Pattern Classification Using Optimum-Path Forest

João Paulo Papa

papa@fc.unesp.br

ICPT First School on Data Science and Machine Learning

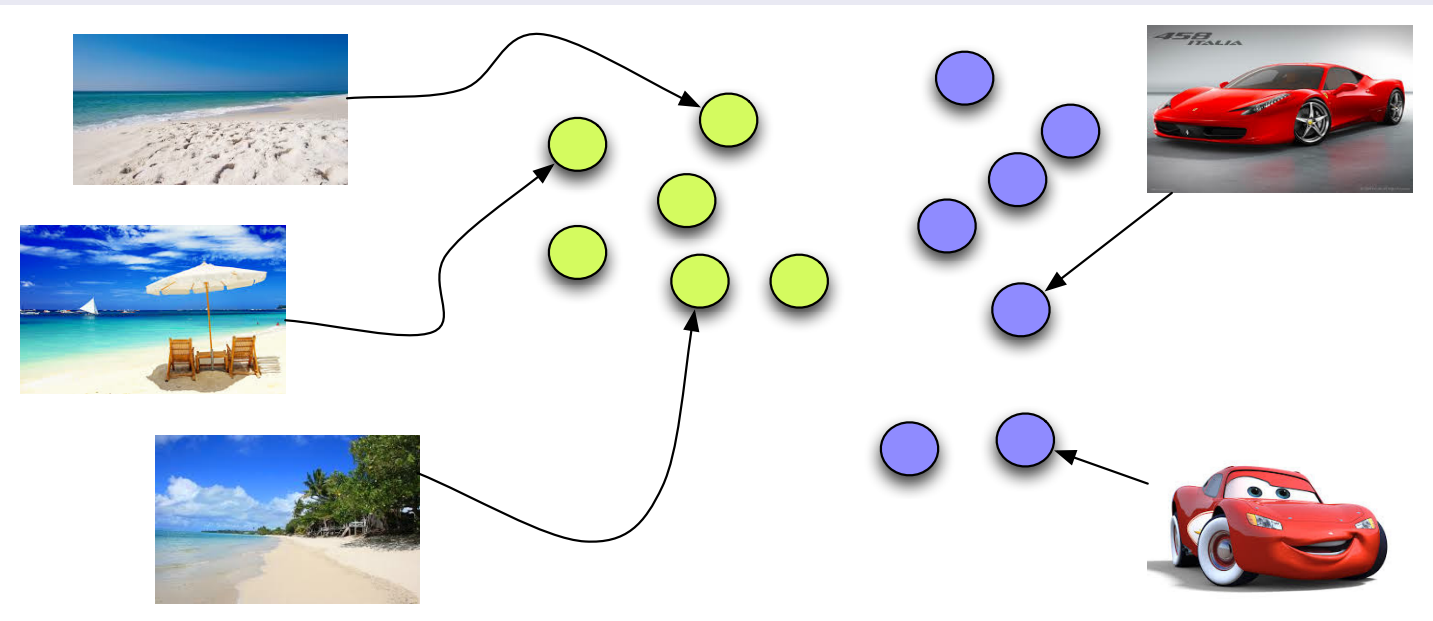
December 15, 2019

Talk Outline

Introduction

Pattern classification

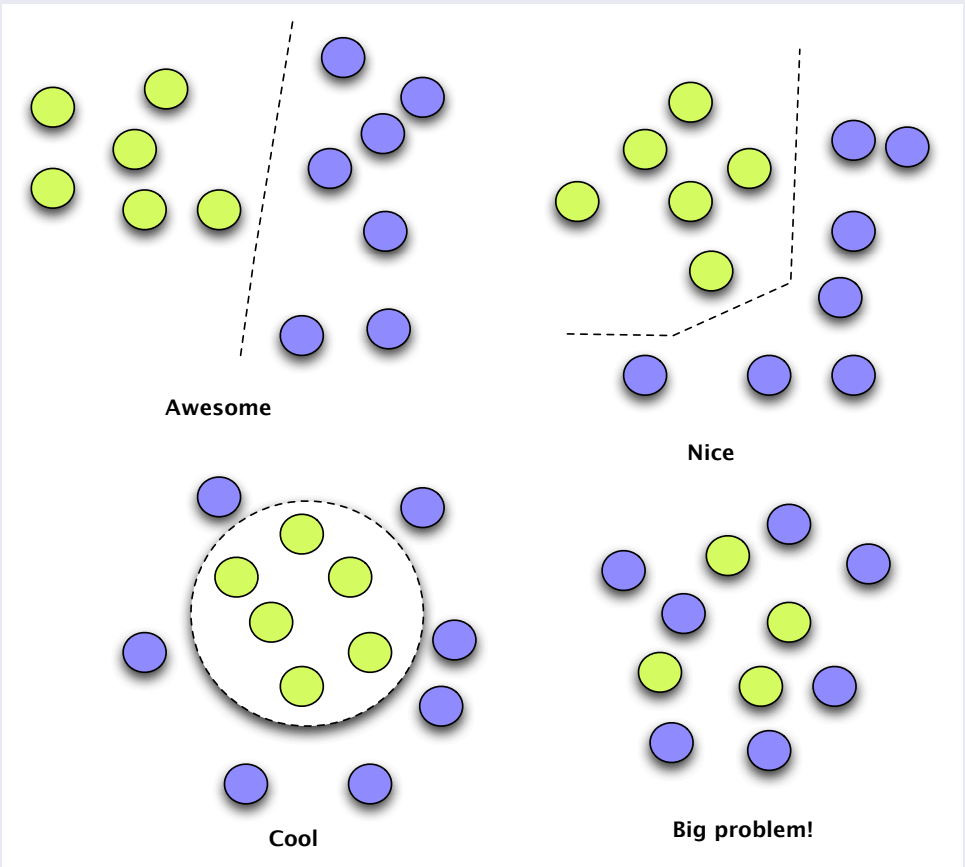
Pattern recognition techniques aim to find decision boundaries for datasets induced in a feature space



Introduction

Pattern classification

