From dynamic classifier selection to dynamic ensemble selection

Albert H.R. Ko, Robert Sabourin, Alceu Souza Britto, Jr

Eider Sánchez

Contenidos

Introduction

- Different classifiers make different errors on different samples
  - By combining classifiers, more accurate decisions
  - Ensemble of Classifiers (EoC): group of classifiers

- Ensemble selection
  - Select adequate classifier group to achieve optimum recognition rates

- Three different schemes for selection and combining classifiers:
  a) static ensemble selection
  b) dynamic classifier selection
  c) proposed dynamic ensemble selection

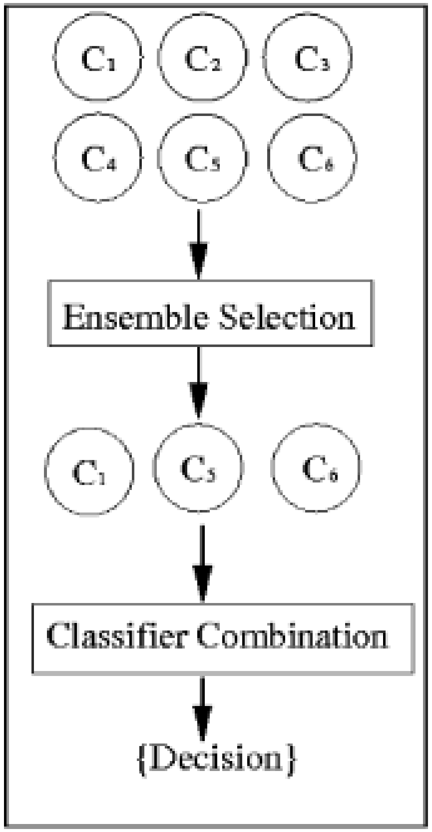
Introduction

a) static ensemble selection

Steps:

- find a **pertinent objective function** for selecting the classifiers
  - most crucial element
  - simple majority voting error (MVE) is one of the best
- use a **pertinent search algorithm** to apply this criterion
  - Genetic Algorithm (GA) considered to have advantage because of its population-based approach

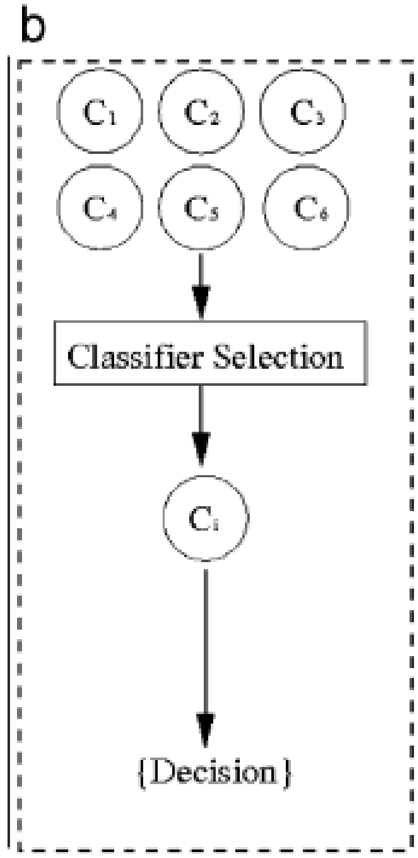
a



Introduction

b) dynamic classifier selection

- Explores the use of different classifiers for different test patterns
  - Based on the different features or different decision regions of each test pattern, a classifier is selected and assigned to the sample
  - Selection methods:
    - A Priori
    - A Posteriori
    - Overal Local Accuracy (OLA)
    - Local Class Accuracy (LCA)



