European Network on New Sensing Technologies for Air Pollution Control and Environmental Sustainability - *EuNetAir*

COST Action TD1105

WGs and MC Meeting at ISTANBUL, 3-5 December 2014

Action Start date: 01/07/2012 - Action End date: 30/06/2016

Year 3: 1 July 2014 - 30 June 2015 (*Ongoing Action*)

AUTOMATED COMPUTATIONAL INTELLIGENCE MODEL SELECTION FOR SENSOR DATA

Roman Neruda, P. Vidnerova, V. Kurkova

Institute of Computer Science

Academy of Sciences of the Czech Republic

Prague, CZ, roman@cs.cas.cz

COST is supported by the EU Framework Programme

ESF provides the COST Office through a European Commission contract

Beyond creating a model from data

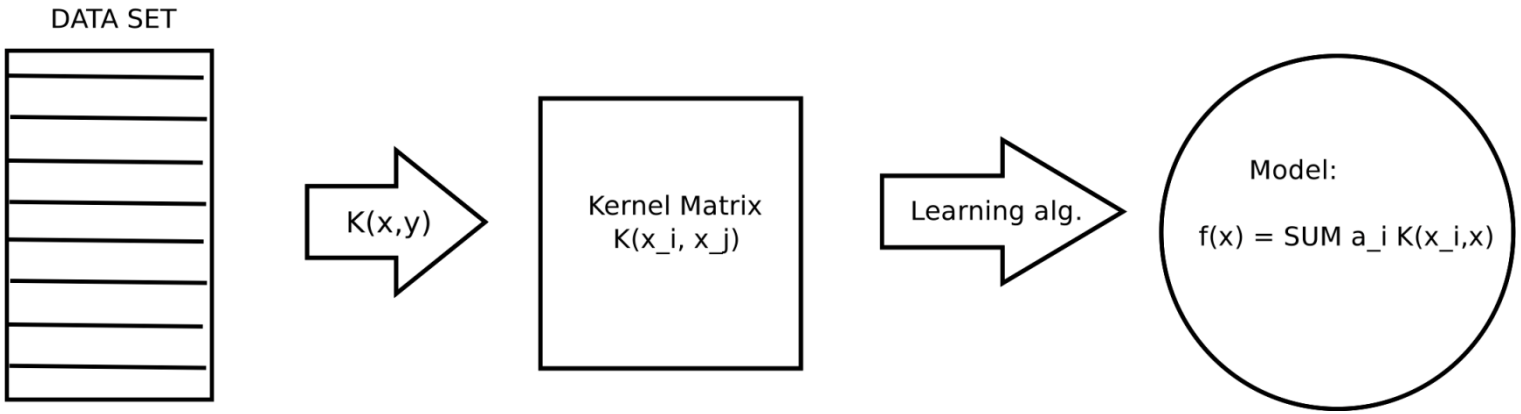
- Numerous data-mining methods have been proposed during the recent decades, and most of them can be used to build model of various datasets.
- Naturally, the performance of different methods is different for each dataset.
- No-free lunch theorem: there is no method which would perform the best on all possible datasets.
- **The need to select the best-performing method for the problem at hand.**

Meta-learning

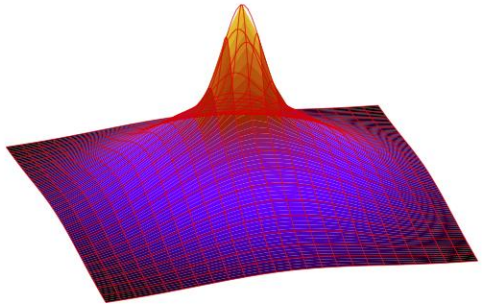
- To augment, or even replace, human expertise in solving the problem of selection of a suitable model (with parameters) for a given data:
  - **method recommendation**,
  - **parameter selection**.
- “Learning to learn”, utilizing:
  - analysis of **meta-data** = dataset properties,
  - **history** of previous computations,
  - sophisticated **search** algorithms in the parameter space of the models.