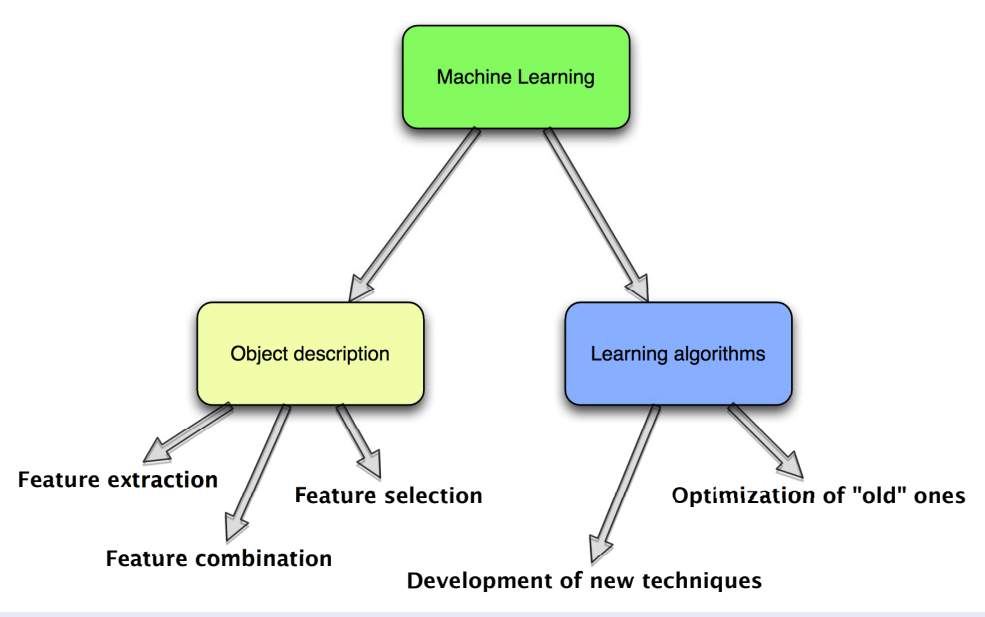
We have been facing ...



Introduction

Pattern classification

Working areas:



Introduction

Pattern classification

We can highlight some of the most widely used techniques, which can be divided in (i) supervised, (ii) semi-supervised and (iii) unsupervised (clustering):

- Artificial Neural Networks.
- Support Vector Machines
- Self-Organizing Maps.
- Bayesian classifiers.

