Prototype and Feature Generation

Mauricio García Limón **Hugo Jair Escalante** Eduardo F. Morales
Instituto Nacional de Astrofísica Óptica y Electrónica
(INAOE)

International Autumn Meeting on Power, Electronics and Computing
2014 IEEE XVI ROPEC

November 5, 2014

Outline

Outline

Introduction

The Classification task consists in associating objects with predefined categories (e.g. Text classification). Among the variety of classification models proposed so far, nearest-neighbors (NN) methods are among the most popular ones.

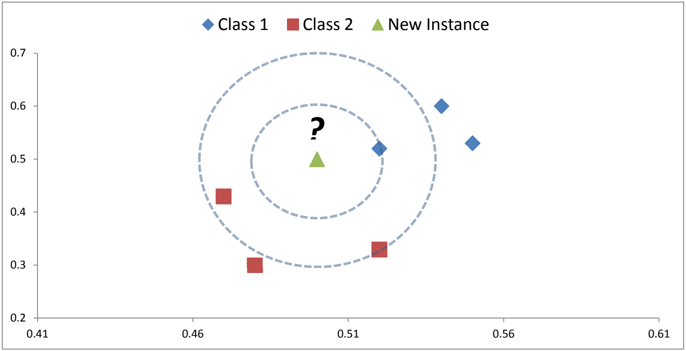

Figure: Example of the kNN method.

Introduction

Altough NN methods are highly effective, and easy to implement, they present some issues that hinder their application to certain types of problems:

- High computational time requirements.
- High storage requirements.
- Sensitivity to noisy (e.g., misclassified instances and outliers).

Introducción

There are different techniques for data reduction.

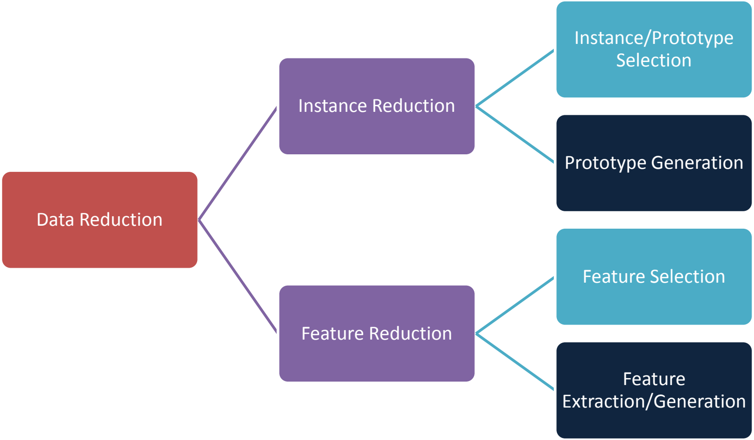

Figure: Techniques for data reduction.

Outline

Related Work

- There are a number of approaches for prototype generation (PG) such as incremental, decremental, mixed (e.g., GENN, PSCSA, PSO methods).
- Triguero et al. review and classify most of the existing PG methods up to 2012, additionally, a taxonomy and experimental comparison of these methods is also reported.
- On the other hand, for the feature extraction there are several approaches such as evolutionary algorithms.

Both tasks, prototype generation and feature extraction, are performed independently due to computational cost. However, there are methods that try to combine both techniques.

Outline

Multi-Objective Approach

- We present a extension of the **Simultaneous Generation of Prototypes and Features through Genetic Programming** (*SGPFGP*) method. The goal is to learn class-specific prototypes and features for pattern classification by NN throgh multi-objective optimization.
- The method is based on NSGA-II algorithm and SGPFGP method.
- We use genetic programming to the generation of prototypes and features by combining instances and features, respectively.
- Our proposal tries to find a good trade-off between accuracy, reduction of instances and features.
- The main difference of the multi-objective approach against the SGPFGP method is the multi-objective optimization.

