Flexible-Hybrid Sequential Search in Feature Selection

Petr Somol and Jana Novovičová

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Prague, Czech Republic

http://ro.utia.cas.cz

Data – Algorithms – Decision Making Int. Workshop
Třešť, December 9 - 12, 2006

Outline

Wrapper, Filter, Hybrid

FS methods broadly fall into three categories:

- Filter Methods:
  - Select feature subset according to a reasonable criterion.
  - The criterion is independent of the learning method.
- Wrapper Methods:
  - Require a predetermined learning algorithm instead of an independent criterion for subset evaluation.
  - Select the subset of features using the performance of the learning algorithm as the evaluation criterion.
- Hybrid Methods:
  - Combinations of the two above.

Motivation for Hybrid Methods 1/2

Comparison: Wrappers vs Filters:

- Wrappers and Filters methods are ultimately heuristic because of the combinatorial barrier.
- Wrappers tend to give superior performance as they find features better suited to the predetermined learning algorithm.
- Filters are general preprocessing algorithms: do not rely on any knowledge of the learning algorithm to be used.
- Wrappers are very time consuming: for each subset of features try to solve the learning problem
- The main advantage of filter methods is their speed and ability to scale to large data sets.

Motivation for Hybrid Methods 2/2

Hybrid Methods:

- Make use of both an independent criteria and a learning algorithm to evaluate feature subsets.
- To goal is to achieve Wrapper-like results in Filter-like time.

Floating Search is suitable for hybridization because:

- It is usable both as Filter and Wrapper.
- Performs well on broad range of problems.

Basic Notion

- Let $Y = \{f_i | i = 1, \ldots, D\}$ be the set of all available D measurements
- Let $X_d = \{f_j | j = 1, \ldots, d; f_j \in Y\}$ be a subset of d features, where $d < D$ and possibly $d << D$
- Let $J(X_d)$ denote the corresponding criterion value
- Let $\mathrm{ADD}(X_d)$ be the operation of adding such feature f^+ to the working set X_d of d features to obtain X_{d+1}, that

$$f^+ = \arg \max_{f \in Y \setminus X_d} J(X_d \cup \{f\}).$$

- Let $\mathrm{REMOVE}(X_d)$ be the operation of removing such feature f^- from the working set X_d to obtain set X_{d-1}, that

$$f^- = \arg \max_{f \in X_d} J(X_d \setminus \{f\}).$$

Basic SFFS Principle

The basic (Forward) Floating Search principle summarized:

1. Initialization: Starting from empty set X_0 call Step 2 twice to obtain set X_2 and $d = 2$.
2. ADD(X_d). $d = d + 1$.
3. Repeat REMOVE(X_d), $d = d - 1$ as long as it improves solutions already known for the lower d and $d > 1$.
4. If $d < D$ go to 2.

hSFFS – the Hybrid Part

In hSFFS the operation ADD() is modified as follows:

When looking for the feature to add, first pre-select c_k^+ most promising candidates by maximizing $J_{Filter}(\cdot)$, then decide according to the best $J_{Wrapper}(\cdot)$ value, i.e.:

$$c_k^+ = \max\{1, \lfloor \lambda(D-k) \rfloor\} \tag{1}$$

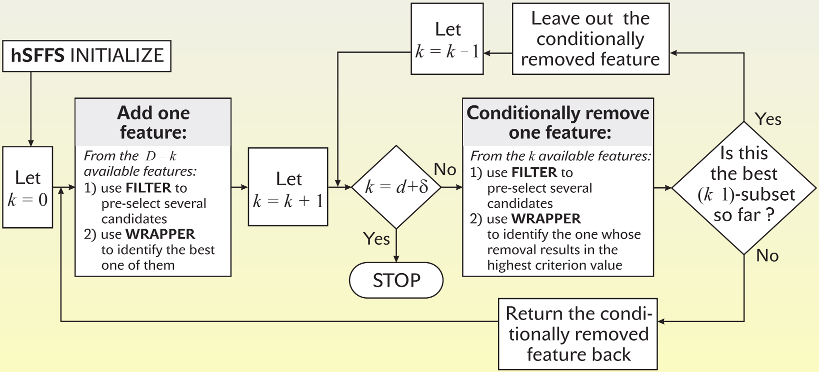
$$C_k^+ = \{x_{i_t}, t = 1, \ldots, c_k^+ : J_{Filter}(X_k \cup \{x_{i_t}\}) \geq J_{Filter}(X_k \cup \{x_j\}) \ \forall j \neq i_t\} \tag{2}$$

$$x^+ = \arg\max_{x \in C_k^+} J_{Wrapper}(X_k \cup \{x\}), \quad X_{k+1} = X_k \cup \{x^+\}. \tag{3}$$

