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:

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad g_i(x) \leq 0 \text{ for } i = \{1,\ldots, l\} \\
& \quad h_j(x) = 0 \text{ for } j = \{1,\ldots, J\} \\
& \quad x^l_k \leq x_k \leq x^u_k \text{ for } k = \{1,\ldots, n\}
\end{align*}
\]
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:

\[
\begin{align*}
\text{min} & \quad f(\mathbf{x}) = \langle f_1(\mathbf{x}), \ldots, f_N(\mathbf{x}) \rangle \\
\text{s.t.} & \quad g_i(\mathbf{x}) \leq 0 \text{ for } i = \{1, \ldots, I\} \\
& \quad h_j(\mathbf{x}) = 0 \text{ for } j = \{1, \ldots, J\} \\
& \quad x^l_k \leq x_k \leq x^u_k \text{ for } k = \{1, \ldots, n\}
\end{align*}
\]
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 $x_1$ dominates $x_2$ iff $x_1$ is better than $x_2$ 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

Trial and error problem solving approach:

1. While not satisfied with solution:
   a. Generate candidate solution(s) for the problem at hand
   b. Evaluate the quality of the candidate solution(s)
      - Return best solution found

EC techniques generate new solutions according to (rough) analogies with biological evolution principles.
NSGA-II : (perhaps) the most used MOEA

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

Require: \( N_{\text{pop}} \), \( f \), \( g \)
\( \{ N_{\text{pop}} \) number of individuals (solutions); \( g \) number of generations \)
\( f = (f_1(P), f_2(P)) \) objectives \\
1: Initialize population \( \mathcal{X}_0 \) \\
2: Evaluate objective functions \( f = (f_1(P), f_2(P)), \forall P : P \in \mathcal{X}_0 \)

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

4: while \( i = 1 < g \) do 
5: Create child population \( \mathcal{Q}_i \) from \( \mathcal{X}_i \) applying evolutionary operators. 
6: Evaluate objective functions \( f, \forall P : P \in \mathcal{Q}_i \)

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

8: \( \mathcal{X}_{i+1} = \emptyset; j = 1; \)
9: while \( |\mathcal{X}_{i+1}| < N_{\text{pop}} \) do 
10: \( \mathcal{X}_{i+1} = \mathcal{X}_{i+1} \cup \mathcal{F}_j; j = j + 1; \)
11: end while 
12: Select the last individuals for \( \mathcal{X}_{i+1} \) from \( \mathcal{F}_j \) using crowding distance 
13: end while

Non-dominated sorting
NSGA-II : (perhaps) the most used MOEA

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

Require: $N_{\text{pop}}, f, g$

- $N_{\text{pop}}$ number of individuals (solutions); $g$ number of generations
- $f = \{f_1(P), f_2(P)\}$ objectives

1: Initialize population $X_0$
2: Evaluate objective functions $f = \{f_1(P), f_2(P)\}$, $\forall P \in X_0$
3: Identify fronts $F_{1,\ldots,g}$ by sorting solutions according to their non-dominance level $\forall P : P \in X_0$
4: while $i = 1 < g$ do
5: Create child population $Q_i$ from $X_i$ applying evolutionary operators.
6: Evaluate objective functions $f$, $\forall P : P \in Q_i$
7: Identify fronts $F_{1,\ldots,g}$ by sorting solutions according to their non-dominance level $\forall P : P \in X_i \cup Q_i$
8: $X_{i+1} = \emptyset$; $j = 1$
9: while $|X_{i+1}| < N_{\text{pop}}$ do
10: $X_{i+1} = X_{i+1} \cup F_j$; $j = j + 1$
11: end while
12: Select the last individuals for $X_{i+1}$ from $F_j$ using crowding distance
13: end while
NSGA-II: (perhaps) the most used MOEA

NSGA-II’s output

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
• 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/
• 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
    • ...
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:

\[
\text{Maximize } \langle \rho(S^0, S), \beta(S) \rangle
\]

• Where:

\[
\rho(S^0, S) = 1 - \frac{6}{n(n^2 - 1)} \sum_i d_{r_i}(S^0, S)^2
\]

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

| 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

<table>
<thead>
<tr>
<th>Rank</th>
<th>List</th>
<th>$S^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Image 181x149 to 224x206" /></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td><img src="image" alt="Image 295x160 to 337x192" /></td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Image 244x160 to 286x192" /></td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td><img src="image" alt="Image 482x155 to 524x211" /></td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td><img src="image" alt="Image 419x160 to 462x192" /></td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td><img src="image" alt="Image 357x149 to 400x206" /></td>
<td>0.16</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

![Initial population](image)
Multi-objective optimization for result diversification

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

\[
\text{Maximize } < \rho(S^0, S) , \beta(S) >
\]

- Where:

\[
\rho(S^0, S) = 1 - \frac{6}{n(n^2 - 1)} \sum dr(i)(S^0, S)^2
\]

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

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:

$$\beta(S) = \sum_{i=2}^{N} \min_{d}(d(I_i, I_1, ..., i-1))$$
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

Table 2: Official results obtained by MORD.

