FRI - Feature Relevance Intervals for Interpretable and Interactive Data Exploration

Lukas Pfannschmidt, Christina Göpfert, Ursula Neumann, Dominik Heider, Barbara Hammer

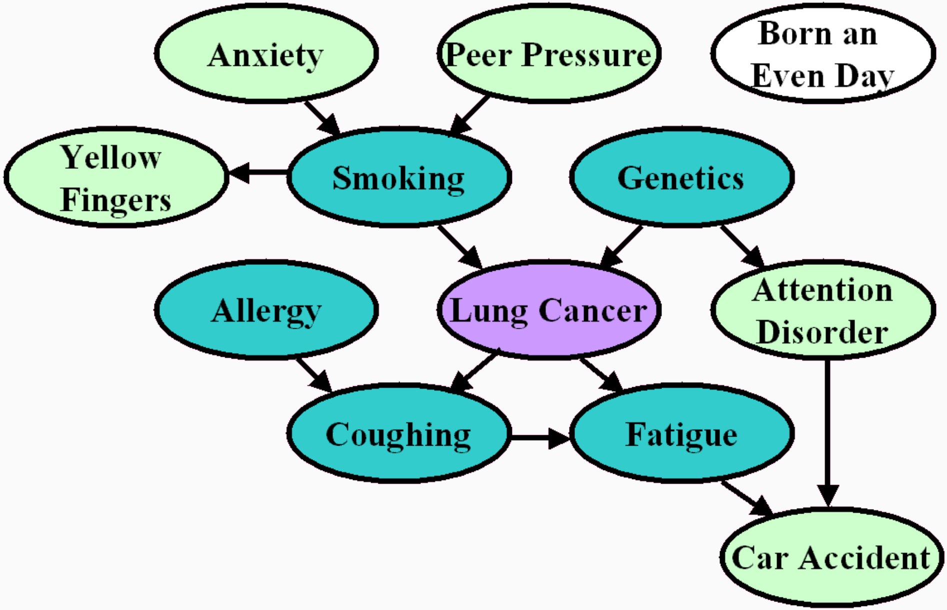
CIBCB19, 10.07.2019

UNIVERSITÄT BIELEFELD

CITEC

Overview

1. Motivation
2. Feature Relevance Intervals
3. This Contribution
4. Evaluation


Anxiety
Peer Pressure
Born an Even Day
Yellow Fingers
Smoking
Genetics
Allergy
Lung Cancer
Attention Disorder
Coughing
Fatigue
Car Accident

Variables / Features in Biomedical Research

Data sources

- serum quantities (chemical composition)
- imaging metrics (shape, size, color)
- DNA sequencing data
- sociodemographic
- ...

Common Paradigm

Machine Learning (in Bioinformatics)

1. preprocessing and feature selection
2. choose supervised learning model
3. use training data to fit model
4. measure performance on validation data

Goals:

- prediction
- understand unknown process by interpreting <u>featureset</u>

Feature Selection

Motivation

- gain insight
- reduce model complexity
- reduce cost
- reduce invasiveness of sample collection for patients

Existing Methods

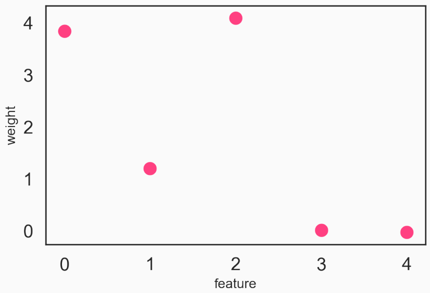
- Wrapper
  - Exhaustive (NP-hard)
  - Greedy (e.g. RFE)
- Embedded (Lasso)
- Filter

Example for Embedded Approach

Linear Model

$$\vec{y} = \vec{w}^T \mathbf{X}$$
$$\mathcal{L}(\vec{w}, \vec{\xi}) = \frac{1}{2}\|\vec{w}\|_1 + \sum_i \xi_i$$

Loss function with sparsity constraint



$$\vec{w} = [3.844, 1.201, 4.096, 0.013, 0.005]$$

All-Relevant Feature Selection

Most existing feature selection approaches (by design) do <u>not</u> give a complete and truthful representation of a features true relevance.

