Robust And Reliable Techniques For Speech-based Emotion Recognition

Christina Brester, Eugene Semenkin, and Maxim Sidorov

Krasnoyarsk, 2014

Outline

- Background and Motivation
  - Problem Definition
  - Corpora Description
  - Feature Selection
  - Inferences #1

- Conventional models
  - Full feature set
  - Reduced feature set
  - Inferences #2

- Collective decision making in emotion recognition
  - Full feature set
  - Reduced feature set
  - Inferences #3

- Conclusions and Future work

Example #1

Human-Human Communication

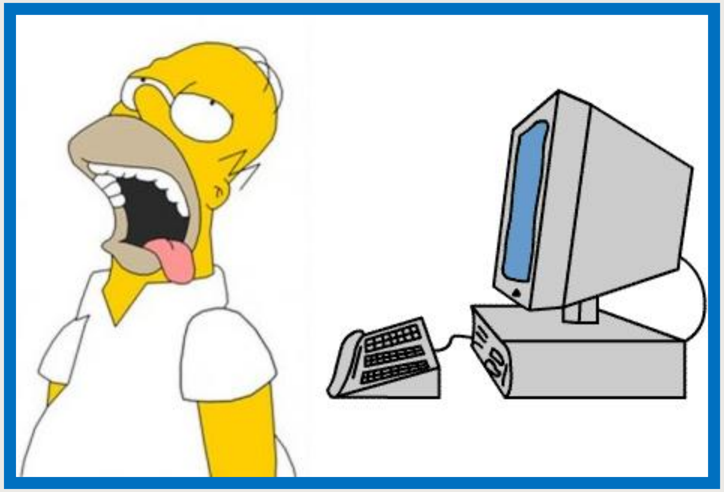
First 30 min

After a while

Human-Machine Communication

First 30 min

After a while



To show regret

or

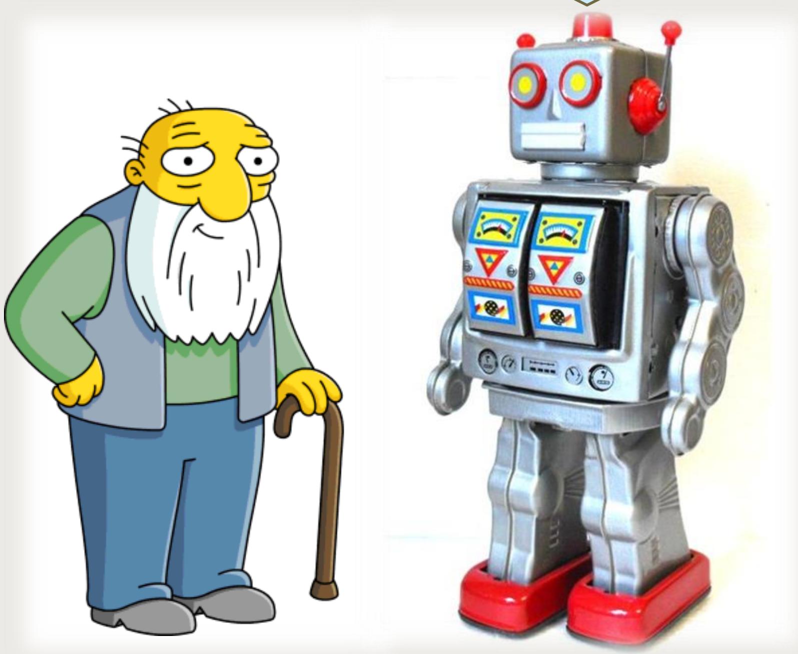
To express happiness

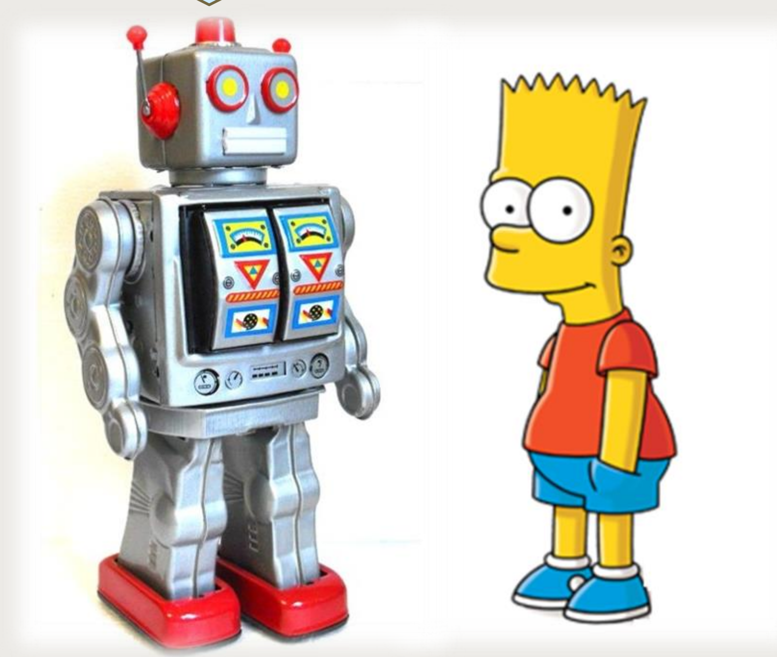
Example #2

To personalize a response

Good morning,
Mister!
Can I help you?

Hey, guy!
What’s up?





Example #3

Quality monitoring of call centres

Are you kidding?
I’ve been waiting for two hours!

Please, wait a minute, Sir!

Consultant

..okay

Speech-based Emotion Recognition Problem

List of extracted features

- General features: Power, Mean, Root mean square, Jitter, Shimmer
- Mel-frequency cepstral coefficients (MFCCs):12 MFCCs
- Formants: 5 Formants
- Pitch, Intensity and harmonicity based features: Mean, Minimum, Maximum, Range, Deviation
- Etc.

Voice

↓

Voice conversion into the digital form

↓

Extraction of numerical characteristics

↓

Classification of sound signals

↓

The **emotion** is detected

Sample

$x_{1,1}$	$x_{1,2}$	…	$x_{1,m}$	y_1
$x_{2,1}$	$x_{2,2}$	…	$x_{2,m}$	y_2
$x_{3,1}$	$x_{3,2}$	…	$x_{3,m}$	y_3
…	…	…	…	…
$x_{n,1}$	$x_{n,2}$	…	$x_{n,m}$	y_n

$\bar{x}_i$ – independent variable,
y_i – dependent variable, $i = \overline{1,n}$,
$y_i \in C$, where $C = \{c_1, c_2, \ldots, c_r\}$ – finite set,
r – the number of classes.

