Machine Learning for the Management of Agricultural Soil Data

Manisha Sanjay Sirsat
Manuel Fernández Delgado

Centro Singular de Investigación en Tecnoloxías da Información
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

citius.usc.es

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL

CiTiUS Centro Singular de Investigación en Tecnoloxías da Información

Index

1. Introduction

- Agriculture in India: conditioned by poor soil fertility, depends on the nutrient levels.
- Soil physical, chemical and biological properties are useful to evaluate its fertility, to design a cultivation plan and to predict the crop productivity.
- Machine Learning (ML), offer new possibilities in the field of agriculture and may help in data evaluation for decision making.
- It is important to develop an effective nutrient management by means of an adequate soil analysis and a proper application of fertilizers.
- The relevance of a research effort to classify soil parameters such as the fertility indices for several nutrients (OC, P2O5, Mn and Fe, among others), soil pH, soil type, preferred crop and levels of several nutrients (N2O, P2O5 and K2O).

2. Soil classification problems

- Village-wise soil fertility indices of OC, P2O5, Mn and Fe: five problems.
- Soil nutrients N2O, P2O5 and K2O: for the recommendation of suitable amounts of fertilizers: three problems.
- Soil pH: The classification of pH into levels is useful to decide suitable crops and pesticides, and to evaluate microbial activity, nutrient levels and soil corrosivity.
- Crop type: it allows to predict which crop is suitable for the next stage of the cropping cycle.
- Soil type: useful to select the best soil for a particular crop.

The 10 classification problems use as inputs chemical measurements from the soil.

3. Data set: inputs, patterns and classes

- EC- Electrical Conductivity
- OC- Organic Carbon
- N2O- Nitrous oxide
- P2O5- Phosphorus pentoxide
- K2O- Potassium oxide
- SO4- Sulfate
- Cu- Copper
- Fe- Iron
- Mn- Manganese
- Zn- Zinc
- B- Boron
- CaCO3-Calcium Carbonate
- pH- potential of Hydrogen

Problems	#Inputs	#Patterns	#classes
OC-F	11	372	2
P2O5-F		372	3
Mn-F		367	2
Fe-F		372	2
N2O P2O5 K2O	10	372	3
pH	9	1137	4
crop	5	2878	4
Soil type	5	1692	2

CiTIUS

4. Classifiers

- Wide and diverse collection of 20 classifiers:
  - Bagging (3)
  - Boosting (1)
  - Decision trees (4)
  - Nearest neighbors (1)
  - Neural networks (5)
  - Random Forests (3)
  - Rule-based (2)
  - Support vector machine (1)
- Performance measures: Cohen kappa, Sensitivity (SE), and Positive Predictivity (PP).
- Experimental methodology: 4-fold validation.

CiTIUS

5. Implementations

- Classifiers implemented in
  - Weka: Version 3.6.11
  - Matlab: Version 7.14.0.739 (R2012a)
  - C++: gcc (Ubuntu 4.9.2-10ubuntu13) 4.9.2
  - R: version 3.1.2 (2014-10-31)
- Hardware features:

  Processor: Intel® Core™ i5-2500 CPU @ 3.30GHz × 4, Ubuntu 15.04

  OS type: 64-bit

6. Execution Time

- Execution time (in s.) of some classifiers for each problem:

	OC-F	P2O5-F	Mn-F	Fe-F	N2O	P2O5	K2O	pH	Crop	Soil	Mean	SD
ab_r	88.6	92.112	93.6	91.6	95.6	96.9	96.5	90.3	92.0	83.5	**92.1**	3.834
bg_r	5.25	5.228	5.08	5.12	5.448	5.648	5.396	8.644	4.264	8.872	5.895	1.473
elm_m	2.664	2.513	2.626	2.612	2.883	2.727	2.671	3.393	8.404	5.758	3.625	1.841
gelm_m	0.003	0.003	0.004	0.004	0.004	0.004	0.003	0.016	0.079	0.0233	*0.0147*	0.0225
knn_r	0.348	0.368	0.364	0.356	0.36	0.368	0.364	0.624	1.24	0.76	*0.5152*	0.276
mlp_m	17.397	18.130	14.947	15.749	17.005	15.770	15.725	15.066	39.844	25.189	**19.482**	7.357
pnn_m	3.086	3.025	3.035	3.049	3.847	3.111	3.086	4.492	8.0487	4.718	3.950	1.496
rf_r	0.74	0.792	0.8	0.748	0.976	0.984	0.968	1.756	3.316	1.46	1.254	0.756
rpt_r	0.292	0.292	0.232	0.268	0.252	0.268	0.276	0.336	0.384	0.264	0.286	0.0418
Mean	13.159	13.607	13.410	13.284	14.048	13.974	13.890	13.844	17.509	14.503		
SD	28.827	29.980	30.432	29.798	31.058	31.482	31.373	29.071	30.534	27.038		

