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

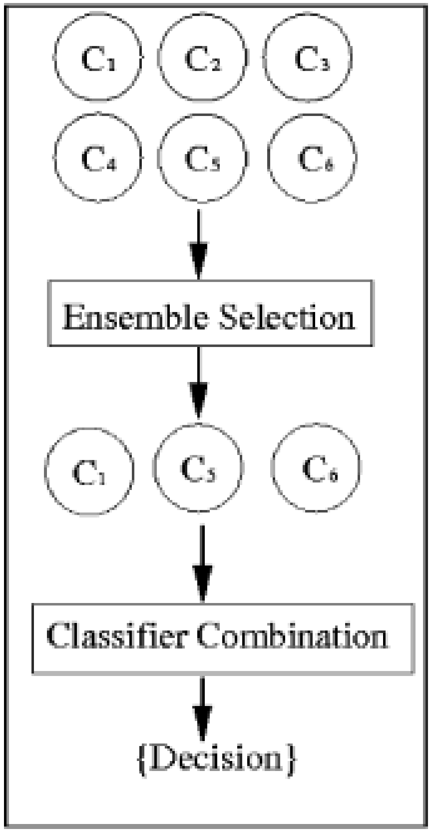
Introduction

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



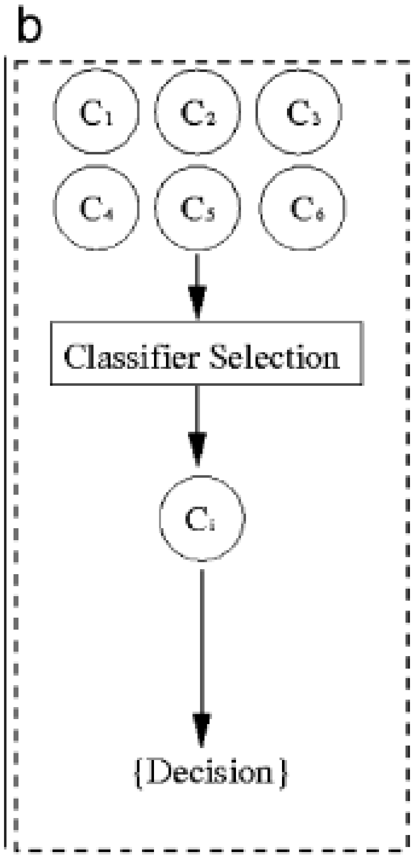
Introduction

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

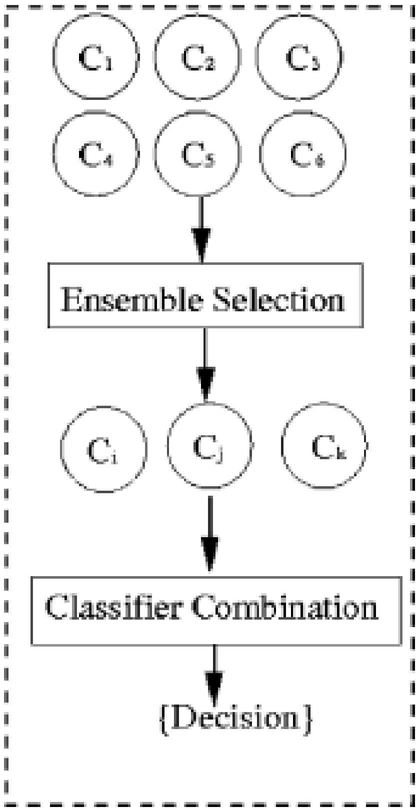


Introduction

c) proposed dynamic ensemble selection

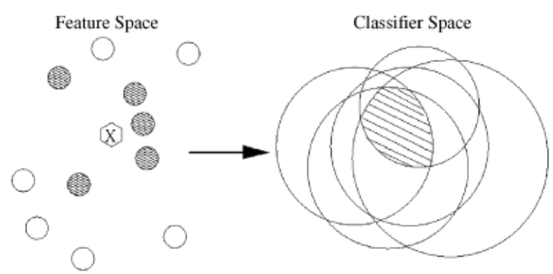
- Dynamic classification selection methods are designed to find the classifier with the greatest possibility of being correct for a sample in a pre-defined neighborhood.
- dynamic ensemble selection is designed to select the most suitable ensemble for each sample.
- Advantage
    - distribute the risk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

c

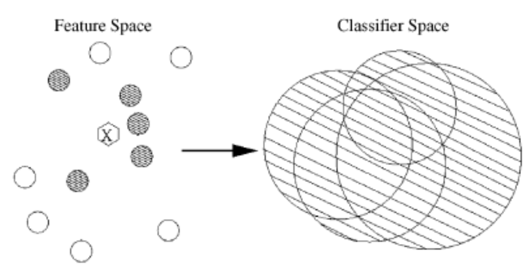


Proposed dynamic ensemble selection - KNORA

- K-nearest-oracles (KNORA)
  - 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
- Schemes
  - KNORA-ELIMINATE



  - KNORA-ELIMINATE-W (vote weighted)
  - KNORA-UNION



  - 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.

