Explicit Control of Feature Relevance and Selection Stability Through Pareto Optimality

Victor Hamer and Pierre Dupont

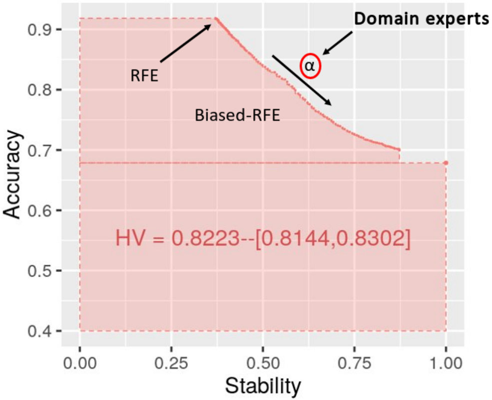
Universite catholique de Louvain

victor.hamer@uclouvain.be

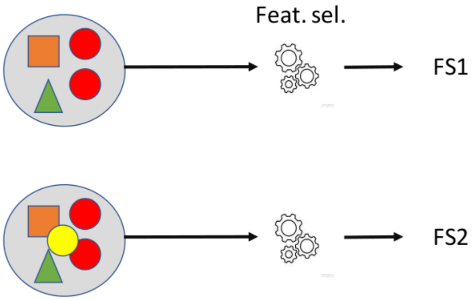
September 16, 2019

Overview

- What is feature selection stability and why is instability a problem ?
- State of the literature
- Contribution: explicit compromise between accuracy and stability



What is feature selection stability ?



$$|FS1 \cap FS2| \approx 0 \Rightarrow \text{stab} \downarrow$$

Instability

- Features can be analyzed by **experts** to gain domain knowledge.
- Instability reduces the interpretability of the predictive models.
- And the trust of domain **experts** towards the selected features.

State of the literature

Increasing stability

- Ensemble feature selection [Saeys et al., 2008, Abeel et al., 2010]
- Instance weighting [Somol and Novovicova, 2010]
- Model selection

$\Rightarrow$ No fine control of the accuracy-stability trade-off.

Stability measure [Nogueira et al., 2017]

$$\phi = 1 - \frac{\frac{1}{d}\sum_{f=1}^{d} p_f(1-p_f)}{\frac{k}{d} * (1-\frac{k}{d})} \qquad \begin{cases} d : \text{number of input features} \\ k : \text{mean number of selected features} \\ p_f : \text{feature f selection frequency} \end{cases}$$

Logistic RFE

$$L = \sum_{i=1}^{n} \log(1 + e^{-y_i * (\mathbf{w} * \mathbf{x_i})}) + \lambda ||\mathbf{w}||_2$$

- Drops a fraction of the least significant features at each step
- Until the desired number of features is met

A feature f with a lower β_f has a higher probability to be selected and vice-versa $\Rightarrow$ control the accuracy-stability tradeoff by tuning β.

Biased logistic RFE

$$L = \sum_{i=1}^{n} \log(1 + e^{-y_i * (\mathbf{w} * \mathbf{x_i})}) + \lambda \beta ||\mathbf{w}||_2$$

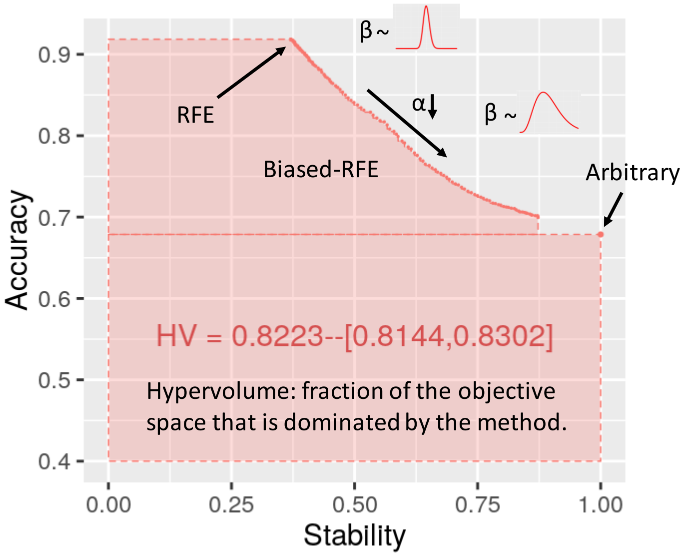
- Drops a fraction of the least significant features at each step
- Until the desired number of features is met

A feature f with a lower β_f has a higher probability to be selected and vice-versa $\Rightarrow$ control the accuracy-stability tradeoff by tuning β.