CiTIUS

7. Results

- The kappa values achieved by each classfier for each classification problems:

Classifier	OC-F	P_2O_5-F	Mn-F	Fe-F	N_2O	P_2O_5	K_2O	pH	Crop	Soil
ab_r	88.50	**85.54**	59.33	67.45	25.40	**35.08**	25.81	44.71	85.27	96.65
bg_r	73.84	74.95	52.72	53.58	28.63	27.42	28.63	26.76	82.02	93.47
bgf_r	85.18	78.77	54.95	64.21	26.61	29.03	29.44	40.06	77.79	95.47
bgp_w	82.42	74.97	54.95	64.09	24.60	24.60	20.97	42.13	86.37	95.49
dj48_w	83.03	80.45	49.44	57.18	31.85	28.63	24.60	42.49	85.15	95.37
dtnb_w	81.33	66.78	47.79	44.69	22.18	22.58	5.24	35.93	85.85	**97.82**
elm_m	75.45	68.99	48.41	49.90	23.39	25.81	12.10	34.61	82.25	91.76
gelm_m	82.41	77.23	57.14	57.85	30.65	30.24	20.16	42.91	85.17	96.20
j48_w	69.25	70.76	37.92	54.42	20.16	21.37	14.92	40.78	82.34	94.76
jrp_w	71.28	70.34	45.02	52.00	19.35	22.18	12.90	40.44	83.12	94.94
knn_r	74.14	72.56	52.77	54.09	25.81	24.19	15.73	41.55	84.32	94.46
mlp_m	56.95	30.25	38.77	41.46	24.60	17.74	12.10	13.73	47.62	86.18
pnn_m	77.43	73.33	55.56	52.82	23.79	22.98	18.15	40.38	84.52	94.90
rbf_m	48.82	40.04	34.40	43.32	2.82	14.92	2.82	13.30	82.99	92.03
rf_r	87.35	83.01	58.78	**69.35**	**33.06**	33.06	**31.85**	**47.32**	**88.13**	96.37
rf_w	**90.65**	79.58	64.27	65.17	30.24	32.26	26.61	46.85	87.64	96.80
rpt_r	67.05	63.81	46.65	36.23	21.37	22.58	21.37	38.92	80.01	92.89
rt_r	69.15	68.93	51.14	36.03	15.32	21.37	19.35	33.49	78.75	93.39
rtf_w	87.39	75.83	57.70	59.01	28.23	30.65	27.42	46.77	86.34	96.80
svm_c	80.2	82.3	**64.8**	60.1	29.0	30.2	18.5	43.0	86.1	95.8

