LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Lehrstuhl für Datenbanksysteme und Data Mining

BAMA Information Event

Information event for all students interested in a bachelor or master thesis

Summer Term 2022

Agenda

1. **Introduction**
   Application for Bachelor and Master Thesis

2. **General Information**
   On Structuring, Grading and General Recommendations

3. **Current Opportunities**
   Brief Overview of Topics and Research Areas

4. **Q&A**

BAMA = Bachelor / Master Thesis

Agenda

1. **Introduction**
   Application for Bachelor and Master Thesis

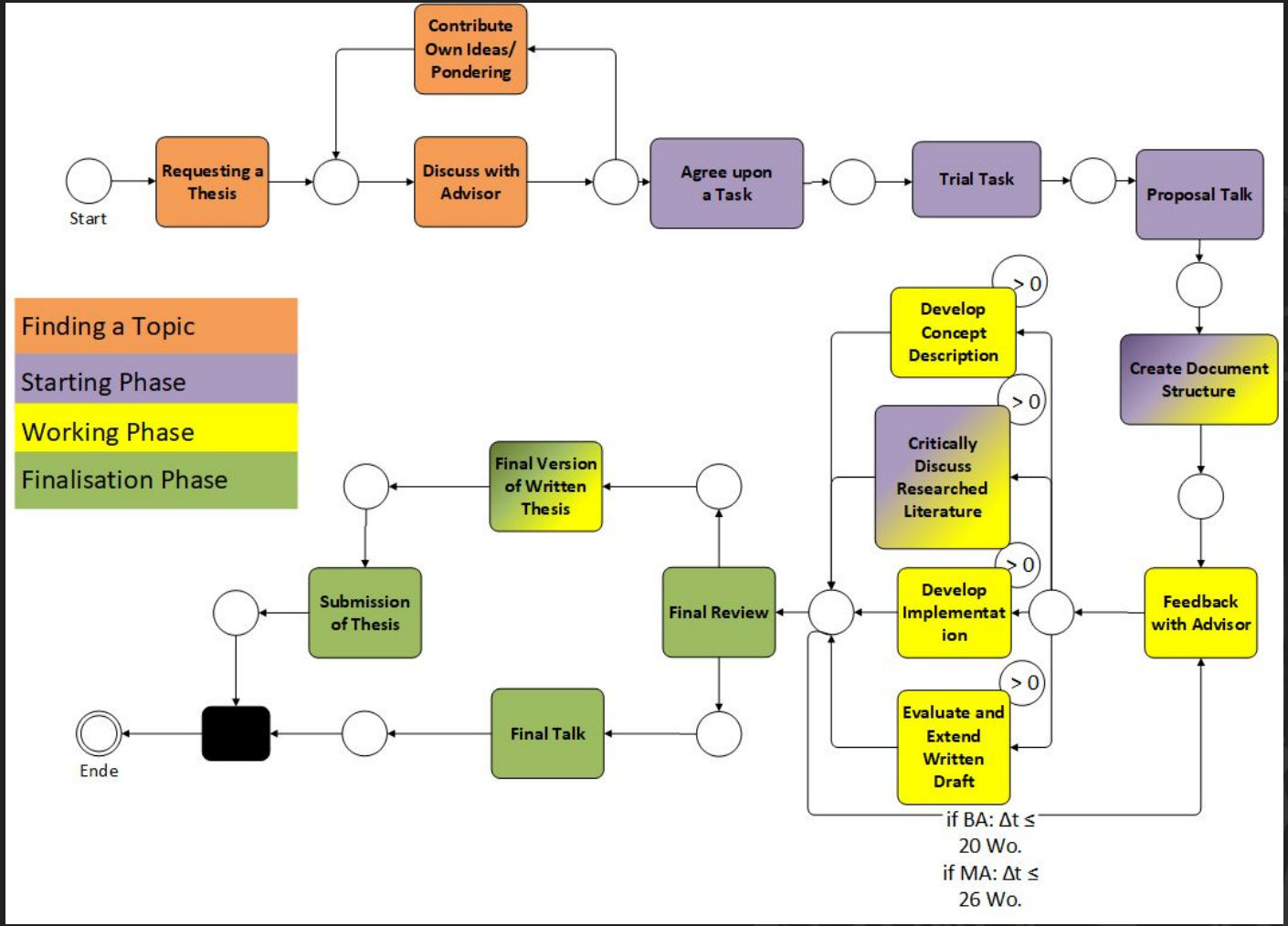
2. **General Information**
   On Structuring, Grading and General Recommendations

