

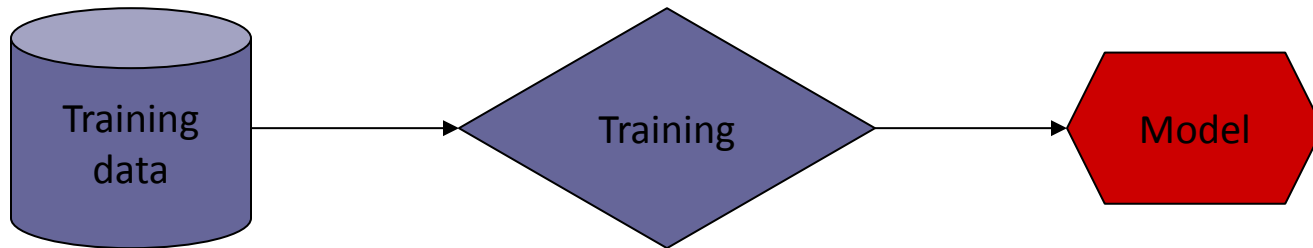
Adaptive Machine Learning

Prof. Dr. Eirini Ntoutsi

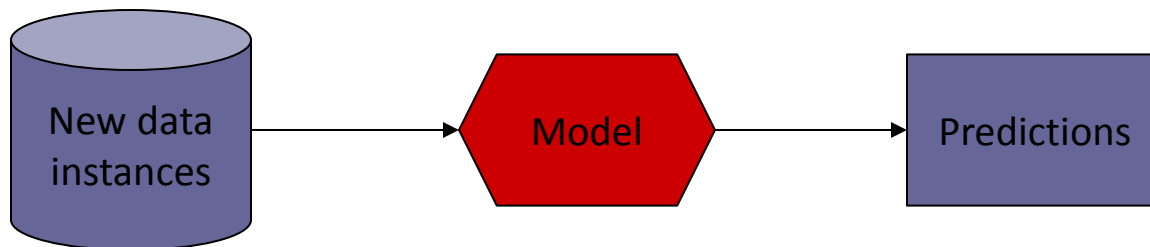
FG Intelligent Systems
Faculty of Electrical Engineering and Computer Science
Leibniz University Hannover

Typical Machine Learning process

- Training a model



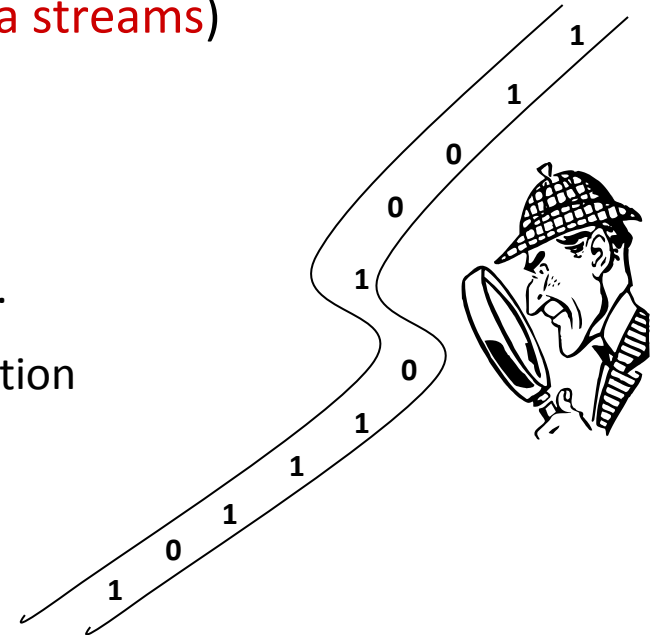
- Testing a model



This is batch/static learning!

Modern data are non-static

- Continuous, massive and rapid data flow (**data streams**)
- Many example applications
 - Production processes: predictive maintenance
 - Network monitoring: early discovery of attacks.
 - Social streams: Sentiment analysis, spam detection
- Such data call for
 - new learning methods
 - new evaluation measures
 - new learning tasks
 - ...



What do we do

- Our research lies in the areas of data mining and machine learning, over complex data and data streams.
- We focus on three areas
 - ML over non-stationary data / Stream mining
 - Stability of models and features
 - Improving data quality for ML

What do we do: Stream mining methods

- Stream Mining methods
 - Stream clustering
 - Stream classification
 - Outlier detection
 - Forgetting data/models
 - Informed/ blind adaptation
 - Adaptive feature space
 - Feature drifts
- Application areas
 - Social streams (sentiment analysis, spam detection, topic classification)
 - Industrial streams: production streams, energy streams
 - Security: network intrusion streams

How can we maintain valid models over the stream?