Experiments: comparison on UCI repository

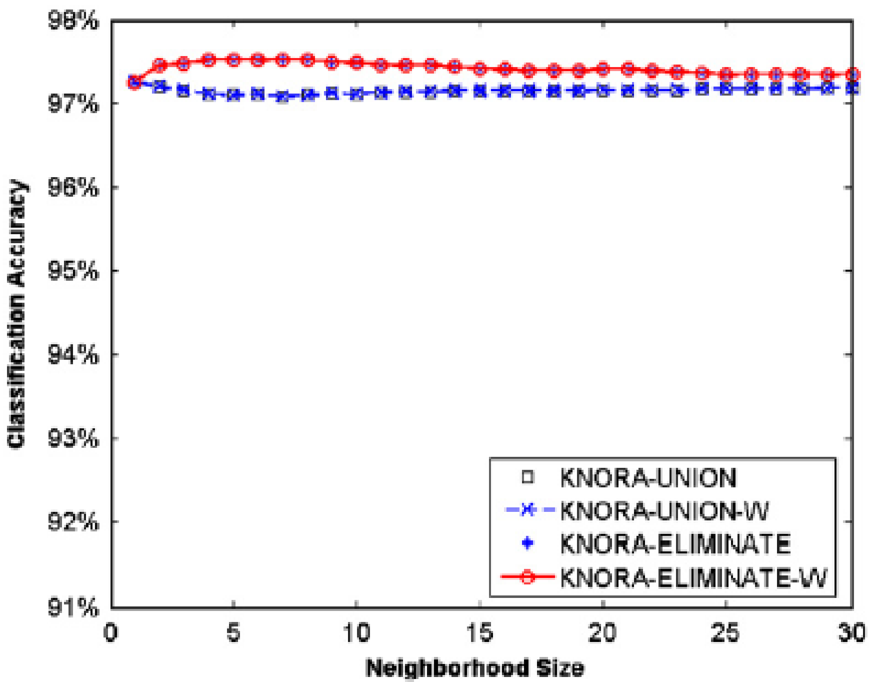
- Random Subspaces
  - KNORA-UNION and LCA have more stable performances than other methods
  - KNORA-UNION-W is not always better than KNORA-UNION
  - KNORA-ELIMINATE-W and KNORA-ELIMINATE have the same performances on Random Subspaces
    - probabilities weighted by the Euclidean distances between 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

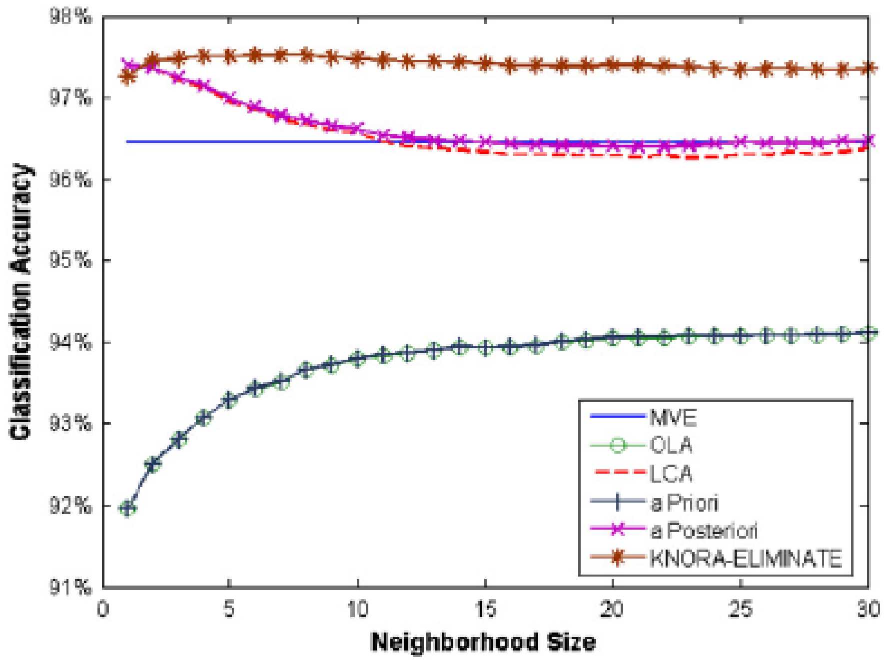
Experiments: comparison on UCI repository

- 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





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

Gracias