3. **Current Opportunities**
   Brief Overview of Topics and Research Areas

4. **Q&A**

BAMA = Bachelor / Master Thesis

Possible Phases of a Thesis



Agenda

1. **Introduction**
   Application for Bachelor and Master Thesis

2. **General Information**
   On Structuring, Grading and General Recommendations

3. **Current Opportunities**
   Brief Overview of Topics and Research Areas

4. **Q&A**

BAMA = Bachelor / Master Thesis

Examination Regulations

“The [Bachelor | Master] thesis shall reveal that the student is able to elaborate

- on a problem from their subject of study
- within a determined period
- autonomously
- by scientific methods”

There are three roles

- you (student) – is responsible for conducting solid research work
- tutor (assistant) – is responsible to help in defining the scope
- examiner (professor) – is responsible overall supervision and grading

Typical structure and expected content

1. Introduction	Motivation, problem statement, objectives, scope, outline
2. Related Work	Literature, state of the art, limitations, differentiations
3. Concept, Methods	Basic idea, preliminaries, ingredients, formalization, derivations, illustrations, proofs, heuristics, analyses
4. Experiments	Practical evaluation, comparison to competitors, parameter sensitivities, discussion of results, limitations
5. Conclusions	Summary, lessons learned, potential for future work

Grading Criteria: (A) Textual Elaboration

1. The thesis is well structured and the overall impression is convincing.
2. Problem statement and research objectives are clearly described. The motivation is engaging.
3. Literature review and discrimination from the state of the art is conclusive.
4. Basic idea, prerequisites and ingredients of the concepts are well elaborated.
5. Formalizations, derivations, and analyses are sound and cover relevant and important aspects.
6. Examples and illustrations are designed coherently and provide appropriate intuition.
7. The text is well phrased and comprehensible, it gets to the heart of the investigations and reveals profound insights and lessons learned.
8. There are no formal deficiencies.

Grading Criteria: (B) Practical Investigations

1. The experiments are clearly and expressively designed.
2. The evaluations cover meaningful real and/or synthetic datasets.
3. The evaluation criteria fit to the research objectives and illustrate the degree of goal attainment with respect to quality and/or efficiency.
4. Meaningful sensitivity analyses cover suitable variations of data (numerosity, dimension, etc.) and methods (parameters, heuristics, design alternatives).
5. The experiments include comparisons with alternative methods and the state of the art in particular.
6. Effects, successes, and limitations are well presented and informatively analysed and discussed.

Grading Criteria: (C) Contributions and Autonomy

1. The research objectives are met and valuable insights are generated.
2. The scope of the thesis is reasonably narrowed down and reaches an appropriate depth.
3. Playgrounds for the development and investigation of variants and alternative design decisions are chosen carefully and yield beneficial insights.
4. Emerging but blowing opportunities are sketched for future work.
5. The developments and investigations are original contributions and were carried out in high autonomy.
6. Discussions with the tutor on scope, directions and results of the thesis were well prepared, efficiently conducted, and profitably incorporated.