<table>
<thead>
<tr>
<th>Expert evaluation</th>
<th>C@10</th>
<th>C@20</th>
<th>P@10</th>
<th>P@20</th>
<th>F@10</th>
<th>F@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3808</td>
<td>0.5699</td>
<td>0.7146</td>
<td>0.7143</td>
<td>0.4795</td>
<td>0.6125</td>
</tr>
<tr>
<td>2</td>
<td>0.3885</td>
<td>0.5732</td>
<td>0.7091</td>
<td>0.7136</td>
<td>0.4801</td>
<td>0.6102</td>
</tr>
<tr>
<td>3</td>
<td>0.3823</td>
<td>0.5690</td>
<td>0.6977</td>
<td>0.7076</td>
<td>0.4728</td>
<td>0.6067</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crowd-sourcing evaluation</th>
<th>C@10</th>
<th>C@20</th>
<th>P@10</th>
<th>P@20</th>
<th>F@10</th>
<th>F@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7194</td>
<td>0.8499</td>
<td>0.6755</td>
<td>0.6745</td>
<td>0.6640</td>
<td>0.7245</td>
</tr>
<tr>
<td>2</td>
<td>0.7503</td>
<td>0.8625</td>
<td>0.6755</td>
<td>0.6898</td>
<td>0.6790</td>
<td>0.7392</td>
</tr>
<tr>
<td>3</td>
<td>0.7479</td>
<td>0.8675</td>
<td>0.6714</td>
<td>0.6918</td>
<td>0.6769</td>
<td>0.7464</td>
</tr>
</tbody>
</table>
Experiments & results

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
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

The two key aspects for the evaluation of PG methods are reduction and accuracy on unseen data. Maximizing reduction may cause accuracy to decrease and vice versa.
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:

\[
\begin{align*}
\text{maximize} & \quad \langle f_1(P), f_2(P) \rangle \\
\text{subject to} & \quad P \in \mathcal{Y}
\end{align*}
\]

• Where:

\[
\begin{align*}
f_1(P) &= \delta(P, D) \\
f_2(P) &= \gamma(P, D)
\end{align*}
\]
MOPG: Representation

Each solution is codified as a matrix of size $P \times d$

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 maintained)
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 from 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
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
Experiments and results

<table>
<thead>
<tr>
<th>Data Set</th>
<th>#Ex.</th>
<th>#Att.</th>
<th>#Num.</th>
<th>#Nom.</th>
<th>#CL.</th>
<th>Data Set</th>
<th>#Ex.</th>
<th>#Att.</th>
<th>#Num.</th>
<th>#Nom.</th>
<th>#CL.</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>4,174</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>28</td>
<td>marketing</td>
<td>8,593</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>appendicitis</td>
<td>106</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>monks</td>
<td>432</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>australian</td>
<td>690</td>
<td>14</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>movement_libras</td>
<td>360</td>
<td>90</td>
<td>90</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>autos</td>
<td>205</td>
<td>25</td>
<td>15</td>
<td>10</td>
<td>6</td>
<td>newthyroid</td>
<td>215</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>balance</td>
<td>625</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>nursery</td>
<td>12,960</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>banana</td>
<td>5,300</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>pageblocks</td>
<td>5,472</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>bands</td>
<td>139</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>penhased</td>
<td>10,992</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>breast</td>
<td>286</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>phoneme</td>
<td>5,404</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>bupa</td>
<td>345</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>pima</td>
<td>768</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>car</td>
<td>1,728</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>ring</td>
<td>7,600</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>chess</td>
<td>3,196</td>
<td>36</td>
<td>0</td>
<td>36</td>
<td>2</td>
<td>solheart</td>
<td>462</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>cleveland</td>
<td>297</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>5</td>
<td>satimage</td>
<td>6,435</td>
<td>36</td>
<td>36</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>coin2000</td>
<td>9,822</td>
<td>85</td>
<td>85</td>
<td>0</td>
<td>2</td>
<td>segment</td>
<td>2,310</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>contraceptive</td>
<td>1,473</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>sonar</td>
<td>208</td>
<td>60</td>
<td>70</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>crx</td>
<td>125</td>
<td>15</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>spam_base</td>
<td>4,597</td>
<td>55</td>
<td>55</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>dermatology</td>
<td>366</td>
<td>33</td>
<td>1</td>
<td>32</td>
<td>6</td>
<td>spectheart</td>
<td>267</td>
<td>44</td>
<td>44</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ecoli</td>
<td>350</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>splice</td>
<td>3,190</td>
<td>60</td>
<td>0</td>
<td>60</td>
<td>3</td>
</tr>
<tr>
<td>flare-solar</td>
<td>1,066</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>tac</td>
<td>151</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>german</td>
<td>1,000</td>
<td>20</td>
<td>6</td>
<td>14</td>
<td>2</td>
<td>texture</td>
<td>5,500</td>
<td>40</td>
<td>40</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>glass</td>
<td>214</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>7</td>
<td>tic-tac-toe</td>
<td>958</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>haberman</td>
<td>306</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>thyroid</td>
<td>7,200</td>
<td>21</td>
<td>6</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>hayes-roth</td>
<td>133</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>titanic</td>
<td>2,201</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>heart</td>
<td>270</td>
<td>13</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>twonorm</td>
<td>7,400</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>hepatitis</td>
<td>155</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>2</td>
<td>vehicle</td>
<td>846</td>
<td>18</td>
<td>18</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>housevotes</td>
<td>435</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>2</td>
<td>vowel</td>
<td>993</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>iris</td>
<td>150</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>lec7digit</td>
<td>500</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>wisconsin</td>
<td>683</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>lymphography</td>
<td>148</td>
<td>18</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>yeast</td>
<td>1484</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>magic</td>
<td>19,020</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>zoo</td>
<td>101</td>
<td>17</td>
<td>0</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>mammographic</td>
<td>961</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiments & results