New examples

$x_{1,1}$	$x_{1,2}$	…	$x_{1,m}$	?
…	…	…	…	…
$x_{l,1}$	$x_{l,2}$	…	$x_{l,m}$	?

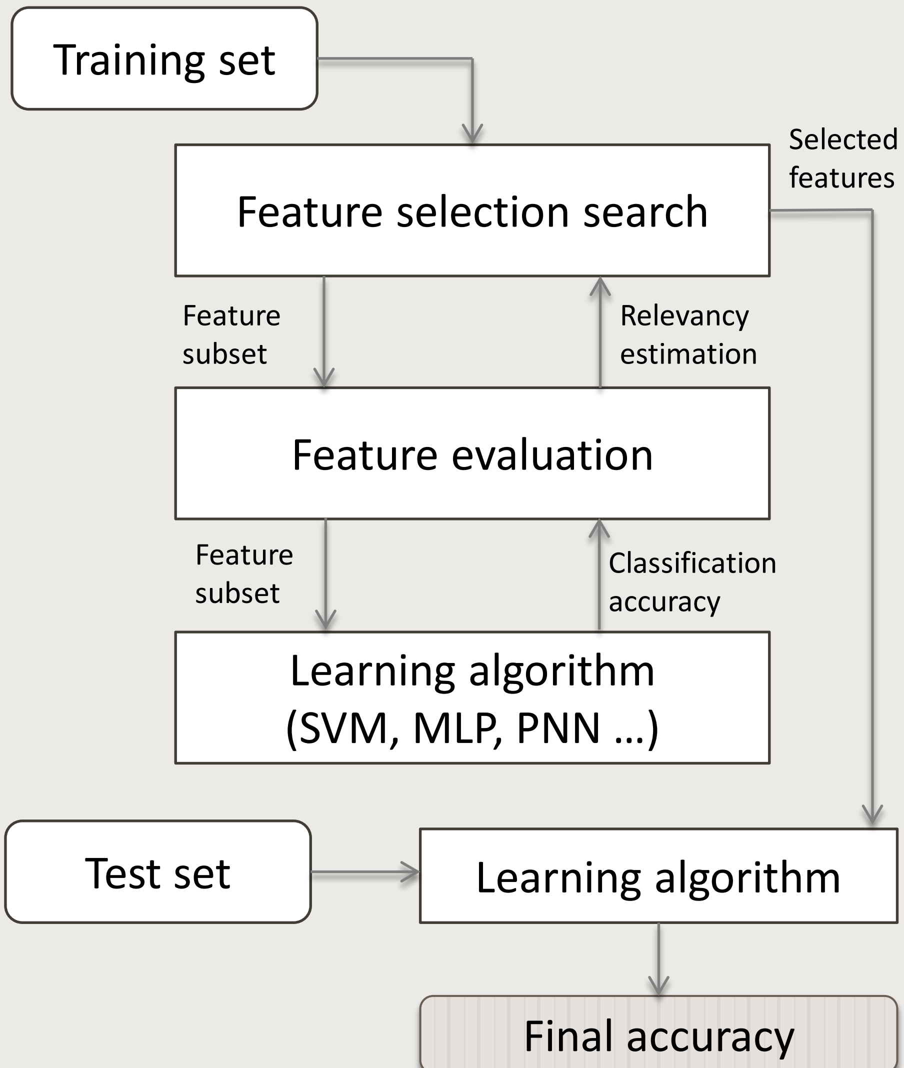
Goal:
To classify new objects based on the sample (supervised learning).

Corpora description

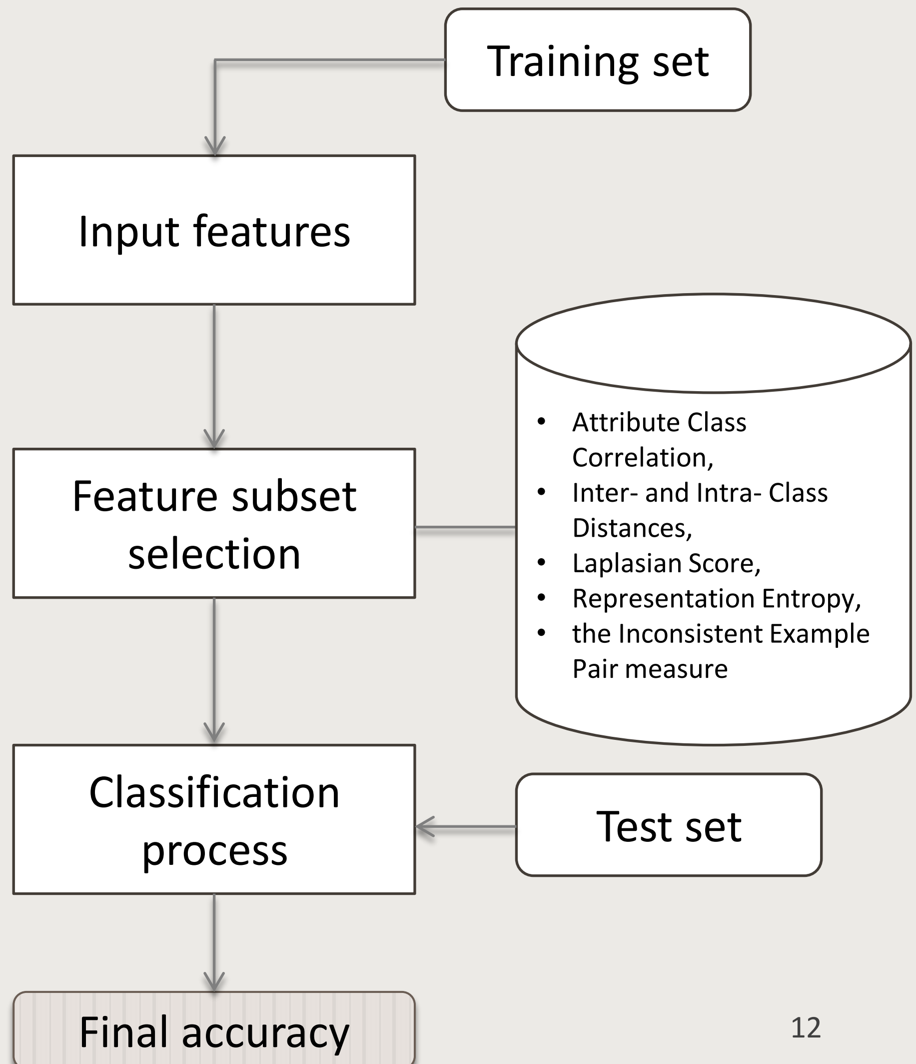
Database	Language	Full length (min.)	Number of emotions	File level duration		Notes
				Mean (sec.)	Std. (sec.)	
Berlin	German	24.7	7	2.7	1.02	Acted
SAVEE	English	30.7	7	3.8	1.07	Acted
LEGO	English	118.2	5	1.6	1.4	Non-acted
VAM	German	47.8	4	3.02	2.1	Non-acted
RadioS	German	278.5	4	6.26	5.17	Non-acted
UUDB	Japanese	113.4	4	1.4	1.7	Non-acted