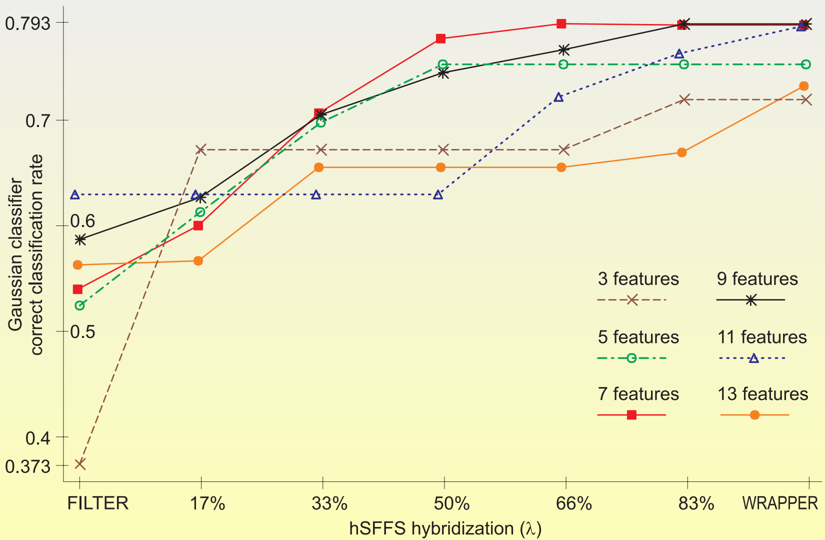
Note: $\lambda \in \langle 0, 1 \rangle$ is a user parameter. λ values closer to 0 (resp. 1) cause more Filter-like (resp. Wrapper-like) hSFFS behaviour.

Remark: REMOVE() is to be modified analogously.

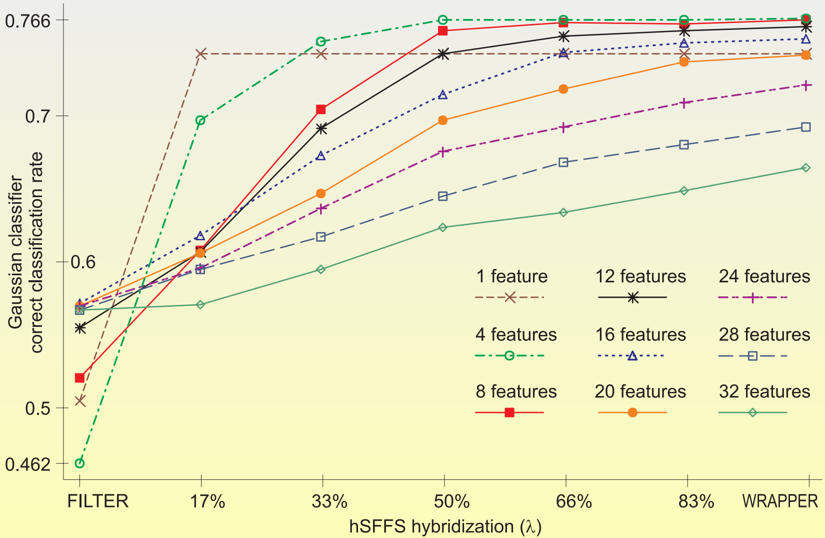
hSFFS Simplified Diagram



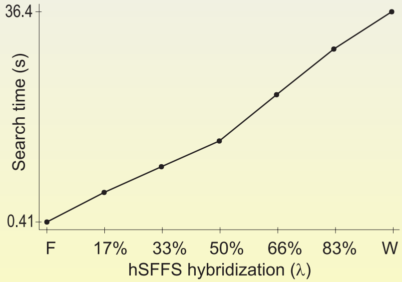
hSFFS Performance - SPEECH dataset



hSFFS Performance - WAVEFORM dataset



hSFFS Performance - Search Time



Conclusions

We have presented:

- Flexible hybrid version of floating search methods for feature selection that
  - can deal flexibly with the quality-of-result vs. computational time trade-off and to enable wrapper based feature selection in problems of higher dimensionality than before;
  - it is possible to trade significant reduction of search time for often negligible decrease of the classification accuracy.

Future Work

In the future we intend to :

- "Hybridize" other search methods in a similar way as presented here.
- To investigate in detail the hybrid behavior of different combinations of various probabilistic measures and learning methods.

Thank you for your attention !