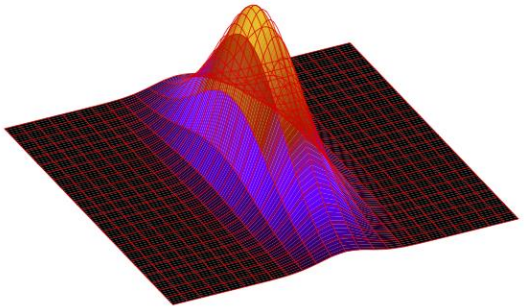
Local (kernel) models

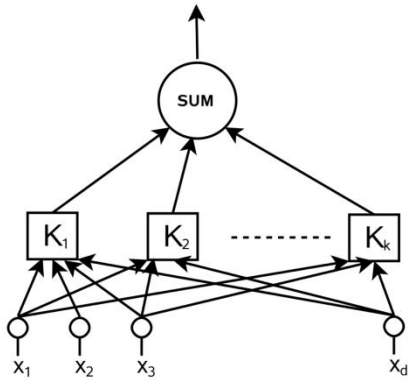
- **Radial Basis Function** network – sum of Gaussians
- **Regularization** networks – sum of other kernels
- Combination of kernels:
  - **Product** kernels
  - Sum/**linear combination** kernels

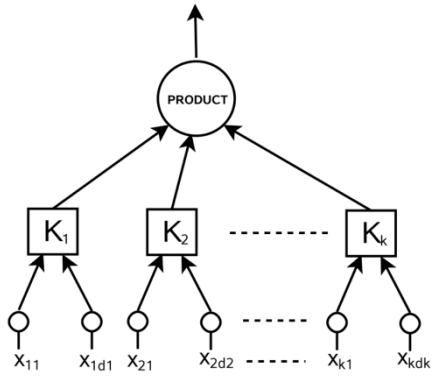


Sum and product kernels

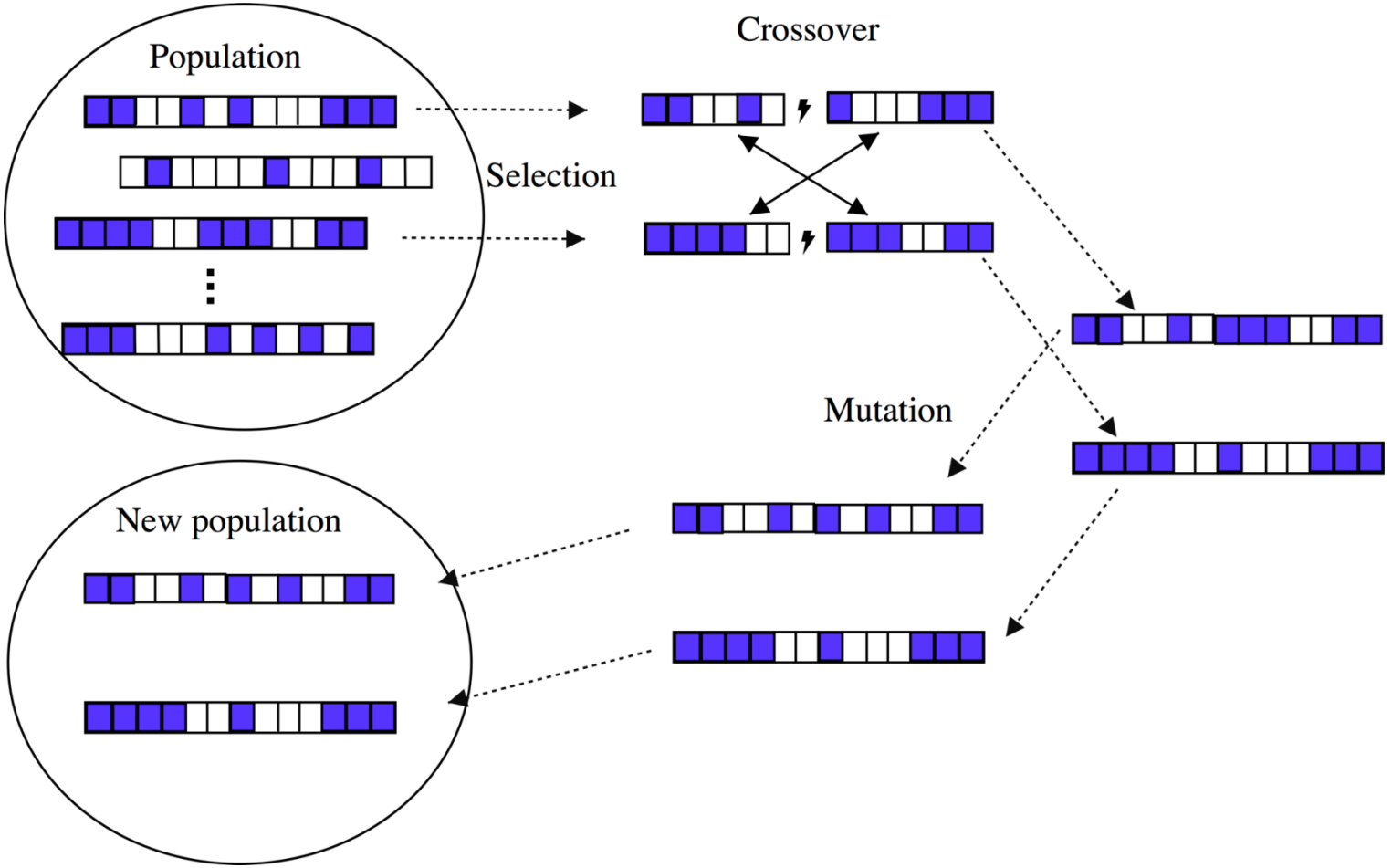




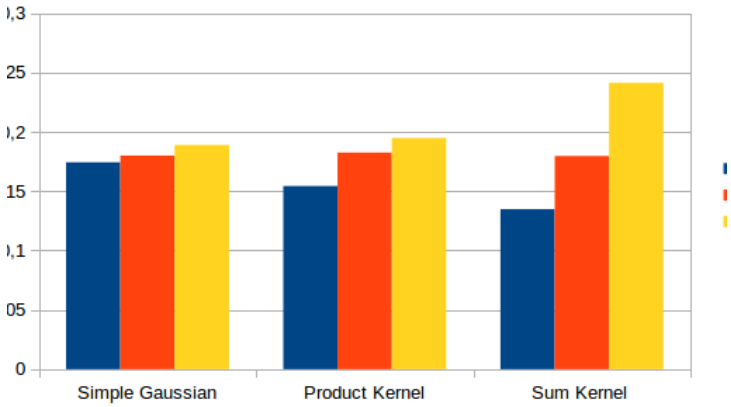




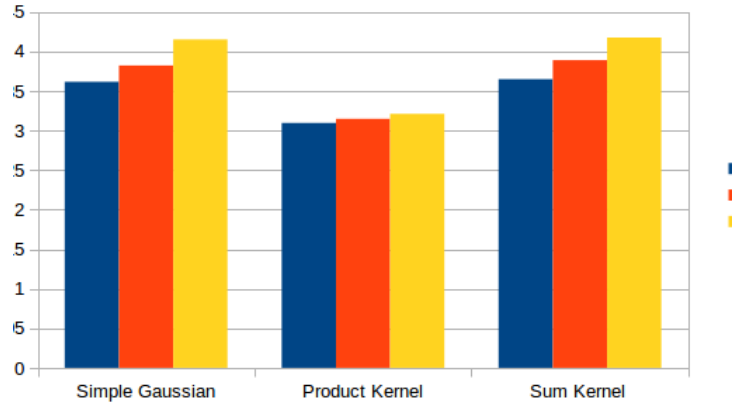
Evolutionary search for parameters



Evolved RNN configurations

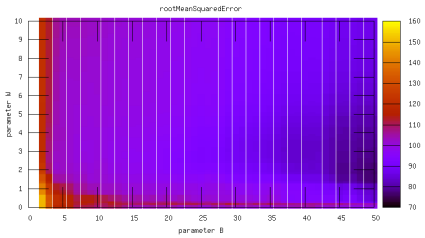


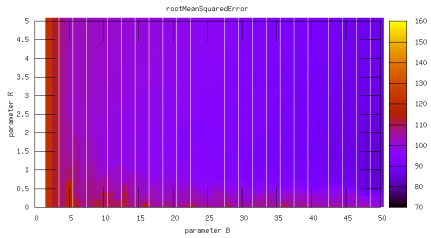
Training data only