Feature selection concepts: *formal models*

Wrapper approach



Filter approach



Feature selection search

Main concepts:

- An optimization model with **binary representation**:

1	0	0	…	1

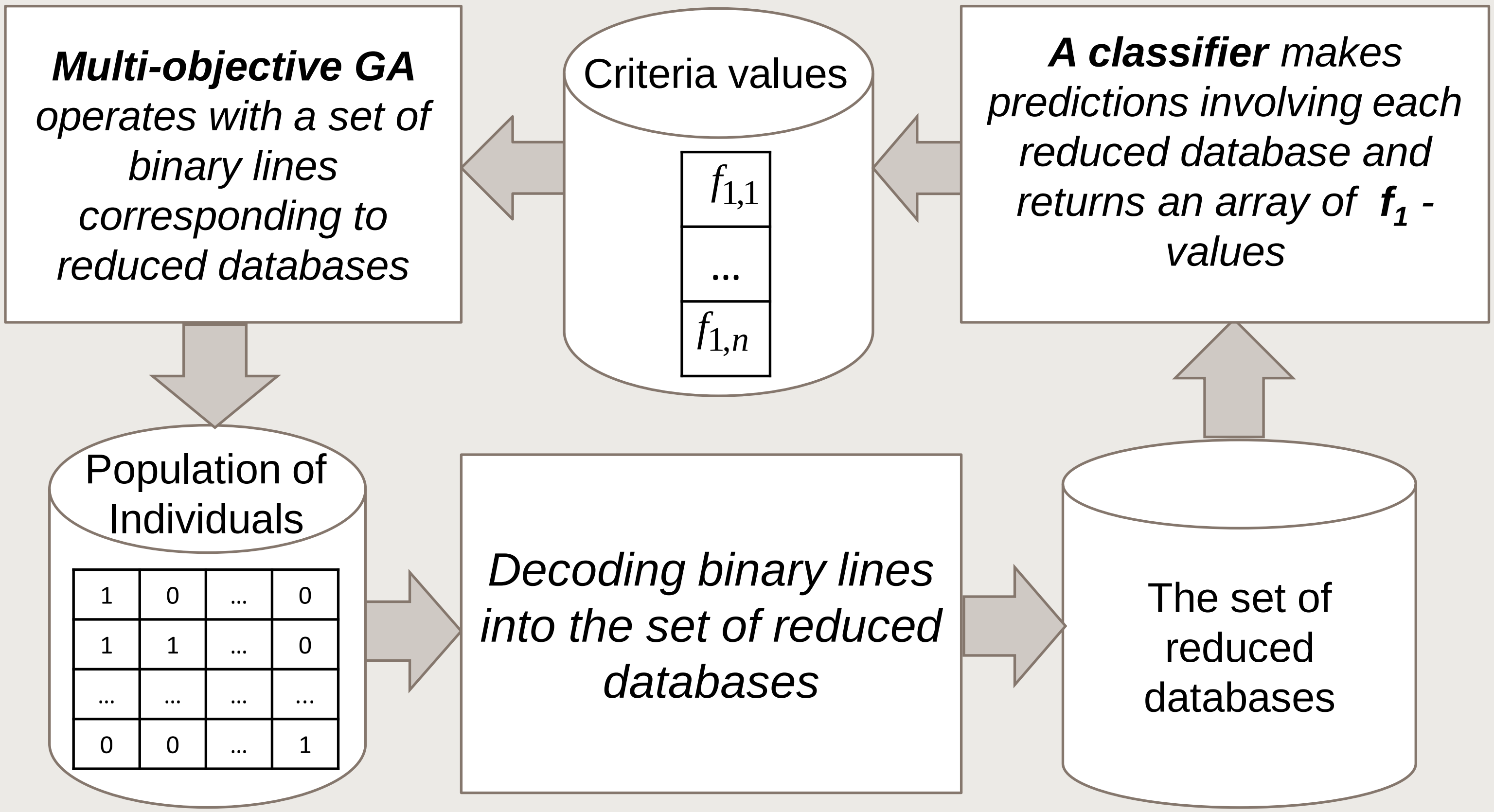
unit corresponds to the relevant attribute;
zero denotes the irrelevant attribute.

- **Evolutionary (genetic) algorithms** as a technique for optimizing both discrete and continuous criteria.

- **The self-adaptation idea** as a strategy to organize the automatic choice of algorithm settings.

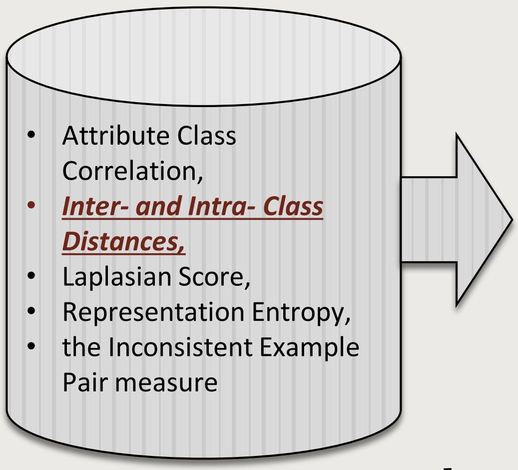
Wrapper approach: *the actual model*

f1 - the relative classification error,
f2 - the number of selected features,
$\boldsymbol{f1 \to min, f2 \to min}$



Filter approach: *the actual model*

f1 – the Intra-Class Distance (IA),
f2 - the Inter-Class Distance (IE),
f1 → min, f2 → max



$$IA = \frac{1}{n}\sum_{r=1}^{k}\sum_{j=1}^{n_r} d(p_j^r, p_r),$$

$$IE = \frac{1}{n}\sum_{r=1}^{k} n_r d(p_r, p),$$

where p_j^r is the j-th example from the r-th class,
p is the central example of the data set,
d(...,...) denotes the Euclidian distance,
p_r and n_r represent the central example and the number of examples in the r-th class.

Inferences #1

Due to the independency of the filter approach from classification models:

- it becomes possible explore the *robustness property* of the filter technique (to consider a number of classification models and check whether this method is effective for most of them or not).