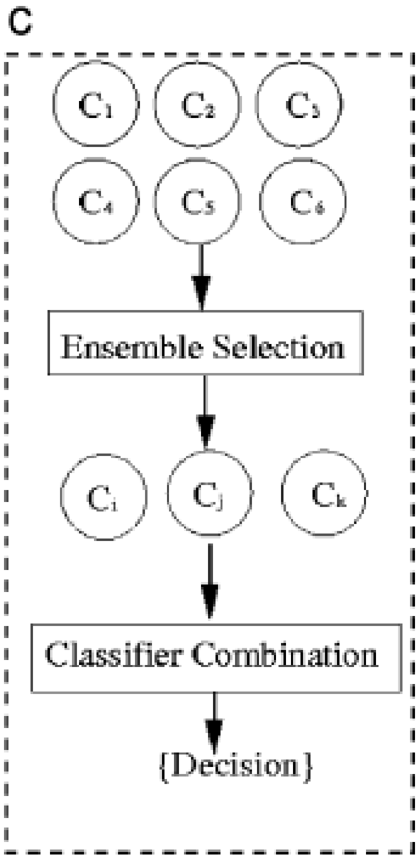
Critical point:

- Choice of one individual classifier over the rest depends on how much we trust the estimate of the generalization of the classifiers

Introduction

c) proposed dynamic ensemble selection

- Dynamic classification selection methods are designed to find the classifier with the greatest possibility of being correct for a sample in a pre-defined neighborhood.
- dynamic ensemble selection is designed to select the most suitable ensemble for each sample.
- Advantage
  - distribute the risk of this over-generalization by choosing a group of classifiers instead of one individual classifier for a test pattern

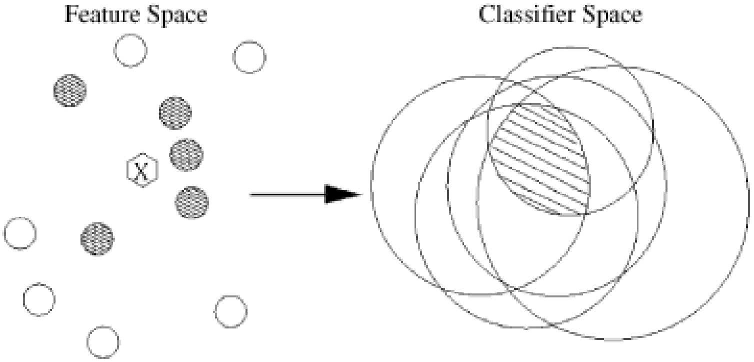


Proposed dynamic ensemble selection - KNORA

- K-nearest-oracles
  - Oracle
    - assigns the correct class label to a pattern if at least one individual classifier from an ensemble produces the correct class label for this pattern
- K-nearest-oracles (KNORA)
  - Instead of finding the most suitable classifier: select the most suitable ensemble for each sample
  - For any test data point
    - finds its nearest *K neighbors in the validation set*
    - *figures out which classifiers* correctly classify those neighbors in the validation set
    - uses them as the ensemble for classifying the given pattern in that test set

Proposed dynamic ensemble selection - KNORA

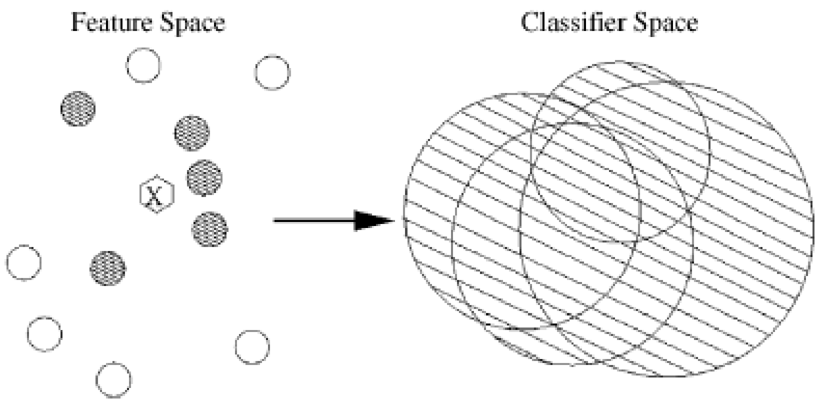
- Schemes
  - KNORA-ELIMINATE
    - Given K neighbors $x_j, 1 \leqslant j \leqslant K$ of a test pattern X
    - Suposing that a set of classifiers $C(j), 1 \leqslant j \leqslant K$ correctly classifies all its K-nearest neighbors
    - then every classifier ci ∈ C(j) belonging to this correct classifier set C(j) should submit a vote on the sample X
    - Case: no classifier can correctly classify all the K-nearest neighbors of the test pattern → decrease value of K until at least one classifier correctly classifies its neighbors