- Evaluation of the selection strategy:

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>All</td>
<td>Small</td>
</tr>
<tr>
<td>Strategy</td>
<td>73.94±18.58</td>
<td>81.29±20.05</td>
</tr>
<tr>
<td>Best</td>
<td>76.64±17.46</td>
<td>82.05±19.77</td>
</tr>
</tbody>
</table>
Experiments & results

- Parameter settings

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Accuracy</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals ((N_{pop}))</td>
<td>50</td>
<td>71.68% ± 18.18</td>
<td>97.24% ± 1.21</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>72.25% ± 17.80</td>
<td>97.26% ± 1.25</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>73.13% ± 18.10</td>
<td>97.23% ± 1.31</td>
</tr>
<tr>
<td>Generations ((g))</td>
<td>50</td>
<td>71.68% ± 18.18</td>
<td>97.24% ± 1.21</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>72.71% ± 18.03</td>
<td>97.53% ± 1.29</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>73.32% ± 18.11</td>
<td>97.62% ± 1.34</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>73.37% ± 18.08</td>
<td>97.70% ± 1.31</td>
</tr>
<tr>
<td>Train set-size ((\eta))</td>
<td>0.1</td>
<td>72.11% ± 18.33</td>
<td>98.84% ± 1.09</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>73.07% ± 18.18</td>
<td>98.19% ± 1.13</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>71.68% ± 18.18</td>
<td>97.24% ± 1.21</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>72.12% ± 18.44</td>
<td>96.85% ± 1.36</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>70.30% ± 20.39</td>
<td>96.38% ± 1.67</td>
</tr>
<tr>
<td>Initial prot. ((I_p))</td>
<td>0.005</td>
<td>71.19% ± 18.18</td>
<td>99.01% ± 1.28</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>71.20% ± 18.31</td>
<td>98.97% ± 1.26</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>72.14% ± 18.74</td>
<td>98.37% ± 1.16</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>71.68% ± 18.18</td>
<td>97.24% ± 1.21</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>73.07% ± 17.76</td>
<td>95.19% ± 1.82</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>73.20% ± 17.72</td>
<td>89.86% ± 3.87</td>
</tr>
</tbody>
</table>
Experiments & results

• Parameter settings

![Graph showing parameter settings and their effects on accuracy and reduction.](image-url)
Experiments & results

• Comparison with related work

<table>
<thead>
<tr>
<th>Measure</th>
<th>Test set accuracy</th>
<th>Training set reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Small</td>
</tr>
<tr>
<td>MOGP</td>
<td>73.94%±18.38</td>
<td>70.93%±16.95</td>
</tr>
<tr>
<td>GENN</td>
<td>78.48%±18.57</td>
<td>75.64%±15.45</td>
</tr>
<tr>
<td>PSCSA</td>
<td>66.94%±20.39</td>
<td>66.82%±18.74</td>
</tr>
<tr>
<td>1NN</td>
<td>77.04%±19.44</td>
<td>73.48%±16.64</td>
</tr>
</tbody>
</table>
Experiments & results