- it might be supposed that this feature selection procedure should be rather effective in combination with various classifiers (it is referred to reliability).

Conventional classification models used to investigate the **robustness** property of the **filter approach**

* Multilayer Perceptron (MLP)
* Support Vector Machine (SVM)
* Linear Logistic Regression (Logit)
* Radial Basis Function network (RBF)
* Naive Bayes
* Decision trees (J48)
* Random Forest
* Bagging
* Additive Logistic Regression (LogitBoost)
* One Rule (OneR)

Experimental results for conventional classifiers

	Berlin			SAVEE			LEGO		
	F_score, %		*Gain, %*	*F_score, %*		*Gain, %*	*F_score, %*		*Gain, %*
	Full	*Reduced*		*Full*	*Reduced*		*Full*	*Reduced*	
MLP	**82.87**	**82.26**	-0.74	**61.72**	**63.58**	3.01	67.53	**71.70**	6.18
SVM	81.71	82.14	0.53	59.22	60.77	2.62	**70.81**	69.88	-1.31
Logit	80.04	82.15	2.64	57.20	63.46	10.95	70.75	69.82	-1.31
RBF	68.93	71.59	3.85	43.27	44.15	2.03	52.61	61.31	16.53
Naive Bayes	66.91	67.45	0.81	43.64	45.53	4.33	57.00	59.43	4.26
J48	50.15	51.96	3.60	42.46	47.79	12.55	57.55	64.90	12.77
Random Forest	54.69	73.43	34.27	38.60	55.73	44.38	65.47	68.47	4.58
Bagging	60.60	63.29	4.43	42.99	52.91	23.07	67.53	68.06	0.79
Logit Boost	66.66	71.21	6.82	49.08	52.22	6.40	67.66	67.04	-0.92
OneR	29.20	29.20	0.00	30.41	30.41	0.00	59.01	59.01	0.00

Experimental results for conventional classifiers

	VAM			RadioS			UUDB		
	F_score, %		*Gain, %*	*F_score, %*		*Gain, %*	*F_score, %*		*Gain, %*
	Full	*Reduced*		*Full*	*Reduced*		*Full*	*Reduced*	
MLP	41.08	**43.05**	4.80	**34.81**	**35.23**	1.21	25.48	34.58	35.71
SVM	43.57	36.92	-15.26	27.26	23.14	-15.13	35.59	33.04	-7.15
Logit	36.88	37.88	2.71	31.91	31.13	-2.44	36.72	36.33	-1.06
RBF	37.87	34.47	-8.97	23.14	23.14	0.00	26.75	23.60	-11.77
Naive Bayes	40.86	42.33	3.60	34.02	33.46	-1.65	36.52	36.45	-0.20
J48	36.20	37.70	4.17	29.81	30.74	3.14	38.70	**42.64**	10.18
Random Forest	**45.66**	37.08	-18.79	30.31	31.87	5.14	40.11	36.56	-8.84
Bagging	37.24	36.67	-1.54	26.37	32.63	23.74	40.94	37.08	-9.42
Logit Boost	40.06	36.14	-9.80	31.24	26.93	-13.79	41.28	37.17	-9.96
OneR	33.34	33.34	0.00	23.94	24.98	4.35	**41.92**	41.56	-0.85

Inferences #2

- There is no classification model which provides a lower F-score value for all of the corpora after the feature selection procedure.

- Obviously, in some cases the dimension reduction is achieved at the detriment of the classifier performance.

- Besides, there is no particular model that is equally effective for all of the databases.

- The random choice of the classifier may lead to significant performance deterioration.

- For the used corpora Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Linear Logistic Regression (Logit) demonstrated rather high performance.