  - KNORA-ELIMINATE-W (vote weighted)

Proposed dynamic ensemble selection - KNORA

- Schemes
  - KNORA-UNION
    - Given K neighbors $x_j, 1 \leqslant j \leqslant K$ of a test pattern X
    - Supposing that the j-nearest neighbor has been correctly classified by a set of classifiers $C(j), 1 \leqslant j \leqslant K$
    - Every classifier ci ∊ C(j) belonging to this correct classifier set C(j) should submit a vote on the sample X
    - Note: a classifier can have more than one vote if it correctly classifies more than one neighbor.



  - KNORA-UNION-W (vote weighted)

Experiments: comparison on UCI repository

- 3 classification algorithms:
  - KNN
  - Parzen windows classifier (PWC)
  - Quadratic discriminant classifier (QDC)

- 3 ensemble creation methods:
  - Random Subspaces
    - Creates diverse classifiers by using different subsets of features to train classifiers
    - Due to the fact that problems are represented in different subspaces, different classifiers develop different borders for the classification
  - Bagging
    - generates diverse classifiers by randomly selecting subsets of samples to train classifiers
  - Boosting
    - uses a part of the samples to train classifiers, but not randomly.
    - difficult samples have higher probability of being selected, and easier samples have less chance of being used for training
    - With this mechanism, most of the classifiers created will focus on hard samples and can be more effective.

vicomtech
visual interaction communication technologies

Experiments: comparison on UCI repository

Table 1
UCI data for ensembles of classifiers

Database	Classes	Tr	Ts	Features	RS-Card.	Bagging (%)	Boosting (%)
Liver-disorders (LD)	2	172	172	6	4	66	66
Pima-diabetes (PD)	2	384	384	8	4	66	66
Wisconsin breast-cancer (WC)	2	284	284	30	5	66	66
Wine (W)	3	88	88	13	6	66	66
Image segmentation (IS)	7	210	2100	19	4	66	66
Letter recognition (LR)	26	10 000	10 000	16	12	66	66

Tr = training samples; Ts = test samples; RS-Card = Random Subspace Cardinality; Bagging = proportion of samples used for Bagging; Boost = proportion of samples used for Boost.

Table 2
Dynamic selection results for Random Subspace using KNN classifiers

Database	KN-E (%)	KN-E-W (%)	KN-U (%)	KN-U-W (%)	A Priori (%)	A Posteriori (%)	OLA (%)	LCA (%)	Oracle (%)	All (%)	Single best (%)
LD	78.47	78.47	80.56	**84.03**	77.78	70.14	79.17	70.83	100.00	76.39	74.31
PD	**97.54**	**97.54**	96.83	96.48	94.37	93.66	96.83	93.66	98.25	96.13	96.83
WC	93.66	93.66	**94.37**	93.66	90.85	80.99	93.31	88.38	99.65	92.61	95.07
W	**97.73**	**97.73**	**97.73**	**97.73**	**97.73**	37.50	**97.73**	**97.73**	97.73	76.14	90.91
IS	78.29	78.29	**78.67**	78.62	75.81	60.90	75.43	59.62	97.29	78.19	84.14
LR	83.33	83.33	83.85	84.20	84.84	87.02	84.84	**87.24**	94.78	83.08	85.32

KN-E = KNORA-ELIMINATE; KN-E-W = KNORA-ELIMINATE-W; KN-U = KNORA-UNION; KN-U-W = KNORA-UNION-W. All = the combined performances of all classifiers. Single best = the performance of the best classifier from the pool. The best classification rates of each method within the neighborhood sizes $1 \leqslant k \leqslant 30$ are shown.
Bold values are the best performances in each row of table.