- minimum redundancy is often enforced (and wanted)
- most subsets only represent one of many feasible feature sets

Definition 1

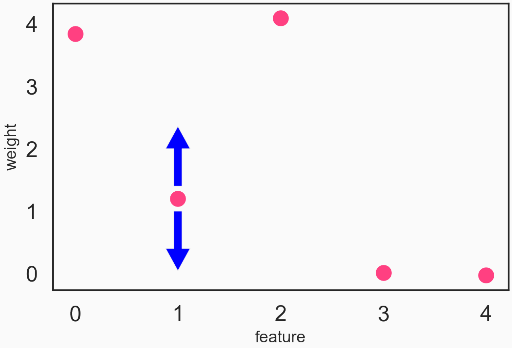
Relevancy classes (Kohavi et al., 1997):

- strongly relevant
- **weakly relevant**
- irrelevant

Overview

Feature Relevance Bounds: Extend to all feasible feature contributions

- Find each features **maximal** and **minimal** use with similar performance
- Based on linear SVM solution
- Computable using LPs



Feature classification

Irrelevant

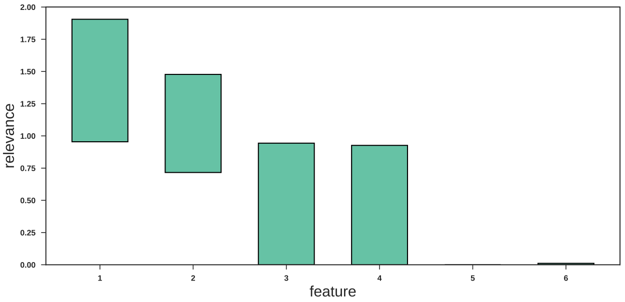
Lower bound $= 0$, Upper bound ≈ 0

Strongly relevant

Lower bound > 0

Weakly relevant

Lower bound $= 0$, Upper bound > 0



Overview

From Method to Usable Implementation

Goal was accessible Python library.

Aspects:

1. handle numerical instabilities (LP solvers)
2. ability to allow user input
3. performance

Aspect 1: Estimate Feature Classification Threshold

Irrelevant

Lower bound $= 0$, Upper bound ≈ 0

Strongly relevant

Lower bound > 0

Weakly relevant

Lower bound $= 0$, Upper bound > 0

Numerical inaccuracies lead to fuzzy values.

Solution: estimate data based threshold

1. Generate probes by permuting real features
2. For each feature i, compute relevance bounds of its probe i_p while excluding i itself
3. Determine threshold according to the distribution of probe relevances

Aspect 2: Interactive Model Manipulation

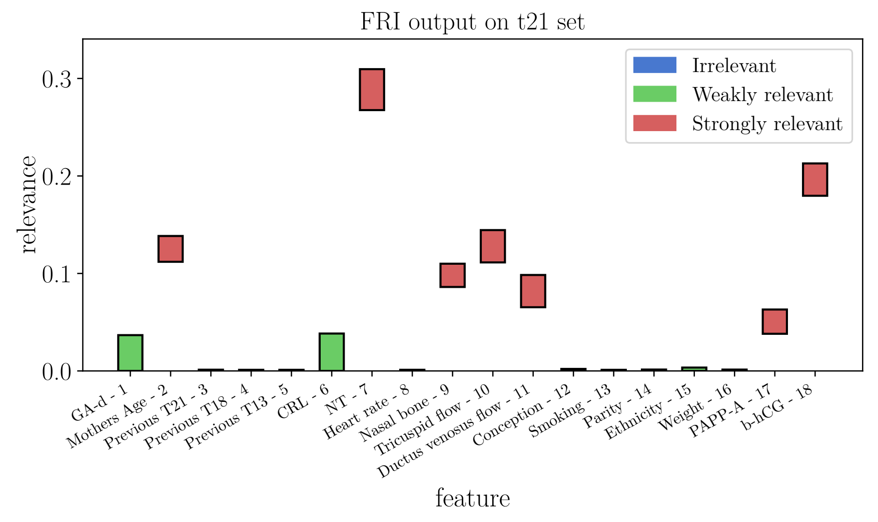
Goals:

- allow user to check own hypotheses by experimenting with the model
- facilitate search for alternative features
- reveal feature dependencies