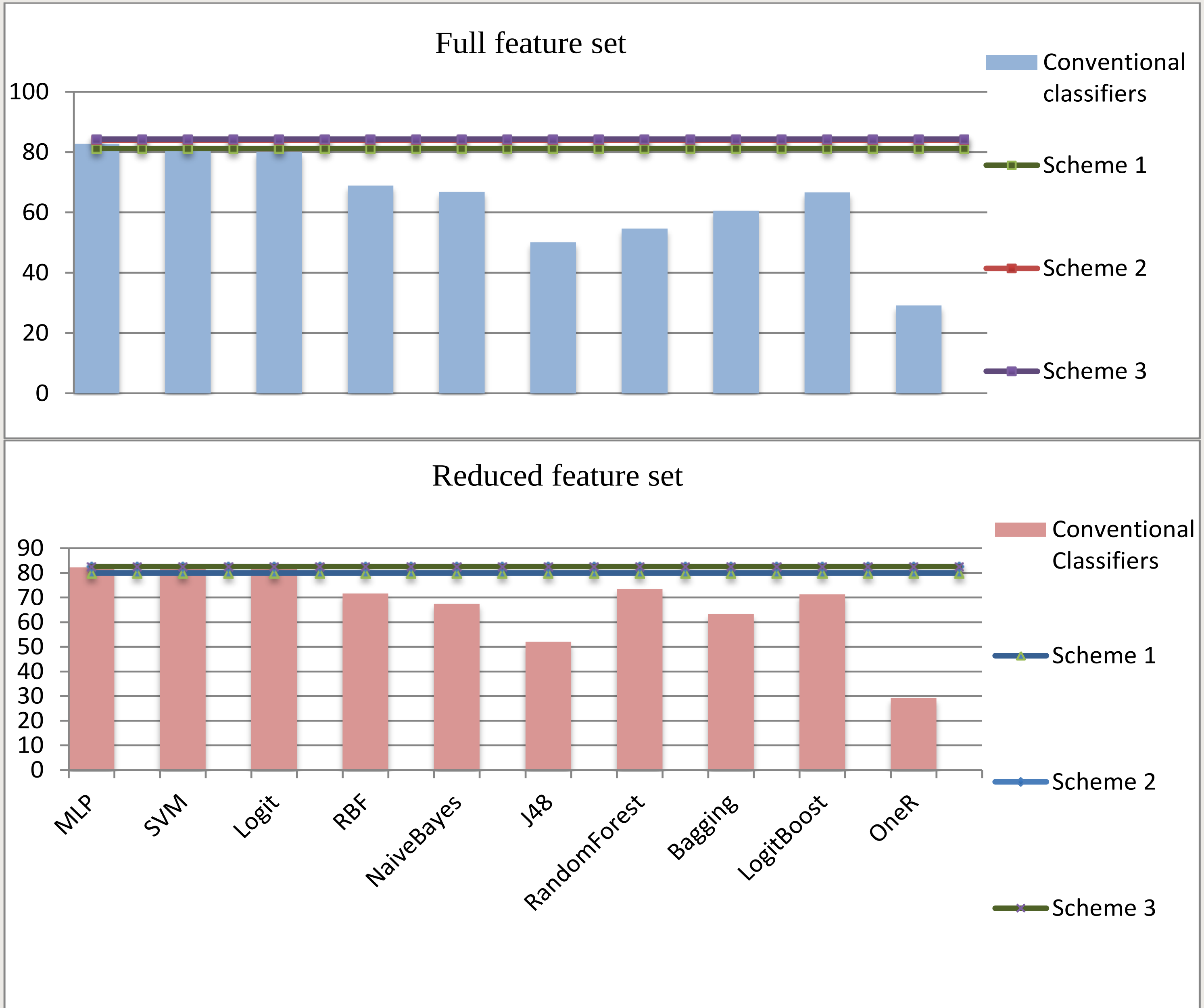
Collective decision making

Concept	Detailed information
Scheme 1. For each test example: *Choose a model that classifies correctly k-nearest neighbours from the training data set.*	1. For each test example it is necessary to determine k-nearest neighbours from the training data set. 2. The prediction of the model that classifies these k-nearest neighbours correctly is used as the final decision. (If several models demonstrate equal effectiveness, choose one of them randomly).
Scheme 2. *Voting procedure is realized with the usage of the majority rule.*	1. For each test example the engaged models vote for different classes according to their own predictions. 2. The final decision is defined as a collective choice based on the majority rule.
Scheme 3. *Combination of Scheme 1 and Scheme 2.*	Combine Schemes 1 and 2 in the following way: - fulfil the voting procedure as it is described in Scheme 2; - if several classes have the maximum number of votes, apply Scheme 1.

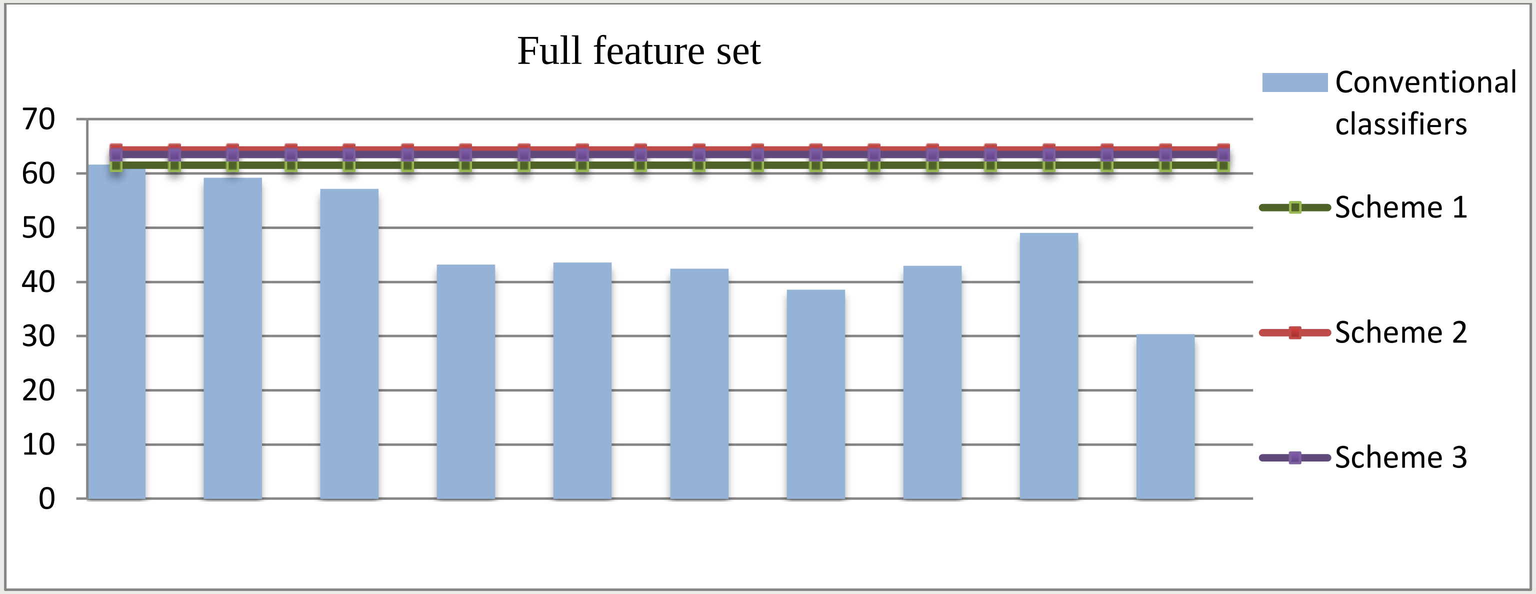
Experimental results for collective decision making schemes