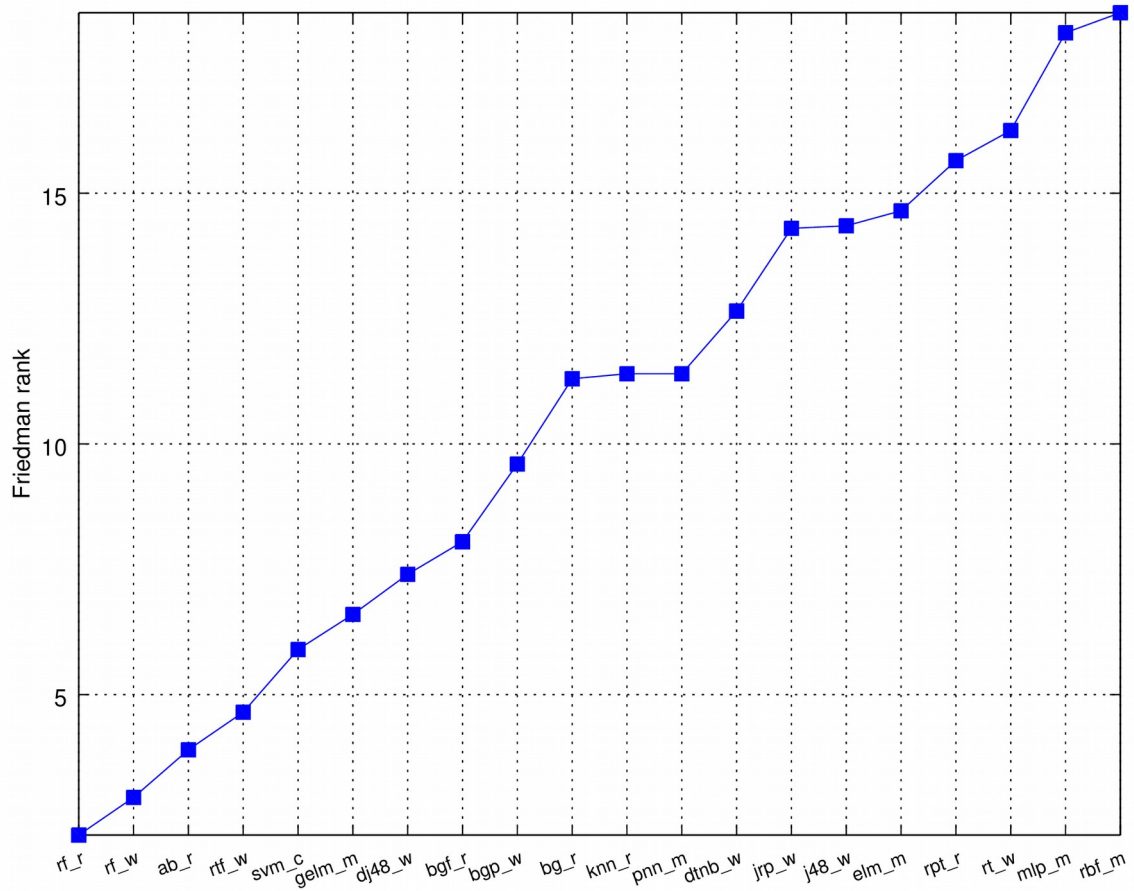
8. Classifier Rank

- Ranking and position of classifier according to Friendman rank test:
  - Classifiers are ordered by their Friendman rank test : rf_r is the best
- Classifiers are ordered by decreasing *p*-value of a wilcoxon signed rank test.

Friedman rank test						Wilcoxon signed rank test					
Pos.	Clasif.	Rank	Pos.	Clasif.	Rank	Pos.	Clasif	*p*-value	Pos.	Clasif	*p*-value
1	rf_r	2.200	11	knn_r	11.400	1	—	—	11	dtnb_w	0.384494
2	rf_w	2.950	12	pnn_m	11.400	2	ab_r	0.969839	12	jrp_w	0.384494
3	ab_r	3.900	13	dtnb_w	12.650	3	rf_w	0.850051	13	elm_m	0.344523
4	rtf_w	4.650	14	jrp_w	14.300	4	rtf_w	0.623046	14	j48_w	0.307308
5	svm_c	5.900	15	j48_w	14.350	5	svm_c	0.520366	15	elm_m	0.34055
6	gelm_m	6.600	16	elm_m	14.650	6	dj48_w	0.495968	16	rt_w	0.272856
7	dj48_w	7.400	17	rpt_r	15.650	7	bgf_r	0.472509	17	bg_r	0.272675
8	bgf_r	8.050	18	rt_w	16.250	8	bgp_w	0.472342	18	rpt_r	0.240968
9	bgp_w	9.600	19	mlp_m	18.200	9	knn_r	0.427181	19	rbf_m	0.10385
10	bg_r	11.300	20	rbf_m	18.600	10	pnn_m	0.427181	20	mlp_m	**0.03114**

