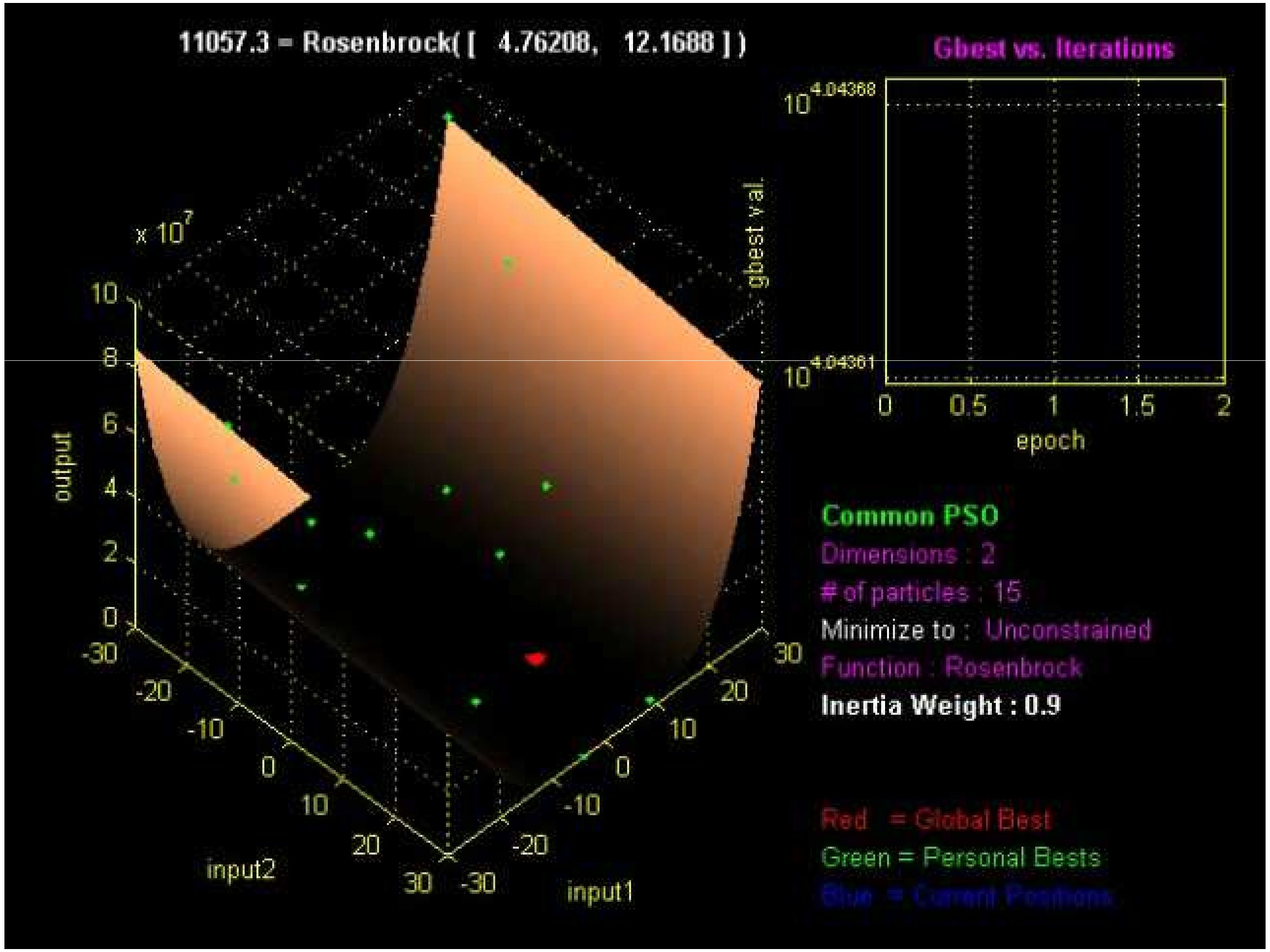
http://en.wikipedia.org/wiki/Mathematical_optimization

# Single-objective optimization

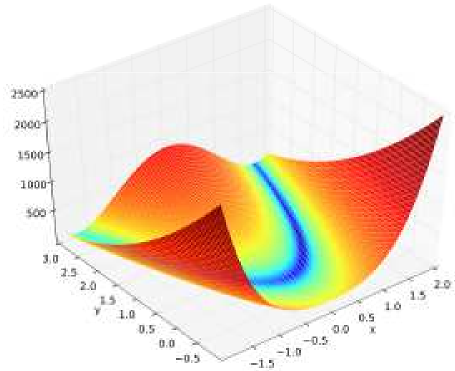
- A single-objective optimization problem can be defined as:

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & g_i(\mathbf{x}) \leq 0 \text{ for } i = \{1,\ldots,\mathrm{I}\} \\
& h_j(\mathbf{x}) = 0 \text{ for } j = \{1,\ldots,\mathrm{J}\} \\
& \mathbf{x}^l_k \leq \mathbf{x}_k \leq \mathbf{x}^u_k \text{ for } k = \{1,\ldots,\mathrm{n}\}
\end{aligned}
$$

# Single-objective optimization




Brian Birge's PSO demo for matlab

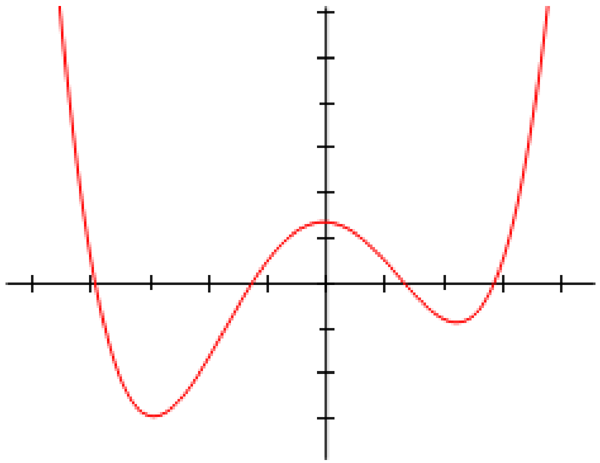


**Función: Rosenbrock**

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2.$$

# Single-objective optimization

- In this type of problems we want to find a solution **x*** associated to an extreme value of *f*. There are different types of methods for approaching this problems (e.g., gradient-based, simplex, heuristic, etc. )

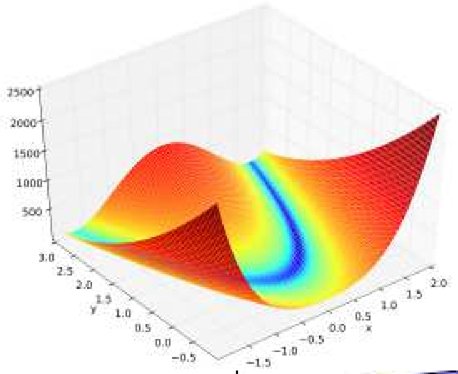


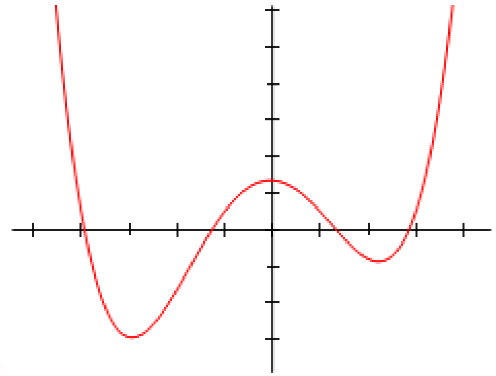
# Multi-objective optimization

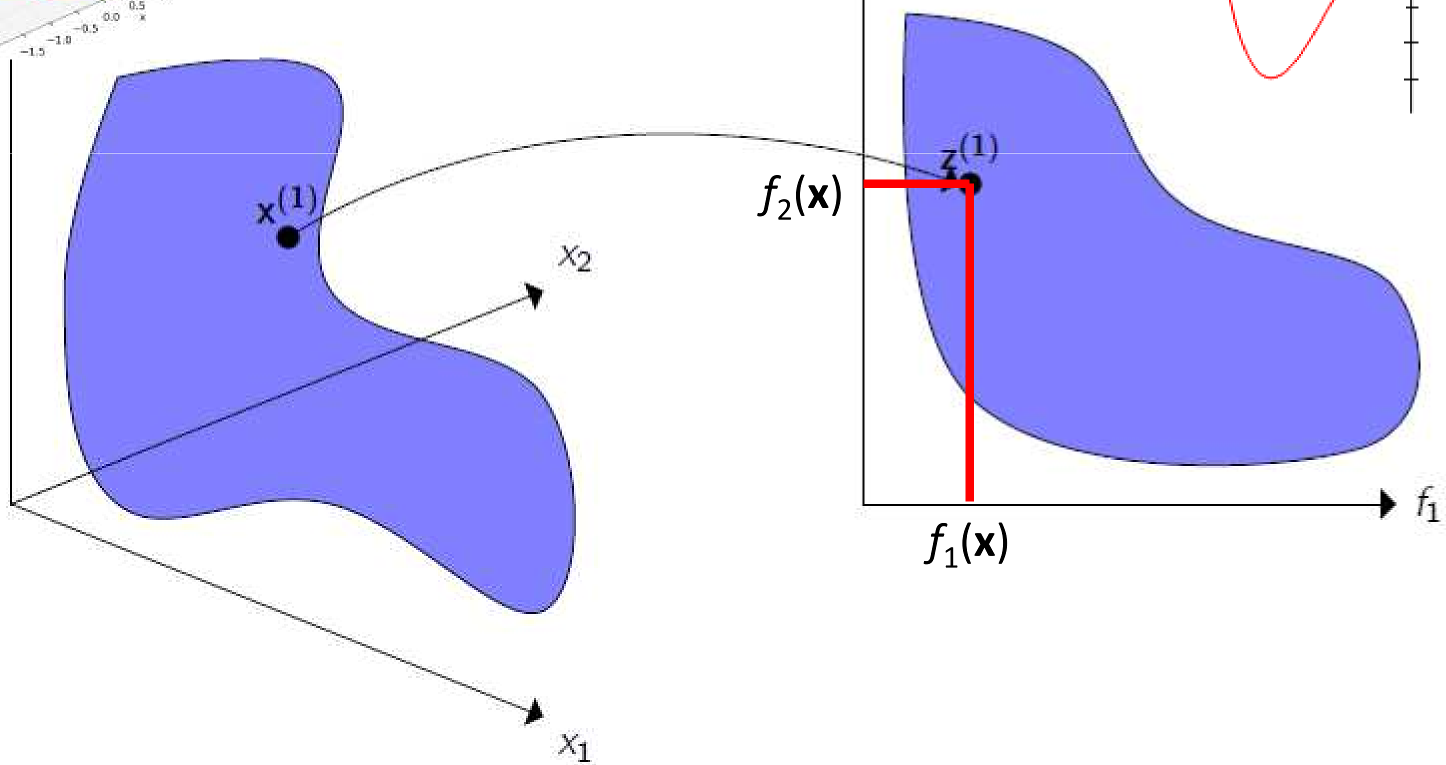
- A multi-objective optimization problem can be defined as:

$$
\begin{aligned}
\text{min} \quad & \boldsymbol{f}(\mathbf{x}) = \langle f_1(\mathbf{x}), \ldots, f_{\mathrm{N}}(\mathbf{x}) \rangle \\
\text{s.t.} \quad & g_i(\mathbf{x}) \leq 0 \text{ for } i = \{1,\ldots,\mathrm{I}\} \\
& h_j(\mathbf{x}) = 0 \text{ for } j = \{1,\ldots,\mathrm{J}\} \\
& \mathbf{x}^l_k \leq \mathbf{x}_k \leq \mathbf{x}^u_k \text{ for } k = \{1,\ldots,\mathrm{n}\}
\end{aligned}
$$

# Multi-objective optimization








**Decision space** **Objectives space**

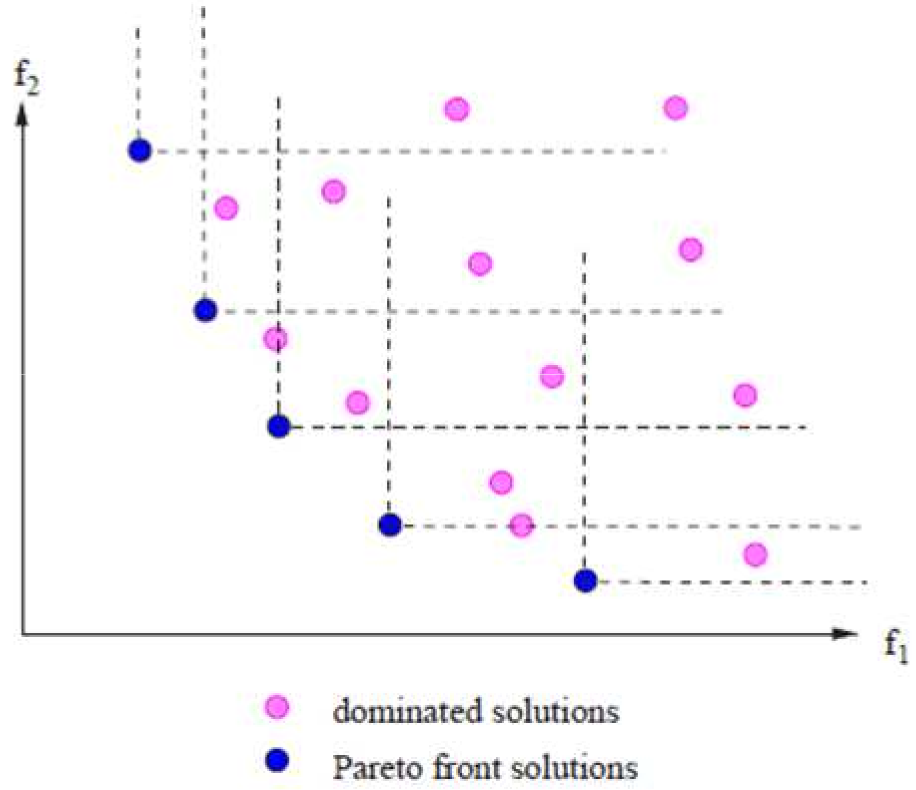
# Multi-objective optimization

- In MOO we deal with problems involving more than one objective. Hence a *good* candidate solution to solve the problem must return *acceptable* values for all of the considered objectives

- **Optimum in MOO:** The solution that represents the best tradeoff among the considered objectives

# Multi-objective optimization

- **Pareto optimality:** one of the most accepted notions of optimum
- (Some) MOO methods are based in the concept of dominance to determine if a solution is better than other

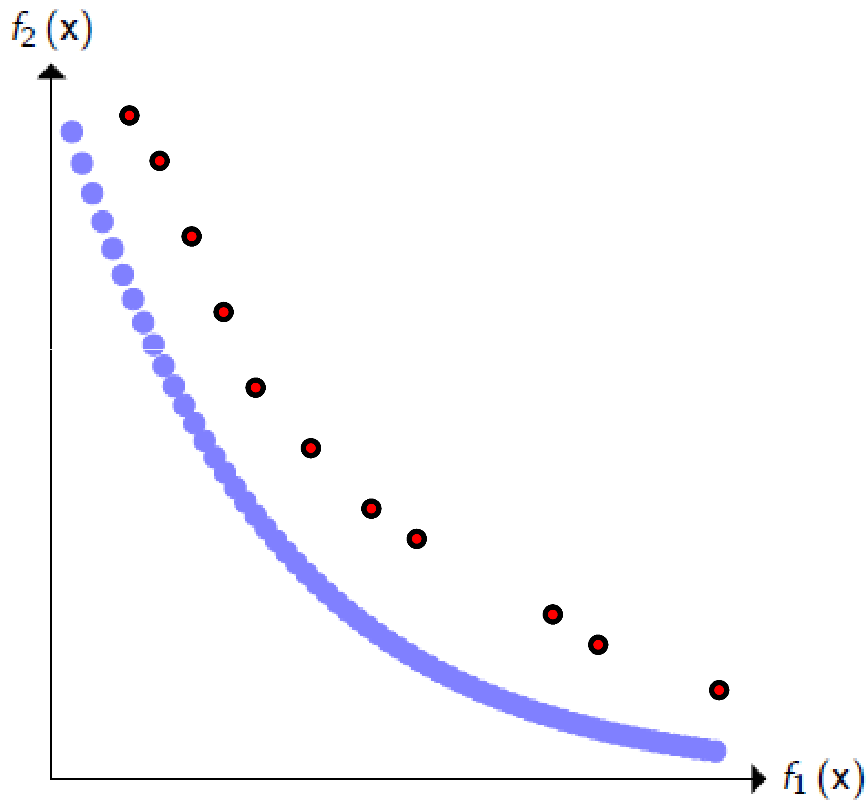



**Pareto dominance:** Solution $\mathbf{x}_1$ dominates $\mathbf{x}_2$ iff $\mathbf{x}_1$ is better than $\mathbf{x}_2$ in at least in one objective and it is not worse in the rest.

# Multi-objective optimization

- A solution **x*** is a Pareto optimum iff does not exist another solution **x´** such that **x´** dominate **x***

- **Problem:** The output of a MOO method is not a single solution but an approximation to the Pareto optimal set




No solution is better than another in the Pareto optimal set. Selecting a single solution is the job of the decision maker.

Multi-objective Evolutionary Algorithms: two applications

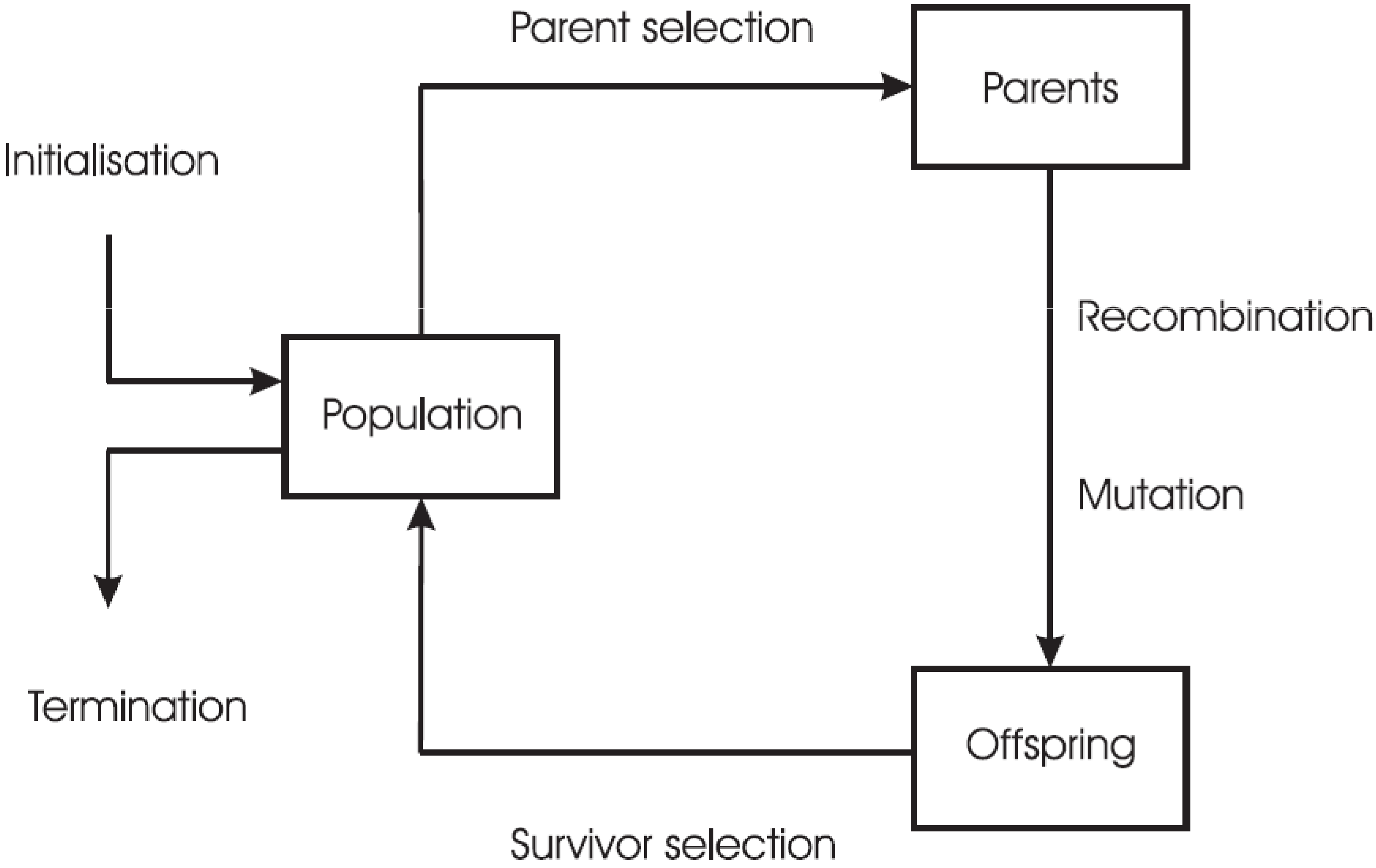
# MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS (NSGA-II)

# Evolutionary Computing

- EC Is the collective name for a range of **problem-solving** techniques based on principles of biological evolution, such as **natural selection** and **genetic inheritance**.