Introduction

Pattern classification

We can highlight some of the most widely used techniques, which can be divided in (i) supervised, (ii) semi-supervised and (iii) unsupervised (clustering):

- Artificial Neural Networks.
- Support Vector Machines
- Self-Organizing Maps.
- Bayesian classifiers.

Introduction

Pattern classification

We can highlight some of the most widely used techniques, which can be divided in (i) supervised, (ii) semi-supervised and (iii) unsupervised (clustering):

- Artificial Neural Networks.
- Support Vector Machines
- Self-Organizing Maps.
- Bayesian classifiers.

Introduction

Pattern classification

We can highlight some of the most widely used techniques, which can be divided in (i) supervised, (ii) semi-supervised and (iii) unsupervised (clustering):

- Artificial Neural Networks.
- Support Vector Machines
- Self-Organizing Maps.
- Bayesian classifiers.

Optimum-Path Forest - OPF

Pattern classification using graph theory

Actually, OPF is a framework to the design of graph-based classifiers. The idea is to perform a competition process among some key nodes (*prototypes*) according to some rules:

- Adjacency relation.
- Methodology to estimate prototypes.
- Path-cost function (function to be minimized/maximized).

Thus, you can choose your itens to build your own OPF classifier.

Optimum-Path Forest - OPF

Pattern classification using graph theory

Nowadays, we have three versions:

- Supervised OPF:
  - Complete graph [a b]
  - k-nn graph [c]
- Unsupervised OPF [d]

---

[a] J.P.Papa, A.X. Falcão, C.T.N. Suzuki, *Supervised pattern classification based on optimum-path forest*, "International Journal on Imaging Systems and Technology", 19(2), 120-131, 2009

[b] J.P.Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, *Efficient supervised optimum-path forest classification for large datasets*, "Pattern Recognition, 45(1), 512-520, 2012

[c] J.P.Papa, S.E. Fernandes A.X. Falcão, *Optimum-Path Forest based on k-connectivity: Theory and Applications*, "Pattern Recognition Letters", 117-126, 2017

[d] L.M.Rocha, F.A.M. Cappabianco, A.X. Falcão, *Data clustering as an optimum-path forest problem with applications in image analysis*, "International Journal on Imaging Systems and Technology", 19(2), 50-68, 2009

Optimum-Path Forest - OPF

Pattern classification using graph theory

Nowadays, we have three versions:

- Supervised OPF:
  - Complete graph [a b]
  - k-nn graph [c]
- Unsupervised OPF [d]

---

[a] J.P.Papa, A.X. Falcão, C.T.N. Suzuki, *Supervised pattern classification based on optimum-path forest*, "International Journal on Imaging Systems and Technology", 19(2), 120-131, 2009

[b] J.P.Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, *Efficient supervised optimum-path forest classification for large datasets*, "Pattern Recognition, 45(1), 512-520, 2012

[c] J.P.Papa, S.E. Fernandes A.X. Falcão, *Optimum-Path Forest based on k-connectivity: Theory and Applications*, "Pattern Recognition Letters", 117-126, 2017

[d] L.M.Rocha, F.A.M. Cappabianco, A.X. Falcão, *Data clustering as an optimum-path forest problem with applications in image analysis*, "International Journal on Imaging Systems and Technology", 19(2), 50-68, 2009

Optimum-Path Forest - OPF

Pattern classification using graph theory

Nowadays, we have three versions:

- Supervised OPF:
  - Complete graph [a][b]
  - k-nn graph [c]
- Unsupervised OPF [d]

---

[a] J.P.Papa, A.X. Falcão, C.T.N. Suzuki, *Supervised pattern classification based on optimum-path forest*, "International Journal on Imaging Systems and Technology", 19(2), 120-131, 2009

[b] J.P.Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, *Efficient supervised optimum-path forest classification for large datasets*, "Pattern Recognition, 45(1), 512-520, 2012

[c] J.P.Papa, S.E. Fernandes A.X. Falcão, *Optimum-Path Forest based on k-connectivity: Theory and Applications*, "Pattern Recognition Letters", 117-126, 2017

[d] L.M.Rocha, F.A.M. Cappabianco, A.X. Falcão, *Data clustering as an optimum-path forest problem with applications in image analysis*, "International Journal on Imaging Systems and Technology", 19(2), 50-68, 2009

Optimum-Path Forest - OPF

Pattern classification using graph theory

The main idea is to model the problem as a graph partition task:

- dataset $\rightarrow$ samples (feature vectors) $\rightarrow$ nodes.
- nodes are connected through edges.
- both edges and nodes can be weighted.
  - edges are weighted by a distan/cematching function.
  - nodes are weighted by a density function.

