Weak supervised and unconventional classification problems

Ronaldo Cristiano Prati
ronaldo.prati@ufabc.edu.br

Classification

- Classification is a common task in machine learning and data mining
- It has a broad range of applications, and literally dozens of different algorithm families

Standard Classification

- Assumes that the learning algorithm has access to a representative, unbiased sample
  - Instances are correctly labeled by an "oracle"
  - Represented by an unique vector of characteristics
  - A single output (class)

Standard Classification

- Many real world problems do not fall in this description
- These problems have motivated the development of new classification paradigms beyond standard classification

Weak supervision

- In many application, manually assigning classe to instances is unfeasible
  - To costly
  - Lack of domain expert
  - Lack of information to properly determine the correct class

Weak supervision

- For many tasks, we can leverage various heuristics and existing datasets as weak supervision.
- We can use these “weak” class labels as a proxy to true class values

Weak supervision

Examples:

- In sentiment analysis, we can use hashtags (e.g. #irony) as an indicator to the class value
- Unreliable non-expert annotators (e.g. crowdsourcing)
- In medical applications, a simpler and cheaper exam can be used instead of more costly procedure

Weak supervision

- This process of weak supervision introduces noisy class labels
- Such noisy labels may produce negative impacts in the learning process

Noisy class labels

- Noisy class labels have been addressed in several ways in the literature:
  - data level approaches (“sanitize the data”)
  - algorithm level approaches (“improve resilience to noise”)

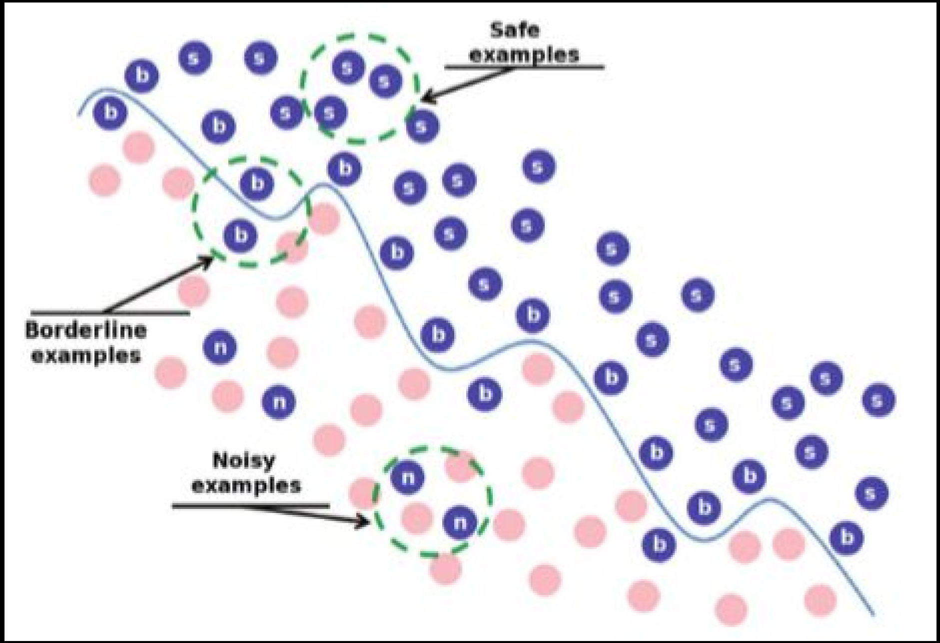
Data level approaches

- Data level approaches are data-driven
- Look at data properties aiming to identify noisy instances
  - Neighborhood
  - Instance “hardness”
  - among others

Data level approaches

- Potential noisy instances identified by these approach may have different causes:
  - Uncertainty due to closeness to class boundaries
  - Sparsity
  - Noise