- These techniques are being increasingly widely applied to a variety of problems, ranging from practical applications in industry and commerce to leading-edge scientific research.

# Evolutionary Computing



# NSGA-II : (perhaps) the most used MOEA

**Algorithm 1** NSGA-II algorithm Deb et al. (2002).

**Require:** $N_{pop}$, $\mathbf{f}$, $g$

{$N_{pop}$ number of individuals (solutions); $g$ number of generations

$\mathbf{f} = \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$ objectives}

Initialize population $\mathcal{X}_0$

Evaluate objective functions $\mathbf{f} = \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

Identify fronts $\mathcal{F}_{1,\dots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

**while** $i = 1 < g$ **do**

  Create child population $\mathcal{Q}_i$ from $\mathcal{X}_i$ applying evolutionary operators.

  Evaluate objective functions $\mathbf{f}$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{Q}_i$

  Identify fronts $\mathcal{F}_{1,\dots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_i \cup \mathcal{Q}_i$

  $\mathcal{X}_{i+1} = \emptyset$; $j = 1$;

  **while** $|\mathcal{X}_{i+1}| < N_{pop}$ **do**

    $\mathcal{X}_{i+1} = \mathcal{X}_{i+1} \cup \mathcal{F}_j$; $j = j + 1$;

  **end while**

  Select the last individuals for $\mathcal{X}_{i+1}$ from $\mathcal{F}_j$ using crowding distance

**end while**

Non-dominated sorting

# NSGA-II : (perhaps) the most used MOEA

**Algorithm 1** NSGA-II algorithm Deb et al. (2002).

**Require:** $N_{pop}$, f, $g$

{$N_{pop}$ number of individuals (solutions); $g$ number of generations

f $= \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$ objectives}

Initialize population $\mathcal{X}_0$

Evaluate objective functions f $= \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

**while** $i = 1 < g$ **do**

  Create child population $\mathcal{Q}_i$ from $\mathcal{X}_i$ applying evolutionary operators.

  Evaluate objective functions f, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{Q}_i$

  Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_i \cup \mathcal{Q}_i$

  $\mathcal{X}_{i+1} = \emptyset$; $j = 1$;

  **while** $|\mathcal{X}_{i+1}| < N_{pop}$ **do**

    $\mathcal{X}_{i+1} = \mathcal{X}_{i+1} \cup \mathcal{F}_j$; $j = j + 1$;

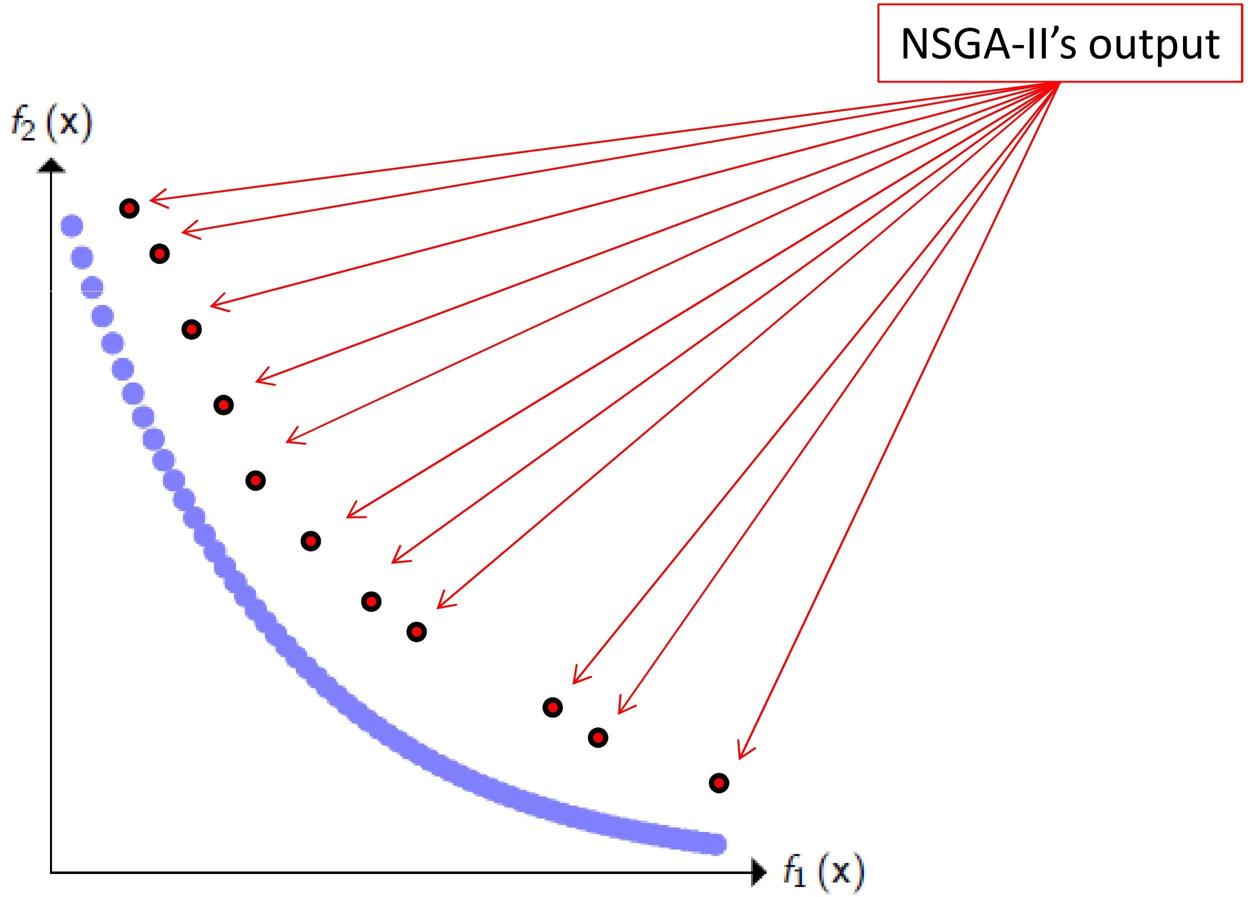
  **end while**

  Select the last individuals for $\mathcal{X}_{i+1}$ from $\mathcal{F}_j$ using crowding distance

**end while**

Crowding distance

# NSGA-II : (perhaps) the most used MOEA



Hugo Jair Escalante, Alicia Morales. **TIA-INAOE's approach for the 2013 Retrieving Diverse Social Images task.** *MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain, CEUR Workshop Proceedings,* Vol. 1043, 2013

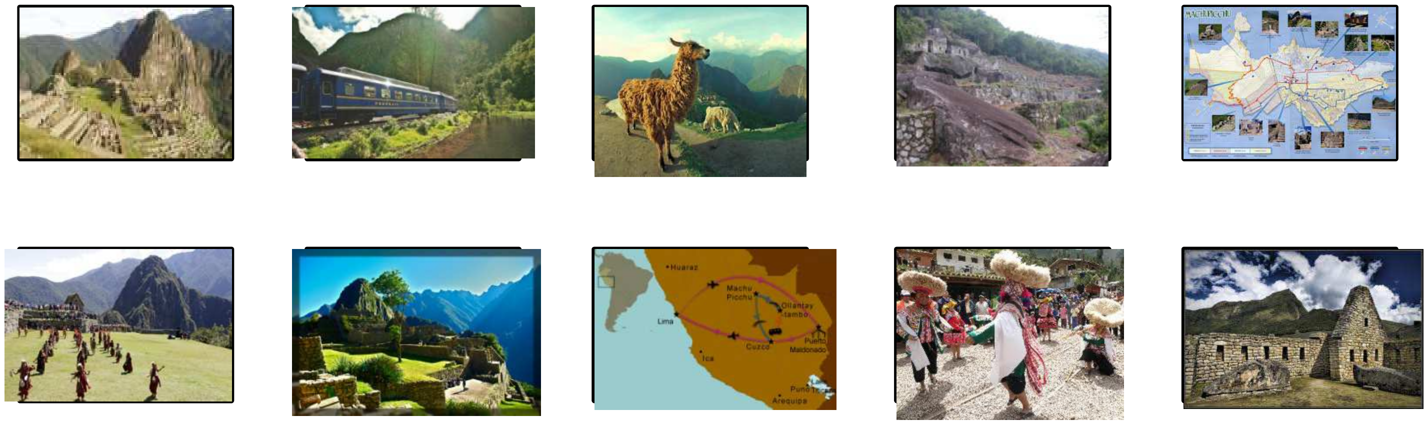
# MAXIMIXING VISUAL DIVERSITY OF IMAGE RETRIEVAL RESULTS

# Diversification of retrieval results in content-based image retrieval

- Given a list of images (relevant to a query), to re-rank the list such that the visual diversity of top-ranked images is maximized

# Diversification of retrieval results in content-based image retrieval

- Given a list of images (relevant to a query), to re-rank the list such that the visual diversity of top-ranked images is maximized



...

MediaEval Benchmark

MediaEval Benchmarking Initiative for Multimedia Evaluation

The "multi" in multimedia: speech, audio, visual content, tags, users, context

- **The 2013 Retrieving Diverse Social Images Task:** *Result diversification in social photo retrieval.* Organizers:
  – Provide data
    - Ranked lists of documents
    - Textual features, visual features, tags, comments, *etc.*
    - Evaluation

  – Evaluate participants

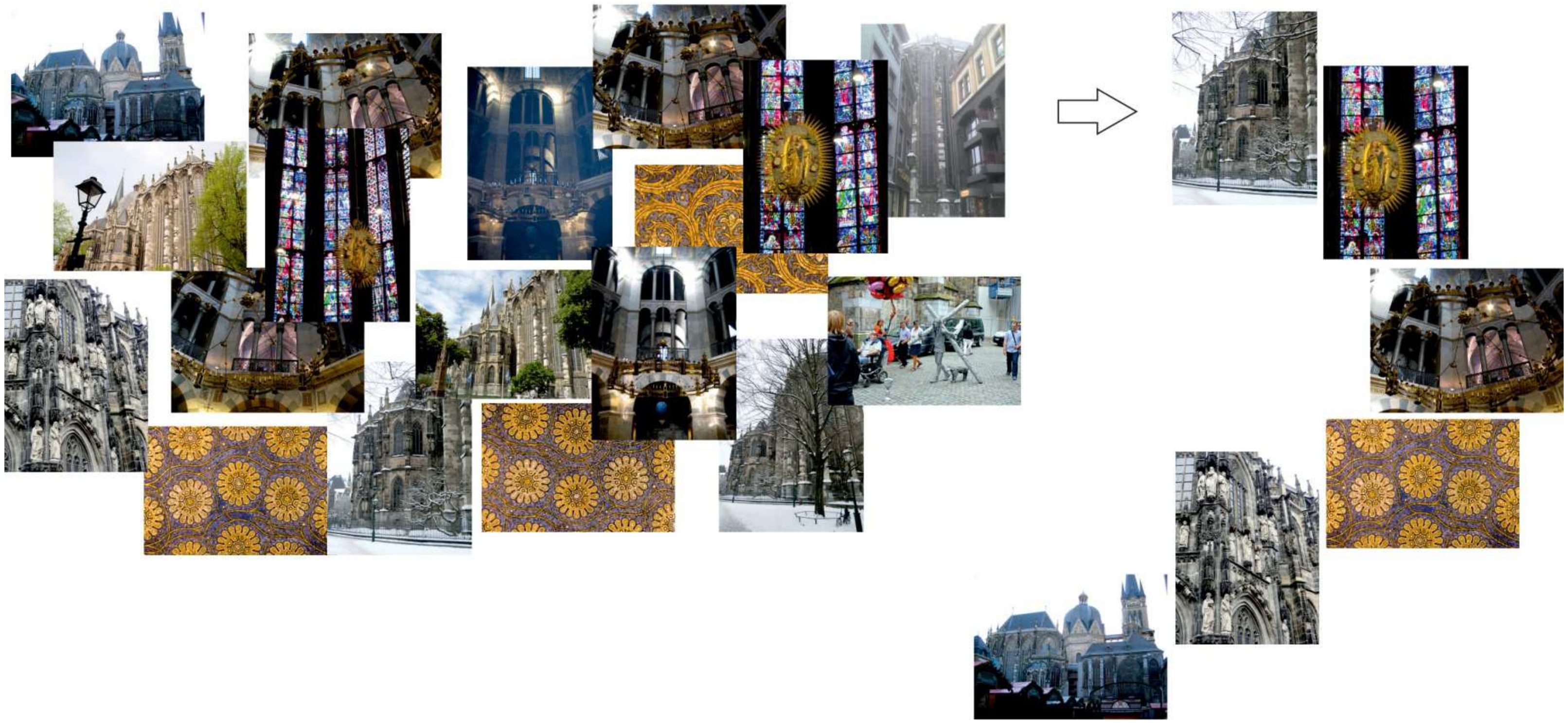
http://www.multimediaeval.org/mediaeval2013/diverseimages2013/

# MediaEval Benchmark

**MediaEval Benchmarking Initiative for Multimedia Evaluation**

The "multi" in multimedia: speech, audio, visual content, tags, users, context

Aachen Cathedral 50° 46′ 29″ N, 6° 5′ 4″ E



MediaEval Benchmark

MediaEval Benchmarking Initiative for Multimedia Evaluation

The "multi" in multimedia: speech, audio, visual content, tags, users, context

- Considered scenario:
  - A user searches for images of a specific location in social media (e.g., Flickr)
  - ***Text*** is used for searching
  - The user wants that images in the first positions of the list are ***visually*** diverse to each other
  - Additionally, all of the images must be relevant:
    - About the searched location (GPS coordinates)
    - No person in the image
    - …

flickr

**Casas Grandes**
**Chihuahua Mexico**

# Multi-objective optimization for result diversification

- **Idea:** to re-rank the list of images such that a tradeoff between relevance and diversity is maximized

# Multi-objective optimization for result diversification

- NSGA-II is used to approach the problem as follows:

Maximize $< \rho(S^0, S)\ ,\ \beta(S) >$

- Where:

$$\rho(S^0, S) = 1 - \frac{6}{n(n^2-1)} \sum_i dr_i(S^0, S)^2$$

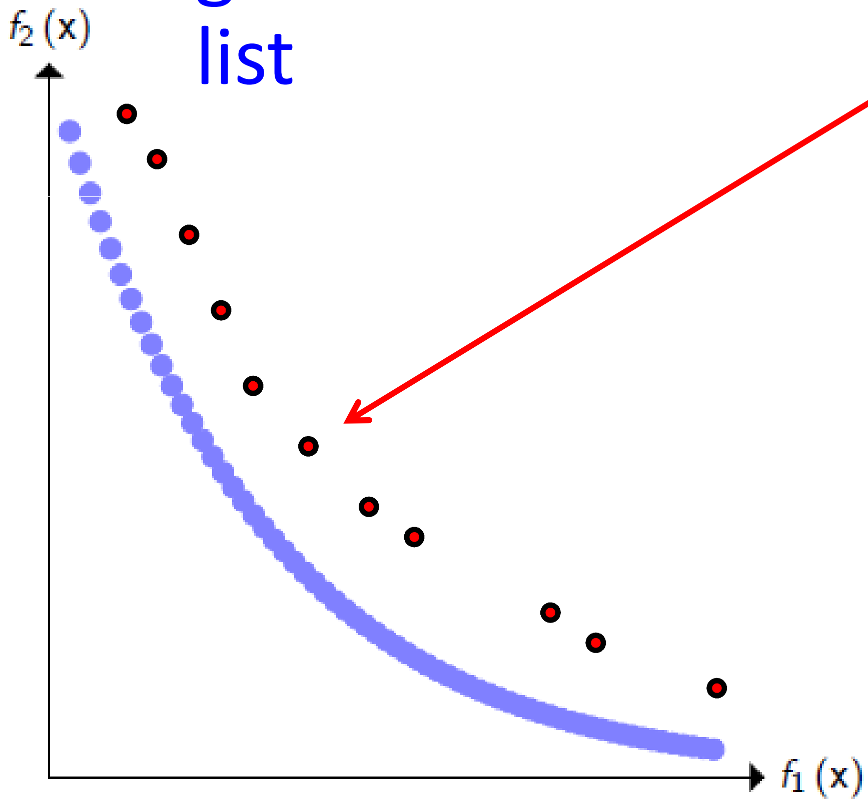
$$\beta(S) = \sum_{i=2}^{N} \min(d_d(I_i, I_{1,\ldots,i-1}))$$

Relevance term

Diversity term

# MORD: Representation

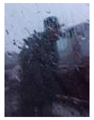
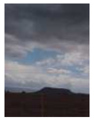
Each solution is the vector of scores to generate the ranked list

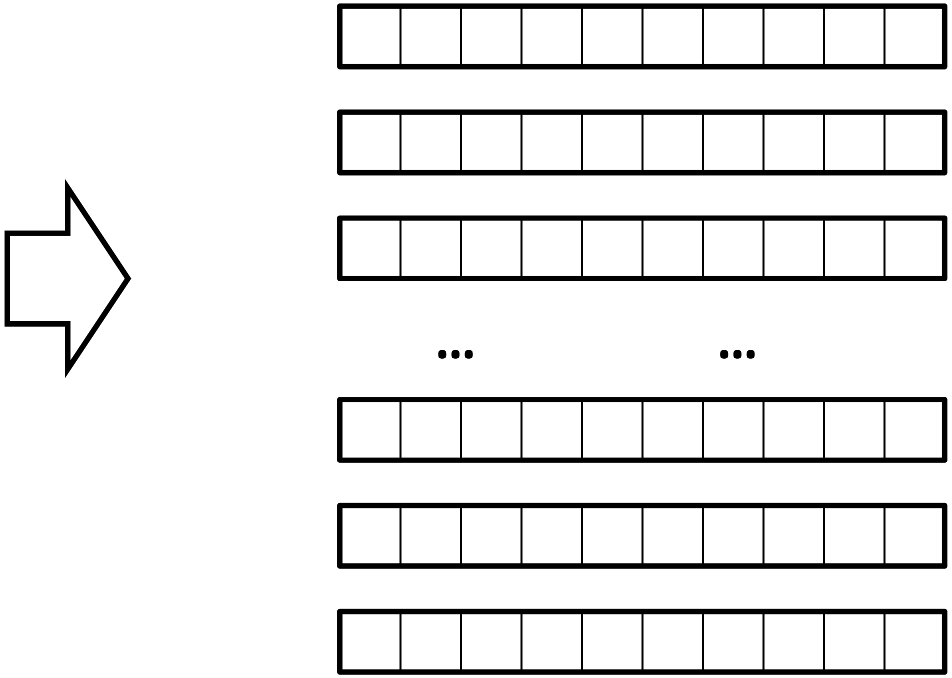




A solution to our problem is a ranked list of images

# MORD: Representation

| Rank | List | $S^0$ |
|---|---|---|
| 1 |  | 1 |
| 2 |  | 0.5 |
| 3 |  | 0.3 |
| 4 |  | 0.25 |
| 5 |  | 0.2 |
| 6 |  | 0.16 |
| ... | ... | ... |




Initial population

# Multi-objective optimization for result diversification

- NSGA-II is used to approach the problem as follows:

Maximize $< \rho(S^0, S)\,,\, \beta(S) >$

- Where:

$$\rho(S^0, S) = 1 - \frac{6}{n(n^2-1)} \sum_i dr_i(S^0, S)^2$$

$$\beta(S) = \sum_{i=2}^{N} \min(d_d(I_i, I_{1,\ldots,i-1}))$$

Relevance term

Diversity term

# Multi-objective optimization for result diversification

- Diversity criterion:

$$\beta(S) = \sum_{i=2}^{N} \min(d_d(I_i, I_{1,\ldots,i-1}))$$

# Multi-objective optimization for result diversification

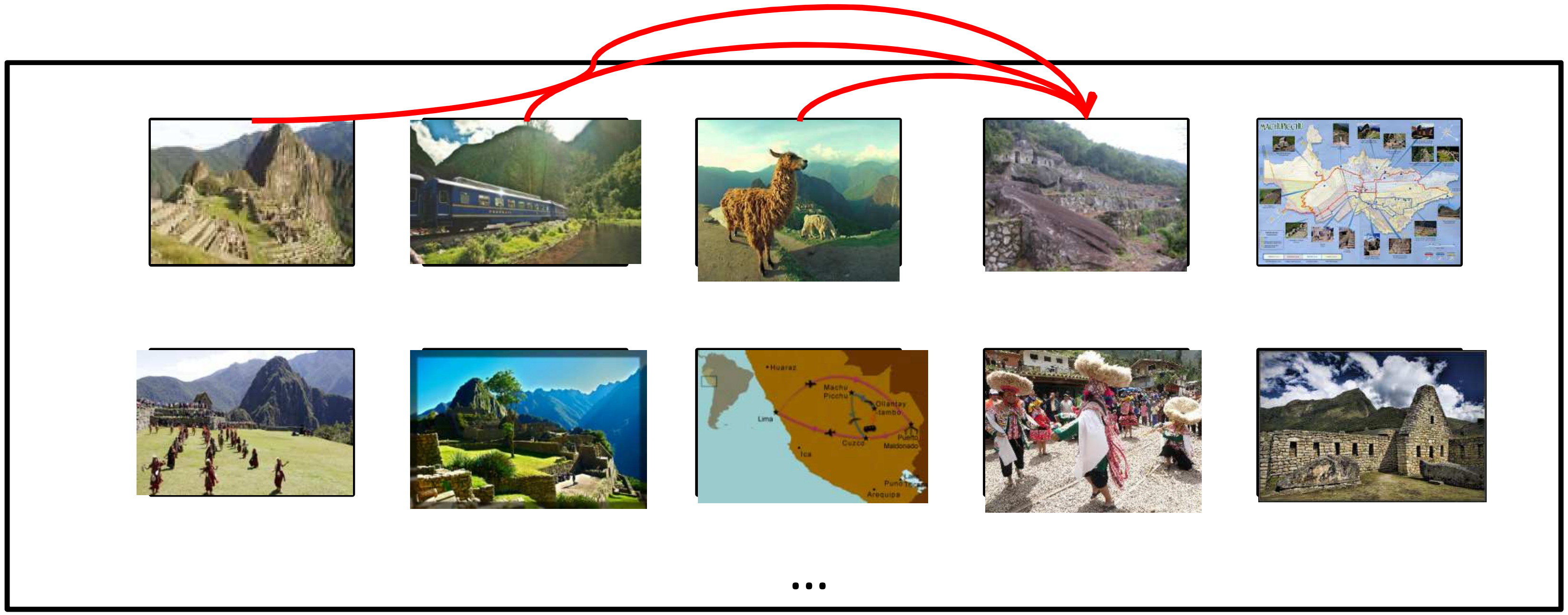
- Diversity criterion:

$$\beta(S) = \sum_{i=2}^{N} \min(d_d(I_i, I_{1,\ldots,i-1}))$$

# Multi-objective optimization for result diversification

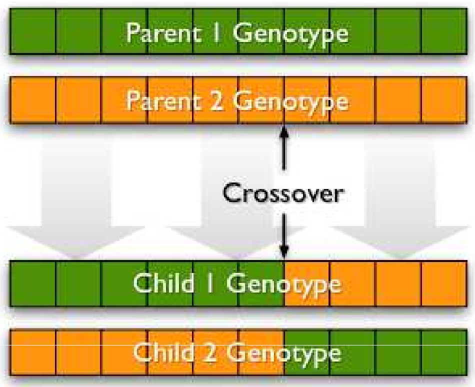
- Diversity criterion:

$$\beta(S) = \sum_{i=2}^{N} \min(d_d(I_i, I_{1,\dots,i-1}))$$

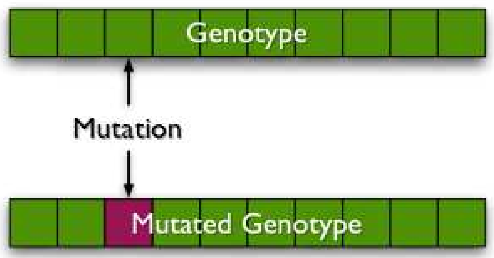


# MORD: Evolutionary stuff

- **Initialization:** Solutions are generated by adding random numbers to the original scores-vector

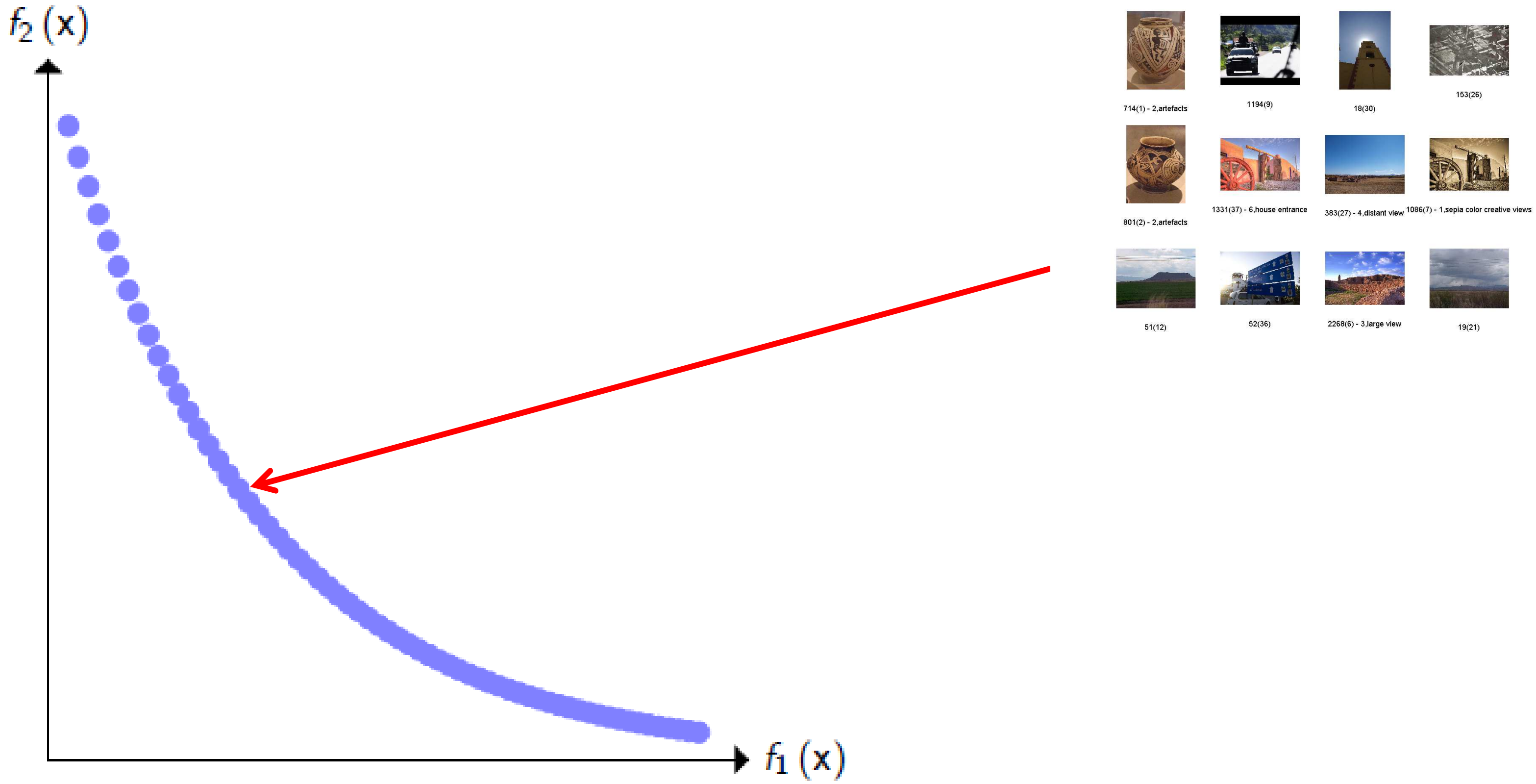



- **Evolutionary operators:** Standard cross-over and mutation operators were used



# MORD: Selection of a single-solution

- We take the solution offering the best tradeoff between both objectives



# Experiments & results

- Three runs were submitted:
  1. Visual
  2. Textual
  3. Visual+Textual

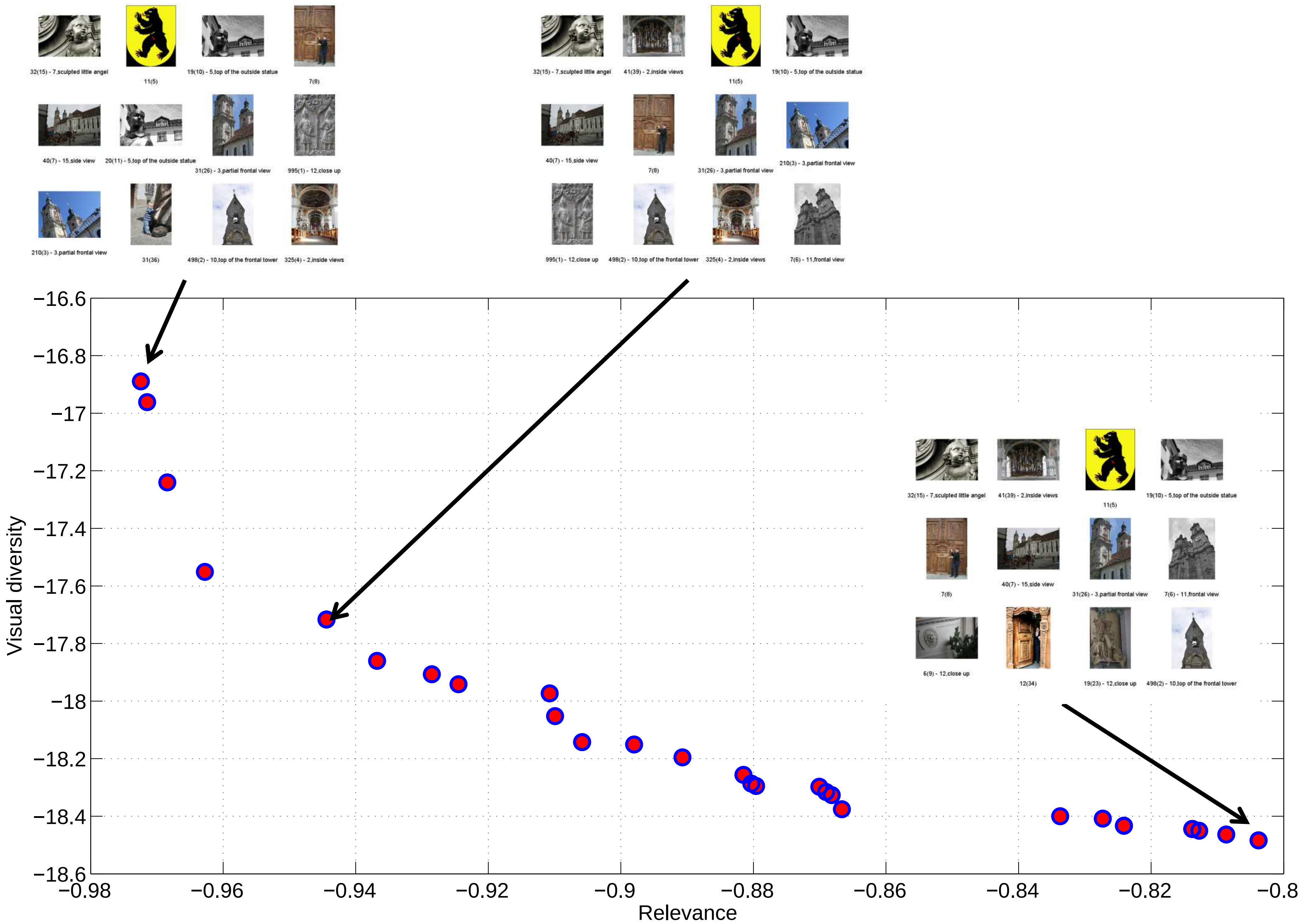
MediaEval Benchmark

MediaEval Benchmarking Initiative for Multimedia Evaluation

The "multi" in multimedia: speech, audio, visual content, tags, users, context

Table 2: Official results obtained by MORD.

| Expert evaluation | | | | | | |
|---|---|---|---|---|---|---|
| | C@10 | C@20 | P@10 | P@20 | F@10 | F@20 |
| 1 | 0.3808 | 0.5699 | 0.7146 | 0.7143 | 0.4795 | 0.6125 |
| 2 | 0.3885 | 0.5732 | 0.7091 | 0.7136 | 0.4801 | 0.6102 |
| 3 | 0.3823 | 0.5690 | 0.6977 | 0.7076 | 0.4728 | 0.6067 |
| **Crowd-sourcing evaluation** | | | | | | |
| | C@10 | C@20 | P@10 | P@20 | F@10 | F@20 |
| 1 | 0.7194 | 0.8499 | 0.6755 | 0.6745 | 0.6640 | 0.7245 |
| 2 | 0.7503 | 0.8625 | 0.6755 | 0.6898 | 0.6790 | 0.7392 |
| 3 | 0.7479 | 0.8675 | 0.6714 | 0.6918 | 0.6769 | 0.7464 |



32(15) - 7.sculpted little angel
11(5)
19(10) - 5.top of the outside statue
7(8)
40(7) - 15.side view
20(11) - 5.top of the outside statue
31(26) - 3.partial frontal view
995(1) - 12.close up
210(3) - 3.partial frontal view
31(36)
498(2) - 10.top of the frontal tower
325(4) - 2.inside views
41(39) - 2.inside views
7(6) - 11.frontal view
6(9) - 12.close up
12(34)
19(23) - 12.close up
Visual diversity
Relevance
−16.6
−16.8
−17
−17.2
−17.4
−17.6
−17.8
−18
−18.2
−18.4
−18.6
−0.98
−0.96
−0.94
−0.92
−0.9
−0.88
−0.86
−0.84
−0.82
−0.8

# Experiments & results

995(1) - 12,close up

498(2) - 10,top of the frontal tower

210(3) - 3,partial frontal view

325(4) - 2,inside views

11(5)

7(6) - 11,frontal view

40(7) - 15,side view

7(8)

6(9) - 12,close up

19(10) - 5,top of the outside statue

20(11) - 5,top of the outside statue

31(12)

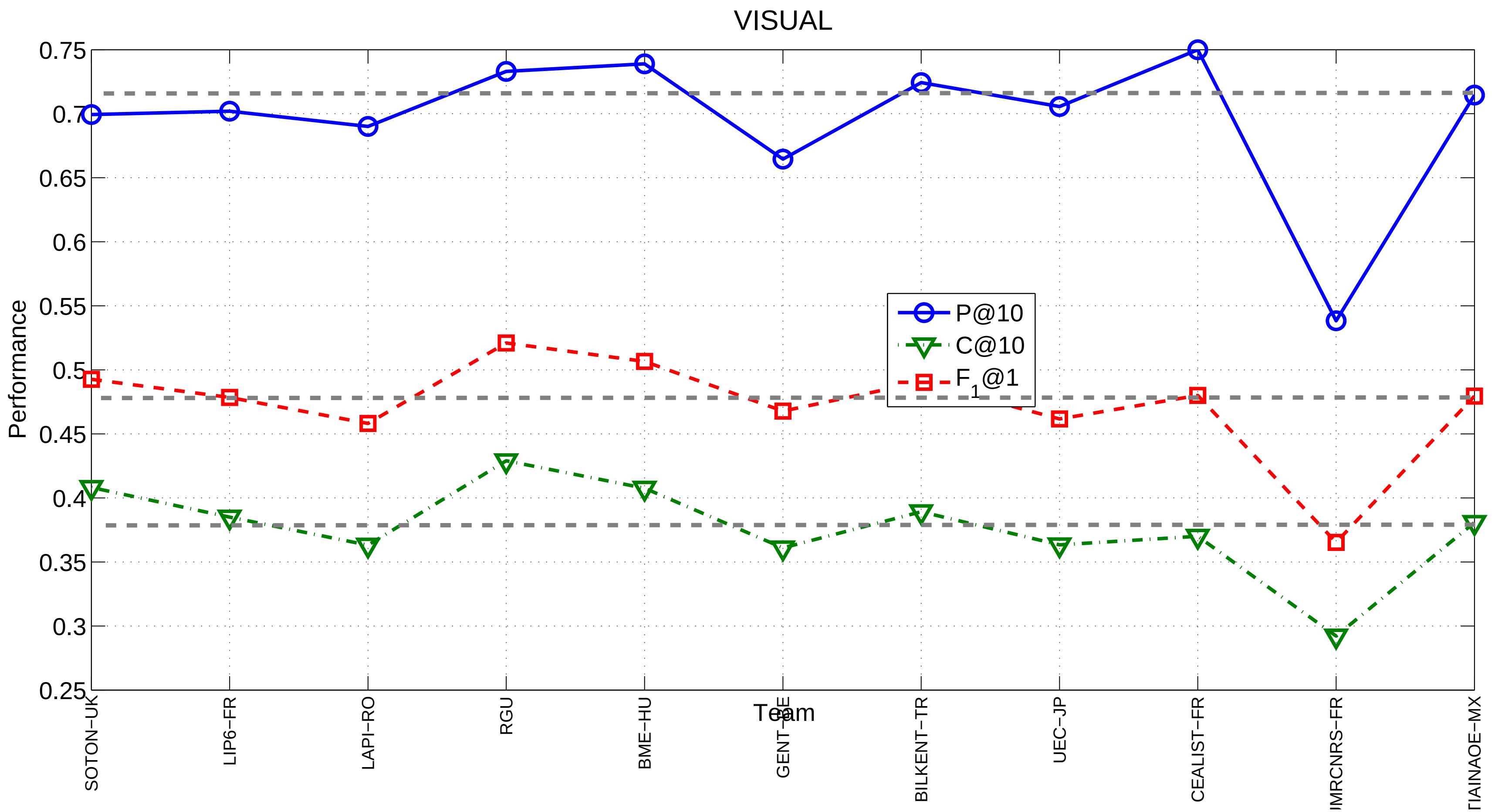
**Initial list (7 topics in top-12 images)**