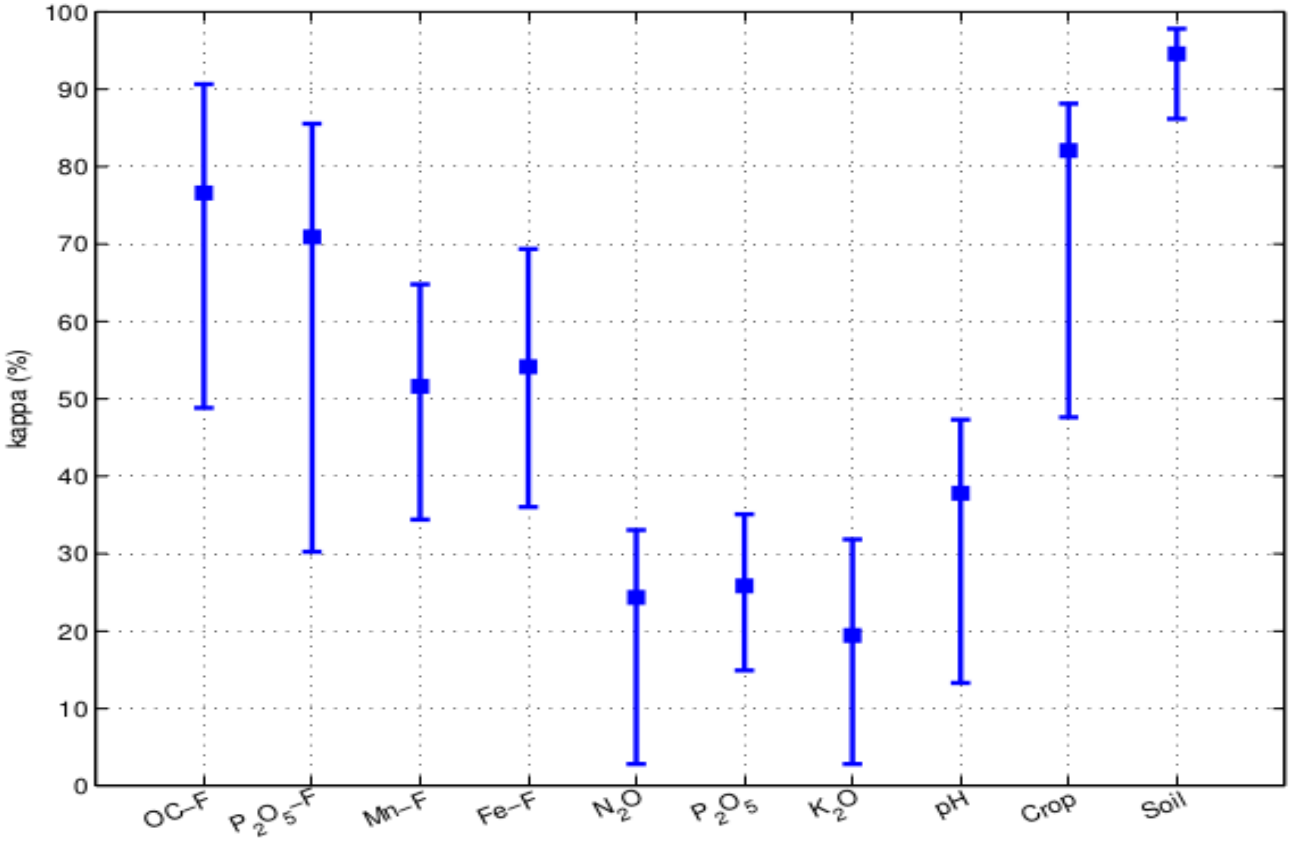
8. Classifier Rank

- Friedman rank test:



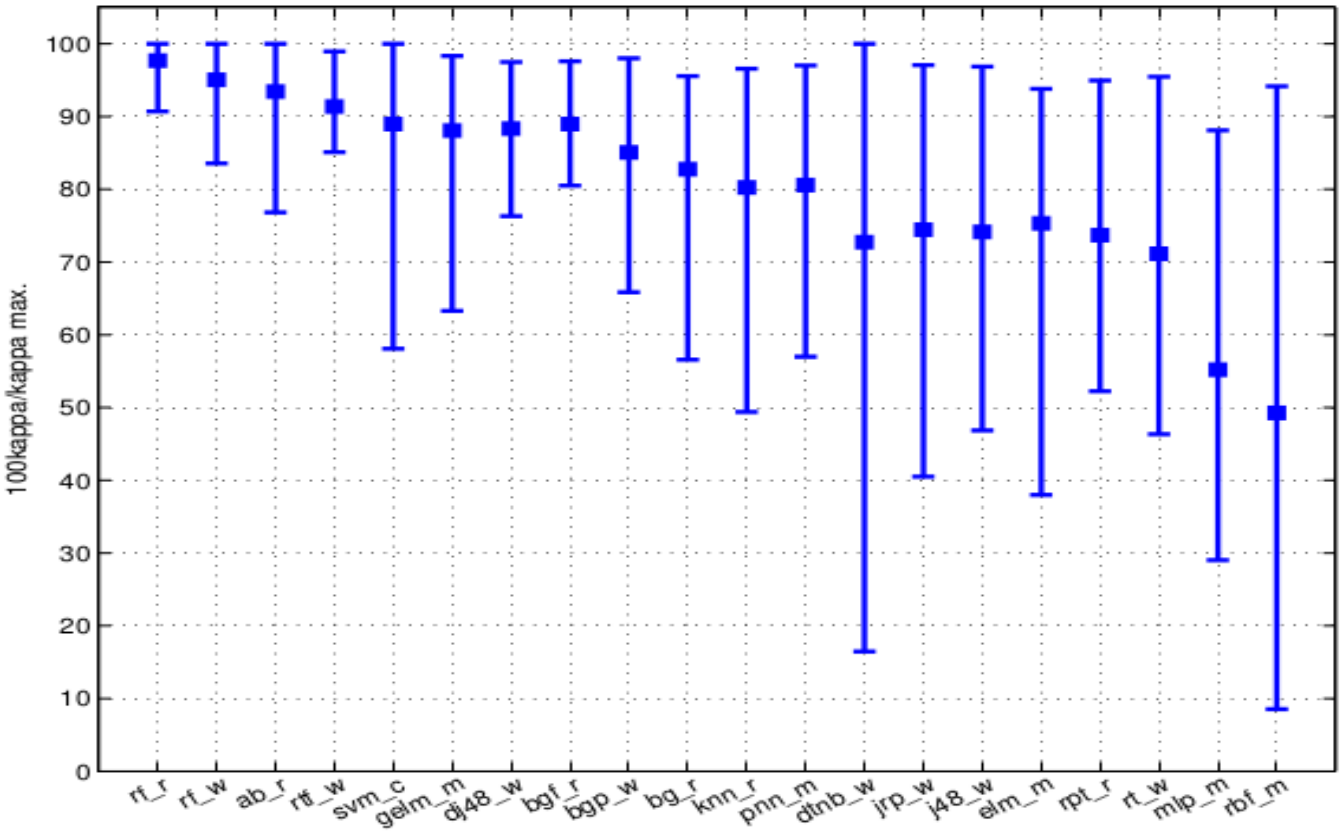
9. Intervals of classifiers and problems