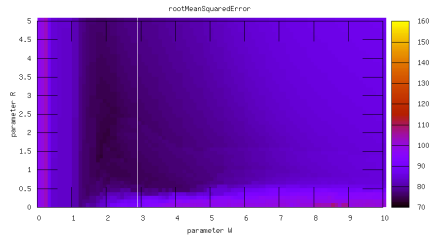


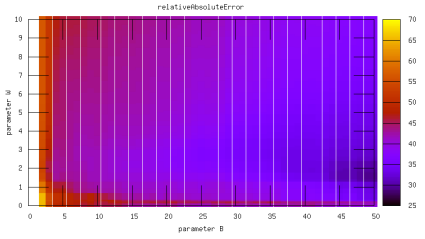
10-fold cross-validation

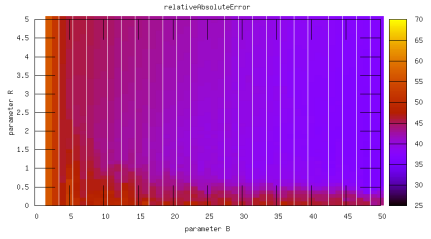
Parameter space exploration

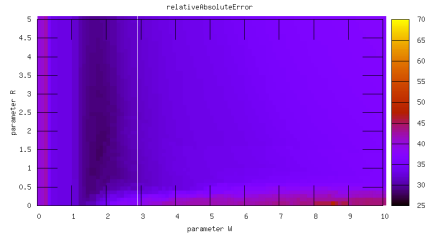


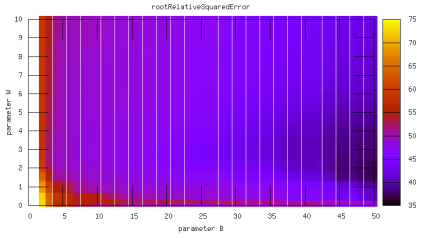


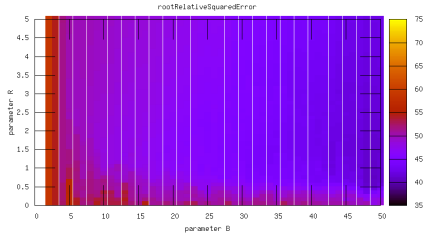


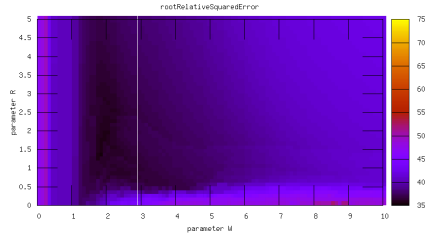












Conclusion

- Sensor data are typically **homogenous** in terms of meta-data indicators, the parameter search and recommendation works rather reliably.
- The quality of method recommendation would improve with **more experiments** utilizing at least dozens of data sets, as well as more methods.
- Need for **methodologies** how to measure model success for specific domains
- Beyond model recommendation?
  - Recommending **complex hybrid** solutions
  - Preprocessing, ensambles, ...

The big picture

Solution of BellKor's Pragmatic Chaos

All developed CF models

BRISMF SVD-Time SBRAMF Split RBM BK3 BK2 MF1 NSVDD RBM day FRBM BK5-SVD++ Movie KNN V Baseline 1/2/3 DRBM SVD++ GTE KNN+time NSVD1 Integrated M. RBM MF2 SVD-AUF Movie KNN CTD/MTD SVDNN User KNN Classif. ModeKNN 1...5 Asym. 1/2/3

Latent User and Movie Features

approx. 500 predictors

Probe Blending

Probe Blending

200 blends

30 blends

Linear Blend 10.09 % improvement

Michael Jahrer / Andreas Töscher – Team BigChaos – September 21, 2009


Thank you
roman@cs.cas.cz