# Experiments & results


**Re-ranked (8 topics in top-12 images)**

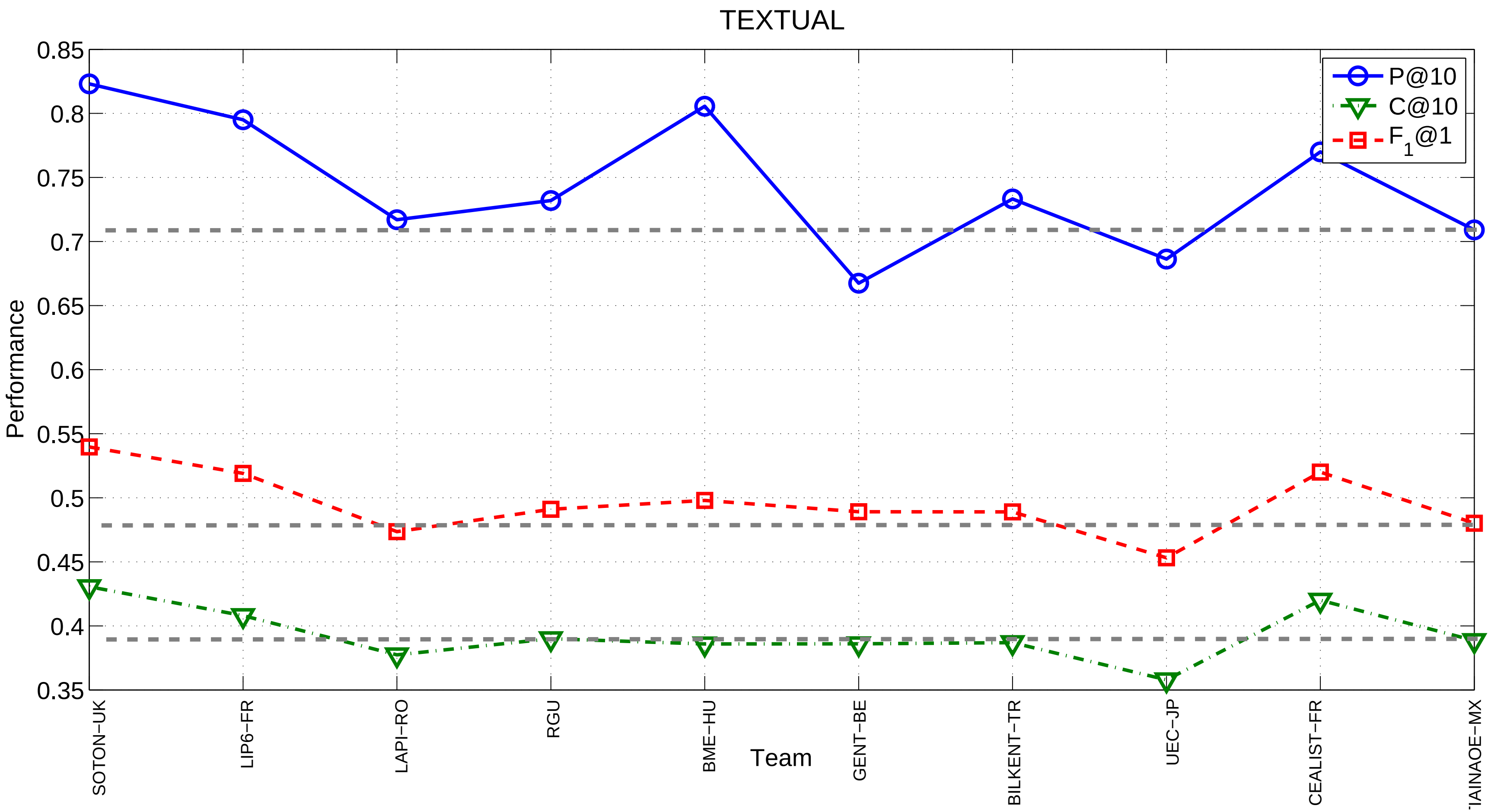
# Experiments & results

- Comparison with other participants: *6th out 11*



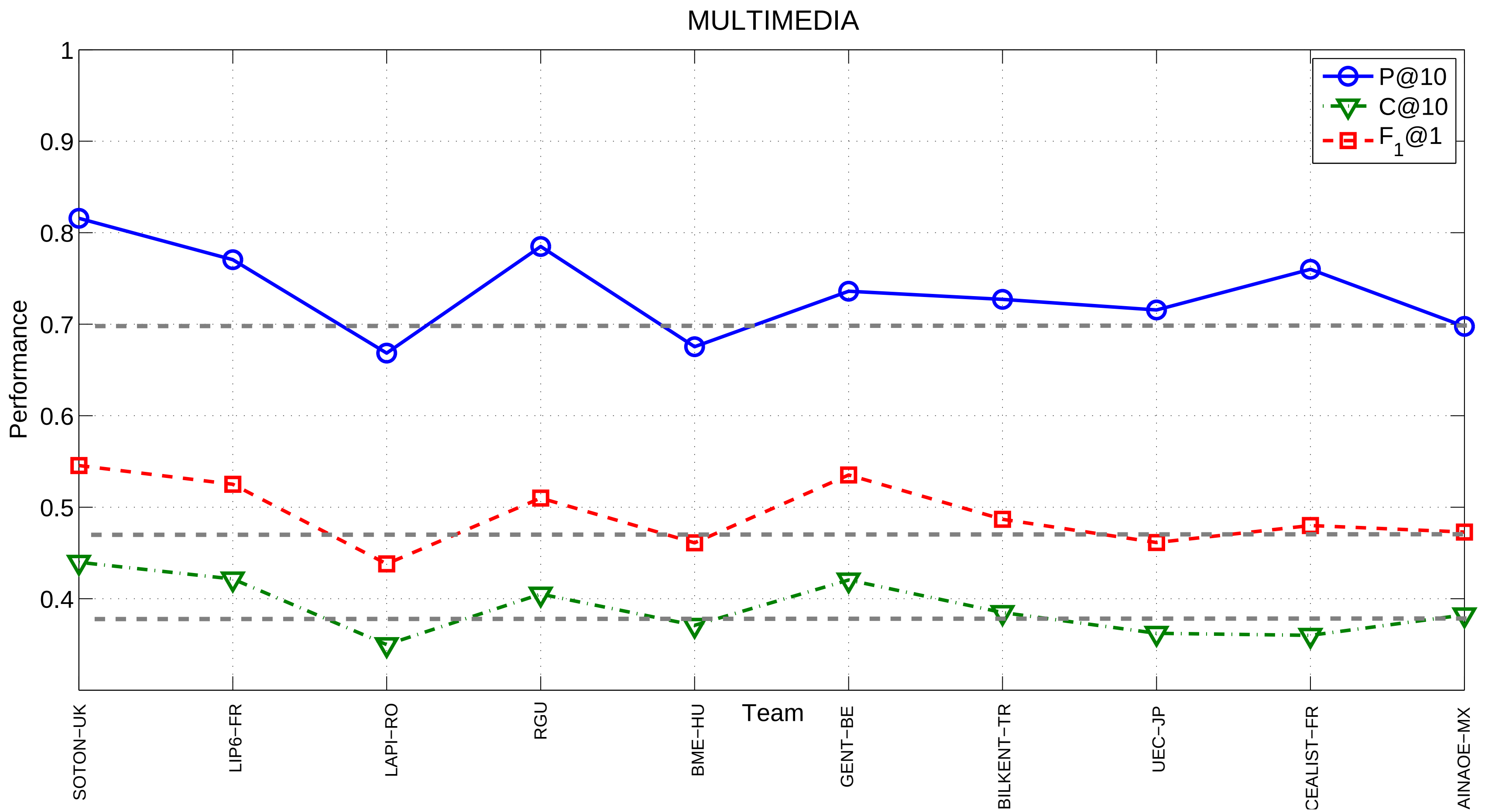
# Experiments & results

- Comparison with other participants: *5th out 11*



# Experiments & results

- Comparison with other participants: *6th out of 11*



# Conclusions

- The multi-objective formulation for RD is promising, but not as effective as we expected
  - The initial ranked list was not too reliable?
  - No feature selection / special processing of features
  - Did not take advantage of meta-data (tags/ comments/ etc.)

- Too many parameters/decisions to fix/take

# Future work

- Alternative objective functions for both relevance and diversity.
- Evaluation of the gains over single-objective combinatoric approaches
- Efficient implementation in GPUs
- Incorporating feature selection into the optimization process

Hugo Jair Escalante, Maribel Marin-Castro, Mario Graff, Alicia Morales-Reyes, Manuel Montes, Alejandro Rosales, Jesús A. González, Carlos A. Reyes. **MOPG: Multi-objective prototype generation for classification.**

*Submitted to Pattern Recognition, October 12, 2013*

# MOPG: MULTI-OBJECTIVE PROTOTYPE GENERATION FOR CLASSIFICATION

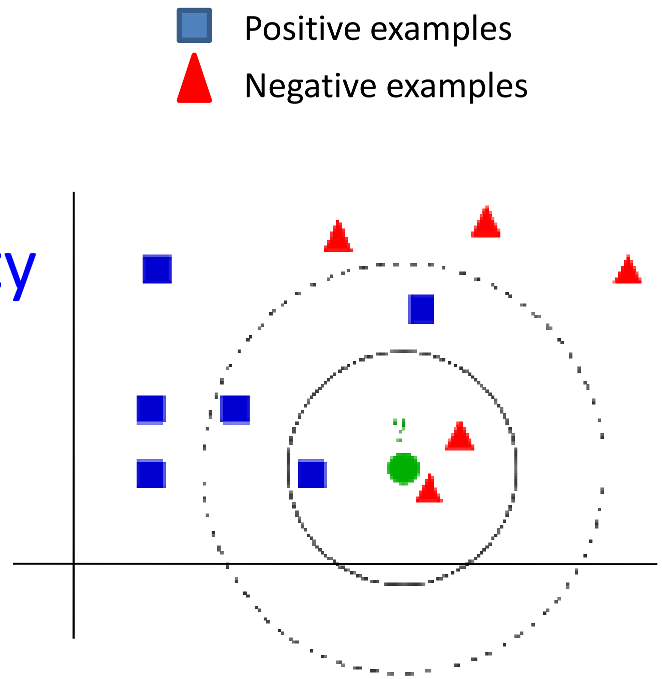
# KNN – classifier

One of the most popular non-parametric classifiers

Easy to implement and very effective

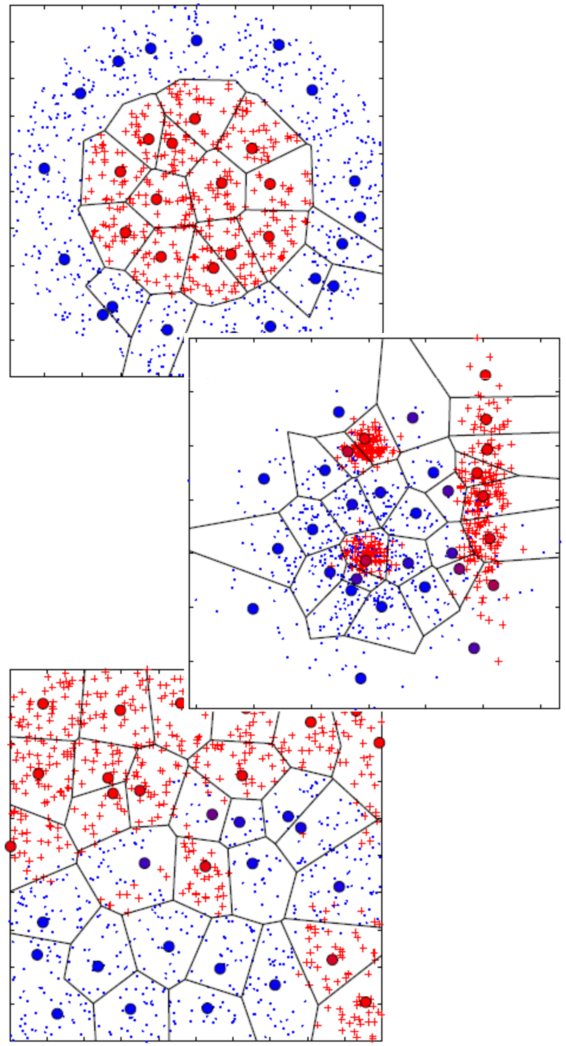
Main issues with KNN:

- The curse of dimensionality
- Efficiency
- Sensibility to noisy data



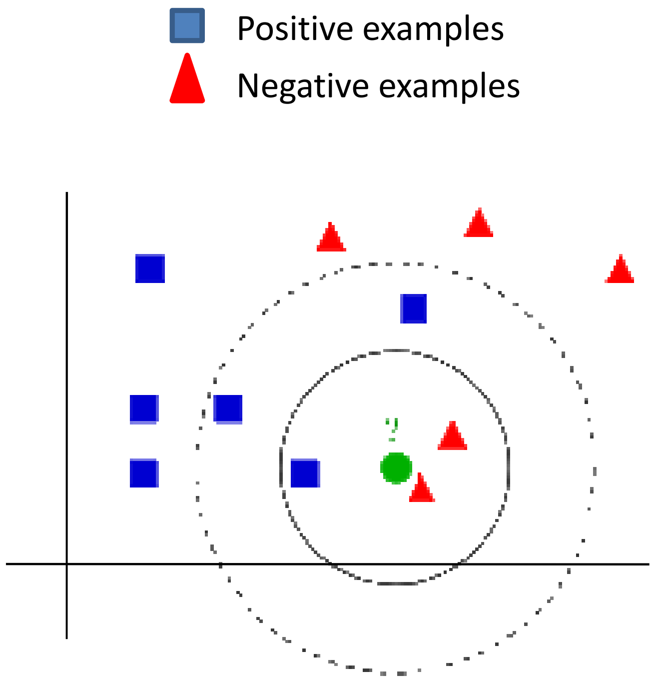
# Prototype-based classification

- KNN classifiers using a subset of the original data
- The goal is to reduce the computational cost of standard KNN, by filtering out noisy/redundant instances and keeping the most informative ones
- **Key issue:** how to select/obtain the set of prototypes for a classification problem?

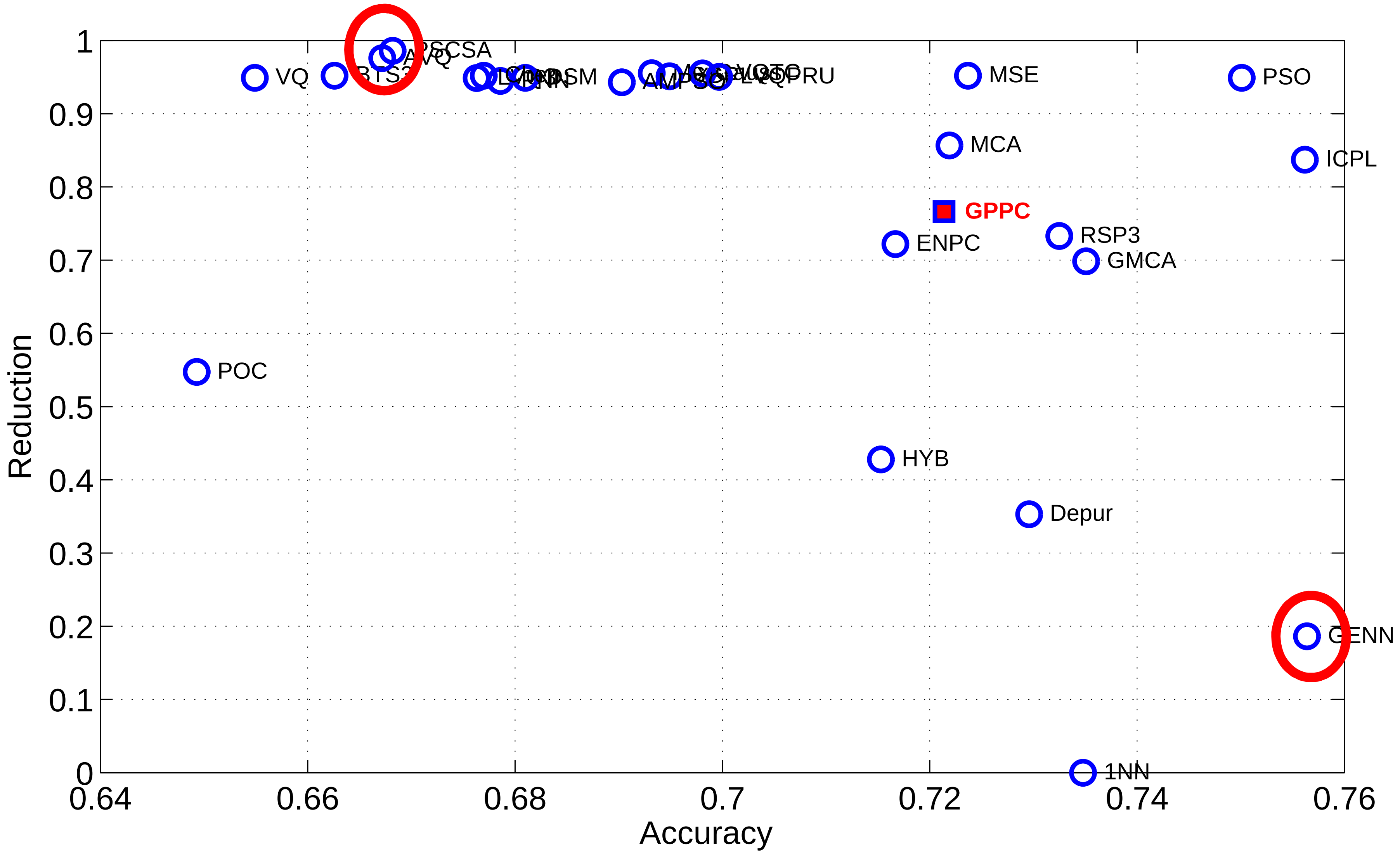


# Prototype generation

- **Problem:** To select a (**small**) subset of instances such that the classification performance of a particular classifier (KNN) is not degraded significantly



# Accuracy vs reduction dilemma



# MOPG: Multi-Objective Prototype Generation

- **Idea:** approaching the PG problem as one of multi-objective optimization, where the objectives are: **reduction** and **accuracy**

- Goal: to obtain solutions that offer a good tradeoff between both objectives, and then select one for classification

# MOPG: Multi-Objective Prototype Generation

- NSGA-II is used to approach the following problem:

$$\underset{\mathcal{P}}{\text{maximize}} \quad \langle f_1(\mathcal{P}), f_2(\mathcal{P}) \rangle$$

$$\text{subject to} \quad \mathcal{P} \in \mathcal{Y}$$

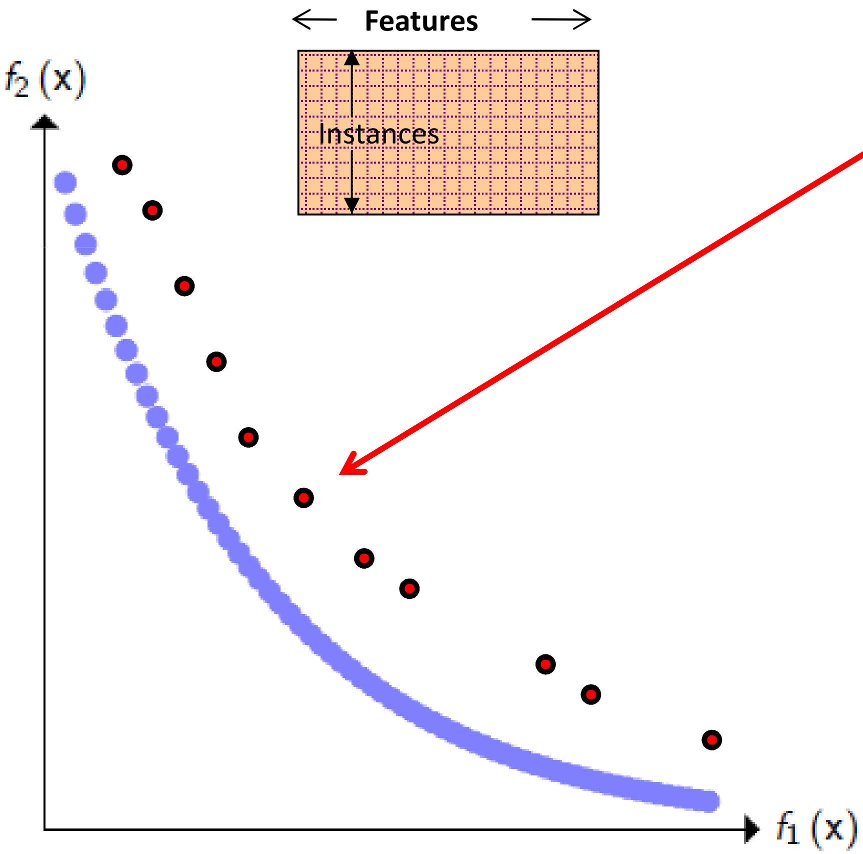
- Where: $\boxed{f_1(\boldsymbol{P}) = \delta(\boldsymbol{P}, \boldsymbol{D});}$ $\boxed{f_2(\boldsymbol{P}) = \gamma(\boldsymbol{P}, \boldsymbol{D})}$

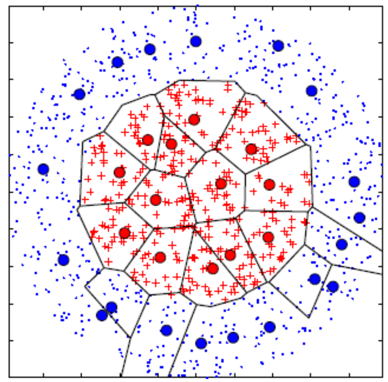
Training set reduction

Hold-out classification performance

# MOPG: Representation

Each solution is codified as matrix of size $P$ x $d$



A solution to our problem is a set of instances (the prototypes)

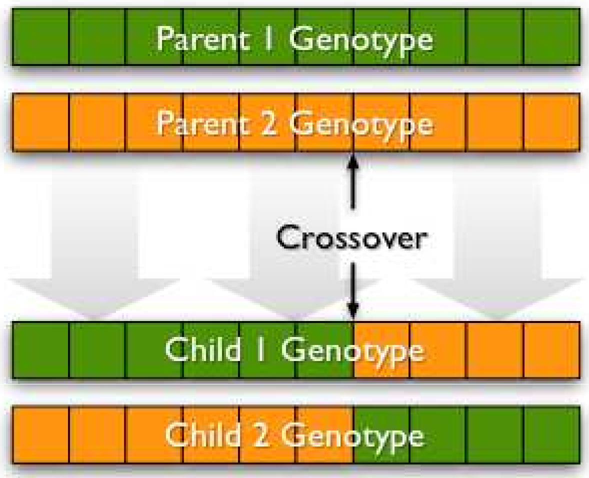
# MOPG: Initialization

- Training data is divided into development and validation partitions
  - *Development:* Instances from which prototypes can be generated
  - *Validation:* Hold-out data set to evaluate solutions
  - The partition is updated every iteration

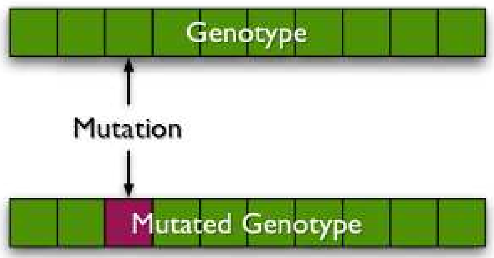
- **Initialization:** For each class we randomly select a set of training instances (class distribution is mantained)

# MOPG: Evolutionary operators

- **Crossover:** with uniform probability either
  - Interchange (same-class) prototypes between solutions
  - Replace a prototype of class *k* in one solution with the average of all prototypes from class k in the other prototype