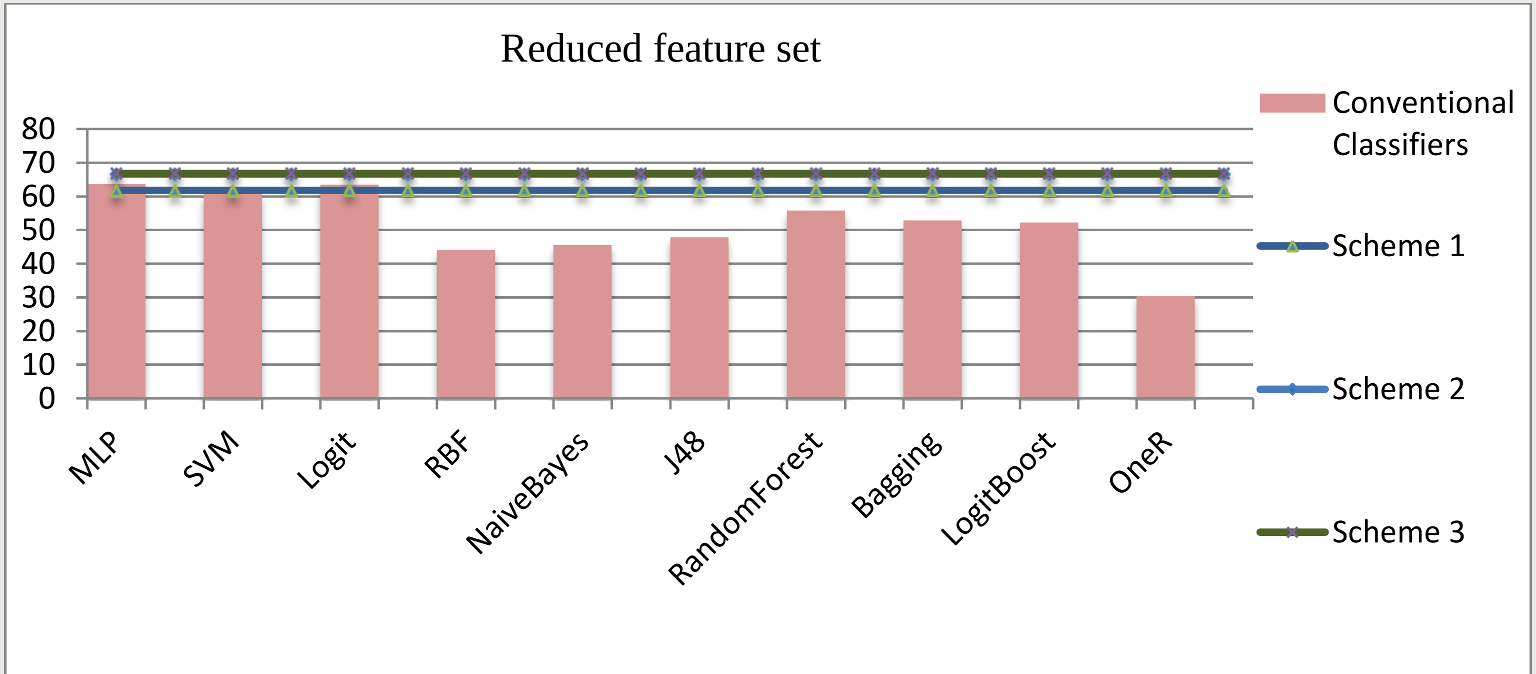
	Full feature set			Reduced feature set		
	Scheme 1	Scheme 2	Scheme 3	Scheme 1	Scheme 2	Scheme 3
Berlin	81.18	84.01	**84.23**	79.91	**82.54**	**82.54**
SAVEE	61.52	**64.33**	63.50	61.78	66.56	**66.80**
LEGO	70.52	**71.19**	71.13	68.57	70.11	**70.22**
VAM	42.29	**50.19**	43.69	37.99	**39.18**	**39.18**
RadioS	**30.68**	26.39	26.39	28.84	**28.92**	**28.92**
UUDB	37.96	36.41	**39.78**	**40.43**	34.99	35.19