Optimum-Path Forest - OPF

OPF with complete graph

We need to configure three itens to design an OPF-based classifier:

- Adjacency relation $\rightarrow$ Complete graph
- Methodology to estimate prototypes ("conquerors") $\rightarrow$ MST.
- Path-cost function $\rightarrow$ f_{max}.

Optimum-Path Forest - OPF

OPF with complete graph

We need to configure three itens to design an OPF-based classifier:

- Adjacency relation $\rightarrow$ Complete graph
- Methodology to estimate prototypes ("conquerors") $\rightarrow$ MST.
- Path-cost function $\rightarrow$ f_{max}.

Optimum-Path Forest - OPF

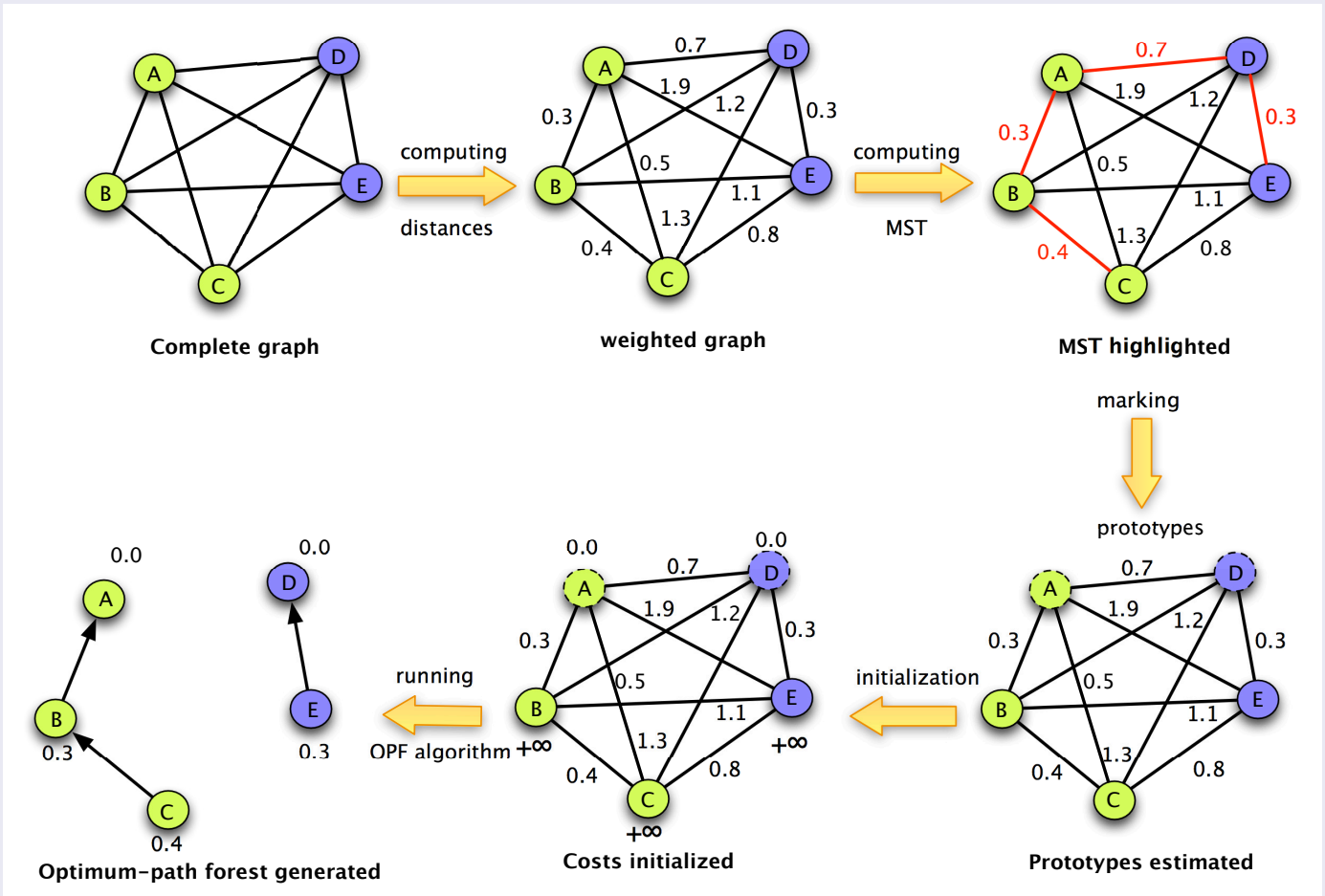
OPF with complete graph

We need to configure three itens to design an OPF-based classifier:

- Adjacency relation $\rightarrow$ Complete graph
- Methodology to estimate prototypes ("conquerors") $\rightarrow$ MST.
- Path-cost function $\rightarrow$ f_{max}.

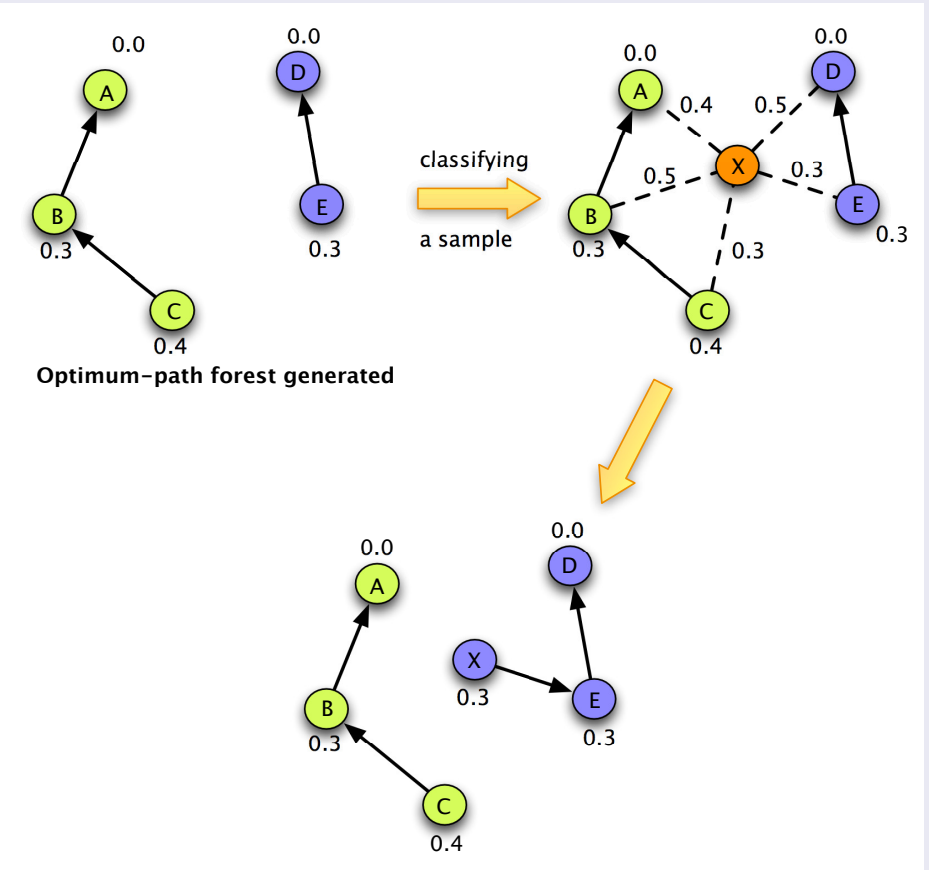
Optimum-Path Forest - OPF

OPF with complete graph - Training step



Optimum-Path Forest - OPF

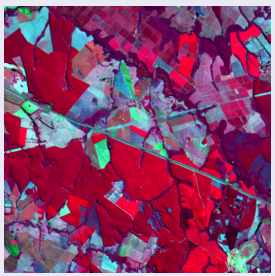
OPF with complete graph - Testing step



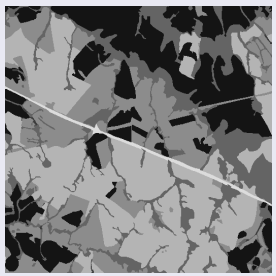
Talk Outline

Applications

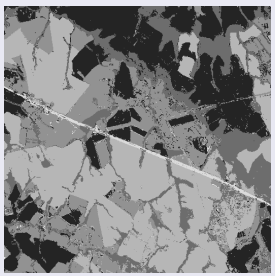
Land use classification in satellite images