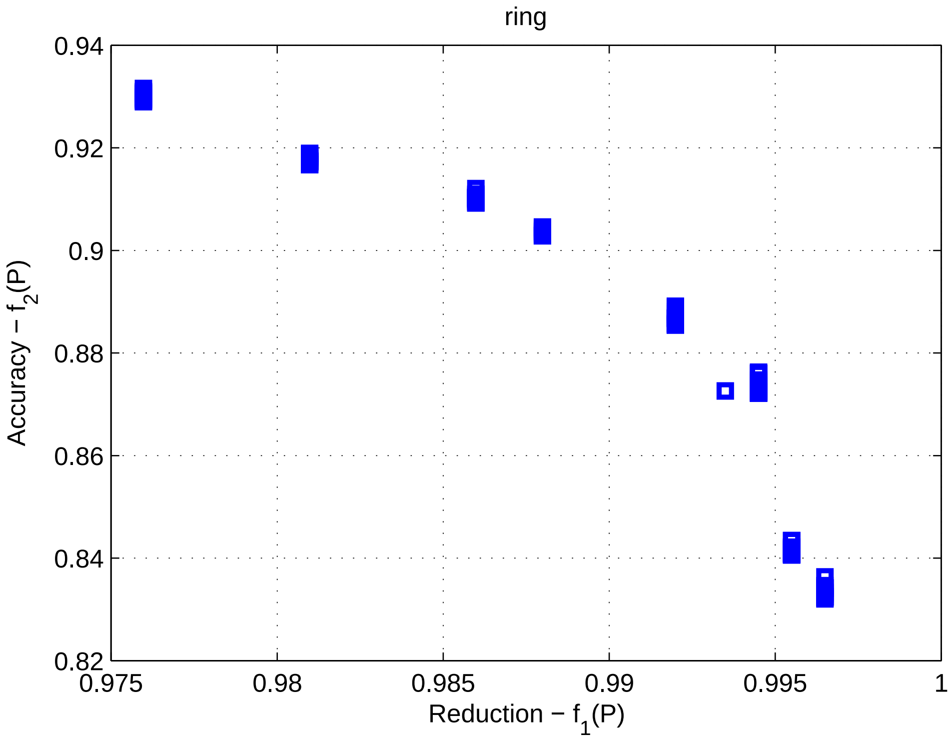


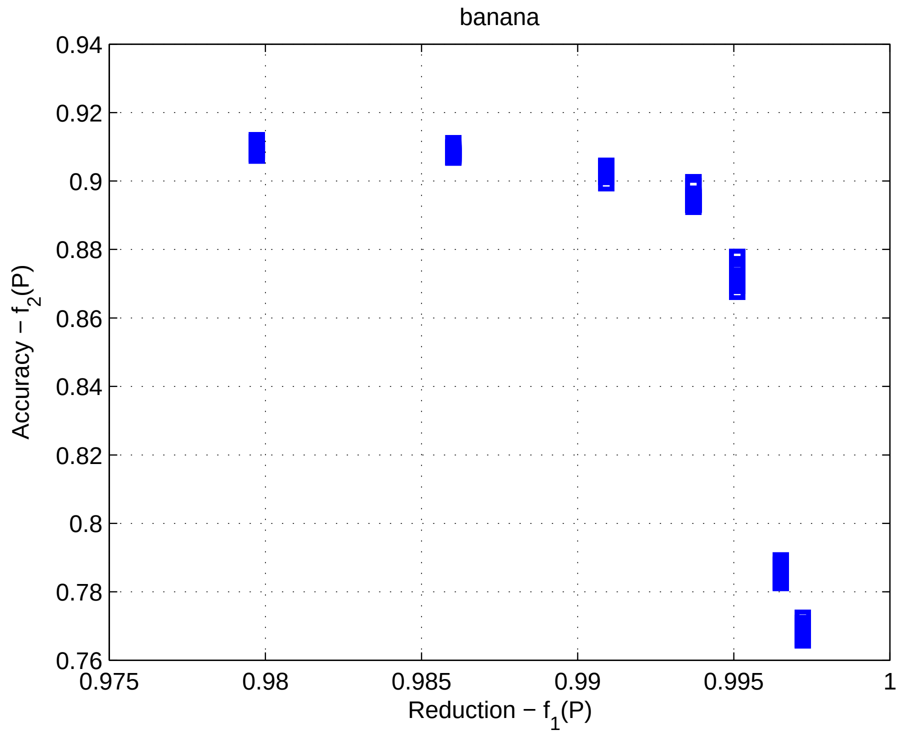

- **Mutation:** with uniform probability either
  - Add a vector of random numbers to a prototype
  - Replace a prototype with another instance frmo the development set



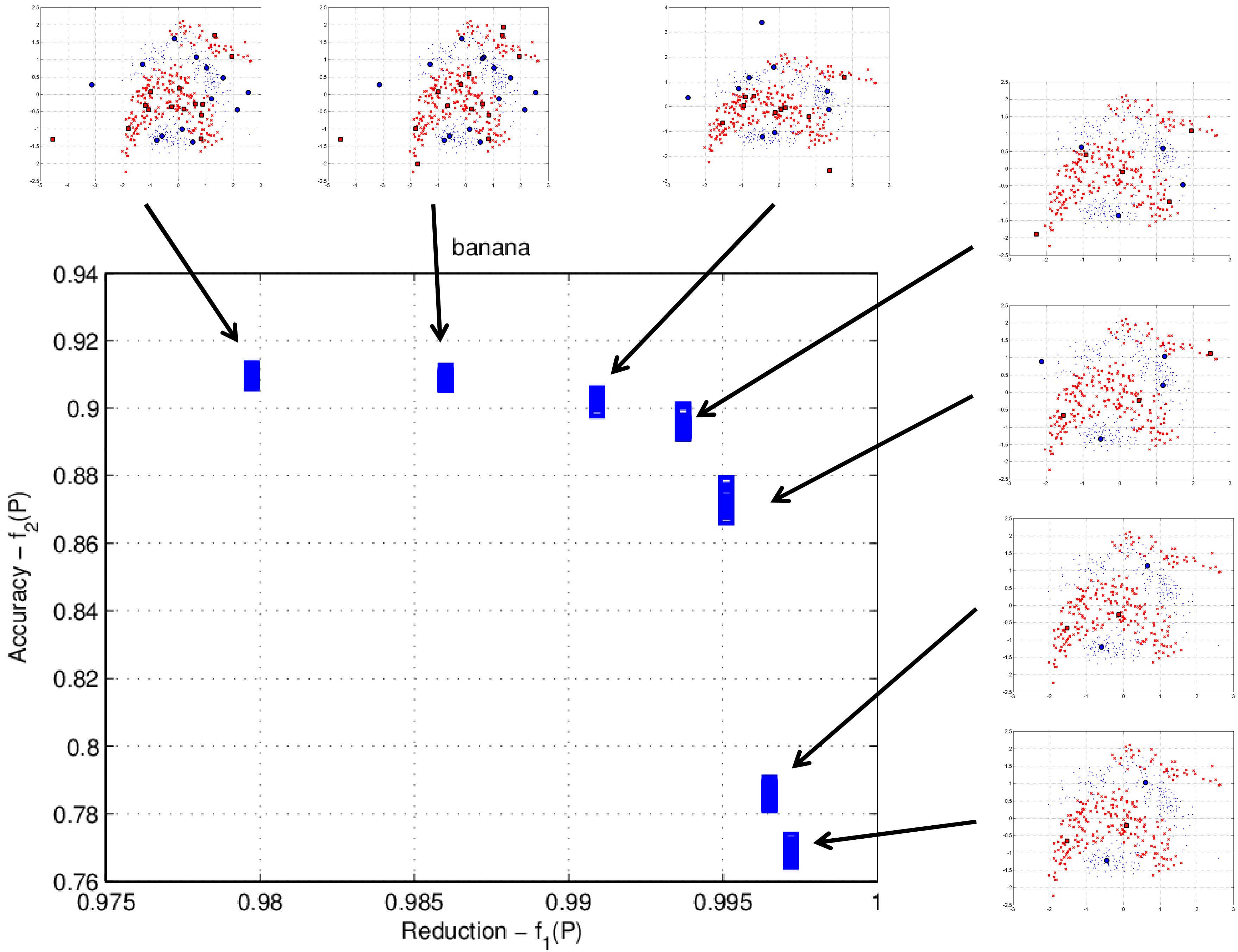
# MOPG: Selection of a single-solution

- We evaluate the performance of each solution in the Pareto front and chose the one with highest accuracy




Pareto front for two sample data sets


banana
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.975
0.98
0.985
0.99
0.995
1
Reduction – $f_1(P)$
Accuracy – $f_2(P)$

# Experiments and results

- We performed experiments over 59 classification problems of diverse characteristics

- Compared the performance of our proposal to that of 25 alternative prototype generation techniques

I. Triguero, J. Derrac, S. García and F.Herrera, **A Taxonomy and Experimental Study on Prototype Generation for Nearest Neighbor Classification .** IEEE Trans. on Systems, Man, and Cybernetics--Part C, 42 (1) (2012) 86-100, 2012

# Experiments and results

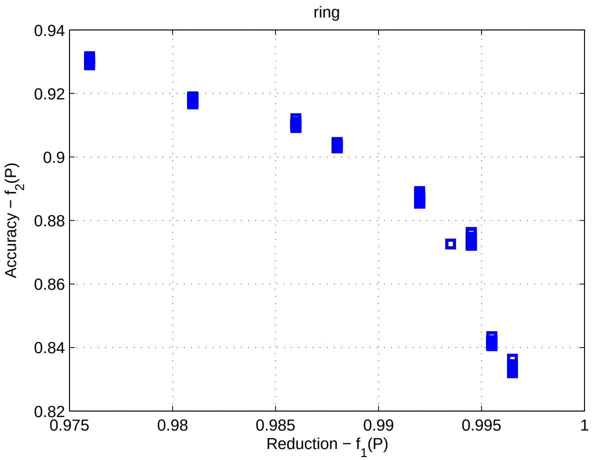
| Data Set | #Ex. | #Atts. | #Num. | #Nom. | #Cl. | Data Set | #Ex. | #Atts. | #Num. | #Nom. | #Cl. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| abalone | 4,174 | 8 | 7 | 1 | 28 | marketing | 8,993 | 13 | 13 | 0 | 9 |
| appendicitis | 106 | 7 | 7 | 0 | 2 | monks | 432 | 6 | 6 | 0 | 2 |
| australian | 690 | 14 | 8 | 6 | 2 | movement_libras | 360 | 90 | 90 | 0 | 15 |
| autos | 205 | 25 | 15 | 10 | 6 | newthyroid | 215 | 5 | 5 | 0 | 3 |
| balance | 625 | 4 | 4 | 0 | 3 | nursery | 12,960 | 8 | 0 | 8 | 5 |
| banana | 5,300 | 2 | 2 | 0 | 2 | pageblocks | 5,472 | 10 | 10 | 0 | 5 |
| bands | 539 | 19 | 19 | 0 | 2 | penbased | 10,992 | 16 | 16 | 0 | 10 |
| breast | 286 | 9 | 0 | 9 | 2 | phoneme | 5,404 | 5 | 5 | 0 | 2 |
| bupa | 345 | 6 | 6 | 0 | 2 | pima | 768 | 8 | 8 | 0 | 2 |
| car | 1,728 | 6 | 0 | 6 | 4 | ring | 7,400 | 20 | 20 | 0 | 2 |
| chess | 3,196 | 36 | 0 | 36 | 2 | saheart | 462 | 9 | 8 | 1 | 2 |
| cleveland | 297 | 13 | 13 | 0 | 5 | satimage | 6,435 | 36 | 36 | 0 | 7 |
| coil2000 | 9,822 | 85 | 85 | 0 | 2 | segment | 2,310 | 19 | 19 | 0 | 7 |
| contraceptive | 1,473 | 9 | 6 | 3 | 3 | sonar | 208 | 60 | 70 | 0 | 2 |
| crx | 125 | 15 | 6 | 9 | 2 | spambase | 4,597 | 55 | 55 | 0 | 2 |
| dermatology | 366 | 33 | 1 | 32 | 6 | spectheart | 267 | 44 | 44 | 0 | 2 |
| ecoli | 336 | 7 | 7 | 0 | 8 | splice | 3,190 | 60 | 0 | 60 | 3 |
| flare-solar | 1,066 | 9 | 0 | 9 | 2 | tae | 151 | 5 | 5 | 0 | 3 |
| german | 1,000 | 20 | 6 | 14 | 2 | texture | 5,500 | 40 | 40 | 0 | 11 |
| glass | 214 | 9 | 9 | 0 | 7 | tic-tac-toe | 958 | 9 | 0 | 9 | 2 |
| haberman | 306 | 3 | 3 | 0 | 2 | thyroid | 7,200 | 21 | 6 | 15 | 3 |
| hayes-roth | 133 | 4 | 4 | 0 | 3 | titanic | 2,201 | 3 | 3 | 0 | 2 |
| heart | 270 | 13 | 6 | 7 | 2 | twonorm | 7,400 | 20 | 20 | 0 | 2 |
| hepatitis | 155 | 19 | 19 | 0 | 2 | vehicle | 846 | 18 | 18 | 0 | 4 |
| housevotes | 435 | 16 | 0 | 16 | 2 | vowel | 990 | 13 | 11 | 2 | 11 |
| iris | 150 | 4 | 4 | 0 | 3 | wine | 178 | 13 | 13 | 0 | 3 |
| led7digit | 500 | 7 | 0 | 1 | 10 | wisconsin | 683 | 9 | 9 | 0 | 2 |
| lymphography | 148 | 18 | 3 | 15 | 4 | yeast | 1484 | 8 | 8 | 0 | 10 |
| magic | 19,020 | 10 | 10 | 0 | 2 | zoo | 101 | 17 | 0 | 17 | 7 |
| mammographic | 961 | 5 | 0 | 5 | 2 | | | | | | |

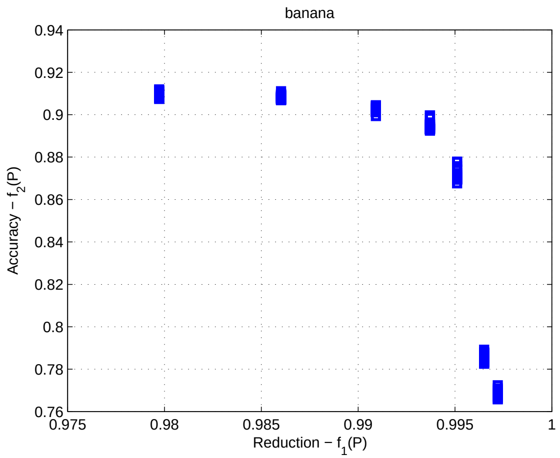
I. Triguero, J. Derrac, S. García and F.Herrera, **A Taxonomy and Experimental Study on Prototype Generation for Nearest Neighbor Classification .** IEEE Trans. on Systems, Man, and Cybernetics--Part C, 42 (1) (2012) 86-100, 2012

# Experiments & results

- Evaluation of the selection strategy:

| | Accuracy | | | Reduction | | |
|---|---|---|---|---|---|---|
| **Method** | **All** | **Small** | **Large** | **All** | **Small** | **Large** |
| Strategy | 73.94±18.58 | 81.29±20.05 | 70.93±16.95 | 98.67±1.15 | 99.41±0.29 | 98.39±1.16 |
| Best | 76.64±17.46 | 82.05±19.77 | 74.51±15.84 | 98.78±1.18 | 99.48±0.29 | 98.52±1.22 |



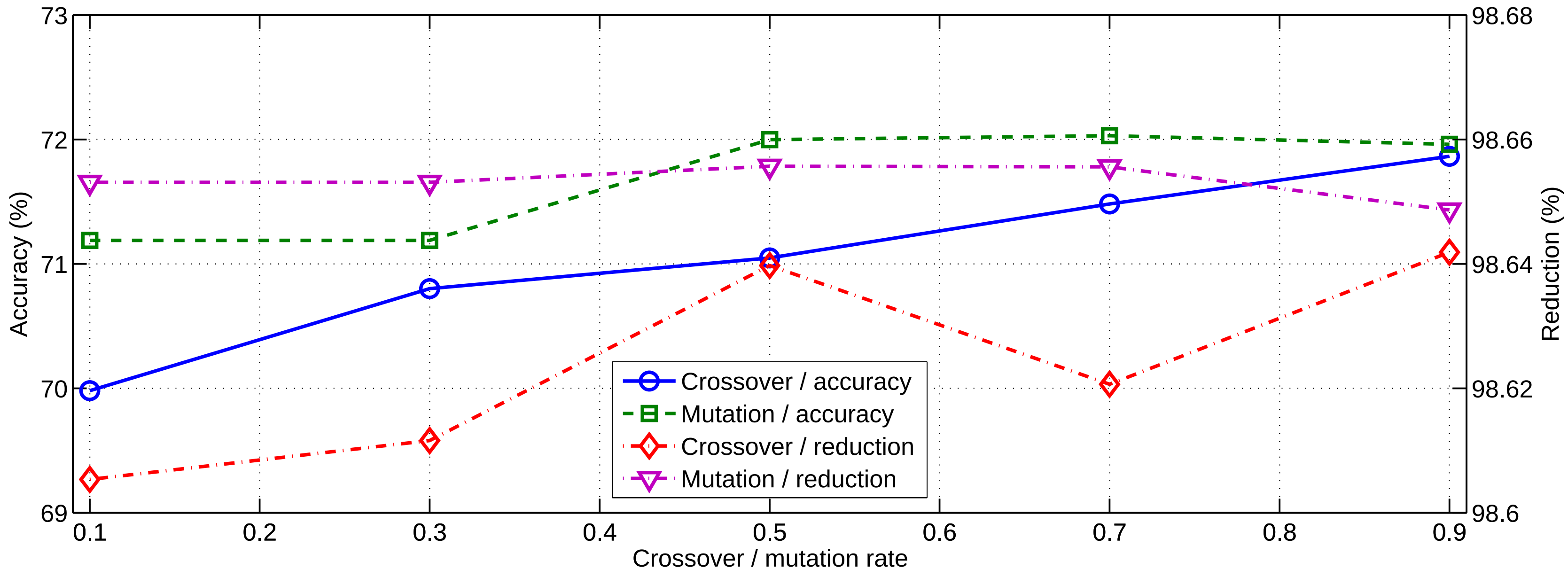
# Experiments & results

- Parameter settings

| Parameter | Value | Accuracy | Reduction |
|---|---|---|---|
| Individuals ($N_{pop}$) | 50 | $71.68\% \pm 18.18$ | $97.24\% \pm 1.21$ |
| | 100 | $72.25\% \pm 17.80$ | $\mathbf{97.26\% \pm 1.25}$ |
| | 250 | $\mathbf{73.13\% \pm 18.10}$ | $97.23\% \pm 1.31$ |
| Generations ($g$) | 50 | $71.68\% \pm 18.18$ | $97.24\% \pm 1.21$ |
| | 100 | $72.71\% \pm 18.03$ | $97.53\% \pm 1.29$ |
| | 250 | $73.32\% \pm 18.11$ | $97.62\% \pm 1.34$ |
| | 500 | $\mathbf{73.37\% \pm 18.08}$ | $\mathbf{97.70\% \pm 1.31}$ |
| Train-set-size ($\eta$) | 0.1 | $72.11\% \pm 18.33$ | $\mathbf{98.84\% \pm 1.09}$ |
| | 0.3 | $\mathbf{73.07\% \pm 18.18}$ | $98.19\% \pm 1.13$ |
| | 0.5 | $71.68\% \pm 18.18$ | $97.24\% \pm 1.21$ |
| | 0.7 | $72.12\% \pm 18.44$ | $96.85\% \pm 1.38$ |
| | 0.9 | $70.30\% \pm 20.39$ | $96.38\% \pm 1.67$ |
| Initial prot. ($I_p$) | 0.005 | $71.19\% \pm 18.18$ | $\mathbf{99.01\% \pm 1.28}$ |
| | 0.01 | $71.20\% \pm 18.31$ | $98.97\% \pm 1.26$ |
| | 0.05 | $72.14\% \pm 18.74$ | $98.37\% \pm 1.16$ |
| | 0.1 | $71.68\% \pm 18.18$ | $97.24\% \pm 1.21$ |
| | 0.2 | $73.07\% \pm 17.76$ | $95.19\% \pm 1.82$ |
| | 0.4 | $\mathbf{73.20\% \pm 17.72}$ | $89.86\% \pm 3.87$ |