Recommendation for Writing

- Proceed bottom-up: start with collection of material, then derive structure.
- Write and formalize early, that allows for maturation.
- Iterate in three phases:

1. Sketch structure	Collect material, create formal document, shape storyline, phrase key statements, complete thesis in rough sentences
2. Elaborate phrasings	Adjust weighting of sections, formulate smooth statements and transitions, complete formal parts (bibliography, declarations, etc.)
3. Polish document	Final revisions, ready for submission

Recommendation for Success

- What we grade is your submission and the disputation
  - not anything around (code, conversations, etc.)
- Limit the scope of your thesis
  - Focus on core problem, don’t try to cover a too broad field
  - Consider deferring additional ideas to future work
  - Stay in touch with your advisor on any question about the scope
- Substantiate discussions of design decisions
  - Investigate variants in depth at selected opportunities
  - Trade-offs are great, elaborate on their effects in regular as well as in degenerate cases

Agenda

1. **Introduction**
   Application for Bachelor and Master Thesis

2. **General Information**
   On Structuring, Grading and General Recommendations

3. **Current Opportunities**
   Brief Overview of Topics and Research Areas

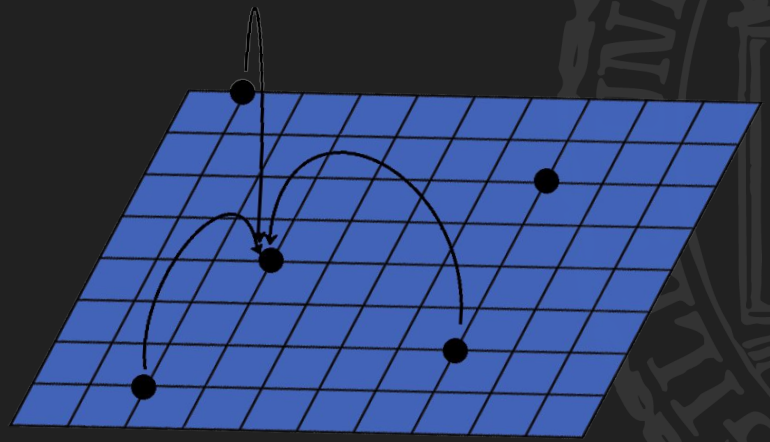
4. **Q&A**

BAMA = Bachelor / Master Thesis

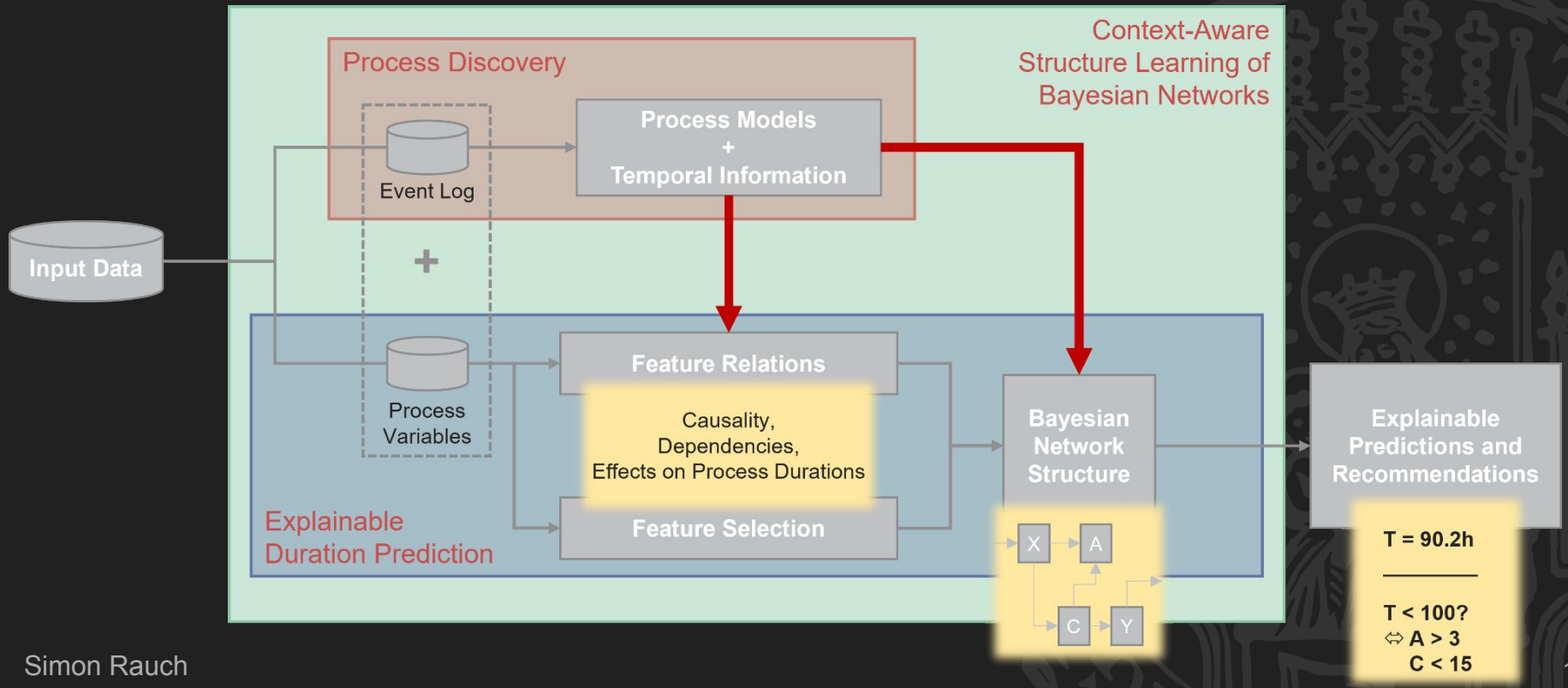
Analysis of Clustering Methods in Subspace Clustering