Small data sets

Accuracy vs. Reduction

- GENN
- Depur
- HYB
- GMCA
- RSP3
- ENPC
- ICPL
- PNN
- BTS3
- MixtGauss
- SGP
- LVQ3
- DSM
- LVQTC
- VQ
- AVQ
- LVQPRU
- Chen
- POC
- AMPSO
- PSCSA
- MSE

Comparison with related work

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENN</td>
<td>0.72</td>
<td>0.2</td>
</tr>
<tr>
<td>Depur</td>
<td>0.74</td>
<td>0.4</td>
</tr>
<tr>
<td>HYB</td>
<td>0.76</td>
<td>0.6</td>
</tr>
<tr>
<td>GMCA</td>
<td>0.64</td>
<td>0.8</td>
</tr>
<tr>
<td>RSP3</td>
<td>0.66</td>
<td>1</td>
</tr>
<tr>
<td>ENPC</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>ICPL</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>PNN</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>BTS3</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>MixtGauss</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>SGP</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>LVQ3</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>DSM</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>LVQTC</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>VQ</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>AVQ</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>LVQPRU</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Chen</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>POC</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>AMPSO</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>PSCSA</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.7</td>
<td></td>
</tr>
</tbody>
</table>
Experiments & results

Large data sets

Accuracy vs Reduction

Algorithms: GENN, Depur, PM, MixtGauss, SGP, MSE, LVQTC, HYB, LVQPRU, RSP3, ENPC, PSO, AMPSO, 1NN, GPPC, MOPG.
Experiments & results

Small data sets

Reduction – Accuracy tradeoff (reduction * accuracy)
Experiments & results

Reduction – Accuracy tradeoff (reduction * accuracy)
Experiments & results

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Ref.</th>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOGP</td>
<td>Ours</td>
<td>80.06</td>
<td>72.33</td>
</tr>
<tr>
<td>SFLSDE/RandtoBest/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>81.54</td>
<td>72.23</td>
</tr>
<tr>
<td>SFLSDE/Rand/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>81.67</td>
<td>71.88</td>
</tr>
<tr>
<td>SSMA+SFLSDE/RandtoBest/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>81.64</td>
<td>74.95</td>
</tr>
</tbody>
</table>

\[
\frac{(2 \times \text{Reduction} \times \text{Accuracy})}{(\text{Reduction} + \text{Accuracy})}
\]

<table>
<thead>
<tr>
<th>Method</th>
<th>Ref.</th>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOGP</td>
<td>Ours</td>
<td>88.98</td>
<td>84.12</td>
</tr>
<tr>
<td>SFLSDE/RandtoBest/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>89.92</td>
<td>84.48</td>
</tr>
<tr>
<td>SFLSDE/Rand/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>89.99</td>
<td>84.25</td>
</tr>
<tr>
<td>SSMA+SFLSDE/RandtoBest/1/Bin</td>
<td>(Triguero et al., 2011)</td>
<td>90.02</td>
<td>86.15</td>
</tr>
</tbody>
</table>
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


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 \((O(N \cdot M))\).
- Alto costo computacional \((O(N^2))\).
- Sensibilidad a ejemplos mal clasificados (Ruido).
Reduciendo el costo de kNN

- Reducción de Datos
  - Reducción de Instancias
    - Selección de Instancias/Prototipos
    - Generación de Prototipos
  - Reducción de Características
    - Selección de Características
    - Generación/Extracción de Características
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)

Simultaneous generation of features and prototypes
Simultaneous generation of features and prototypes

- A different feature space for each class
Algorithm 1 NSGA-II algorithm Deb et al. (2002).

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

\{ $N_{pop}$ number of individuals (solutions); $g$ number of generations \}
\begin{align*}
f & = (f_1(P), f_2(P)) \text{ objectives} \\
1: & \text{ Initialize population } X_0 \\
2: & \text{ Evaluate objective functions } f = (f_1(P), f_2(P)), \forall P : P \in X_0 \\
3: & \text{ Identify fronts } F_{1,...,F} \text{ by sorting solutions according to their non-dominance level } \forall P : P \in X_0 \\
4: & \text{ while } i = 1 < g \text{ do} \\
5: & \text{ Create child population } Q_i \text{ from } X_i \text{ applying evolutionary operators.} \\
6: & \text{ Evaluate objective functions } f, \forall P : P \in Q_i \\
7: & \text{ Identify fronts } F_{1,...,F} \text{ by sorting solutions according to their non-dominance level } \forall P : P \in X_i \cup Q_i \\
8: & X_{i+1} = \emptyset; j = 1; \\
9: & \text{ while } |X_{i+1}| < N_{pop} \text{ do} \\
10: & X_{i+1} = X_{i+1} \cup F_j; j = j + 1; \\
11: & \text{ end while} \\
12: & \text{ Select the last individuals for } X_{i+1} \text{ from } F_j \text{ using crowding distance} \\
13: & \text{ end while}
\end{align*}
NSGA-II : (perhaps) the most used MOEA