Experiments: comparison on UCI repository

- Random Subspaces
  - KNORA-UNION and LCA have more stable performances than other methods
  - KNORA-UNION-W is not always better than KNORA-UNION
  - KNORA-ELIMINATE-W and KNORA-ELIMINATE have the same performances on Random Subspaces
    - probabilities weighted by the Euclidean distances between the test pattern and validation patterns do not affect the decisions of KNORA-ELIMINATE on Random Subspaces.
- Bagging
  - KNORA-ELIMINATE, KNORA-UNION and LCA have good performances.
  - KNORAUNION-W is not always better than KNORA-UNION
    - the probabilities weighted by the Euclidean distances between the test pattern and validation patterns do not always contribute to higher classification rates for either dynamic classifier selection or dynamic ensemble selection.
- Boosting
  - KNORA-ELIMINATE, KNORA-UNION and LCA seem to be quite stable
  - KNORA-UNION-W is not always better than KNORA-UNION
  - KNORA-ELIMINATE-W and KNORA-ELIMINATE have the same performances

Experiments: comparison on UCI repository

- dynamic ensemble selection can marginally improve the accuracy, but not always performs better than dynamic classifier selection
- But: problems extracted from the UCI machine learning repository usually consist of a small number of samples with few features.
- need to carry out a larger scale experiment on a problem with more features and larger classifier pools

Experiments: handwritten numerals

- Experiment: 10-class handwritten-numeral problem with 132 features and 100 classifiers
  - Nearest neighbor classifiers *(K = 1) for KNN,*
    - *each KNN classifier* having a different feature subset of 32 features extracted from the total of 132 features

- Used
  - Training set with 5000 samples (hsf_{0 − 3}) to create 100 KNN in Random Subspaces.
  - The optimization set containing 10,000 samples (hsf _{0 − 3})
  - Test set containing 60,089 samples *(hsf_{7}) to evaluate EoC accuracies*

Experiments: handwritten numerals

- Most of the dynamic schemes have so far proved better than all the tested objective functions for static ensemble selection

Table 11
The recognition rates on test data of ensembles searched by GA with the mean classifier error, majority voting error

OF	Min (%)	Q_L (%)	Median (%)	Q_U (%)	Max (%)
ME	94.18	94.18	94.18	94.18	94.18
MVE	96.32	96.41	96.45	96.49	96.57

ME = mean classifier error; MVE = majority voting error; OF = objective functions. Min = minimum accuracy of ensembles; Max = maximum accuracy of ensembles; Q_U = upper quarter accuracy of ensembles; Q_L = lower quarter accuracy of ensembles; Median = median accuracy of ensembles.

Table 12
The best recognition rates of each dynamic ensemble selection methods within the neighborhood sizes $1 \leqslant k \leqslant 30$

	KN-E	KN-E-W	KN-U	KN-U-W	OLA	LCA	A Priori	A Posteriori
RR	97.52%	97.52%	97.25%	97.25%	94.11%	97.40%	94.12%	97.40%
K-value	7,8	7,8	1	1	30	1	30	1

RR = recognition rates.

Experiments: handwritten numerals

- Performance declined with an increase in the value k
  - KNORE-ELIMINATE stable with k

