Shobeir Fakhraei, Hamid Soltanian-Zadeh, Farshad Fotouhi, Kost Elisevich

Effect of Classifiers in Consensus Feature Ranking for Biomedical Datasets

Wayne State University
Henry Ford Health System

Dimension Reduction

- Prediction accuracy of practical machine learning algorithms **degrades** when faced with many features that are **not necessary** for predicting the desired output.

- Feature Construction / Extraction
  - Construct new features based on the original data
    e.g. PCA and ISOMAP.
- Feature Selection / Ranking
  - Choose features from the original feature set.
    e.g. Filter and Wrapper methods.

WAYNE STATE UNIVERSITY

Henry Ford HEALTH SYSTEM

Feature Selection / Ranking

- Improves the prediction performance.
- Eases understanding of the underlying process that generated the data.
- Reduces measurement and storage requirements.
- Facilitates data visualization.
- Reduces training and utilization times.

Feature Ranking

- The output of the process is a ranked list of features according to a criteria.
- $f_{R1}, f_{R2}, \dots, f_{Rn}$

- Variable ranking is not necessarily used to build predictors:
  - Understanding of the underlying data.
  - e.g. which medical test is more accurate or reliable than the others in a diagnosis.

WAYNE STATE UNIVERSITY

Henry Ford HEALTH SYSTEM

Consensus Feature Ranking

- Ensemble (consensus) methods have been used to mitigate the problems of traditional methods such as poor accuracy, bias, and stability.

- $$FinalScore(f_i) = Combination(score_1(f_i), \ldots, score_n(f_i))$$

Motivation

- $FinalScore(f_i) = Combination(score_1(f_i), \dots, score_n(f_i))$

- $score_i$ is a Single Variable Classifier
- Feature score is the predictive performance of a classifier build based on only that single feature.

Motivation

- $FinalScore(f_i) = Combination(score_1(f_i), \ldots, score_n(f_i))$

- The effect of inclusion of classifiers in the combination (ensemble function) has been studies to see which classifier plays a positive/negative role.
  - Logistic-Regression
  - Support Vector Machines (SVM)
  - K-nearest Neighbors
  - Naïve Bayes
  - Bagging

Wayne State University

Henry Ford Health System

Biomedical Datasets

- When applying Feature Ranking methods on medical datasets, one has to consider the common characteristics of medical datasets:
  - Class-imbalanced data
  - Missing values

Missing Value / Class-Imbalance

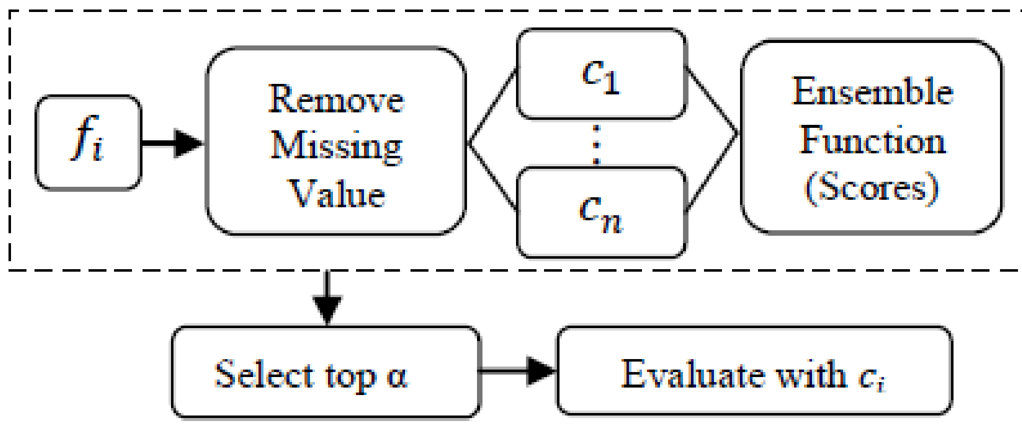
- Missing value estimation and imputation negatively affects the reliability of the model.
- We performed the study only based on properly recorded values and missing values were **eliminated**.
  - Adversely affecting the imbalance distribution
- We used the area under receiver operating characteristic (ROC) curve (AUC) as a performance evaluator for individual features, to address the balance problem.

WAYNE STATE UNIVERSITY

Henry Ford HEALTH SYSTEM

Framework

- Experimental Framework:



Evaluation

- **α features from the top** of the ranked features were selected and the predictive power of this feature subset was tested with a classifier via cross validation.
- To use the maximum possible instances for each feature subset, we used the samples that have all the values for **only the features in the subset** being evaluated.
- The **number of instances varies** for each feature subset, making the comparison of the ranking methods with different feature subsets difficult.