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

Require: \( N_{\text{pop}}, f, g \)

- \( N_{\text{pop}} \) number of individuals (solutions); \( g \) number of generations
- \( f = (f_1(P), f_2(P)) \) objectives

1: Initialize population \( X_0 \)

2: Evaluate objective functions \( f = (f_1(P), f_2(P)), \forall P : P \in X_0 \)

3: Identify fronts \( F_{1,...,F} \) by sorting solutions according to their non-dominance level \( \forall P : P \in X_0 \)

4: while \( i = 1 < g \) do

5: Create child population \( Q_i \) from \( X_i \) applying evolutionary operators.

6: Evaluate objective functions \( f, \forall P : P \in Q_i \)

7: Identify fronts \( F_{1,...,F} \) by sorting solutions according to their non-dominance level \( \forall P : P \in X_i \cup Q_i \)

8: \( X_{i+1} = \emptyset; \ j = 1; \)

9: while \( |X_{i+1}| < N_{\text{pop}} \) do

10: \( X_{i+1} = X_{i+1} \cup F_j; \ j = j + 1; \)

11: end while

12: Select the last individuals for \( X_{i+1} \) from \( F_j \) using crowding distance

13: end while
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:

<table>
<thead>
<tr>
<th></th>
<th>Propuesta Media</th>
<th>Std</th>
<th>Sólo Prototipos Media</th>
<th>Std</th>
<th>Sólo Características Media</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exactitud Promedio</td>
<td>80.25</td>
<td>19.93</td>
<td>78.37</td>
<td>20.24</td>
<td>82.82</td>
<td>21.91</td>
</tr>
<tr>
<td>Reducción de Instancias Promedio</td>
<td>99.43</td>
<td>0.09</td>
<td>99.40</td>
<td>0.11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reducción de Características Promedio</td>
<td>41.45</td>
<td>12.76</td>
<td>-</td>
<td>-</td>
<td>36.78*</td>
<td>14.65</td>
</tr>
</tbody>
</table>
Simultaneous generation of features and prototypes

• Some results:

<table>
<thead>
<tr>
<th></th>
<th>Multi-Objetivo</th>
<th>Mono-Objetivo</th>
<th>EMOPG+FS</th>
<th>MOPG</th>
<th>1NN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Media</td>
<td>Std.</td>
<td>Media</td>
<td>Std.</td>
<td>Media</td>
</tr>
<tr>
<td>Exactitud</td>
<td>70.84</td>
<td>17.74</td>
<td>71.97</td>
<td>15.76</td>
<td>74.26*</td>
</tr>
<tr>
<td>Red. de Instancias</td>
<td>98.43</td>
<td>1.65</td>
<td>98.39*</td>
<td>1.37</td>
<td>97.32*</td>
</tr>
<tr>
<td>Red. de Características</td>
<td><strong>88.65</strong></td>
<td>7.52</td>
<td>42.62*</td>
<td>5.13</td>
<td>60.47*</td>
</tr>
</tbody>
</table>

* Diferencia estadísticamente significante.

<table>
<thead>
<tr>
<th></th>
<th>Multi-Objetivo</th>
<th>Mono-Objetivo</th>
<th>EMOPG+FS</th>
<th>MOPG</th>
<th>1NN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Media</td>
<td>Std.</td>
<td>Media</td>
<td>Std.</td>
<td>Media</td>
</tr>
<tr>
<td>Exactitud</td>
<td>76.19</td>
<td>21.25</td>
<td>80.25*</td>
<td>19.93</td>
<td><strong>81.82</strong></td>
</tr>
<tr>
<td>Red. de Instancias</td>
<td><strong>99.82</strong></td>
<td>0.16</td>
<td>99.43*</td>
<td>0.09</td>
<td>98.42*</td>
</tr>
<tr>
<td>Red. de Características</td>
<td><strong>89.74</strong></td>
<td>12.36</td>
<td>41.45*</td>
<td>12.76</td>
<td>49.09*</td>
</tr>
</tbody>
</table>
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?