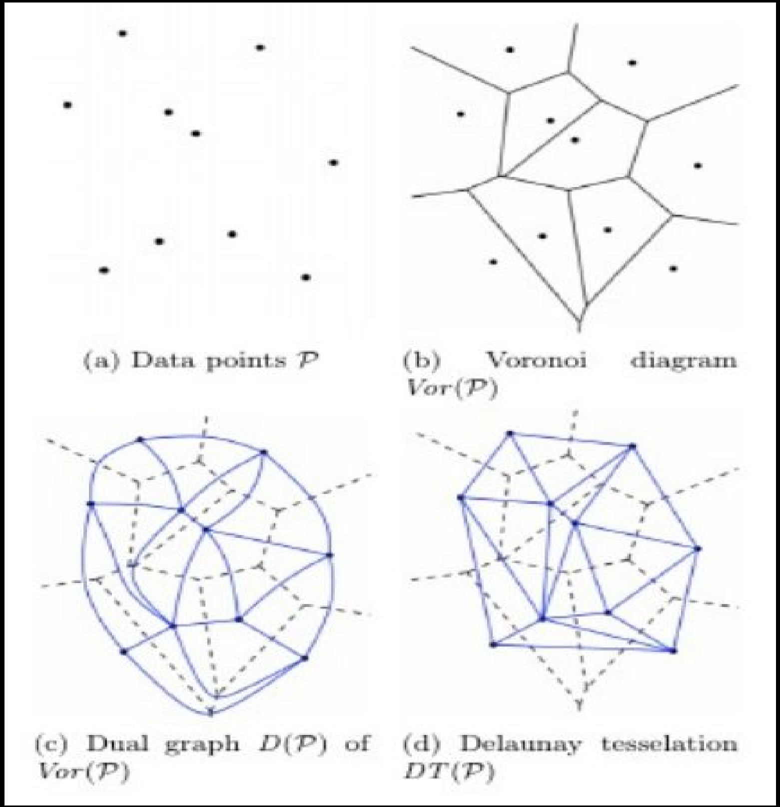
Data level approaches



Our approach

- We developed a meta-heuristic approach based on graph embedding for handling possible noise instances
- The approach is based on Delaunay tessellation, a well known algorithm in computer vision

Delaunay tessellation



DTFE

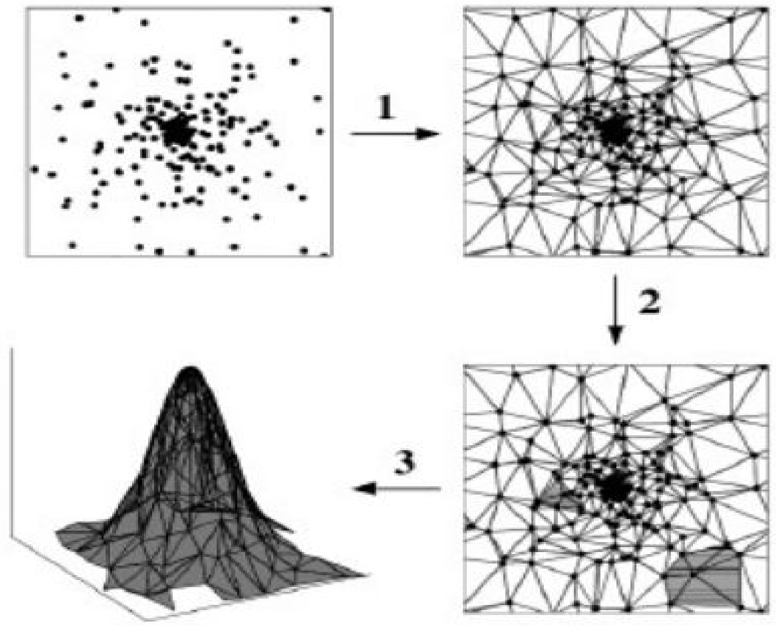


Figure 3: Overview of Delaunay Tesselation field estimator (Public Domain, PublicDomain,https://en.wikipedia.org/w/index.php?curid=13860944

Our approach

- By using the graph embedding, we can have both local and global information
  - Local neighborhood
  - Density
- The hypothesis is both local and global information are better baselines for noise clearance.