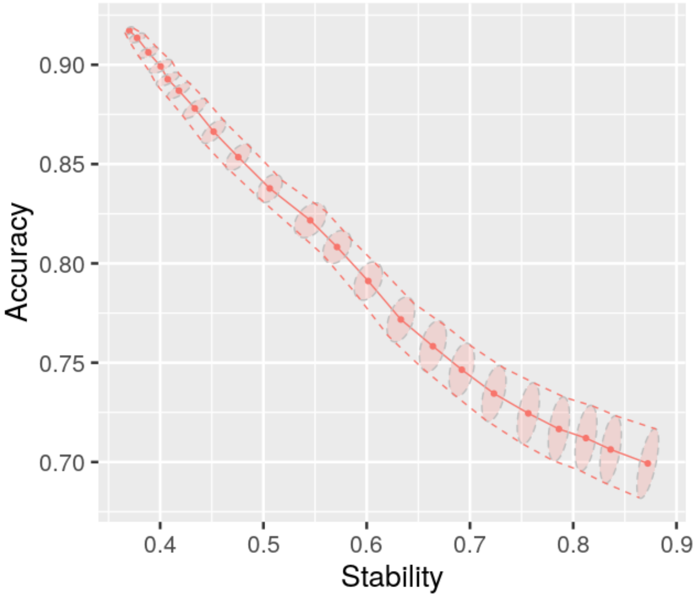
Results (prostate dataset, d=12600, n=102)

Paper

$\beta_f \sim \Gamma(\alpha, 1)$

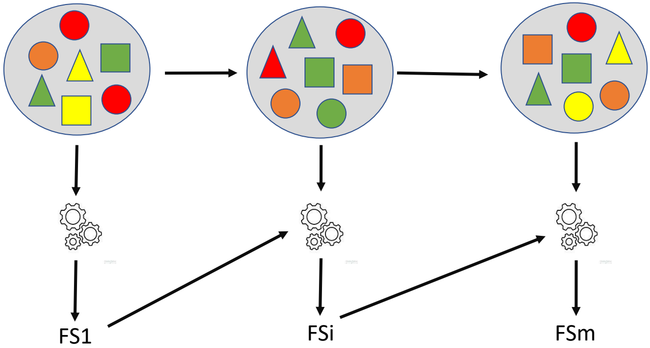


Confidence intervals



Transfer learning

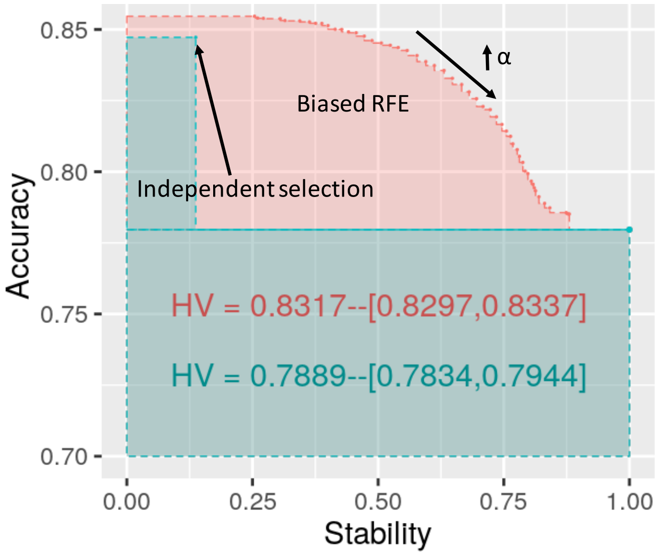
Sometimes, one wants to find similar feature subsets for different tasks.



Paper weighting scheme

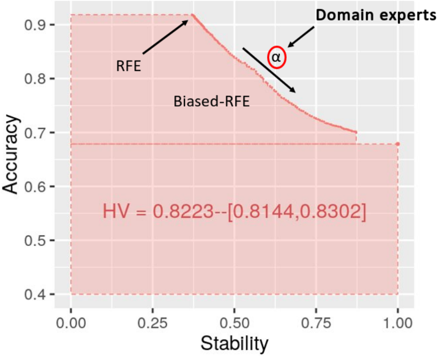
Stability increase if f is taken: $2p_f - 1 \Rightarrow \beta_f \propto \exp(-\alpha * p_f)$

Transfer learning: results



Conclusion

Domain experts can tune the accuracy-stability tradeoff at will.



Future work

- Extension to multi-task selection.
- Apply differential shrinkage to other losses or regularizations (Elastic Net penalty, deep feature selectors, ...).

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., and Saeys, Y. (2010).
Robust biomarker identification for cancer diagnosis with ensemble feature selection methods.
Bioinformatics, 26(3):392–398.

Nogueira, S., Sechidis, K., and Brown, G. (2017).
On the stability of feature selection algorithms.
The Journal of Machine Learning Research, 18(1):6345–6398.

Saeys, Y., Abeel, T., and Van de Peer, Y. (2008).
Robust feature selection using ensemble feature selection techniques.
In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 313–325. Springer.

Somol, P. and Novovicova, J. (2010).
Evaluating stability and comparing output of feature selectors that optimize feature subset cardinality.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(11):1921–1939.