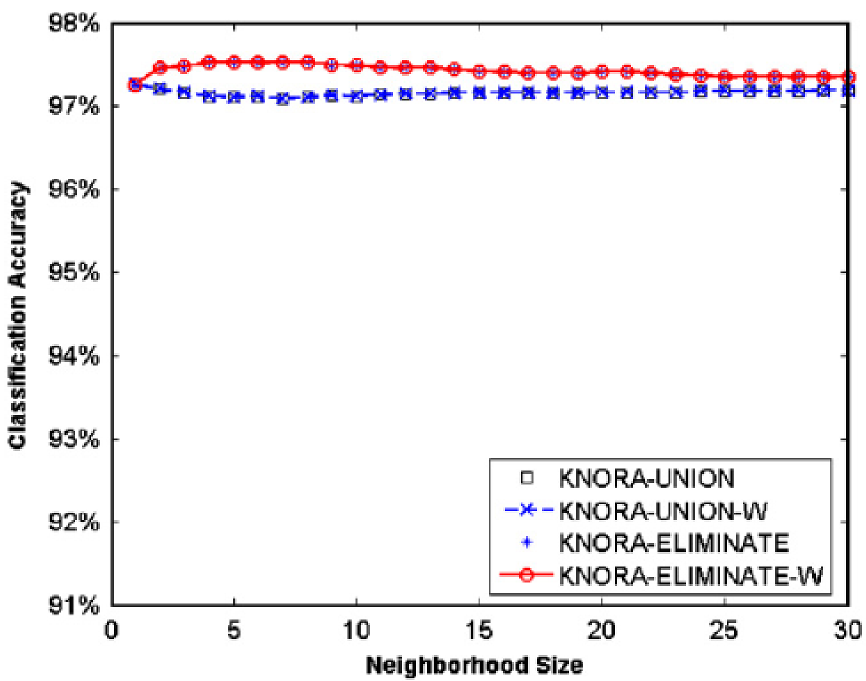


Fig. 4. The performances of proposed dynamic ensemble selection schemes based on different neighborhood sizes $1 \leqslant k \leqslant 30$ on NIST SD19 database. In the figure KNORA-ELIMINATE overlaps with KNORA-ELIMINATE-W, and KNORA-UNION overlaps with KNORA-UNION-W.

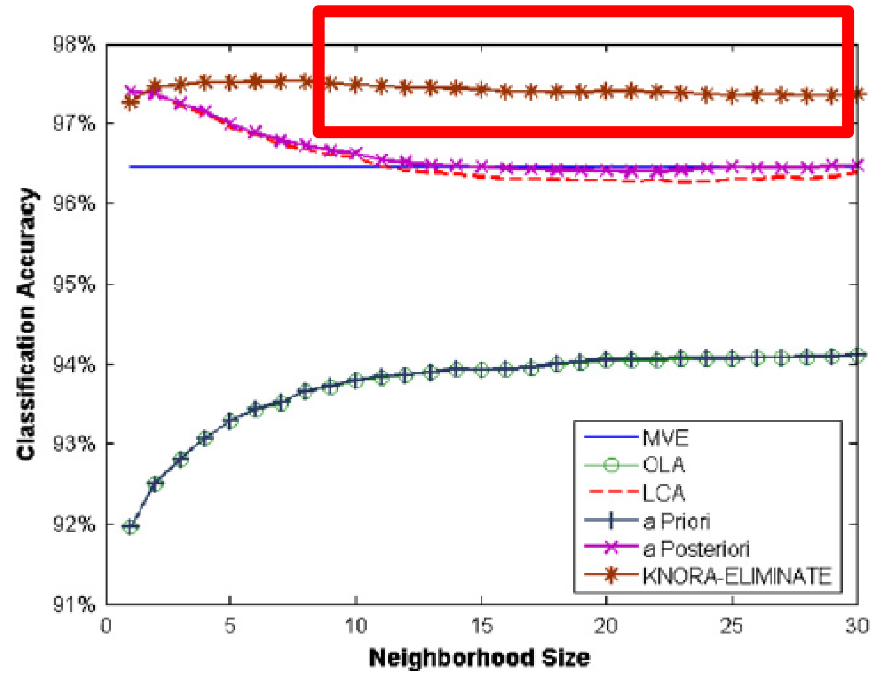
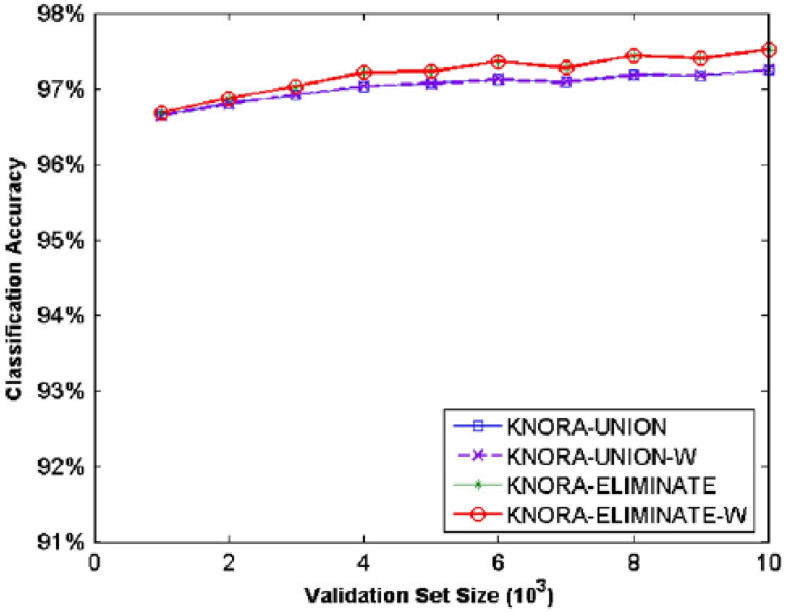
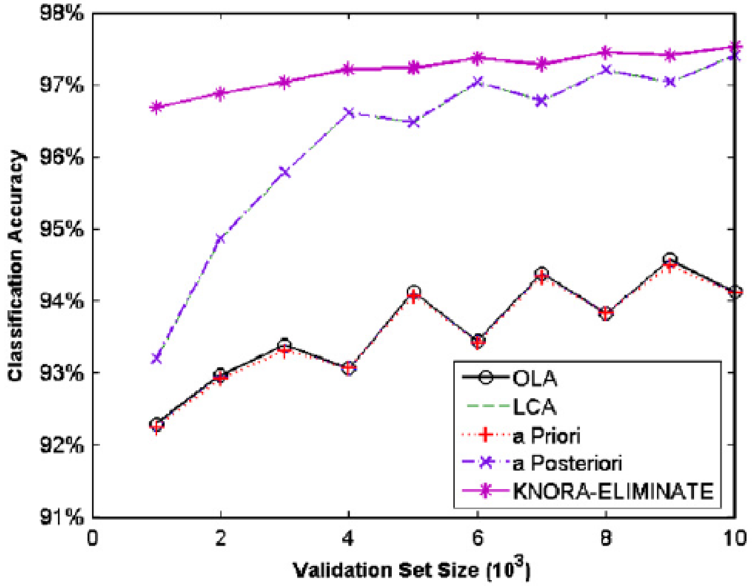