Meta-heuristic main idea

- Use the local neighborhood to analyze the “local” class distribution of instances
- Use the instance density to analyze the “quality” of instance information

Some preliminary results

alg	filter	0	5	10	15	20	25	30	35	40	45	50
5NN	ENN	3.132	6.362	6.374	4.955	4.306	2.791	2.247	1.442	0.975	1.527	1.136
J48	ENN	3.231	2.928	1.907	1.067	0.451	-0.590	-1.768	-2.395	-2.767	-2.273	-2.113
SVM	ENN	3.032	1.976	1.499	0.916	0.338	0.137	-0.088	-0.468	-1.893	-2.464	-3.008
5NN	HARF	1.740	5.232	5.489	4.973	4.973	3.766	2.783	1.967	1.255	1.143	0.844
J48	HARF	2.678	0.579	0.027	-1.629	-2.079	-3.202	-3.259	-4.192	-4.683	-5.321	-5.156
SVM	HARF	2.022	0.757	0.249	-0.098	-0.512	-0.668	-1.233	-2.800	-4.895	-6.061	-5.215
5NN	IPF	6.106	8.187	7.976	7.463	7.174	6.976	6.485	4.869	3.515	5.352	6.688
J48	IPF	9.082	7.565	6.848	5.750	4.857	3.775	3.116	1.806	1.432	0.995	1.665
SVM	IPF	5.638	4.736	4.719	4.387	4.081	4.328	3.811	3.020	1.360	0.125	-0.908

Multi-instance

- In standard classification, we often assume that each object is represented by a single vector of characteristic
- This may not be adequate in some applications involving complex objects (e.g. images, texts, molecules, etc.)

Multi-instance

- An alternative approach is to represent the complex objects as a “Bag” of instances
- Class labels are associated to the bag, not individual instances

Multi-instance

Example: in text classification, each paragraph may be mapped to an instance. However, the entire text is associated to document class, and not all paragraphs may related to that class

Bag “sanitization”

- We develop a methodology aiming to identify “weak” instances within each bag
- The hypothesis is that removing these instances may lead to better classification performance

Our approach

- The main idea is to “break” the bags, associating the bag label to each instance
- Such labels are assumed to be “weak” labels that needs to be sanitized

Some preliminary results

Simple-MI			
	Sin filtro	ST-IPF	NN-IPF
0%	81,3605	80,5469	*78,8171
5%	80,3384	80,1162	78,3623
10%	78,3608	*79,8879	78,4198
15%	78,4201	79,2590	77,1075
20%	75,9204	*78,3870	74,8545

MI-Boost			
	Sin filtro	ST-IPF	NN-IPF
0%	77,9908	**81,7616	78,8491
5%	78,7612	**80,9387	79,3059
10%	77,8180	**80,6538	78,6762
15%	77,2428	**80,4784	78,6054
20%	77,3866	**80,1285	76,9014

Multi-label Learning (MLL)

- Another aspect not covered in standard classification is the possibility to assign multiple outputs to the same instance
- In case of multiple classes, we have multi-label classification

Multi-label

Examples:

- A movie may belong to the genre action and adventure at the same time
- A piece of news may refer to economy and politics at the same time

A fuzzy decision tree MLL

- Decision trees are a common approach in machine learning
- Fuzzy logic is a widely used approach for incorporating uncertainty learning

A fuzzy decision tree MLL

- The method is based on the Generalized Fuzzy Entropy
- We extend a recently proposed method for the MLL case
- It induces a single, overall model
- Leaves may contain partial label sets