What do we do: Stream mining methods

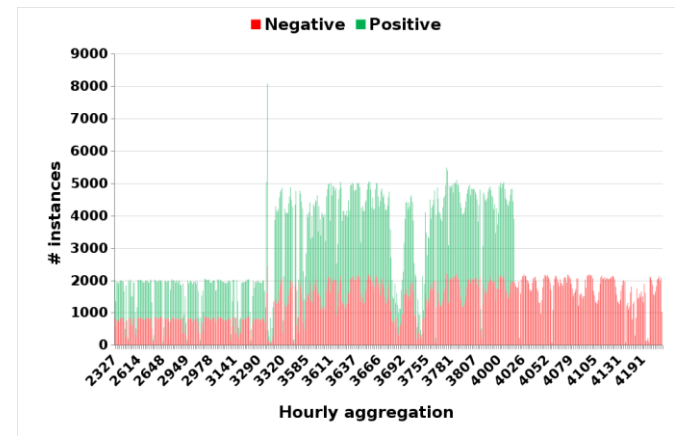
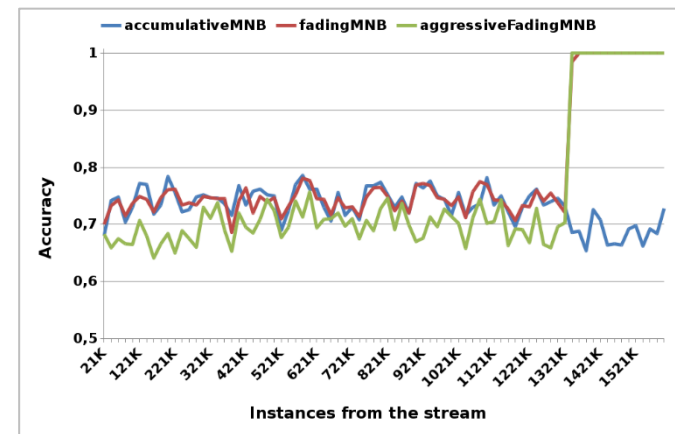
■ Stream Mining methods

- Stream clustering
- Stream classification
- Outlier detection
- Forgetting data/models
- Informed/ blind adaptation
- Adaptive feature space
 - Feature drifts

■ Application areas

- Social streams (sentiment analysis, spam detection, topic classification)
- Industrial streams: production streams, energy streams
- Security: network intrusion streams

Example: Sentiment Analysis



What do we do: Stream mining methods

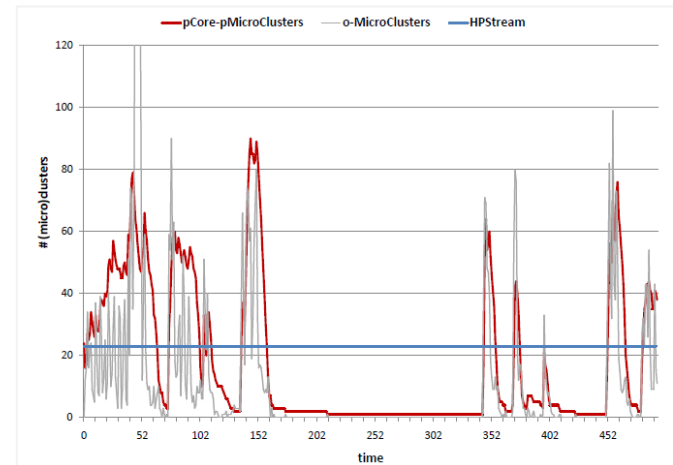
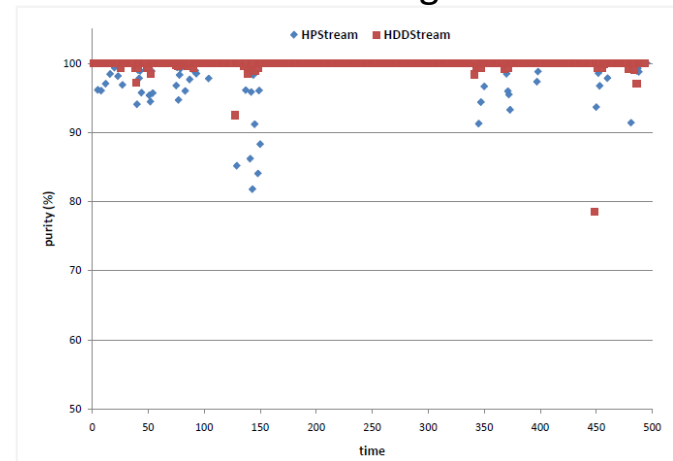
■ Stream Mining methods