Fig. 5. The performances of various ensemble selection schemes based on different neighborhood sizes $1 \leqslant k \leqslant 30$ on NIST SD19 database. In the figure OLA overlaps with A Priori selection.

Experiments: handwritten numerals

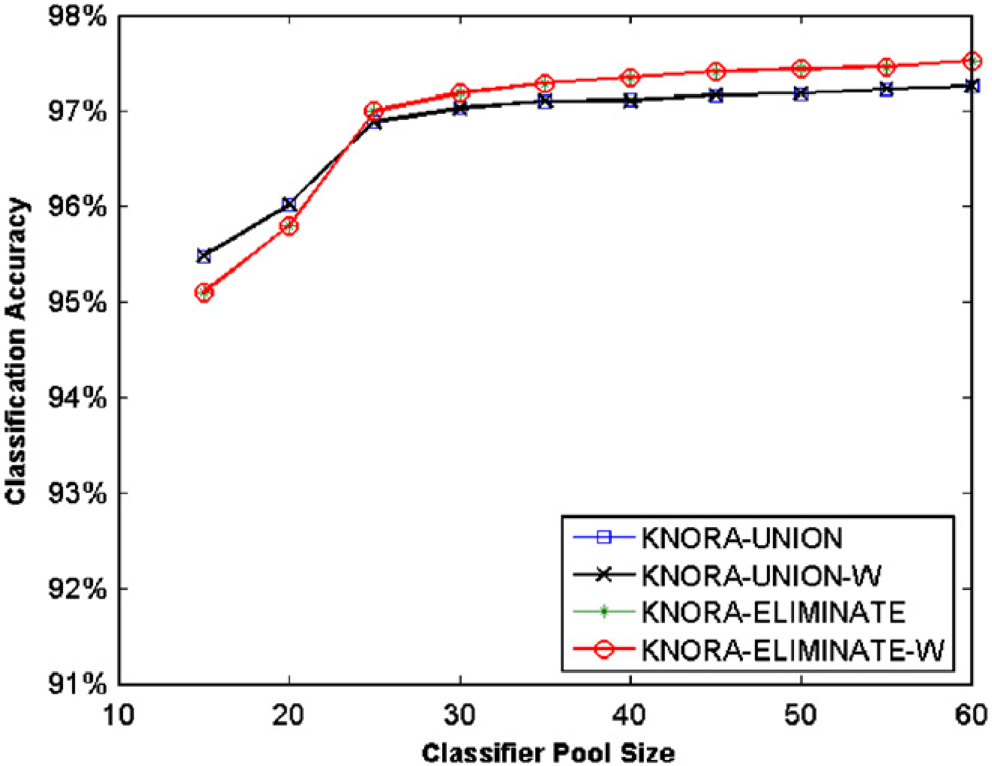
- Increase of validation samples
  - KNORA: More likely that test pattern has better nearest neighbors
  - Traditional dynamic selection - fluctuations