WAYNE STATE UNIVERSITY

Henry Ford HEALTH SYSTEM

Performance Index

- To mitigate the mismatching number of instances.

$$PI(n,c) = \sum_{i=1}^{n}\left(\frac{F_i\_ins}{i}.AUC\left(c(F_i)\right)\right)/\sum_{i=1}^{n}\left(\frac{F_i\_ins}{i}\right)$$

- ***n*** is the number of features considered in the calculation.
- ***c*** is the evaluating classifier.
- F_i is the set of ***i*** features with the highest score
- F_i_ins is the numbers of instances that have all the values for features in F_i.
- ***AUC(c(F_i))*** represents the average AUC of ROC for evaluation of on ***c***, using the leave-one-out technique.

Performance Index

$$PI(n,c) = \sum_{i=1}^{n}\left(\frac{F_i\_ins}{i}.AUC\left(c(F_i)\right)\right)/\sum_{i=1}^{n}\left(\frac{F_i\_ins}{i}\right)$$

- A consideration in this formula is that the ranking methods that achieve a **higher accuracy with fewer features and more instances** are preferable.

- For this reason, the number of features appears in the weight factor as ***1/i*** and the number of instances as ***F_i_ins*** .

Experiments Environment

- The dataset used in the experiments is from Human Brain Image Database System (HBIDS), developed in the Radiology Department of Henry Ford Health System (Detroit, Michigan USA).
- The main task in this dataset is a binary classification that predicts the patients' lateralization (side of abnormality).
- The database contains 197 medical features and 145 patients.

Some features in HBIDS

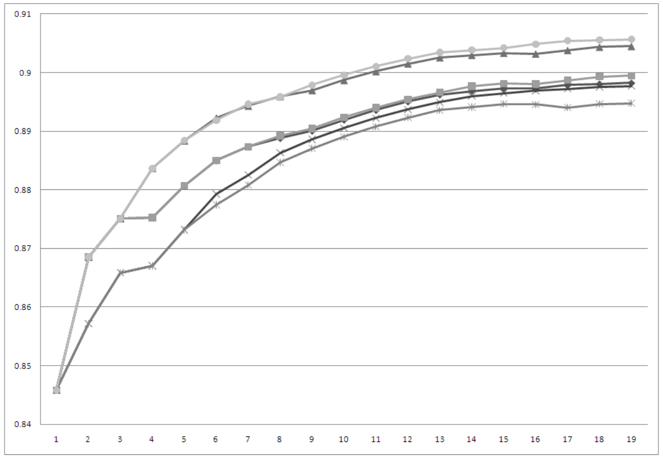
- Semiology,
- Pre- and postoperative neuropsychological profiles
- Location of surgery,
- Surgery outcome according to the Engel classification.
- Interictal waveforms, their location and predominance as well as ictal onset location.
- Both magnetic resonance (MR) and single photon emission computed tomography (SPECT) (ictal and interictal) imaging is included with the provision for quantitative semi-automated assessment of compartmental volume, fluid-attenuated inversion recovery (FLAIR) mean signal and standard deviation and texture analysis
- Compartmentalized ictal SPECT subtraction image analysis is also available.

Wayne State University
Henry Ford Health System

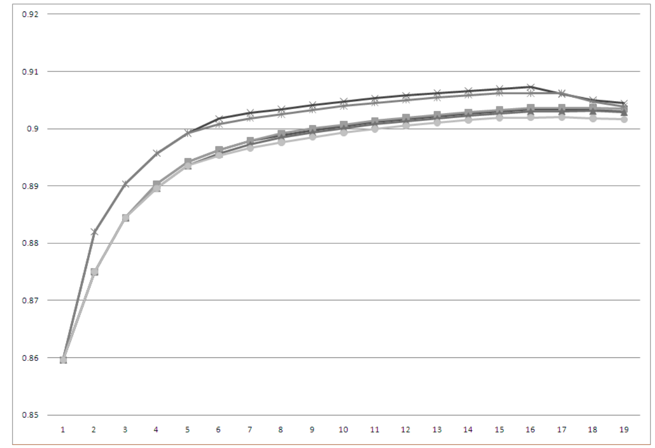
HBIDS Missing Values

- Missing values were identified for:
  - EEG features in 21% of cases
  - Wada studies in 31% of cases
  - Imaging features in 46% of cases
  - The remaining features in about 20% of cases on average.