Classification results for Berlin

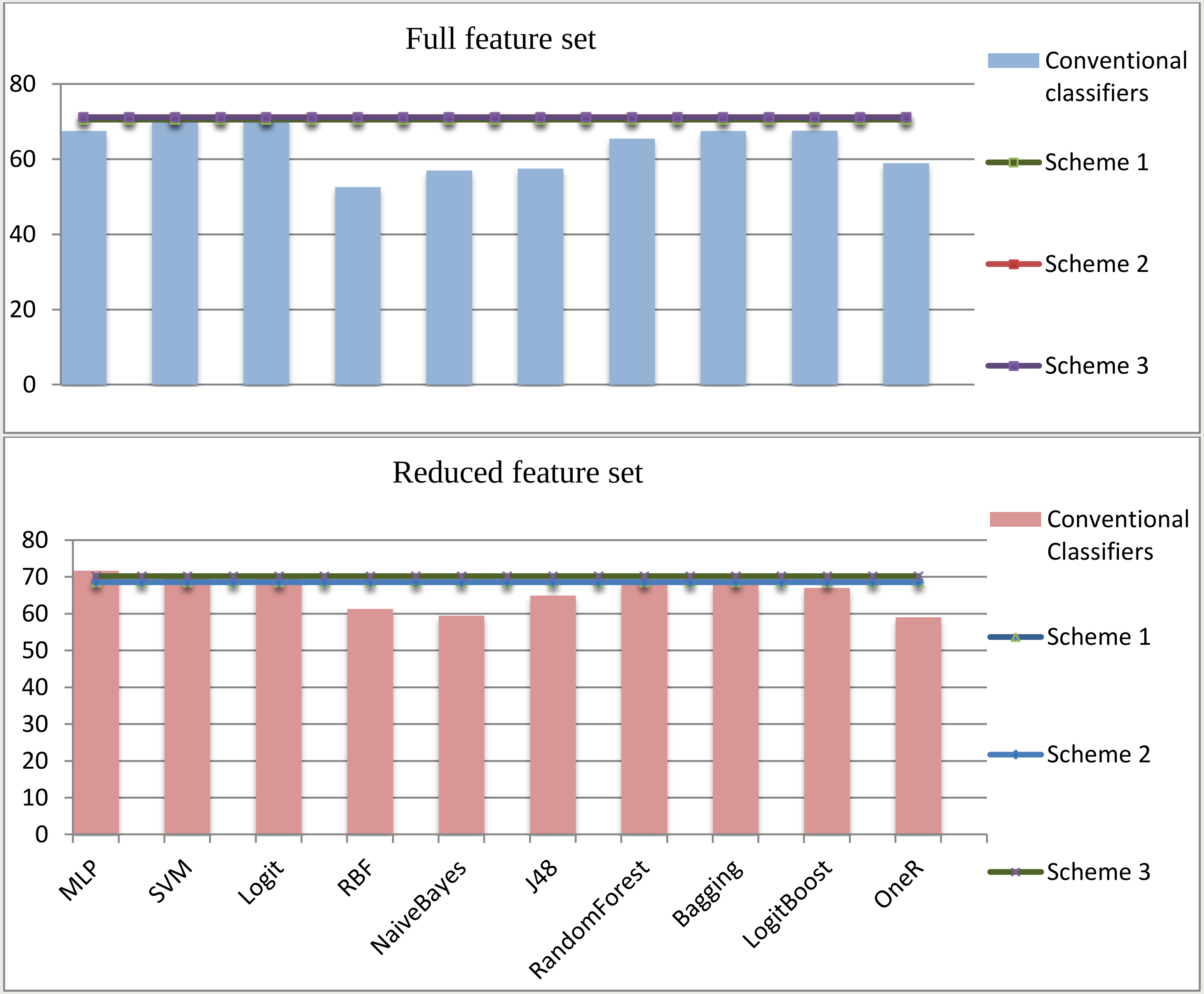


Classification results for SAVEE



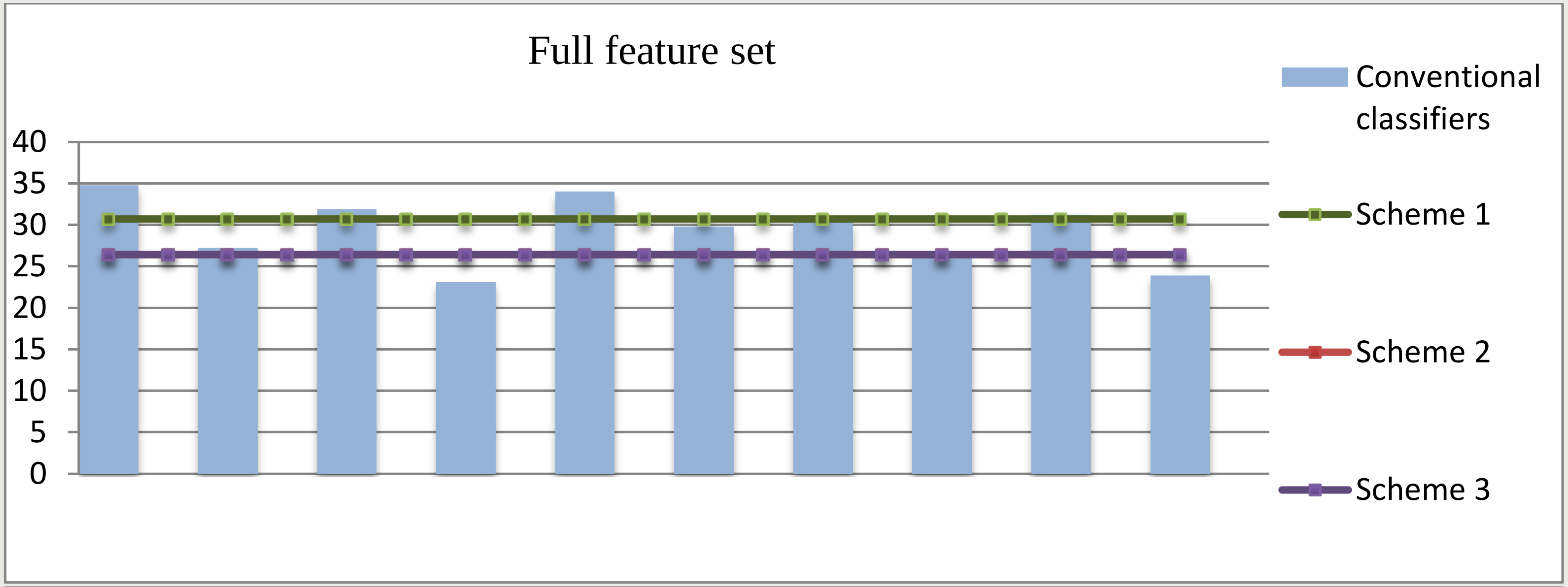


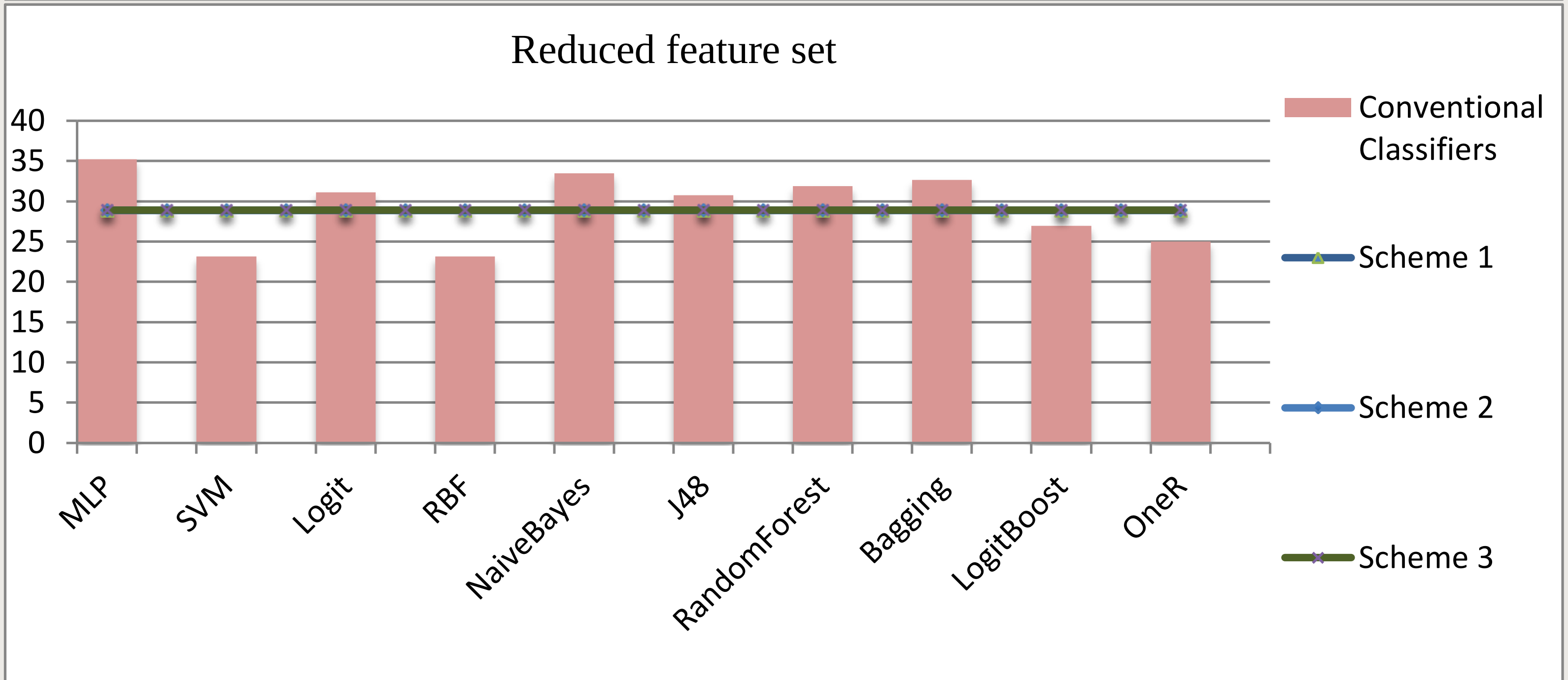
Classification results for LEGO



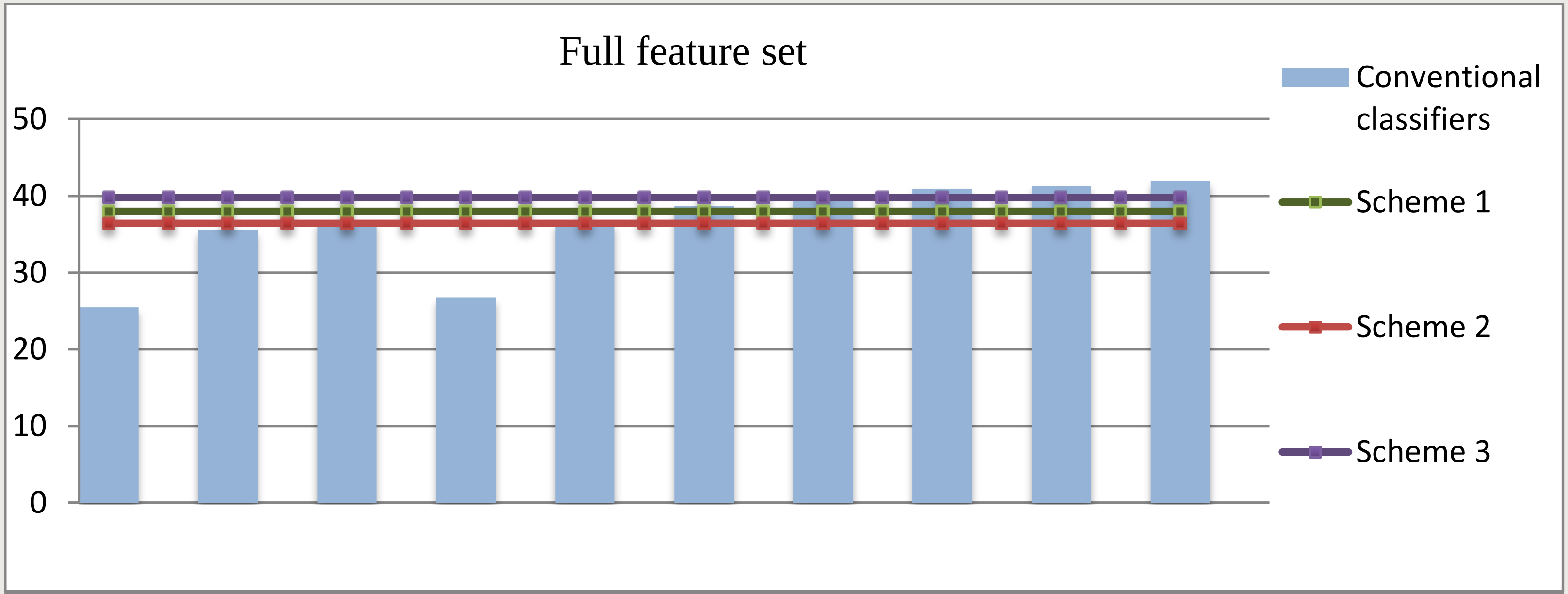
Classification results for VAM

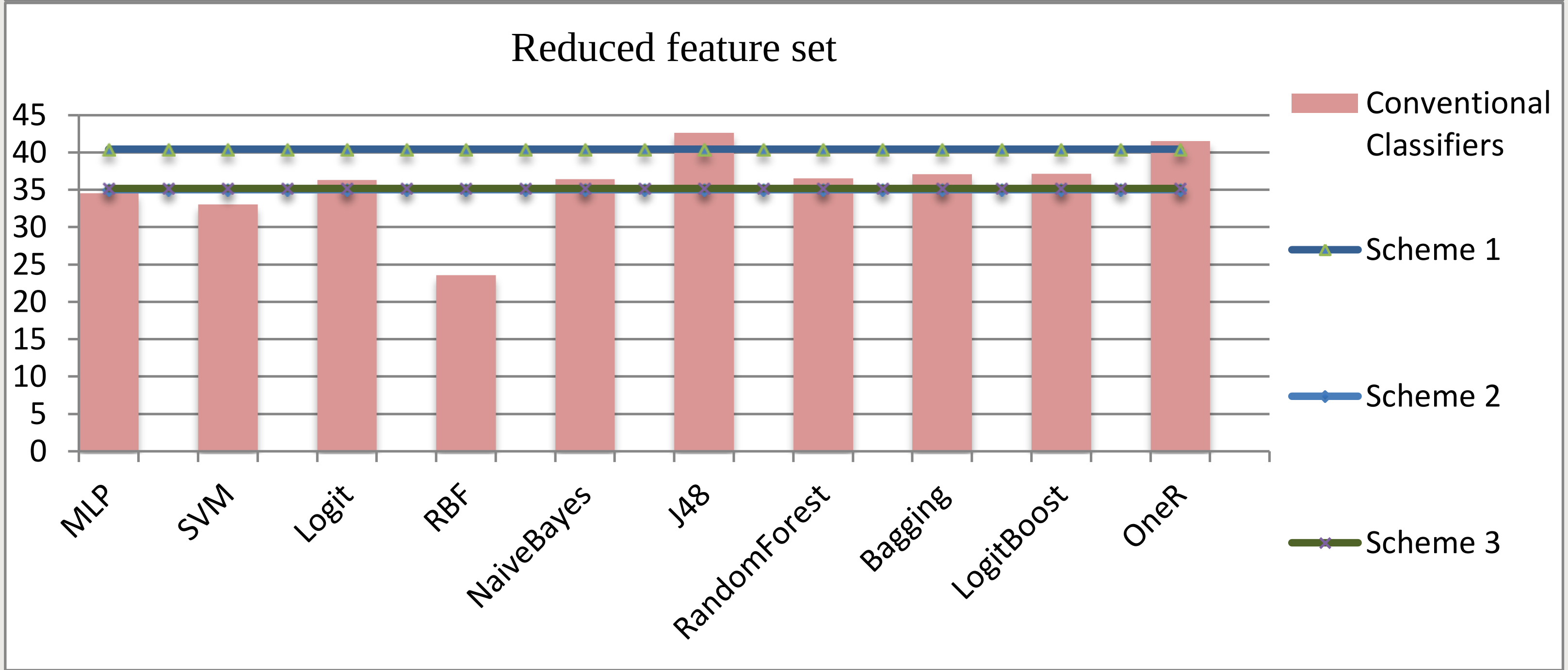
Classification results for RadioS





Classification results for UUDB





Inferences #3

- Due to the usage of the proposed techniques it became possible to improve the classification results for most of the corpora (in some cases even by up to 9.93% relative improvement).

- The conducted experiments also exposed that the proposed schemes of collective choice might be effectively applied to the full data set as well as to the reduced one (after feature selection).

Conclusions and Future work

Although we managed to achieve some good results, there are a number of questions:

- The first one is related to the feature selection technique, in particular, to the introduced criteria:

Whether it is reasonable to take into consideration other criteria (Laplasian Score, Representation Entropy and the Inconsistent Example Pair measure) or not? Should we engage the information about the classifier performance into the heuristic search on the stage of feature selection or ignore it totally to maintain the robustness of this approach?

- Other questions pertain to the classification models involved in the collective decision making process:

How many classifiers should we use to provide the most reliable scheme? What kind of models should it be compulsory to include in the ensemble of classifiers?

Thanks a lot