# Experiments & results

- Parameter settings

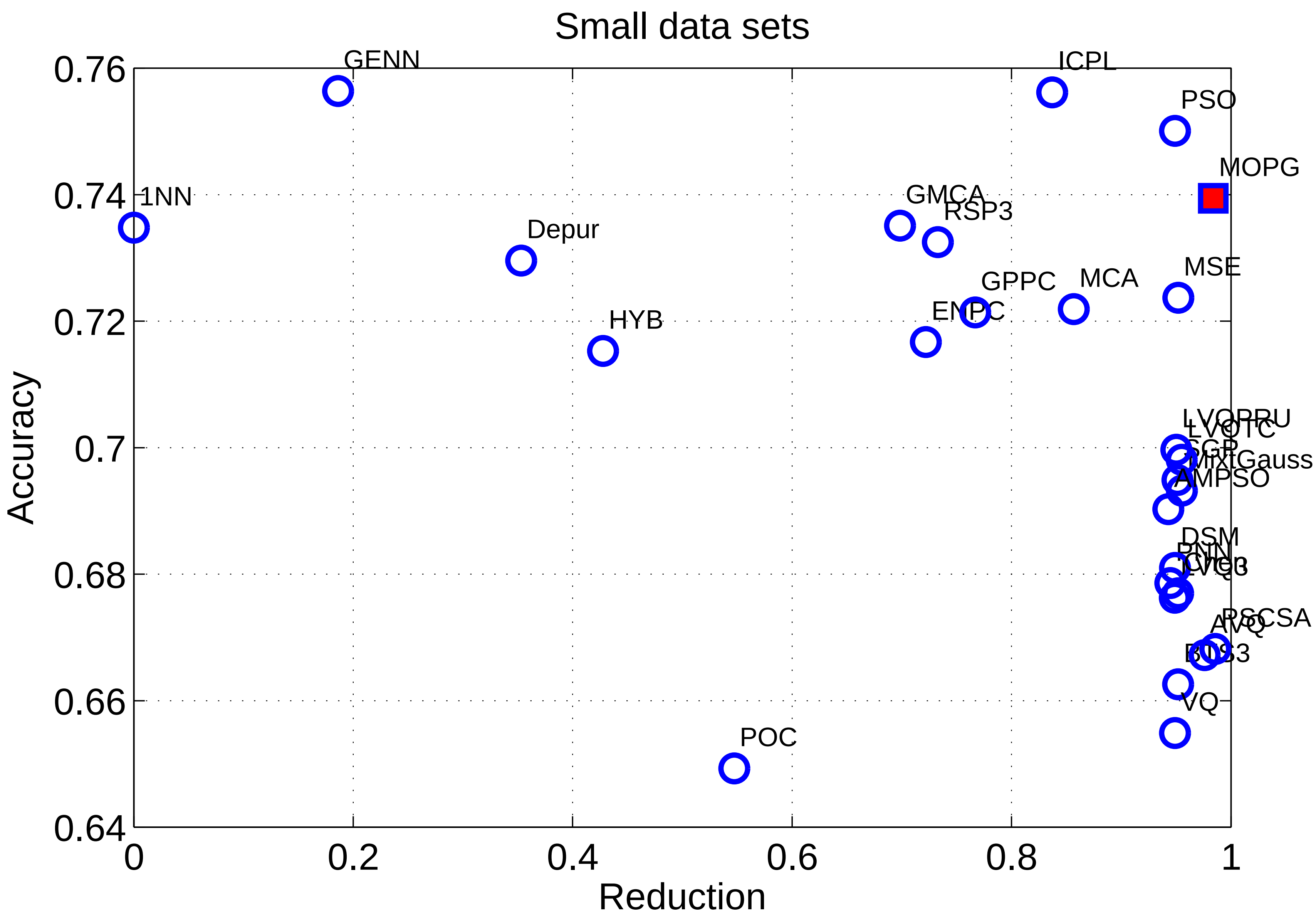


# Experiments & results

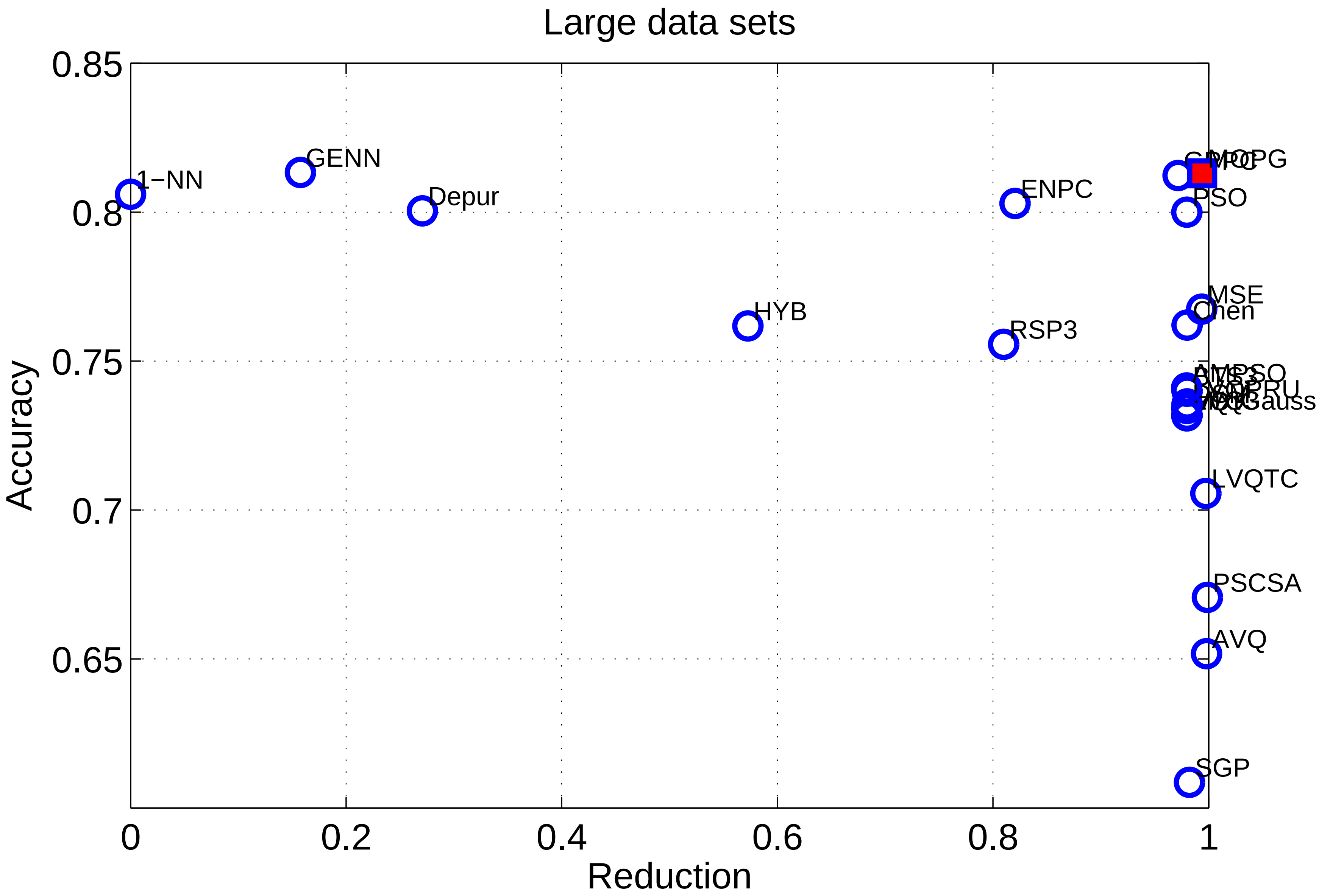
- Comparison with related work

| | Test set accuracy | | | Training set reduction | | |
|---|---|---|---|---|---|---|
| Measure | All | Small | Large | All | Small | Large |
| MOGP | 73.94%±18.58 | 70.93%±16.95 | 81.30%±20.05 | 98.67% | 98.39% | 99.41% |
| GENN | 78.48%±18.57 | 75.64%±15.45 | 81.33%±21.70 | 17.19% | 18.62% | 15.76% |
| PSCSA | 66.94%±20.39 | 66.82%±18.74 | 67.07%±22.05 | 99.23% | 98.58% | 99.88% |
| 1NN | 77.04%±19.44 | 73.48%±16.64 | 80.60%±22.24 | 0% | 0% | 0% |

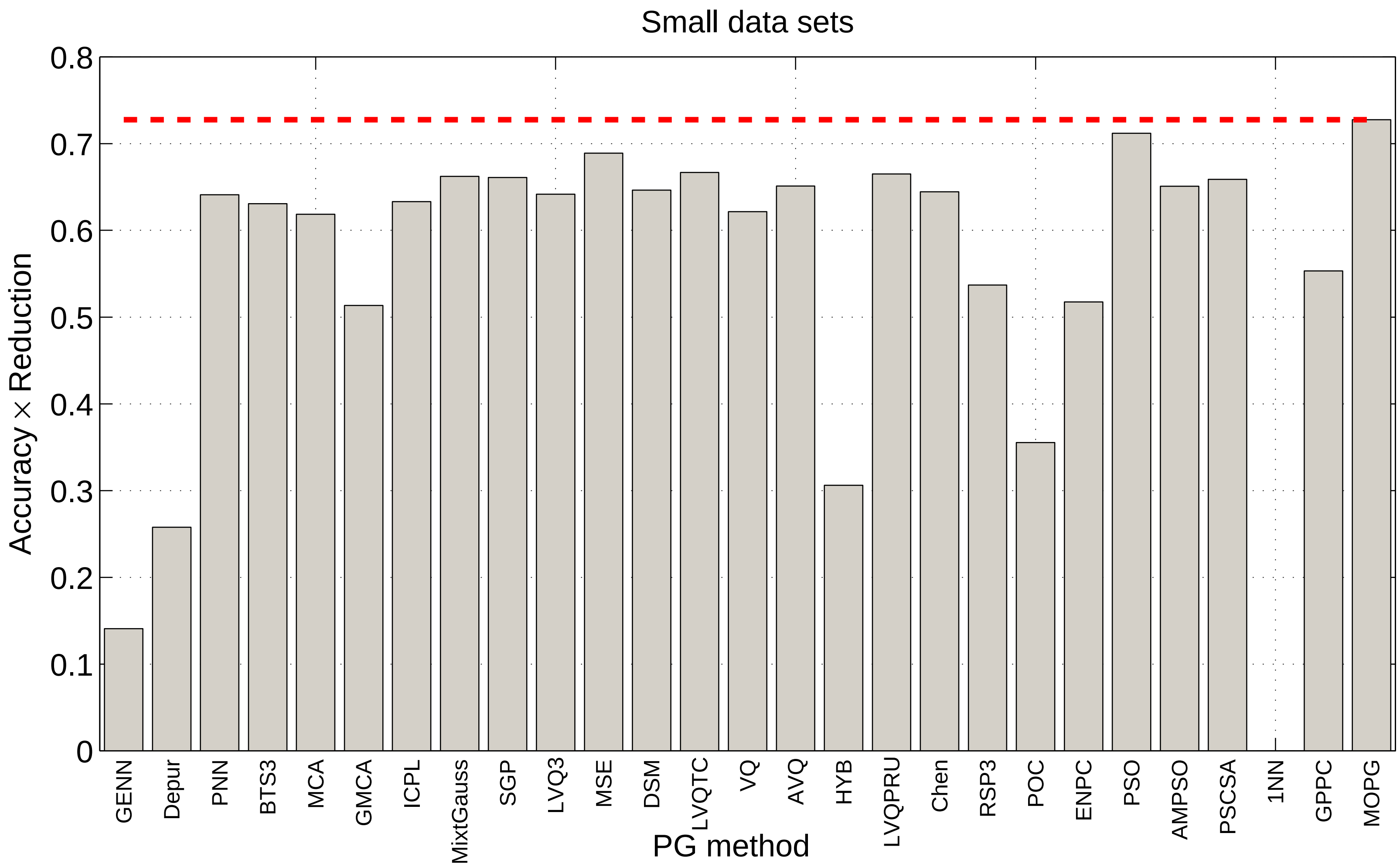
# Experiments & results



# Experiments & results

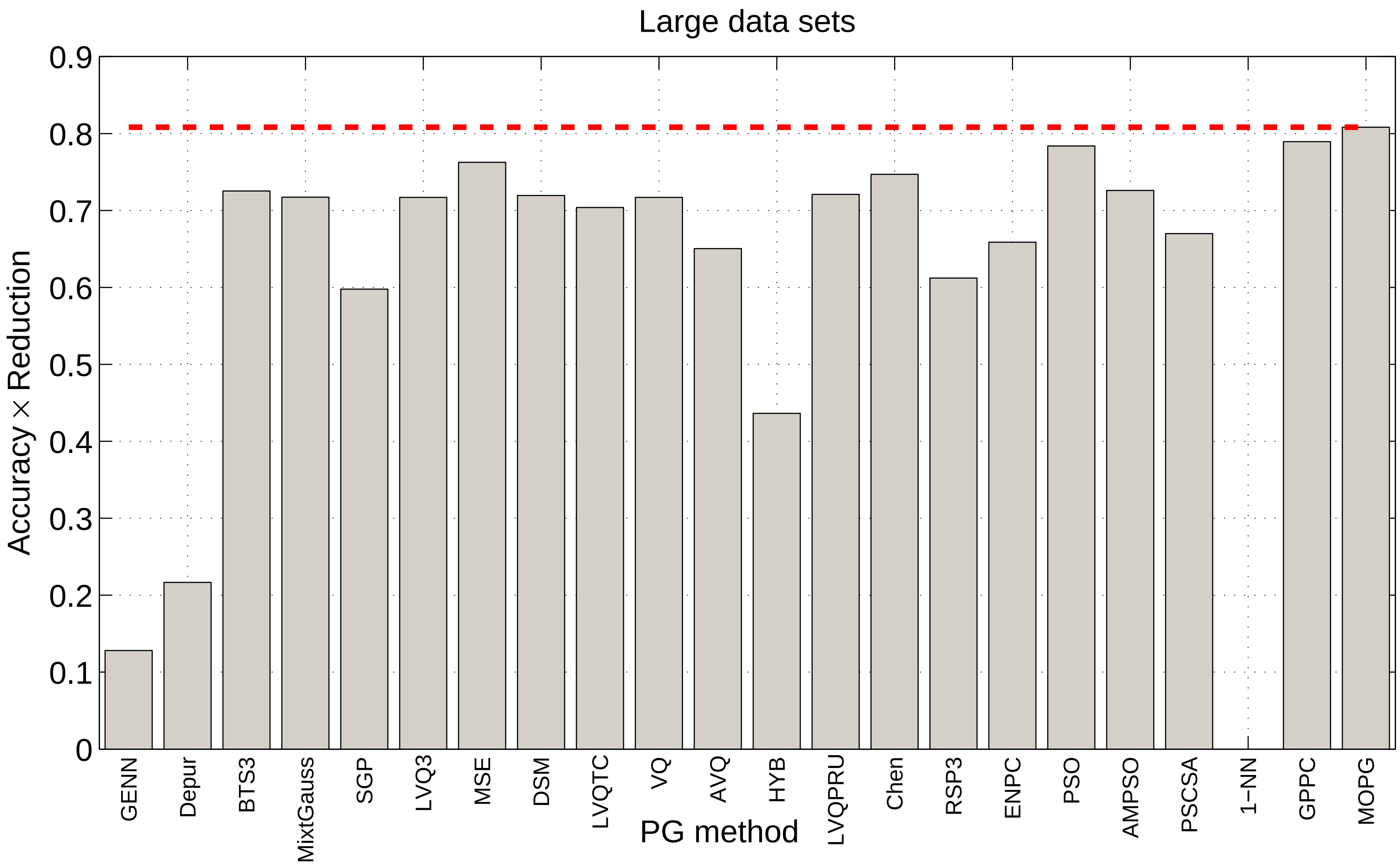


# Experiments & results




**Reduction – Accuracy tradeoff (reduction * accuracy)**

# Experiments & results




**Reduction – Accuracy tradeoff (reduction * accuracy)**

# Experiments & results

- Comparison with the best* methods (so far) for PG

| Reduction × Accuracy | | | |
|---|---|---|---|
| **Method** | **Ref.** | **Small** | **Large** |
| MOGP | Ours | 80.06 | 72.33 |
| SFLSDE/RandtoBest/1/Bin | (Triguero et al., 2011) | 81.54 | 72.23 |
| SFLSDE/Rand/1/Bin | (Triguero et al., 2011) | 81.67 | 71.88 |
| SSMA+SFLSDE/RandtoBest/1/Bin | (Triguero et al., 2011) | 81.64 | 74.95 |
| **(2 × Reduction × Accuracy)/(Reduction + Accuracy)** | | | |
| MOGP | Ours | 88.98 | 84.12 |
| SFLSDE/RandtoBest/1/Bin | (Triguero et al., 2011) | 89.92 | 84.48 |
| SFLSDE/Rand/1/Bin | (Triguero et al., 2011) | 89.99 | 84.25 |
| SSMA+SFLSDE/RandtoBest/1/Bin | (Triguero et al., 2011) | 90.02 | 86.15 |

# Conclusions

- The multi-objective formulation for PG is a promising alternative to mono-objective approaches
  - We hope our work can foster the development of other multi-objective optimization methods for PG.

- We showed evidence supporting the hypothesis that our proposal, MOPG, is very competitive in terms of both objectives reduction and accuracy
  - MOPG outperforms most PG methods proposed so far

# Future work

- Devising better ways to select the best solution from the Pareto front
- Efficient implementation of MOPG to deal with big-data problems (GPUs)
- Adapt MOPG for the generation of visual vocabularies

M. García-Limón, H. J. Escalante, E. Morales, A. Morales. Simultaneous **Generation of Prototypes and Features through Genetic Programming.** GECCO '14 Proceedings of the 2014 conference on Genetic and evolutionary computation, pp. 517-524, (Full paper, Oral presentation), Vancouver, Canada, July, 12-17, 2014.

M. Alfonso García, H. J. Escalante, E. Morales. **Towards Simultaneous Prototype and Feature Generation.** Proc. of the XVI IEEE Autumn Meeting of Power, Electronics and Computer Science ROPEC 2014 INTERNACIONAL, pp. 393—398, 2014.

# SIMULTANEOUS GENERATION OF PROTOTYPES AND FEATURES

# Generación de prototipos y características mediante programación genética multi-objetivo

**Mauricio A. García Limón**
Tesis de Maestría
Asesores: Dr. Hugo Jair Escalante.
Dr. Eduardo Morales F.

Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)
Coordinación de Ciencias Computacionales

Diciembre 1, 2014

**Best MS Thesis on Artificial Intelligence 2015, (SMIA)**

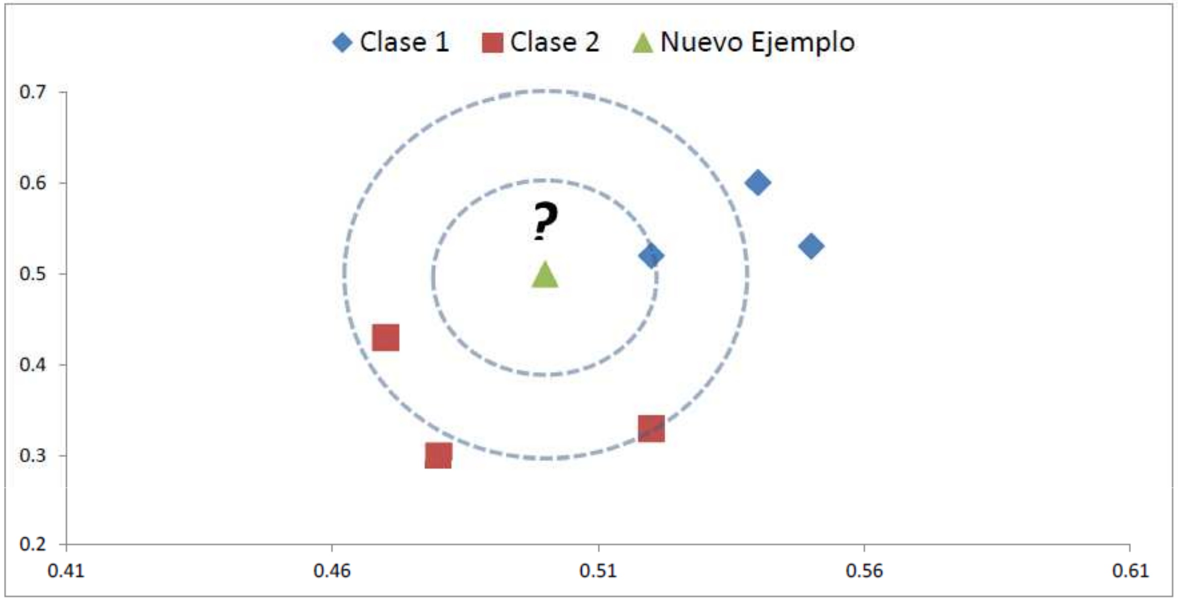



Figura: Clasificador de vecinos más cercanos (kNN).