- Stream clustering
- Stream classification
- Outlier detection
- Forgetting data/models
- Informed/ blind adaptation
- Adaptive feature space
 - Feature drifts

■ Application areas

- Social streams (sentiment analysis, spam detection, topic classification)
- Industrial streams: production streams, energy streams
- Security: network intrusion streams

Example: Network Intrusion monitoring



What do we do: Stability analysis methods

- Stability analysis
 - Cluster(-ing) stability
 - Classification stability
 - Feature space stability
 - Feature importance

How stable our models are?

- Application areas
 - Bioarchaelogy
 - Textual streams
 - Numerical streams
 - ...

What do we do: Stability analysis methods

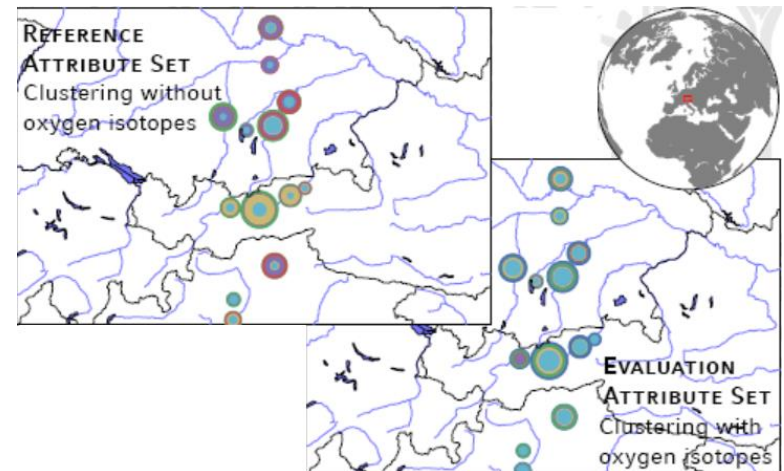
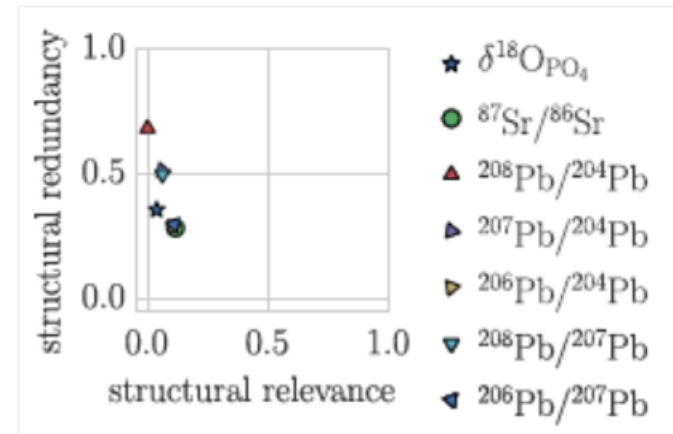
■ Stability analysis

- Cluster(-ing) stability
- Classification stability
- Feature space stability
- Feature importance

■ Application areas

- Bioarchaeology
- Textual streams
- Numerical streams
- ...

Example: Bioarchaeology

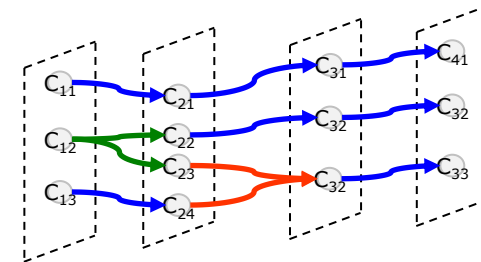


What do we do: Stability analysis methods

■ Stability analysis

- Cluster(-ing) stability
- Classification stability
- Feature space stability
- Feature importance

Example: Textual streams



■ Application areas

- Bioarchaeology
- Textual streams
- Numerical streams
- ...