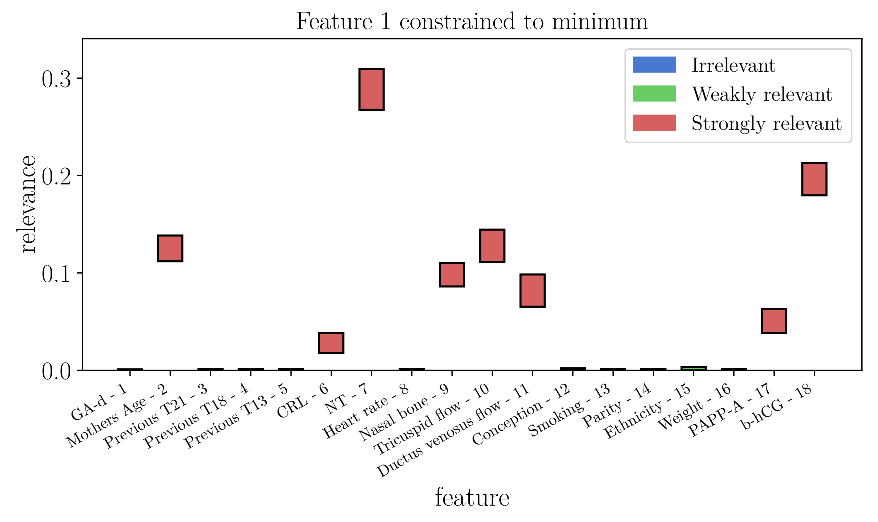
Solution:

- LPs allow simple addition of feature constraints
- simple user defined values based on relevance intervals
- direct connection between visualization and manipulation -> intuitive

Unconstrained



One Feature Disabled



Aspect 3: Parallel Processing

Interactive workflow requires fast results:

- relevance bounds can be computed independently per feature
- structure program for parallel processing (Joblib library)
- use frameworks for distributed computing

-> ability to use all CPUs on machine or nodes in cluster (Dask framework)

Overview

How to evaluate feature selection performance?

Compare selected features sets per method:

- Are all relevant features included?
- Is the featureset compact?
- Is it computationally feasible?

Experimental setup:

- Test on data with known ground truth (toy sets)
- Test on real biomedical data
- Repeat tests over 50 bootstrap iterations

Toy Sets

	Strongly relevant	Weakly relevant	Irrelevant
Sim1	4	4	22
Sim2	12	8	10
Sim3	4	0	26
Sim4	18	0	12
Sim5	0	20	10

Related work: all-relevant feature selection

- **Elastic Net**: weighted sum of L1 and L2 regularization scheme, preserves redundancies
  (Zou and Hastie, 2004)
- **Boruta**: wrapper around Random Forest, using random contrast variables and statistical tests
  (Kursa and Rudnicki, 2010)
- **Ensemble Feature Selection**: combination of multiple other feature relevancy scores
  (Neumann et al., 2016)
- **Stability Selection**: aggregation of multiple noisy bootstrap samples of original data to compute stability score
  (Meinshausen and Bühlmann, 2010)

Model training accuracy

	Boruta	EFS	ElasticNet	FRI	SS
Sim1	0.99	-	1.00	0.92	1.00
Sim2	0.97	-	1.00	0.96	1.00
Sim3	0.99	-	1.00	0.96	1.00
Sim4	0.97	-	1.00	0.93	1.00
Sim5	1.00	-	1.00	0.91	1.00
colp.	1.00	-	0.99	0.97	0.99
flip	1.00	-	0.90	0.82	0.90
spectf	1.00	-	0.99	0.92	0.98
t21	1.00	-	0.98	0.93	0.98
wbc	1.00	-	1.00	0.98	1.00

Performance of Feature Sets used in Logistic Regression Model

	Boruta	EFS	EN	FRI	SS
colposcopy	0.56	0.58	0.64	**0.66**	0.62
flip	0.80	0.65	**0.81**	0.74	0.70
spectf	0.87	0.87	0.86	**0.88**	**0.88**
t21	0.97	0.97	0.97	0.97	0.97
wbc	0.99	0.99	0.99	0.99	0.99

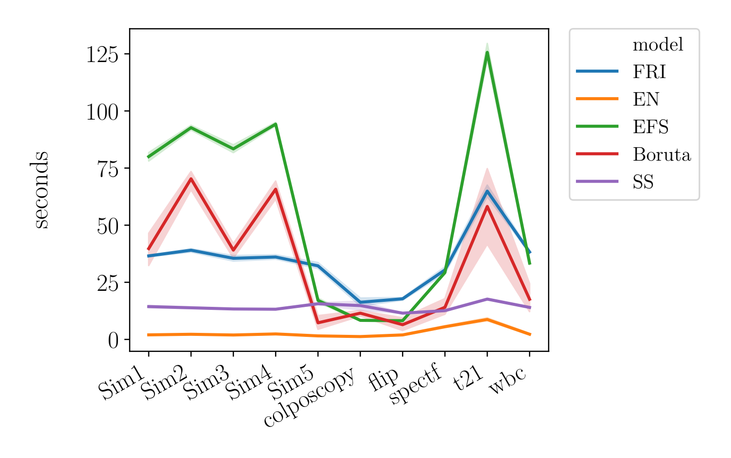
All-Relevant Feature Selection Sensitivity

score	F1				
data	Sim1	Sim2	Sim3	Sim4	Sim5
Boruta	**0.98**	0.82	0.91	0.82	0.98
EFS	0.96	0.76	0.71	0.84	0.94
ElasticNet	0.62	0.84	0.44	0.82	0.80
FRI	**0.98**	**0.98**	0.99	**0.99**	**0.99**
StabilitySelection	0.77	0.75	**1.00**	0.91	0.27