Topic for a Bachelor Thesis

- Self-Expressive Subspace Clustering as foundation
- Investigation of substitution of Spectral Clustering in the clustering step
- Primary goal: comparison of different clustering methods

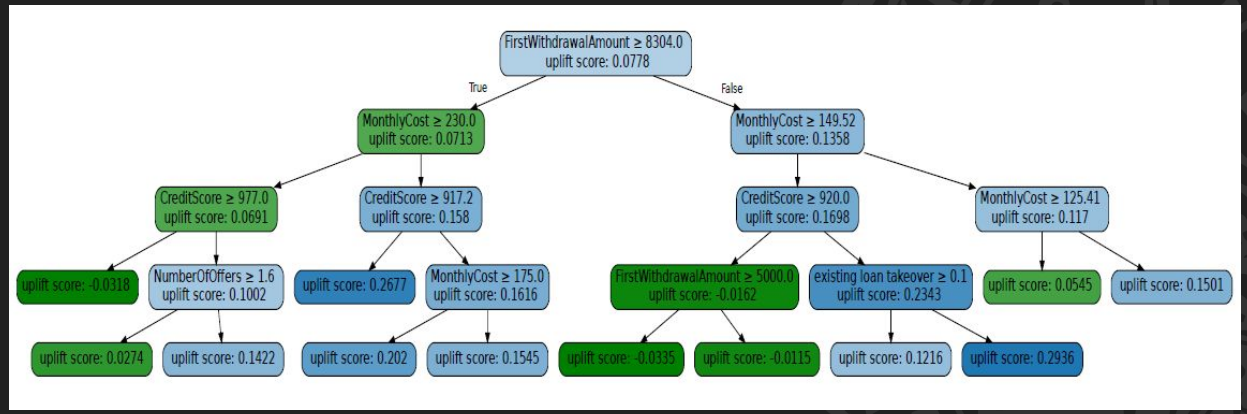


Context-Aware and Automated Structure Learning of Bayesian Networks



Causal Rules from Event Logs

- Adaptation of Causal Rule Mining for Event Logs to identify causal relationships between process variables and process durations (or other process KPIs)



Bozorgi et al. (2020) Process Mining Meets Causal Machine Learning: Discovering Causal Rules from Event Logs