An example

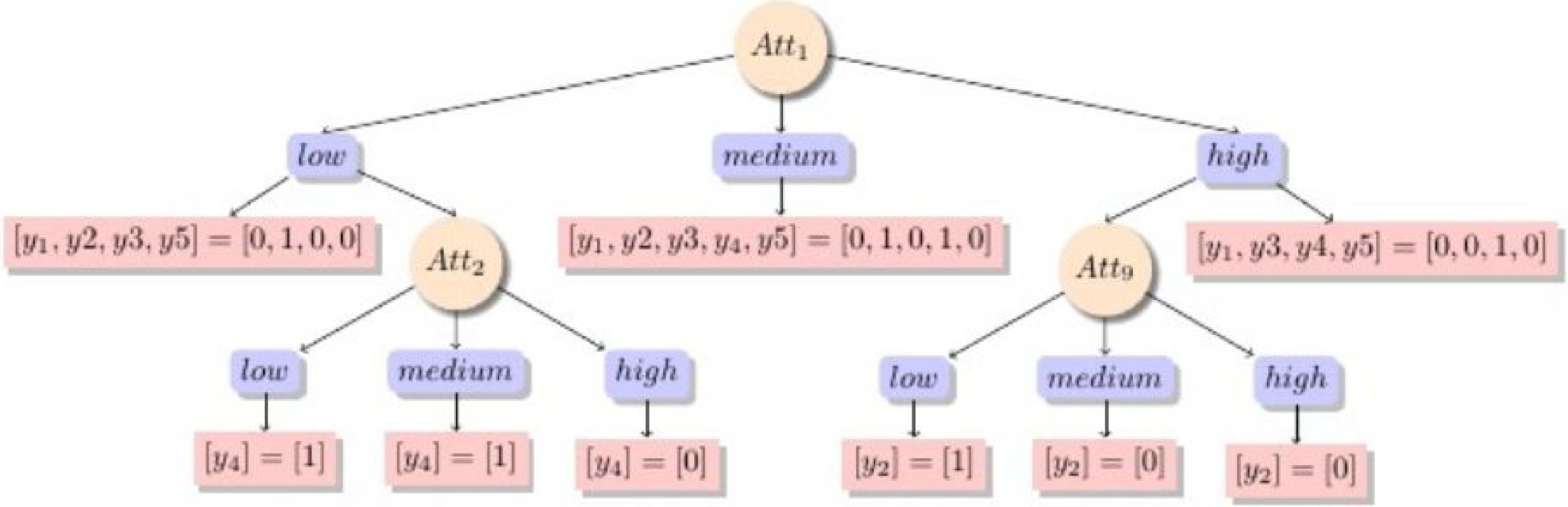


Fig. 1. An example of a MLC fuzzy decision tree induced by $\text{FuzzDT}_{\text{ML}}$

Some results

TABLE II
AVERAGE HAMMING LOSS ↓

dataset	BR(J48)	LPS(J48)	MLC45	FuzzDT_{ML}
cal500	0.1610	0.2014	0.1371	**0.1367**
emotions	0.2497	0.2734	**0.2421**	0.2490
scene	**0.1311**	0.1494	0.1341	0.1573
yeast	0.2467	0.2778	0.2250	**0.2244**
genbase	0.0484	0.0660	**0.0090**	0.0463
medical	**0.0104**	0.0131	0.0229	0.0216
slashdot	**0.0422**	0.0548	0.0497	0.0525
tmc2007	**0.0550**	0.0706	0.0721	0.0827
flags	**0.2577**	0.2861	0.2661	0.3076

TABLE III
AVERAGE RANKING LOSS ↓

dataset	BR(J48)	LPS(J48)	MLC45	FuzzDT_{ML}
cal500	0.2968	0.6550	**0.1807**	0.1811
emotions	0.2977	0.3330	0.2624	**0.2087**
scene	0.2362	0.2199	**0.1862**	0.2409
yeast	0.3130	0.4015	0.2033	**0.1952**
genbase	0.6040	0.6039	**0.0062**	0.3797
medical	**0.0663**	0.1364	0.1119	0.1122
slashdot	**0.1389**	0.2586	0.1930	0.1876
tmc2007	0.1099	0.3230	**0.0954**	0.1401
flags	0.2463	0.4910	**0.1998**	0.2517

Concluding remarks

- I've presented some recent (and ongoing work) related to weak supervised and and unconventional classification problems
-

Acknowledgements

- FAPESP (visitor research grant)
- These are joint work with collaborators from University of Granada