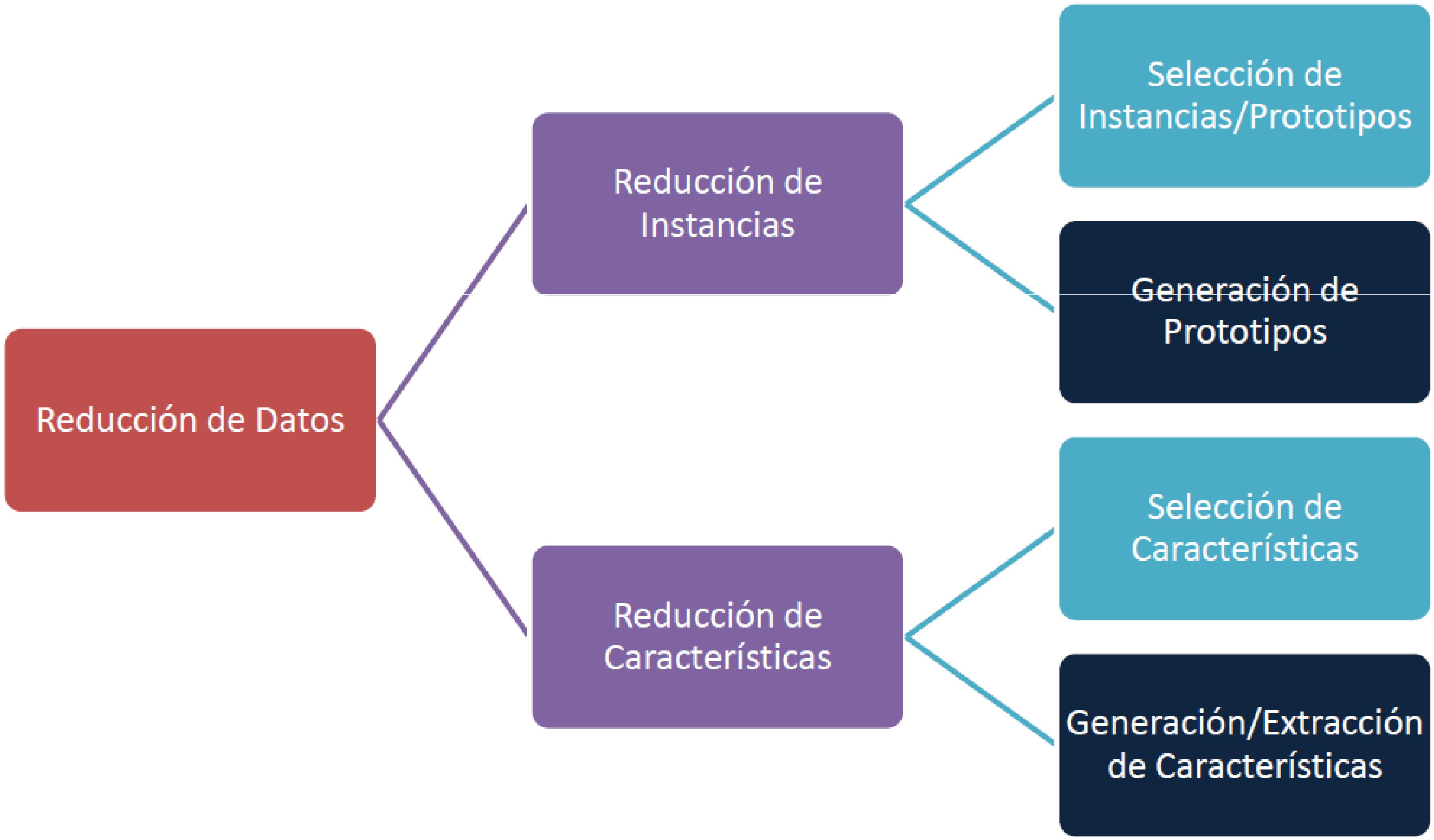
**Ventajas**

* Sencilla implementación
* No necesita etapa de entrenamiento.
* Manejo de un gran número de clases.

**Desventajas**

* Alto costo en almacenamiento ($O(N * M)$).
* Alto costo computacional ($O(N^2)$).
* Sensibilidad a ejemplos mal clasificados (Ruido).

# Reduciendo el costo de kNN



# Simultaneous generation of features and prototypes

- Is it possible to apply the same approach to generate features?
- Is it possible to perform both feature and prototype generation simultaneously?
- A multi-objective formulation would further help?

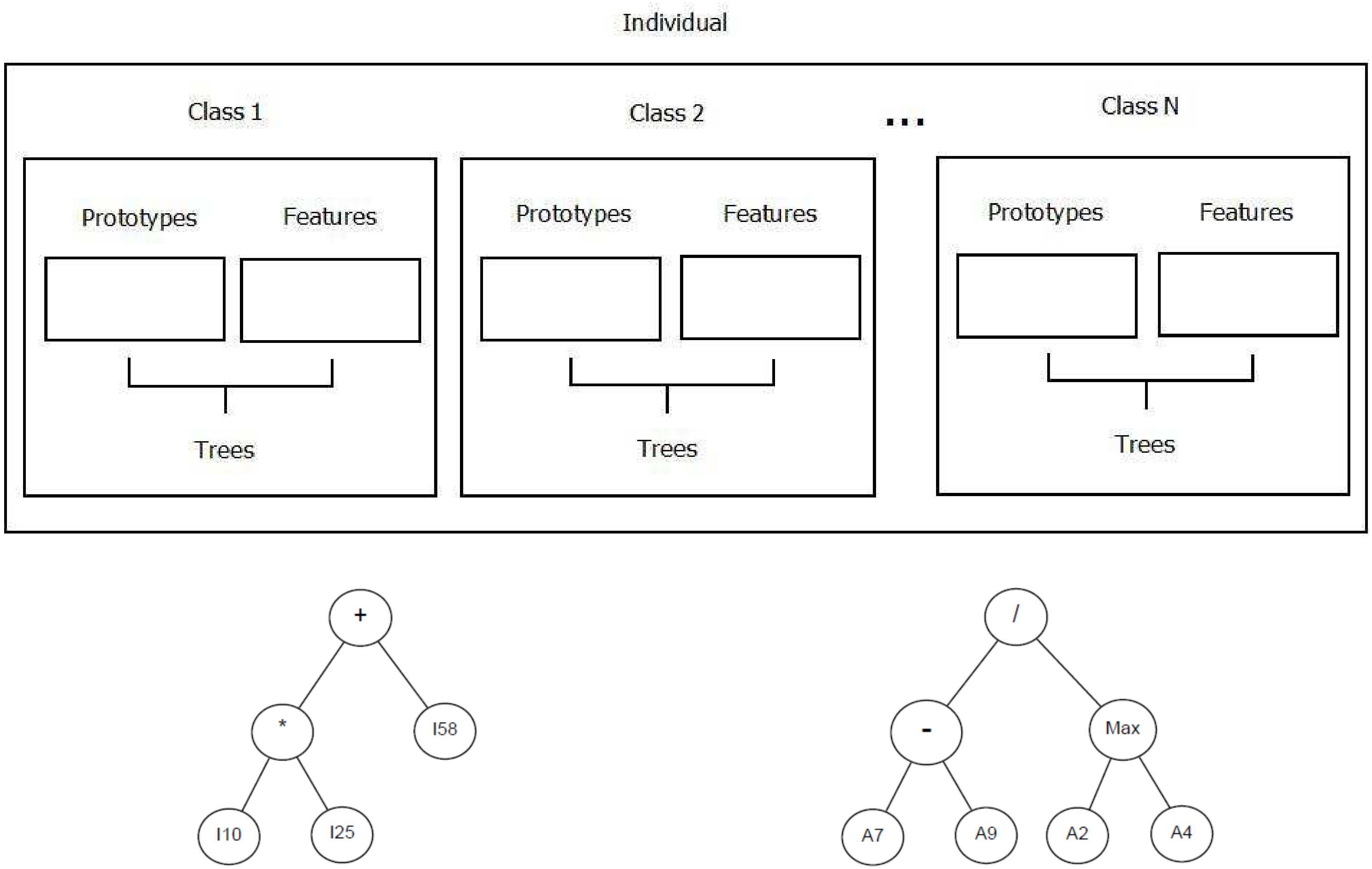
# Simultaneous generation of features and prototypes

- We aim to find a set of prototypes and features such that:
  - Accuracy is maximized
  - Number of instances reduced
  - Number of features is kept low

- Proposed solution: **Multi-objective GP**
  - Same idea: combine instances/features to generate prototypes/features.
  - Multiobjective implementation (NSGA-II)

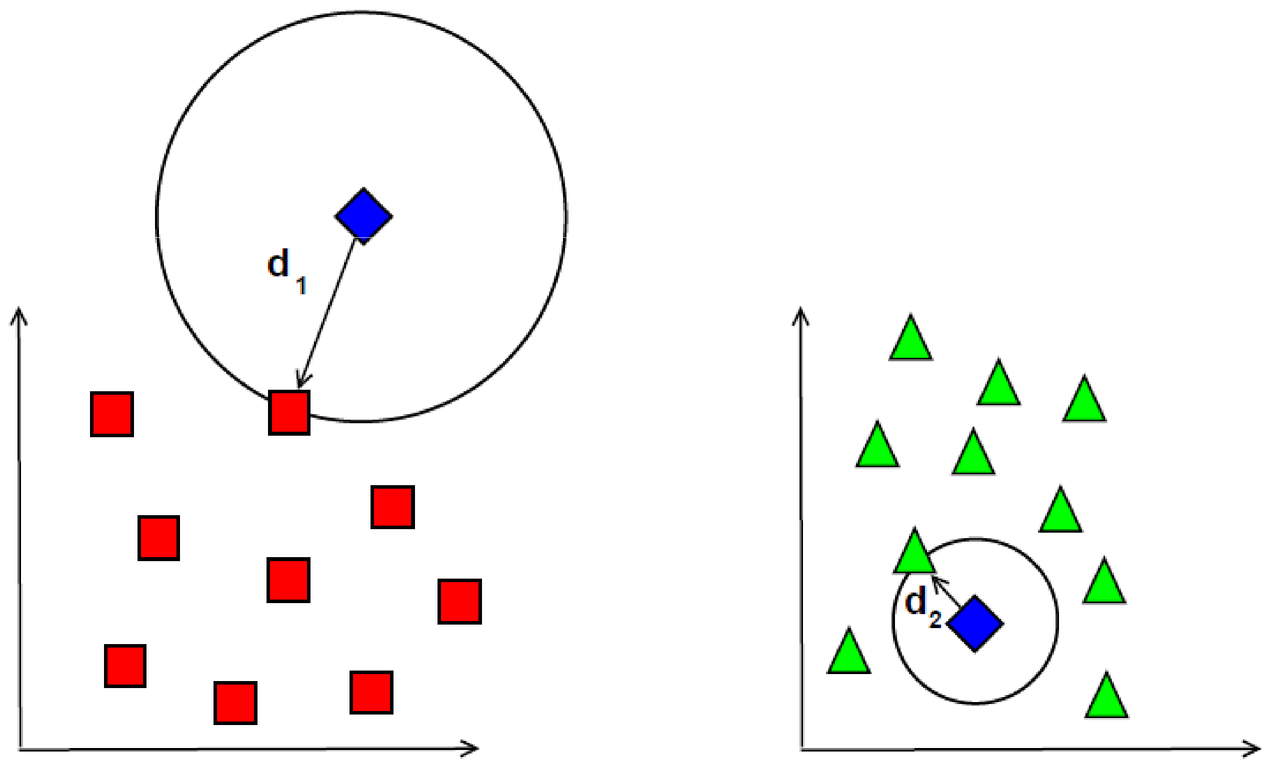
M. García-Limón, H. J. Escalante, E. Morales, A. Morales. Simultaneous **Generation of Prototypes and Features through Genetic Programming.** GECCO '14 Proceedings of the 2014 conference on Genetic and evolutionary computation, pp. 517-524, (Full paper, Oral presentation), Vancouver, Canada, July, 12-17, 2014.

# Simultaneous generation of features and prototypes



# Simultaneous generation of features and prototypes

- A different feature space for each class



# NSGA-II : (perhaps) the most used MOEA

**Algorithm 1** NSGA-II algorithm Deb et al. (2002).

**Require:** $N_{pop}$, $\mathbf{f}$, $g$

$\{N_{pop}$ number of individuals (solutions); $g$ number of generations

$\mathbf{f} = \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$ objectives$\}$

Initialize population $\mathcal{X}_0$

Evaluate objective functions $\mathbf{f} = \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

**while** $i = 1 < g$ **do**

  Create child population $\mathcal{Q}_i$ from $\mathcal{X}_i$ applying evolutionary operators.

  Evaluate objective functions $\mathbf{f}$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{Q}_i$

  Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_i \cup \mathcal{Q}_i$

  $\mathcal{X}_{i+1} = \emptyset$; $j = 1$;

  **while** $|\mathcal{X}_{i+1}| < N_{pop}$ **do**

    $\mathcal{X}_{i+1} = \mathcal{X}_{i+1} \cup \mathcal{F}_j$; $j = j + 1$;

  **end while**

  Select the last individuals for $\mathcal{X}_{i+1}$ from $\mathcal{F}_j$ using crowding distance

**end while**

Non-dominated sorting

# NSGA-II : (perhaps) the most used MOEA

**Algorithm 1** NSGA-II algorithm Deb et al. (2002).

**Require:** $N_{pop}$, f, $g$

$\{N_{pop}$ number of individuals (solutions); $g$ number of generations

f $= \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$ objectives$\}$

Initialize population $\mathcal{X}_0$

Evaluate objective functions f $= \langle f_1(\mathcal{P}), f_2(\mathcal{P})\rangle$, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_0$

**while** $i = 1 < g$ **do**

  Create child population $\mathcal{Q}_i$ from $\mathcal{X}_i$ applying evolutionary operators.

  Evaluate objective functions f, $\forall \mathcal{P} : \mathcal{P} \in \mathcal{Q}_i$

  Identify fronts $\mathcal{F}_{1,\ldots,F}$ by sorting solutions according to their non-dominance level $\forall \mathcal{P} : \mathcal{P} \in \mathcal{X}_i \cup \mathcal{Q}_i$

  $\mathcal{X}_{i+1} = \emptyset$; $j = 1$;

  **while** $|\mathcal{X}_{i+1}| < N_{pop}$ **do**

    $\mathcal{X}_{i+1} = \mathcal{X}_{i+1} \cup \mathcal{F}_j$; $j = j + 1$;

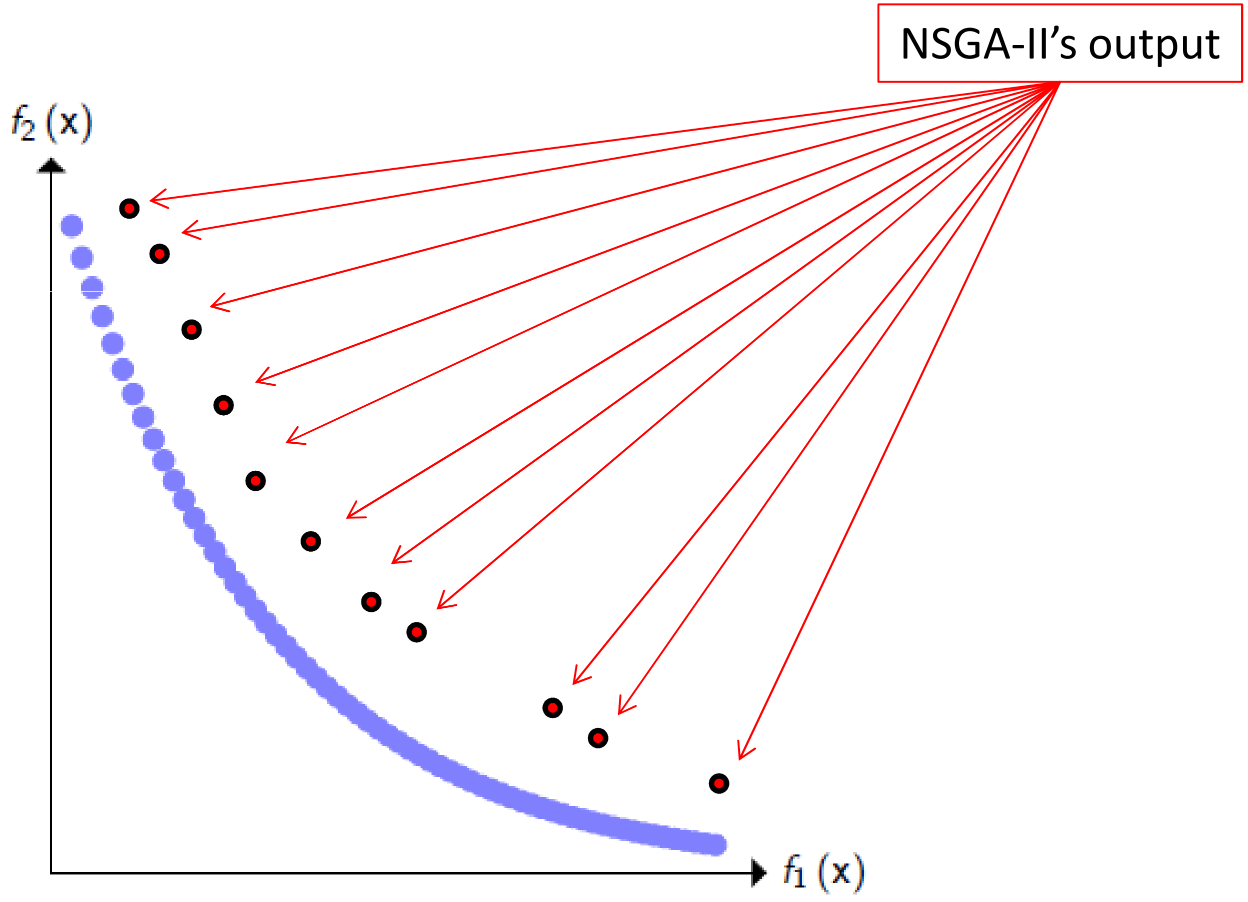
  **end while**

  Select the last individuals for $\mathcal{X}_{i+1}$ from $\mathcal{F}_j$ using crowding distance