τ	1999	2000	2001	2002	2003	2004
0.45	4	7	7	1	5	4
0.50	4	5	7	1	3	4
0.55	3	3	3	0	2	3
0.60	3	2	3	0	1	1
0.65	3	0	1	0	0	1
0.70	2	0	1	0	0	0

What do we do: Improving data quality for ML

How to improve our data for learning?

- Improving data quality for ML
 - dealing with label scarcity
 - dealing with class-imbalance
 - dealing with biased data
- Methods
 - semi-supervised learning
 - data augmentation for machine learning
 -
- Application areas
 - Energy streams
 - ...

What do we do: Improving data quality for ML

■ Improving data quality for ML

- dealing with label scarcity
- dealing with class-imbalance
- dealing with biased data

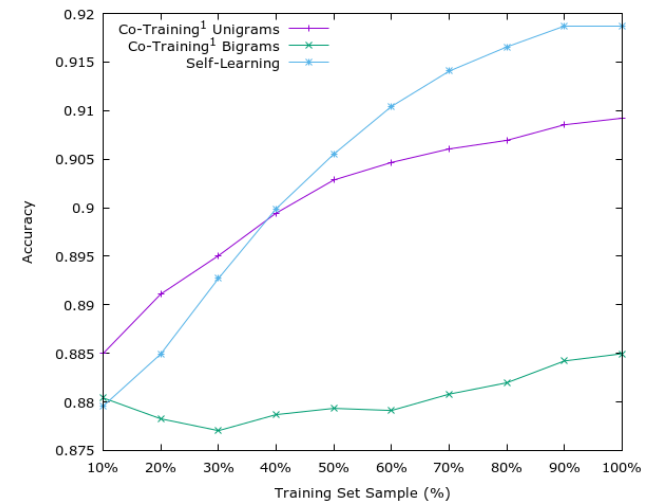
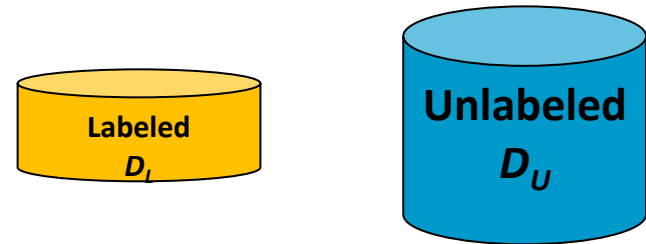
■ Methods

- semi-supervised learning
- data augmentation for machine learning
-

■ Application areas

- Energy data
- text
- ...

Example: Textual streams



What do we do: Improving data quality for ML

■ Improving data quality for ML

- dealing with label scarcity
- dealing with class-imbalance
- dealing with biased data

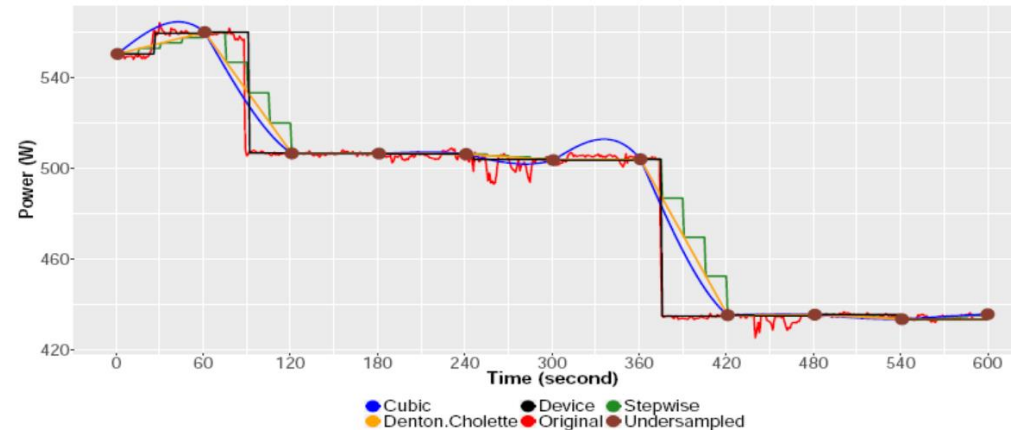
■ Methods

- semi-supervised learning
- data augmentation for machine learning
-

■ Application areas

- Energy data
- text
- ...

Example: Energy disaggregation



Thank you for your attention!

Questions? Ideas?

Contact:

<http://www.kbs.uni-hannover.de/~ntoutsi/>

Email: ntoutsi@l3s.de