Multi-Objective Approach

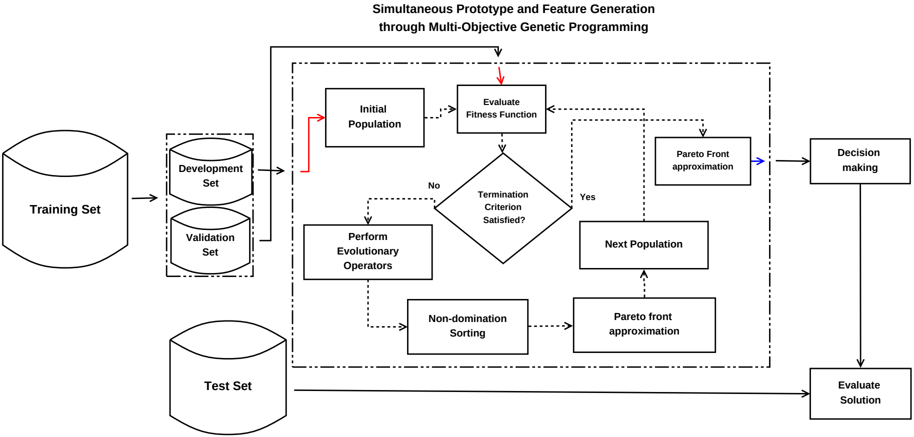

Figure: Flow chart of the multi-objective approach.

Multi-Objective Approach

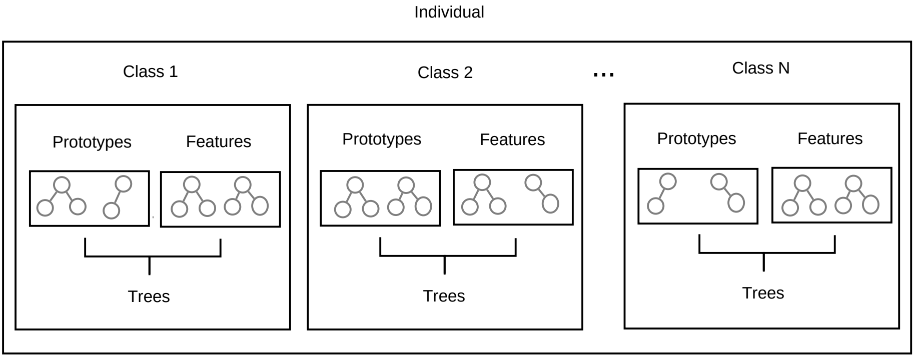

Figure: Representation of an individual: class-specific prototype-feature trees.

Multi-Objective Approach

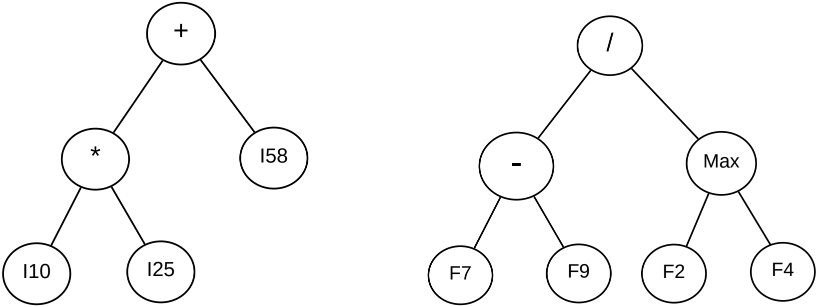

Figure: Prototype repersentation (left): instances are combined, feature representation (right): features are combined

- The function set for combining instances and features is the following set of operators: {+,-,*,/(Protected), Min, Max}.
- The Fitness function was the classification performance obtained by a 1NN when classify in all instances in a validation set.

Outline

Experimental Framework

Table: Summary of features for the 59 data sets of the considered benchmark.

	Small	Large
# data sets	40	19
# Classes	[2, 15]	[2, 28]
# Instances	[101, 1,728]	[2,201, 19,020]
# Attributes	[3, 90]	[2, 85]

- Number of generations is 100.
- Population size is set to 200.
- Maximum tree depth is fixed to 3.
- Statistically test: Wilcoxon signed-rank, with confidence level of 95 %.

Outline

Experimental Results

Table: Average classification accuracy for SGPFGP and reference methods.