**end while**

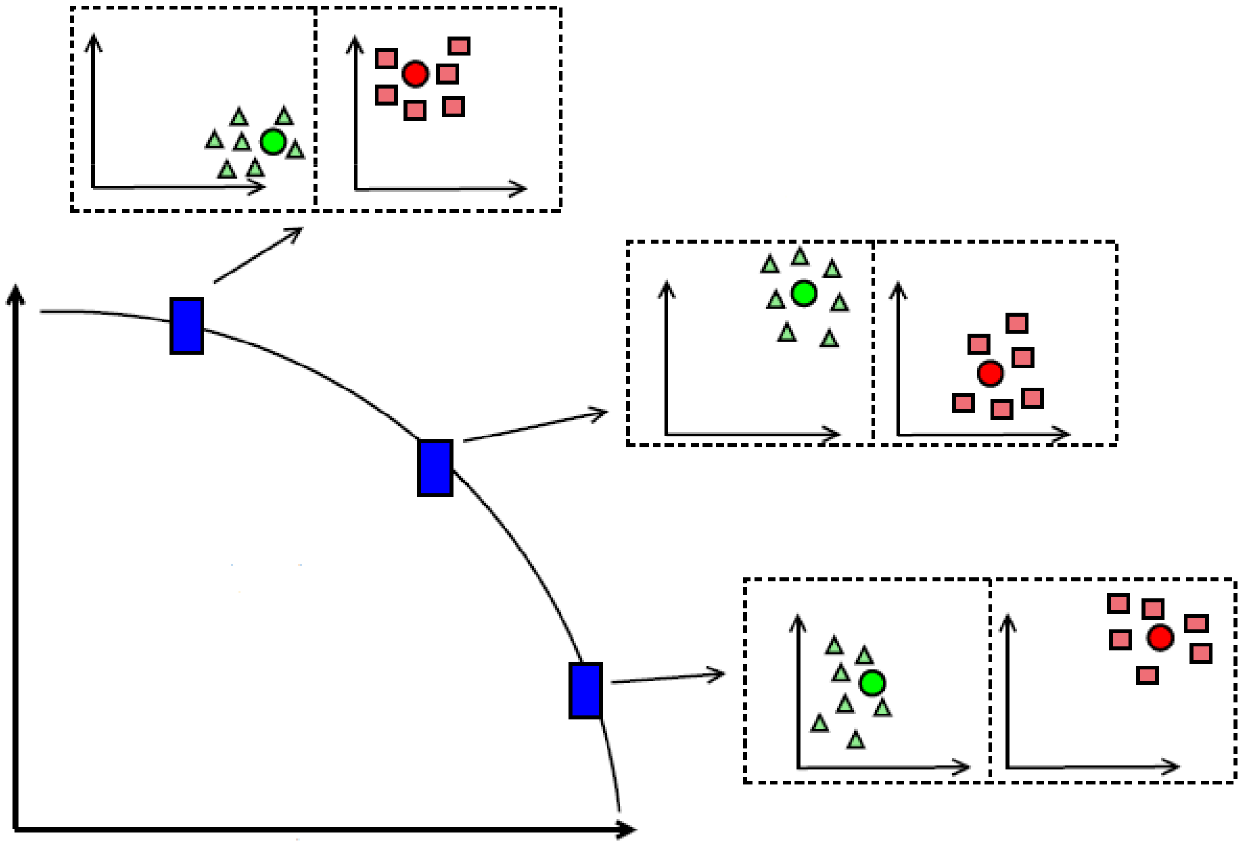
Crowding distance

# NSGA-II : (perhaps) the most used MOEA



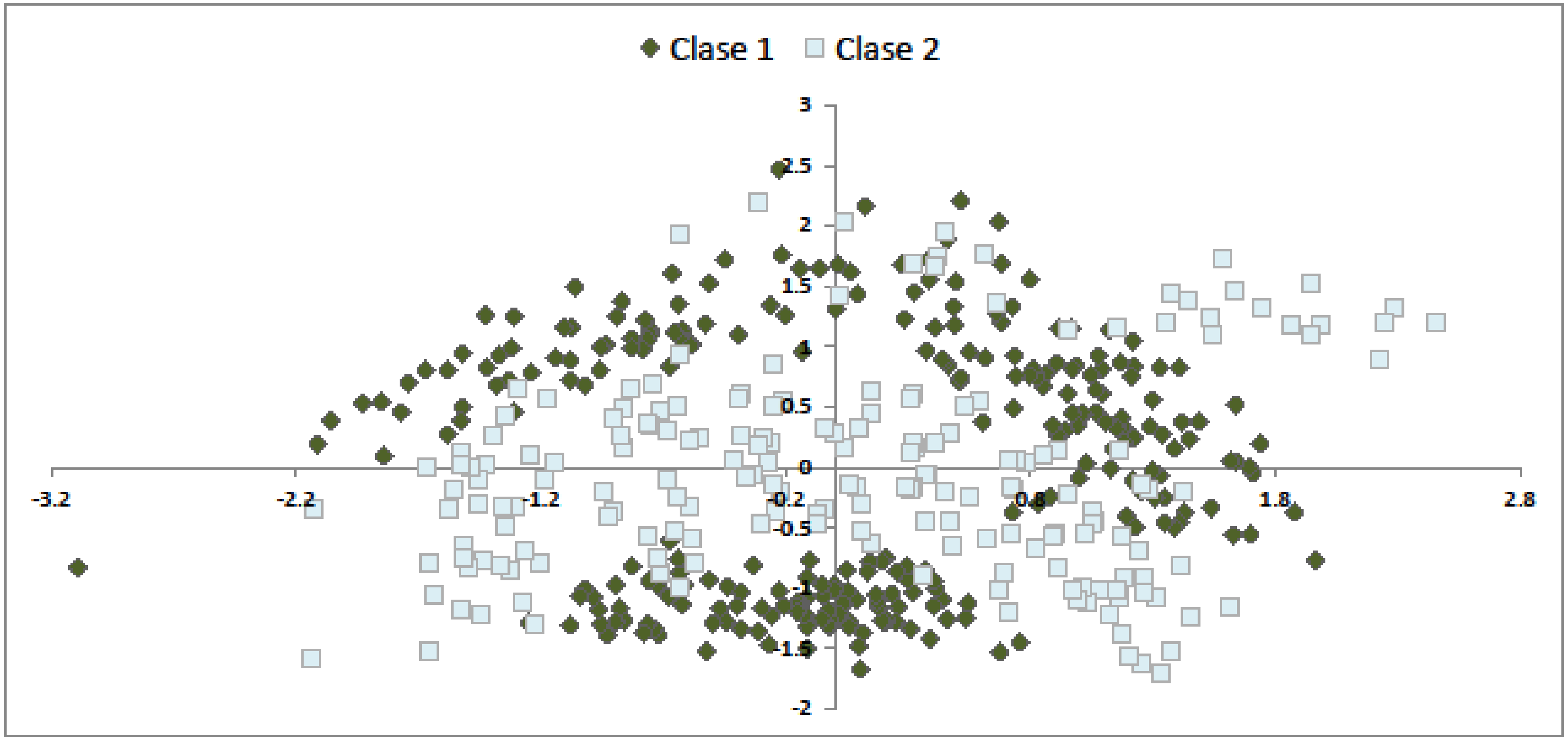
# Simultaneous generation of features and prototypes

- We select a solution by looking at accuracy only



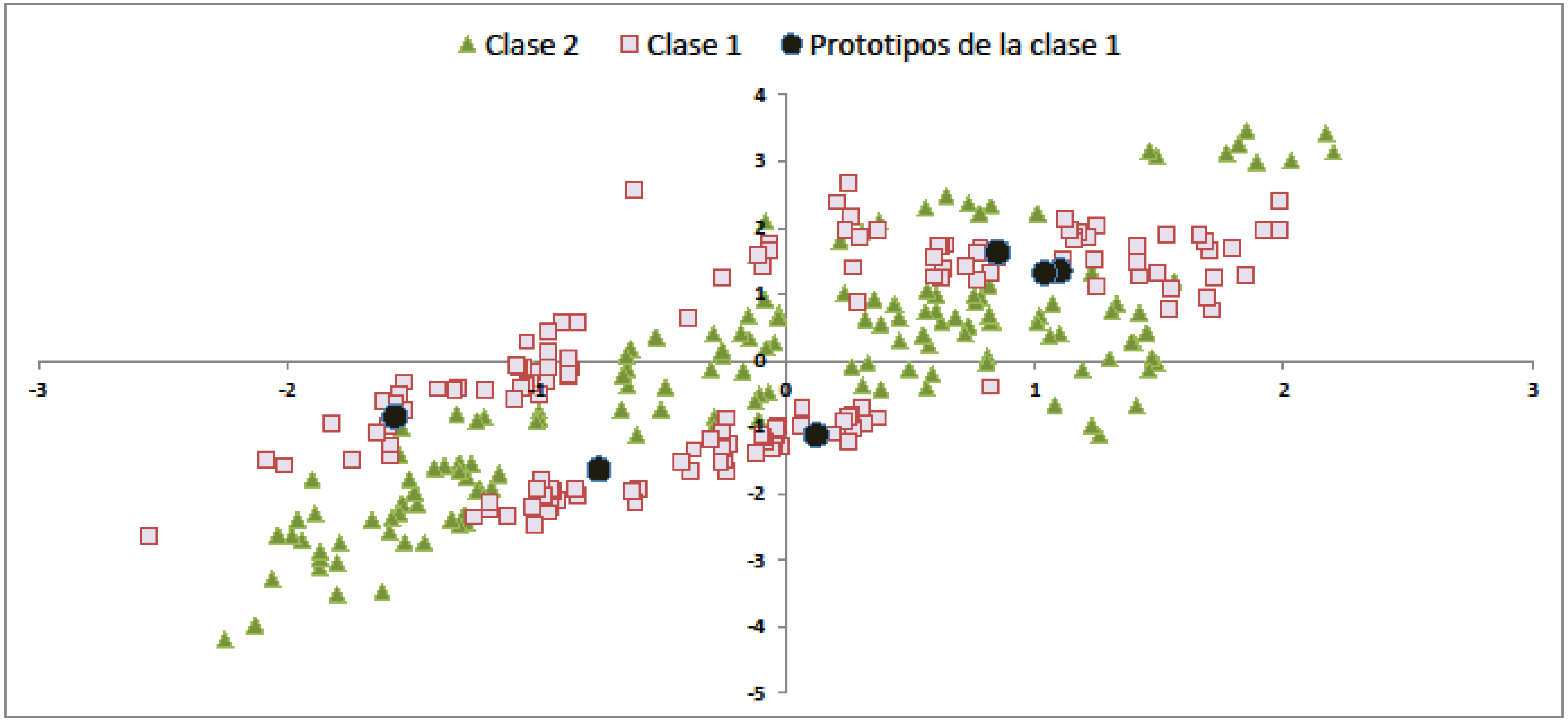
# Simultaneous generation of features and prototypes

- Example:
  - Original data set (initial instances and input space)



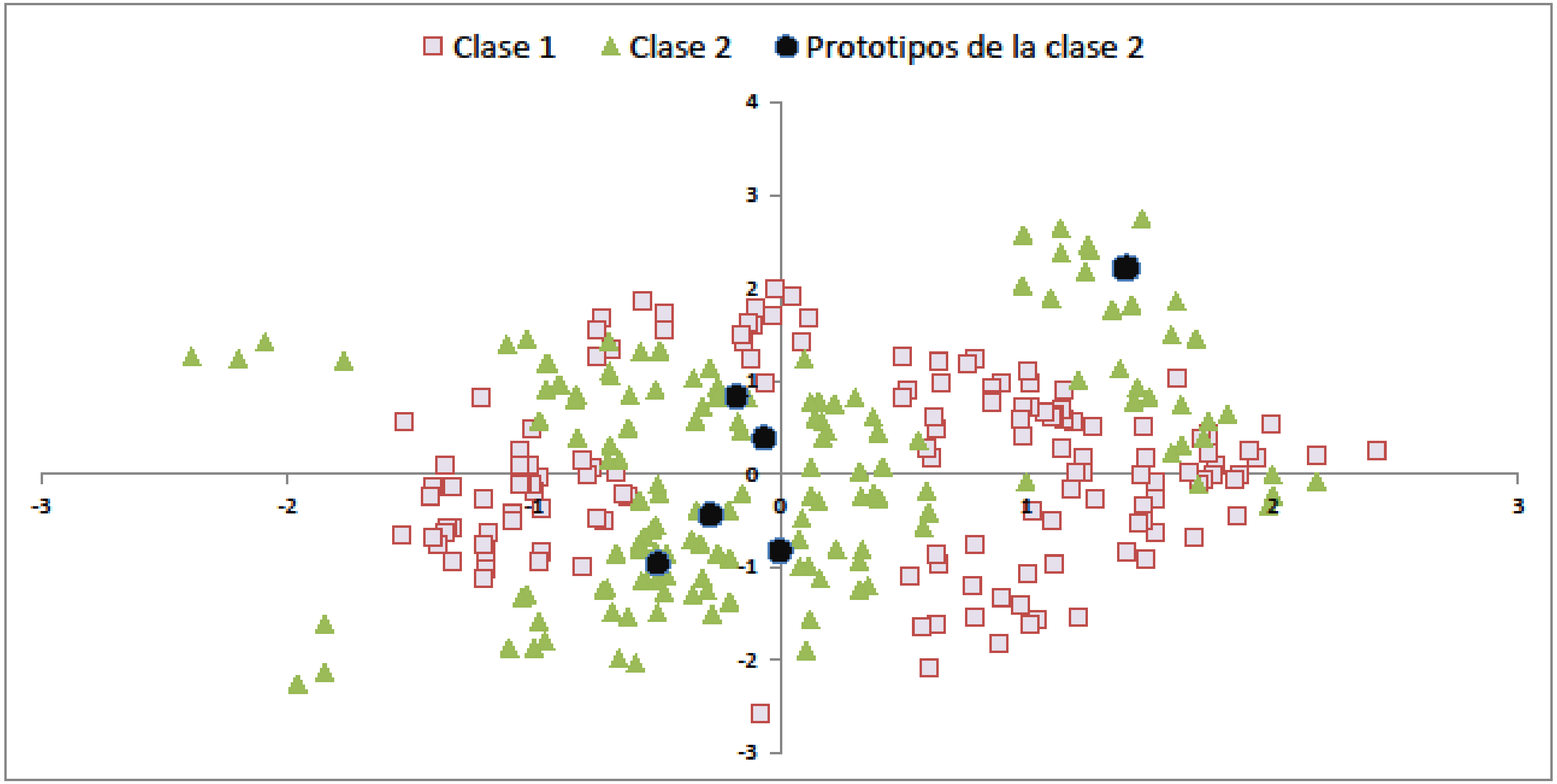
# Simultaneous generation of features and prototypes

- Example:
  - Prototypes and input space for class 1



# Simultaneous generation of features and prototypes

- Example:
  - Prototypes and input space for class 2



# Simultaneous generation of features and prototypes

- Some results:

Tabla: Método mono-objetivo vs generación independiente de prototipos y características.

| | Propuesta | | Sólo Prototipos | | Sólo Características | |
|---|---|---|---|---|---|---|
| | Media | Std | Media | Std | Media | Std |
| Exactitud Promedio | 80.25 | 19.93 | 78.37 | 20.24 | **82.82** | 21.91 |
| Reducción de Instancias Promedio | **99.43** | 0.09 | 99.40 | 0.11 | - | - |
| Reducción de Características Promedio | **41.45** | 12.76 | - | - | 36.78⋆ | 14.65 |

# Simultaneous generation of features and prototypes

- Some results:

Small data sets

| | Multi-Objetivo | | Mono-Objetivo | | EMOPG+FS | | MOPG | | 1NN | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Media | Std. | Media | Std. | Media | Std. | Media | Std. | Media | Std. |
| Exactitud | 70.84 | 17.74 | 71.97 | 15.76 | 74.26⋆ | 14.70 | **74.82** | 18.35 | 73.48 | 16.64 |
| Red. de Instancias | 98.43 | 1.65 | 98.39⋆ | 1.37 | 97.32⋆ | 1.49 | **98.60** | 1.32 | - | - |
| Red. de Características | **88.65** | 7.52 | 42.62⋆ | 5.13 | 60.47⋆ | 19.14 | - | - | - | - |

⋆ Diferencia estadísticamente significante.

Large data sets

| | Multi-Objetivo | | Mono-Objetivo | | EMOPG+FS | | MOPG | | 1NN | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Media | Std. | Media | Std. | Media | Std. | Media | Std. | Media | Std. |
| Exactitud | 76.19 | 21.25 | 80.25⋆ | 19.93 | **81.82**⋆ | 20.24 | 72.10 | 19.46 | 80.60 | 22.24 |
| Red. de Instancias | **99.82** | 0.16 | 99.43⋆ | 0.09 | 98.42⋆ | 0.45 | 98.82⋆ | 0.70 | - | - |
| Red. de Características | **89.74** | 12.36 | 41.45⋆ | 12.76 | 49.09⋆ | 29.00 | - | - | - | - |

# Simultaneous generation of features and prototypes

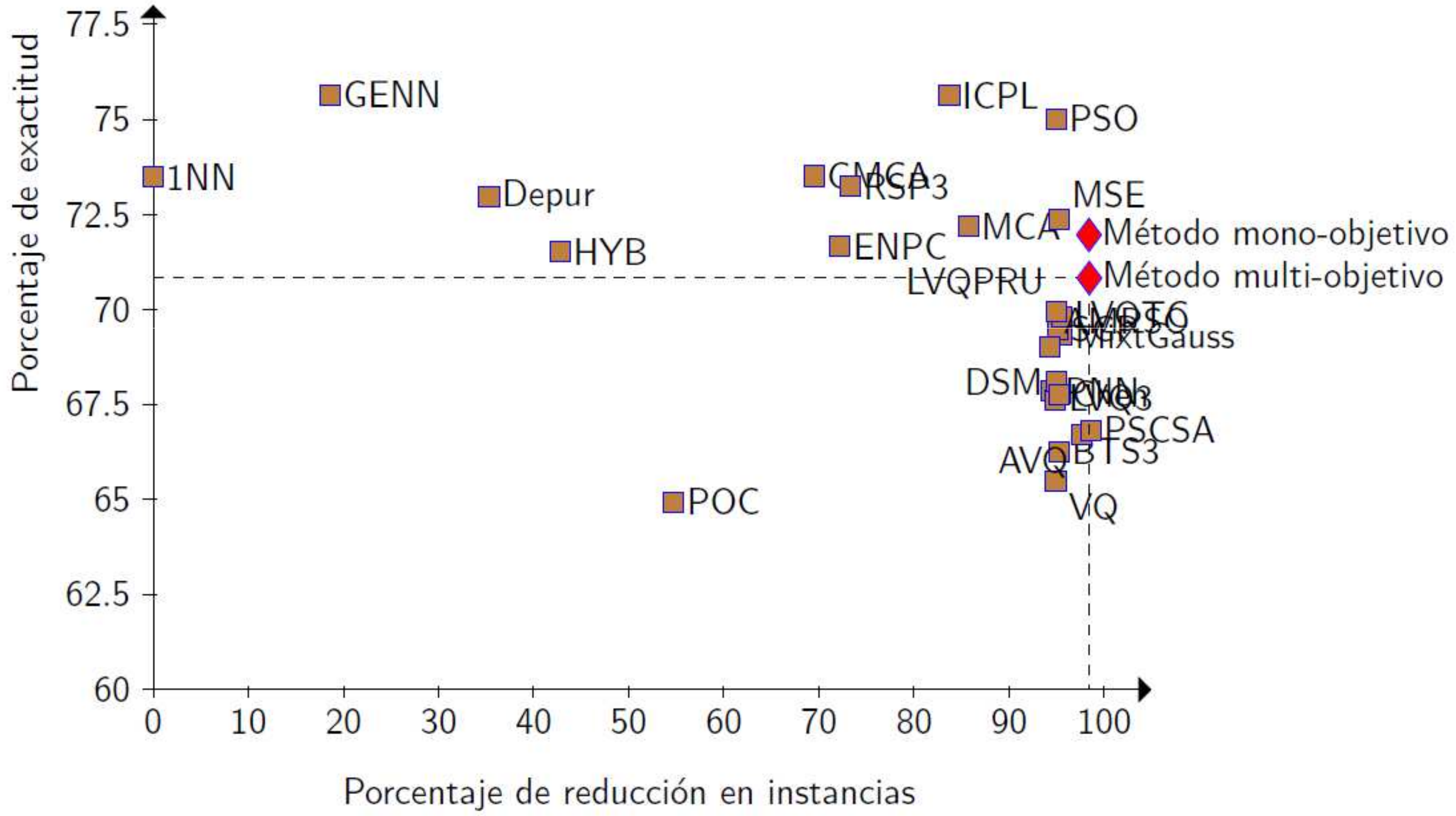



Figura: Reducción promedio vs. exactitud para bases de datos pequeñas

# Simultaneous generation of features and prototypes

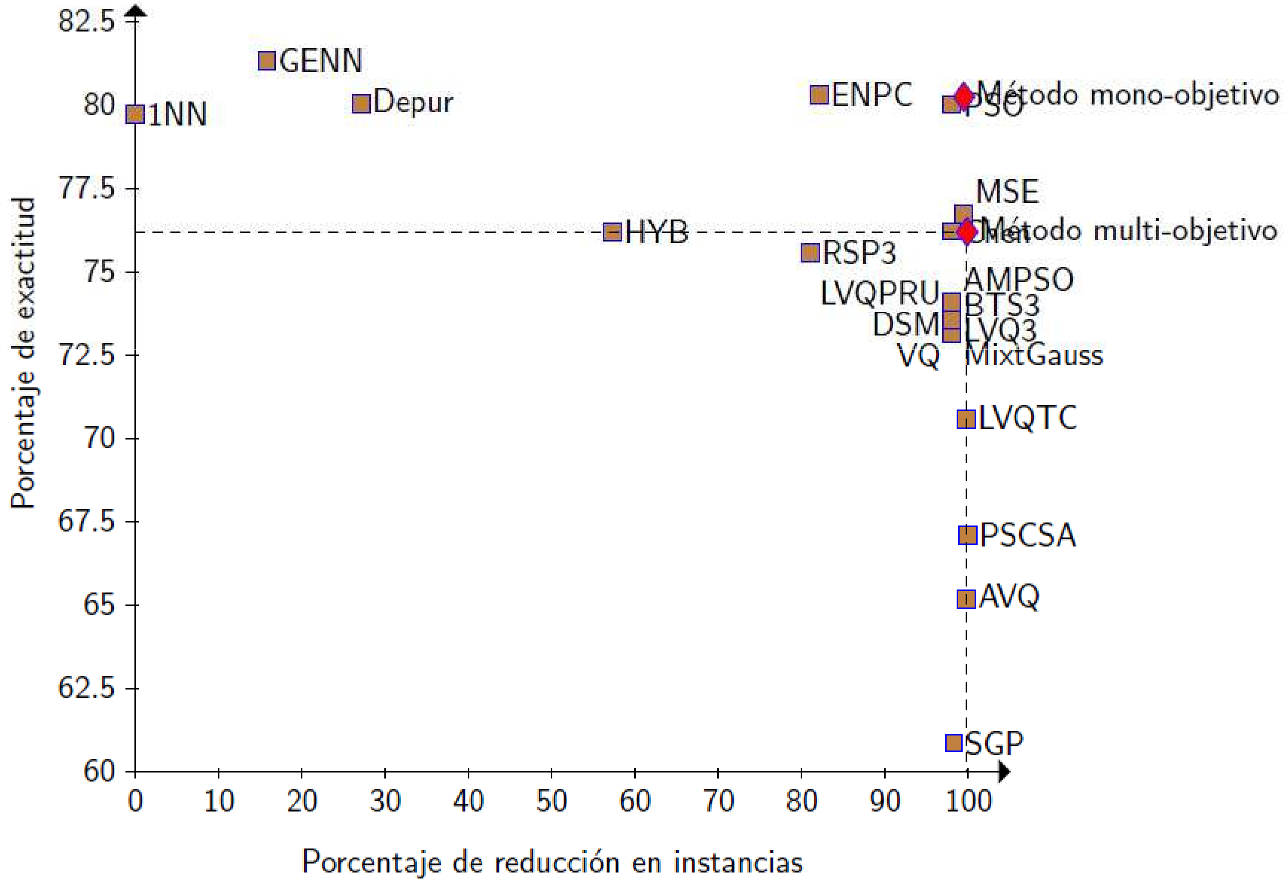



Figura: Reducción promedio vs. exactitud para bases de datos grandes

# Simultaneous generation of features and prototypes

- Competitive performance on generation of both prototypes and features

- Class-specific input spaces

- Other uses: oversampling, data embedding, visualization,

- **Issue:** not scalable to large data sets

Questions?