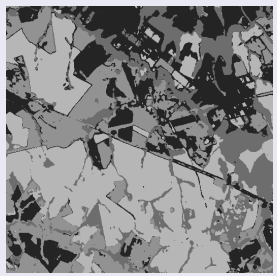
(a) Landsat



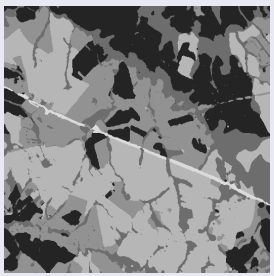
(b) Ground truth



(c) OPF[a]



(d) SVM



(e) OPF-MRF[b]

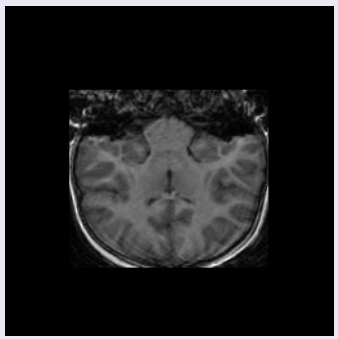
---

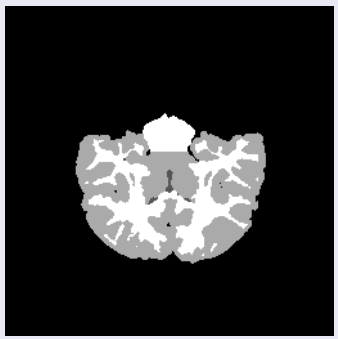
[a] R.J.Pisani, R.Y.M. Nakamura, P.S. Riedel, C.R.L. Zimback, A.X. Falcão, J.P. Papa, *Toward Satellite-Based Land Cover Classification Through Optimum-Path Forest*, "IEEE Transactions on Geoscience and Remote Sensing", 52 (10), 6075-6085, 2014.

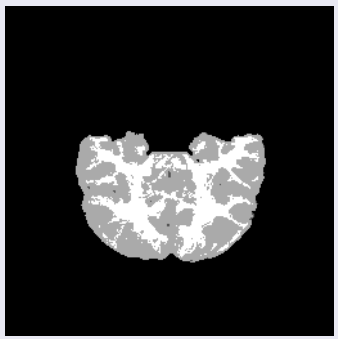
[b] D. Osaku, R.Y. M. Nakamura, A.L.M. Levada, R.J. Pisani, A.X. Falcão, J.P.Papa, *Improving Land Cover Classification Through Contextual-based Optimum-Path Forest*, "Information Sciences", 324 (10), 60-87, 2016.

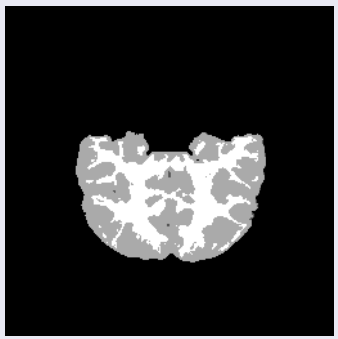
Applications

GM/WM classification


(a) MR image


(b) Ground truth


(c) OPF[a]


(e) OPF-MRF[b]

---

[a] R.Y.M. Nakamura, D. Osaku, A.L.M. Levada, F.A.M. Cappabianco, A.X. Falcão, J.P. Papa:, *OPF-MRF: Optimum-Path Forest and Markov Random Fields for Contextual-Based Image Classification*, "15th International Conference on Computer Analysis of Images and Patterns", 233–240, 2013.

[b] D. Osaku, R.Y.M. Nakamura, J.P. Papa, A.L.M. Levada, F.A.M. Cappabianco, A.X. Falcão, *Optimizing Contextual-Based Optimum-Forest Classification through Swarm Intelligence*, "15th International Conference on Advanced Concepts for Intelligent Vision Systems ", 203–214, 2013.

Applications

Ongoing works

- OPF-ML (Multilabel classification)
- cudaOPF
- OPF-MIL
- Active OPF
- LibOPF $\rightarrow$ check it out at `http://www.ic.unicamp.br/~afalcao/libopf`

Applications

Ongoing works

- PyOPF
- MatOPF
- OPFWeb
- R-OPF