Decision Snippet Features

Pascal Welke, Fouad Alkhoury, Christian Bauckhage, Stefan Wrobel

ICPR 2021

UNIVERSITÄT BONN

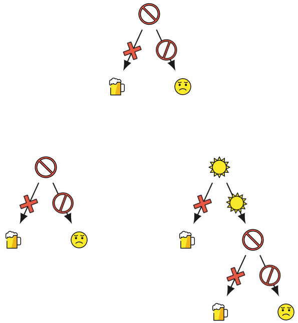
A Small Dataset

lockdown	rain	cold	sunny	drink outside
❌	🌧	❄	❌	😟
❌	❌	❌	☀	🍺
🚫	❌	❌	❌	😟
❌	🌧	❌	❌	😟
❌	❌	❌	❌	🍺
🚫	❌	❄	❌	😟
🚫	❌	❄	☀	😟

Basically, I drink outside whenever there is no lockdown and it is not raining.
We see only a random training subset, so an algorithm might come to a different conclusion.

Motivation

- Decision trees are great
  - interpretable by humans
  - fast to train and apply
  - tend to overfit
- Ensembles (i.e. *Random Forests*) reduce variance
  - larger model size
  - less interpretable (due to larger size)
- How can we retain the benefits of random forests and decision trees?
  - the trees in a random forest are not independent
  - arguably, common structures might result from the underlying learning problem

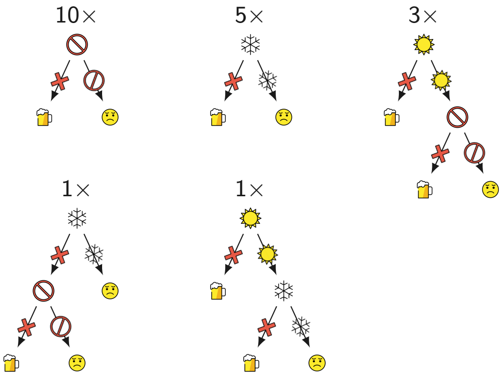


Let's learn from random forests to identify a relevant smaller trained random forest

A Random Forest on Our Dataset

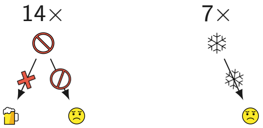
lockdown	rain	cold	sunny	drink outside
✗	🌧	❄	✗	☹
✗	✗	✗	☀	🍺
🚫	✗	✗	✗	☹
✗	🌧	✗	✗	☹
✗	✗	✗	✗	🍺
🚫	✗	❄	✗	☹
🚫	✗	❄	☀	☹

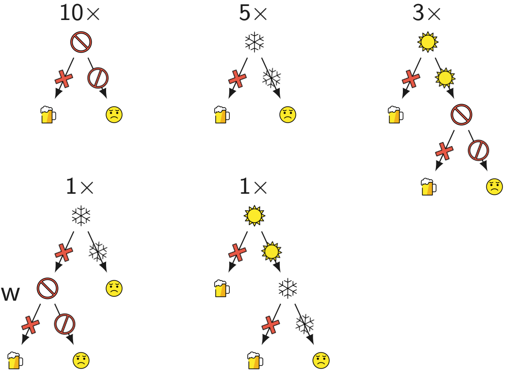
- Let's train a random forest with 20 trees on this training data



Let's Look at these Trees

- Three trees are found multiple times
- Substructures occur even more frequently

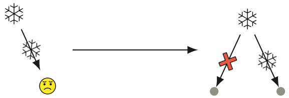




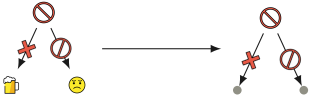
We will use frequent subtrees to build new (smaller) ensemble models.

Technical Issues

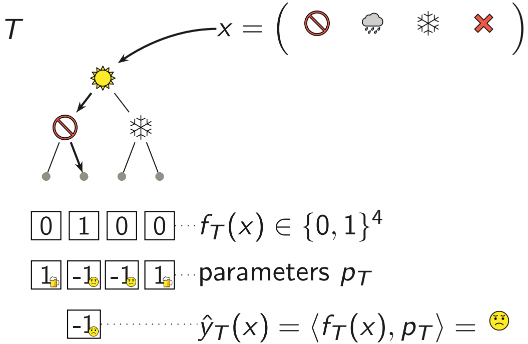
- Substructures may be incomplete
  - We need to add leaves



- Substructures see different data
  - We cannot use the leaf labels



Random Forests are Representation Learning + Linear Model



$f_T(x) \in \{0,1\}^4$

parameters p_T

$\hat{y}_T(x) = \langle f_T(x), p_T \rangle =$ 😟

Decision Snippet Features
Training Process

1. Train Random Forest on Data
2. Mine Decision Snippets
3. Transform Data to Decision Snippet Feature space
4. Train a linear classifier

Conclusion

- Decision Snippet Features are based on regularities in random forests
- They work well
  - Size reductions up to orders of magnitude
  - comparable predictive performance

Check out our paper!

☀ 🍺