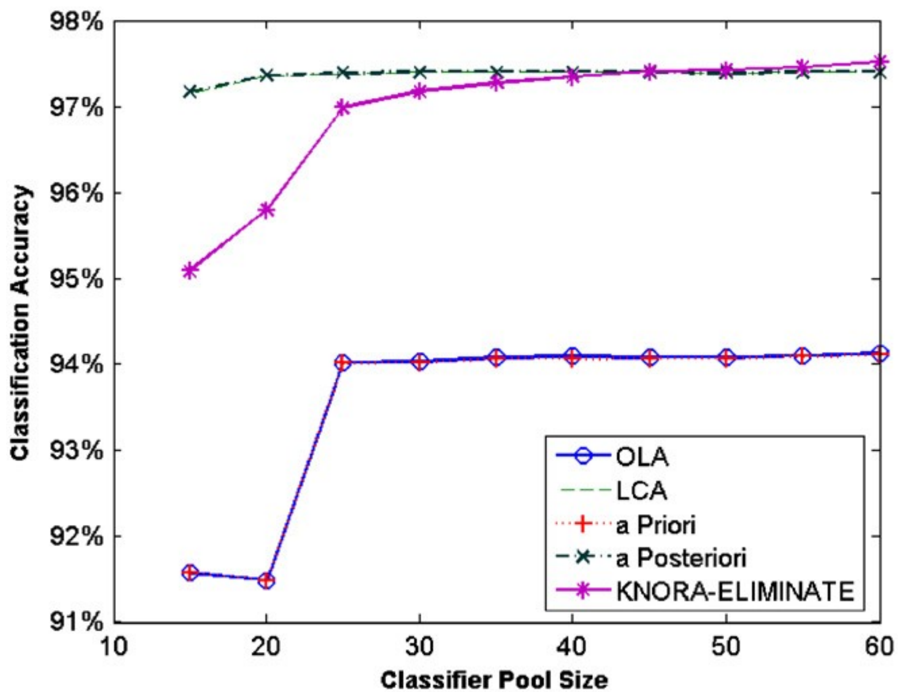




Experiments: handwritten numerals

- Classifier Pool-Size
- KNORA methods are better suited to large classifier pools





Conclusions

- OLA and A Priori dynamic selection schemes were not as good as the static GA selection scheme with the MVE
- KNORA-UNION and KNORAUNION-W perform less well than KNORA-ELIMINATE or KNORA-ELIMINATE-W
- KNORA-ELIMINATE also performs slightly better than the other dynamic selection schemes
- However, the performance of KNORAELIMINATE is still far from the oracle

vicomtech
visual interaction
communication
technologies

Gracias

INI-GraphicsNet

ik4 research alliance

20