	Multi-Objective	SGPFGP	1NN
Small	$70{,}84 \pm 17{,}24$	$71{,}97 \pm 15{,}76$	$73{,}48 \pm 16{,}64$
Large	$76{,}19 \pm 21{,}25$	$80{,}25 \star \pm 19{,}93$	$80{,}60 \pm 22{,}24$

$\star$ statistically significant difference.

Table: Average instance reduction rates obtained by the considered methods.

	Multi-Objective	SGPFGP
Small	$98{,}43 \pm 1{,}65$	$98{,}39 \star \pm 1{,}37$
Large	$99{,}82 \pm 0{,}16$	$99{,}43 \star \pm 0{,}09$

$\star$ statistically significant difference.

Experimental Results

Table: Average feature reduction rates obtained by the considered methods.

	Multi-Objective	SGPFGP
Small	$88{,}65 \pm 7{,}52$	$42{,}62 \star \pm 5{,}13$
Large	$89{,}74 \pm 12{,}36$	$41{,}45 \star \pm 12{,}76$

$\star$ statistically significant difference.

Experimental Results

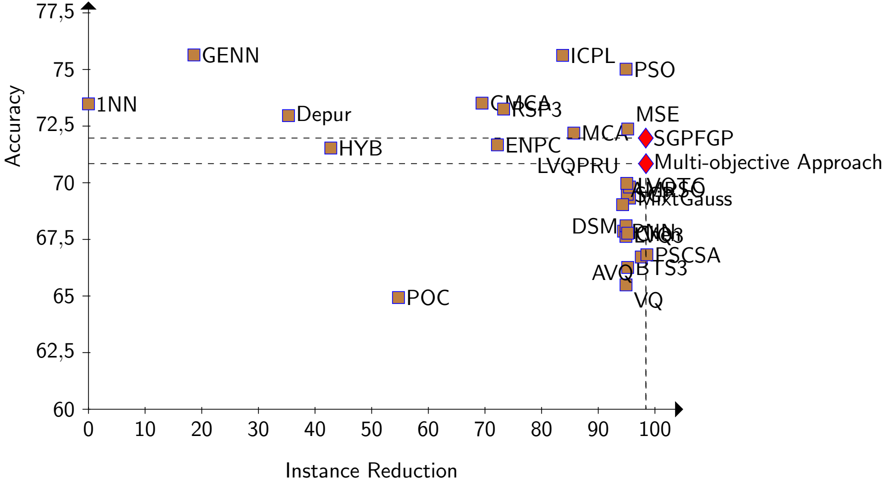

Figure: Average reduction ($x-$axis) vs. accuracy ($y-$axis) for small data sets.

Experimental Results

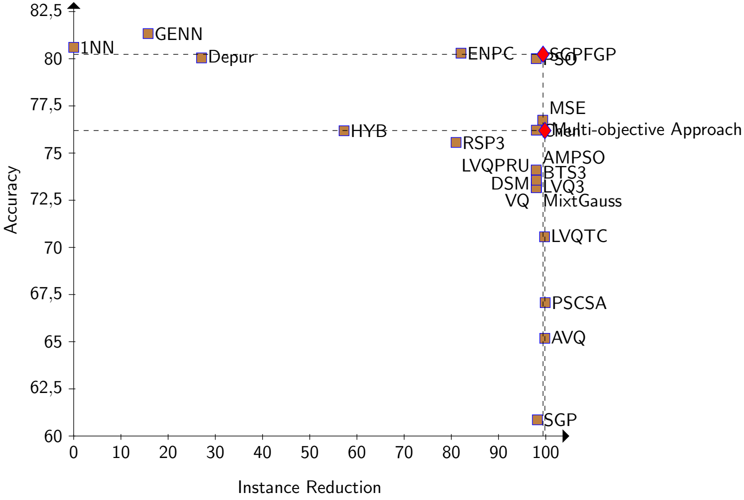

Figure: Average reduction ($x-$axis) vs. accuracy ($y-$axis) for large data sets.

Experimental Results

The Chess data set is associated to a binary classification task (2-classes) that contains 3,196 instances described by 36 attributes.