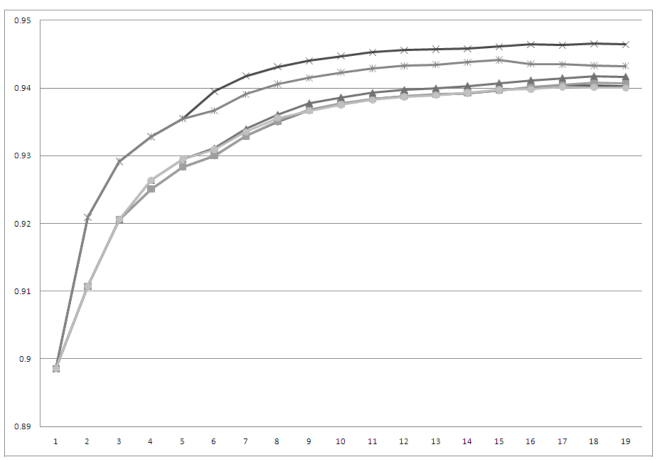
Experimental Results



Evaluation with SVM



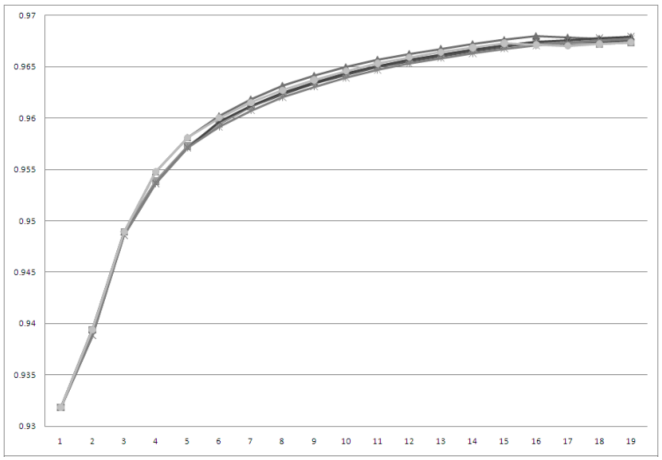
Evaluation on Bagging



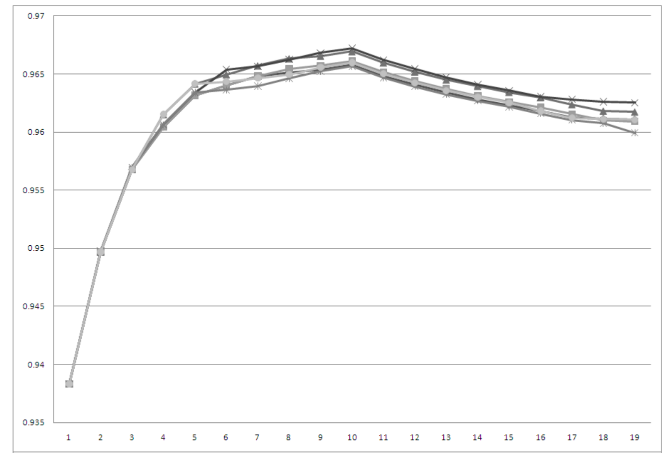
Evaluation on K-Nearest-Neighbors

Consensus Ranking
Excluding SVM
Excluding KNN
Excluding NaiveBayes
Excluding Logistic Reg.
Excluding Bagging

Experimental Results



Evaluation on Logistic-Regression



Evaluation on Naïve-Bayes

Consensus Ranking
Excluding SVM
Excluding KNN
Excluding NaiveBayes
Excluding Logistic Reg.
Excluding Bagging

Observations

- Evaluation with SVM:
  - SVM: Neutral
  - Naïve-Bayes: Positive
  - K-Nearest Neighbors: Negative
  - Bagging: Negative
  - Logistic Regression: Positive
- Evaluation with Bagging:
  - SVM: Neutral
  - Naïve-Bayes: Negative
  - K-Nearest Neighbors: Neutral
  - Bagging: Neutral
  - Logistic Regression: Negative
- Evaluation with K-NN:
  - SVM: Neutral
  - Naïve-Bayes: Negative
  - K-Nearest Neighbors: Neutral
  - Bagging: Neutral
  - Logistic Regression: Negative
- Evaluation with Naïve-Bayes:
  - SVM: Neutral
  - Naïve-Bayes: Negative
  - K-Nearest Neighbors: Negative
  - Bagging: Neutral
  - Logistic Regression: Neutral

Observations

- Performance of the consensus feature ranking with a classifier is **not highly dependent on inclusion of that classifier** itself in the fusion.
- Therefore, features ranked based on ensemble of scores from multiple classifiers are likely to perform well on **unseen classifiers**.
- This ranking plays an important role in **data-warehousing**, where data are gathered with the possibility to be used with **new emerging classifiers** in the future.

Refrences

1. Y. Saeys, I. Inza, P. Larranaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23, p. 2507, 2007.
2. I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-82, 2003.

WAYNE STATE UNIVERSITY

Henry Ford HEALTH SYSTEM

Thank you

If you are interested to get more details about this research please contact

Shobeir Fakhraei {shobeir@wayne.com}