- We developed 20x10=200 experiments. Interval of kappa (in %) of the different classifiers for each problem:



9. Intervals of classifiers and problems

- Interval of percentages of the maximum kappa for each problem, for all the problems and for each classifier:



10. Classification comparison among regions

- Validation of classifiers: Trained using data from one region, to test data from different regions.
- Classification of village-wise soil OC, P2O5, Mn and Fe fertility indices and pH level.
- Three regions: Marathwada, North Maharashtra and Paschim Maharashtra.

	OC-F		P_2O_5-F		Mn-F		Fe-F		pH	
Regions	Best	κ	Best	κ	Best	κ	Best	κ	Best	κ
M-NM	rpt_r	17.21	bgp_r	15	bg_r	65.98	rpt_r	66.78	rtf_w	23.97
M-PM	elm_m	9.66	rf_r	100	bg_r	66.62	rpt_r	70.52	rtf_w	66.22
NM-M	rtf_w	32.60	rf_r	34.10	rt_w	45.70	svm_c	42.50	–	–
NM-PM	pnn_m	14.99	bg_r	100	bg_r	100	j48_w	100	–	–
PM-M	rt_w	12.73	bgf_r	34.57	bgf_r	45.15	elm_m	44.33	knn_r	69.61
PM-NM	pnn_m	11.13	bg_r	100	bg_r	100	svm_c	100	rf_r	48.23

CiTUS

11. Conclusions (I)

- RF achieves the best kappa for six of ten problems, overcoming 90% of the maximum kappa in all the cases followed by adaboost, SVM and Gaussian extreme learning machine.
- We achieved kappa above 90% for village-wise OC fertility index and soil type classification and above 85% for P2O5 fertility index and crop classification, while Mn and Fe fertility indices achieved nearly 65% kappa value.
- We studied data validity across three regions, finding compatibilities between North Maharshtra and Paschim Maharashtra for village-wise P2O5, Mn and Fe fertlity indices classification.

11. Conclusions (II)

- This results might contribute to design agricultural strategies of the Indian Government to manage the soil fertility degradation, crop productivity and usage of fertilizers.
- This research helps to avoid expensive chemical measurements of OC, N2O, P2O5 and K2O.
- Saves time of specialized technicians in developing expensive chemical analysis.
- Work can supports Indian government to make decisions about improving soil quality and crop production.

Other projects

- We are also working in some other projects which do require computational resources.
- Exhaustive algorithm comparison:
  - About 60 different algorithms, implemented in R.
  - Over a collection of about 50 data sets from the UCI Machine Learning repository.
  - Data sets range from 100 to 500,000 patterns.
  - Number of inputs (pattern width) ranging from 3 to 500,000.

CiTIUS