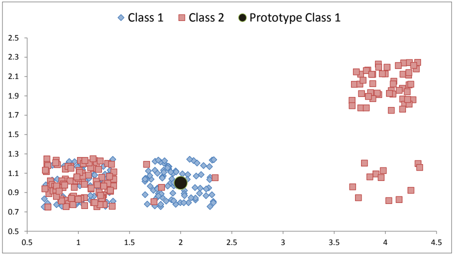

Figure: Feature space and prototype of class 1.

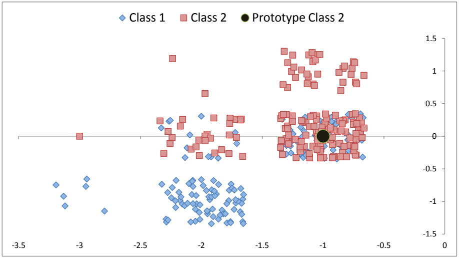

Figure: Feature space and prototype of class 2.

Experimental Results

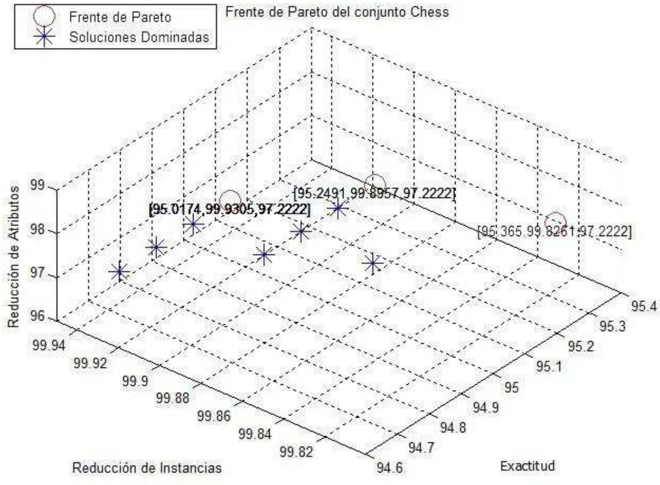

Figure: Pareto front of Chess data set (3-D)

Experimental Results

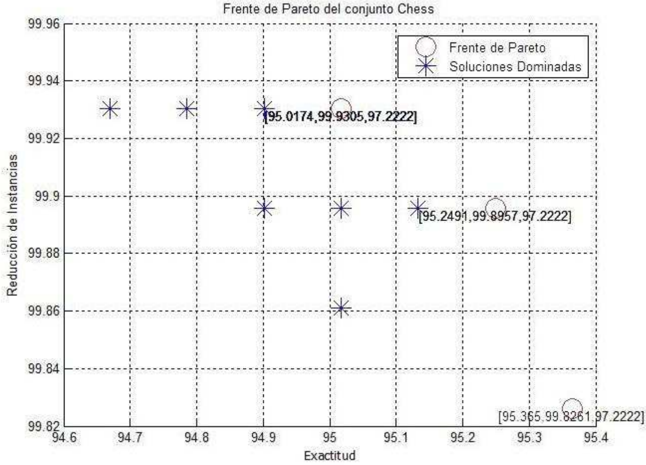

Figure: Pareto front of Chess data set (2-D) (Accuracy, Instance reduction)

Outline

Conclusions

- The multi-objective is able to reduce data dimensionality and the number of training instances without compromising the classification performance.
- Instance-reduction performance of multi-objective approach is very close to the optimal reduction rates (i.e., one instance per class).
- The multi-objective reduces about **90 %** data sets dimensionality. Building new features that are more informative for designing NN classifiers.
- The multi-objective method **outperforms statistically** in terms of instance and feature reduction rate to SGPFGP method.

Future Work

- Explore the use of distances for different types of data such as Value Difference Metric or Heterogeneous Value Difference Metric.
- Use other metrics to evaluate the classification performance such as the area under the curve.
- Mixing the different solutions obtained for each Pareto front, in order to exploit each of the solutions generated.

Questions

Questions ?

Thank you for your attention!