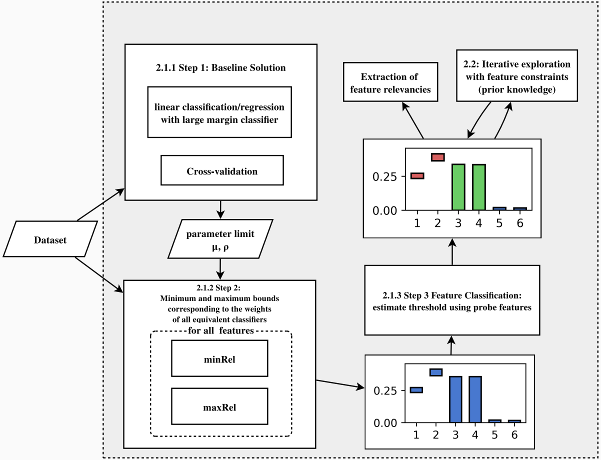
Average selected feature set size

	Boruta	EFS	EN	SS	FRI	FRI_s	FRI_w
Sim1	8.1	8.7	17.8	5.0	8.1	5.1	3.0
Sim2	14.3	12.3	26.6	12.1	19.4	12.4	7.0
Sim3	4.6	7.2	14.8	4.0	4.1	4.0	0.1
Sim4	12.6	13.2	26.2	15.0	17.9	17.9	0.0
Sim5	19.1	17.9	29.7	3.2	19.9	0.0	19.9
colp.	35.1	25.4	46.5	41.5	20.3	5.9	14.4
flip	18.8	8.1	16.9	9.1	8.9	8.8	0.1
spectf	44.0	20.3	43.1	5.9	19.9	5.9	14.0
t21	15.5	7.9	14.2	9.6	9.6	6.6	3.0
wbc	29.9	12.5	26.9	4.7	15.6	4.0	11.6

Single Thread Runtime Comparison



FRI Python Library



- available as Python library
  $ pip install fri
  or
  github.com/lpfann/fri
- batch processing API
- interactive workflow functions

Conclusion

Feature selection

- we conserve all relevant features
- sparse and interpretable
- competitive performance

Interactive exploration

- intuitive way to manipulate the model
- visual feedback

Thank you for your attention!

Lower relevance bound

$$
\text{minRel}((x_i, y_i)_{i=1}^n, j) : \min_{\omega, b, \xi} |\omega_j|
$$

$$
\begin{aligned}
&\text{s.t.} \\
&\quad y_i(\omega^\top x_i - b) \geq 1 - \xi_i \\
&\xi_i \geq 0 \\
&\sum_{i=1}^n \xi_i \leq \rho \\
&\|\omega\|_1 \leq \mu.
\end{aligned}
$$

ρ and μ are the upper limits from an initial baseline L_1 model.

Manipulate the Model

Adding constraints

$$
\begin{aligned}
\text{minRel}\mathbf{C}(\mathbb{D}, j, \text{fc}) : \min_{\omega,b,\xi} &\ |\omega_j| \\
\text{s.t.} &\\
&\ y_i(\omega^\top x_i - b) \geq 1 - \xi_i \\
&\ \xi_i \geq 0, \\
&\ \sum_{i=1}^{n} \xi_i \leq \rho \\
&\ \|\omega\|_1 \leq \mu. \\
&\ \text{fc}^k_{min} \geq |\omega_k| \geq \text{fc}^k_{max}, \forall k \neq j
\end{aligned}
$$

All-Relevant Feature Selection Sensitivity - Precision/Recall

score	precis					recall			
data	Sim1	Sim2	Sim3	Sim4	Sim5	Sim1	Sim2	Sim3	Sim4
Boruta	0.99	1.00	0.87	1.00	1.00	1.00	0.72	0.98	0.70
EFS	0.93	1.00	0.57	1.00	1.00	1.00	0.62	0.98	0.73
ElasticNet	0.46	0.74	0.28	0.69	0.67	1.00	0.98	1.00	1.00
FRI	0.98	1.00	0.98	0.99	1.00	0.99	0.97	1.00	0.98
StabilitySelection	1.00	1.00	1.00	1.00	1.00	0.62	0.60	1.00	0.83