Subspace Clustering on Process Data

Janina Sontheim

Process Occurances → **Process Log** → **Data Mining on Process Data**

time	case	activity
2018-6-6-6:29	732	a
2018-6-6-6:32	744	a
2018-6-6-6:33	732	b
2018-6-6-6:34	728	a
2018-6-6-6:35	732	d
2018-6-6-6:37	744	b
2018-6-6-6:38	728	c
2018-6-6-6:39	751	a
2018-6-6-6:42	744	d
2018-6-6-6:43	732	d
2018-6-6-6:44	744	e
2018-6-6-6:45	751	c
2018-6-6-6:47	732	e
2018-6-6-6:48	744	g
2018-6-6-6:59	751	d
2018-6-6-7:02	751	e
2018-6-6-7:03	728	e
2018-6-6-7:04	768	a
2018-6-6-7:05	751	h
2018-6-6-7:07	768	c
2018-6-6-7:08	728	h
2018-6-6-7:09	732	g
2018-6-6-7:12	768	d
2018-6-6-7:13	779	a
2018-6-6-7:14	768	e
2018-6-6-7:15	779	b
2018-6-6-7:17	768	h
2018-6-6-7:18	779	d

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

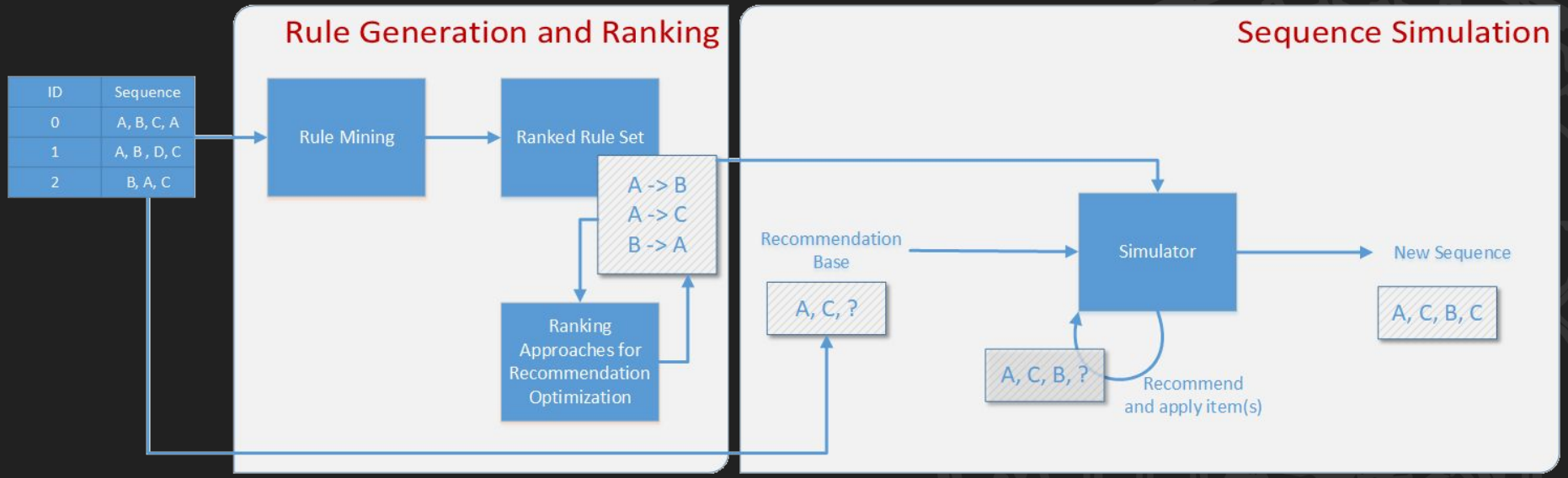
One visualization of the traces, along the time axis

Issues for Clustering on Process Data:

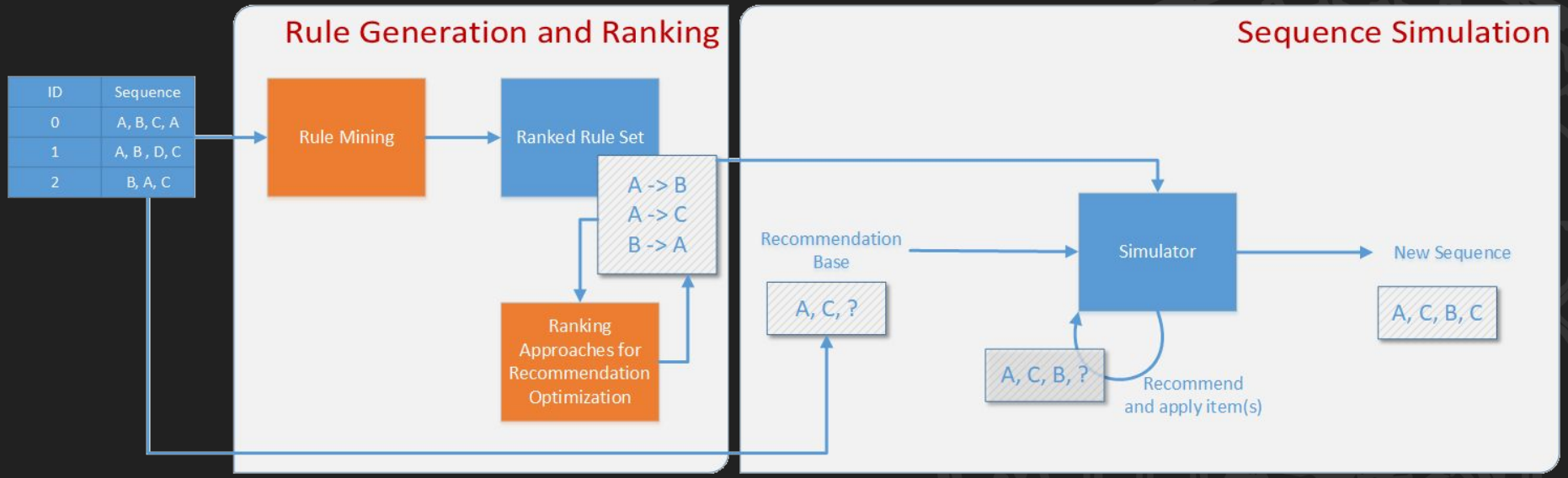
- Traces may have different lengths.
  → Clustering with missing data
- How to cluster on a combination of activity and timestamp?
- Not all activites might be included in the cluster
  → How to find clusters on just a subset of the trace?

BAMA-Thesis:
Grid based Subspace Clustering on Process Data.

Rule Mining for Recommender Systems



Rule Mining for Recommender Systems



Analysing (Pre-)Filtering Approaches on Partially-Ordered Sequential Rule Mining (BA)

- What impact do different rule filtering approaches have on the quality of recommendations?
- Implementation and comparison of filtering by
  - Closed/Maximal Patterns
  - Generator Patterns
  - Combination of aforementioned approaches
  - Introduction of negative items
    - Reward rules in which popular items do not occur in the antecedent
- Requirements:
  - Familiar with Python or **willingness to learn it**
  - Optional/Beneficial: KDD 1 (Lecture)

ID	Rule
0	A, B → E
1	B, C → D
2	B → E
3	A,B,C → E

Analysing (Pre-)Filtering Approaches on Partially-Ordered Sequential Rule Mining (BA)

- What impact do different rule filtering approaches have on the quality of recommendations?
- Implementation and comparison of filtering by
  - Closed/Maximal Patterns
  - Generator Patterns
  - Combination of aforementioned approaches
  - Introduction of negative items
    - Reward rules in which popular items do not occur in the antecedent
- Requirements:
  - Familiar with Python or **willingness to learn it**
  - Optional/Beneficial: KDD 1 (Lecture)

ID	Rule
~~0~~	~~A, B → E~~
1	B, C → D
~~2~~	~~B → E~~
3	A,B,C → E

Video & Multimodal Learning with Text

Typical Tasks:

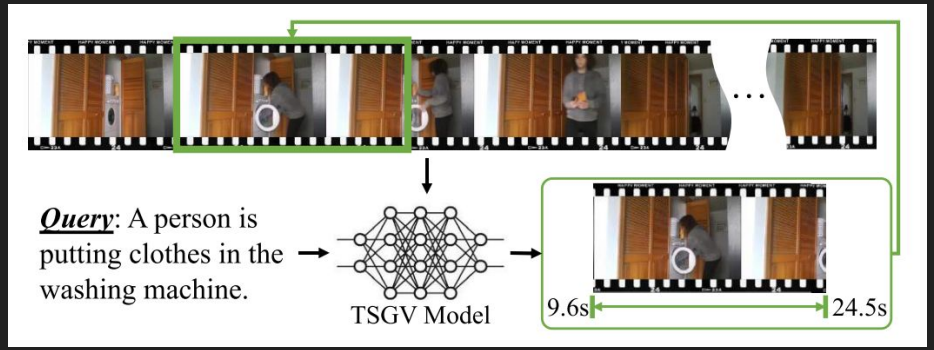
1. Segmentation & Tracking
2. Natural Language Queries
3. Question Answering
4. Summarization
5. . . .

Research Focus:

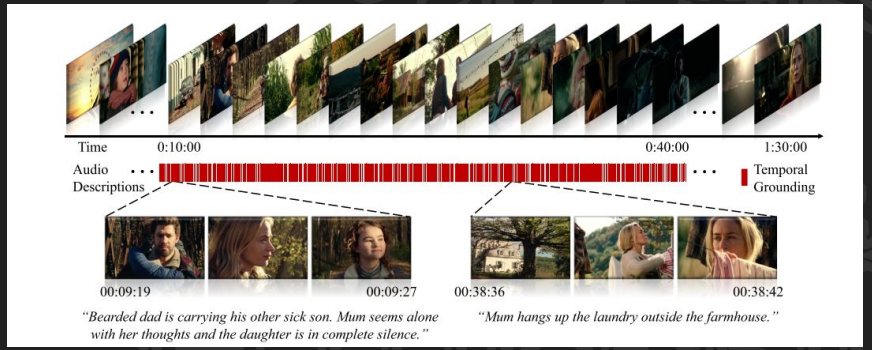
1. Resource Constraints
2. Long Video
3. Temporal Property
4. Multimodal Fusion
5. . . .

An Example Theses

Natural Language Queries in Long Videos

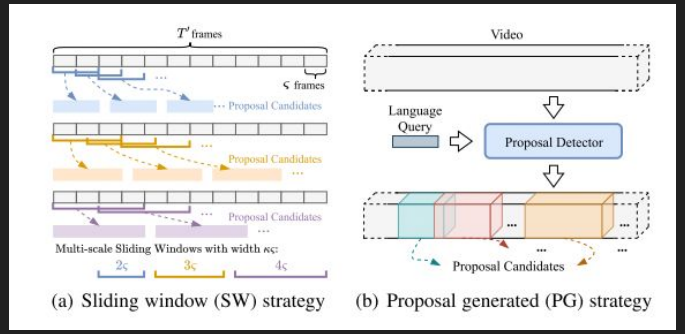


Short Video



Long Video - Movie Dataset

Natural Language Queries in Long Videos



Proposal Based
Exponential Computation

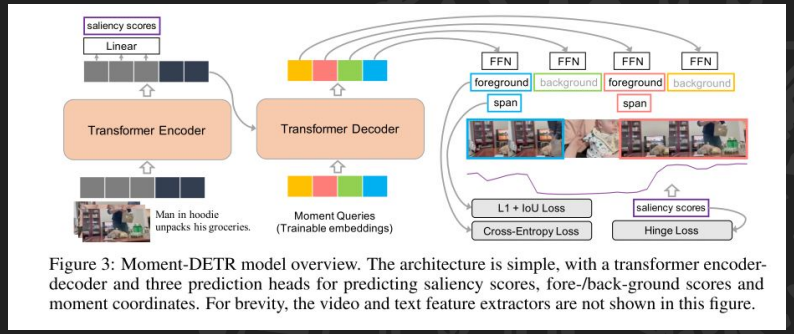


Figure 3: Moment-DETR model overview. The architecture is simple, with a transformer encoder-decoder and three prediction heads for predicting saliency scores, fore-/back-ground scores and moment coordinates. For brevity, the video and text feature extractors are not shown in this figure.

Proposal Free
Exponential Memory

Existing Solutions on Short Videos

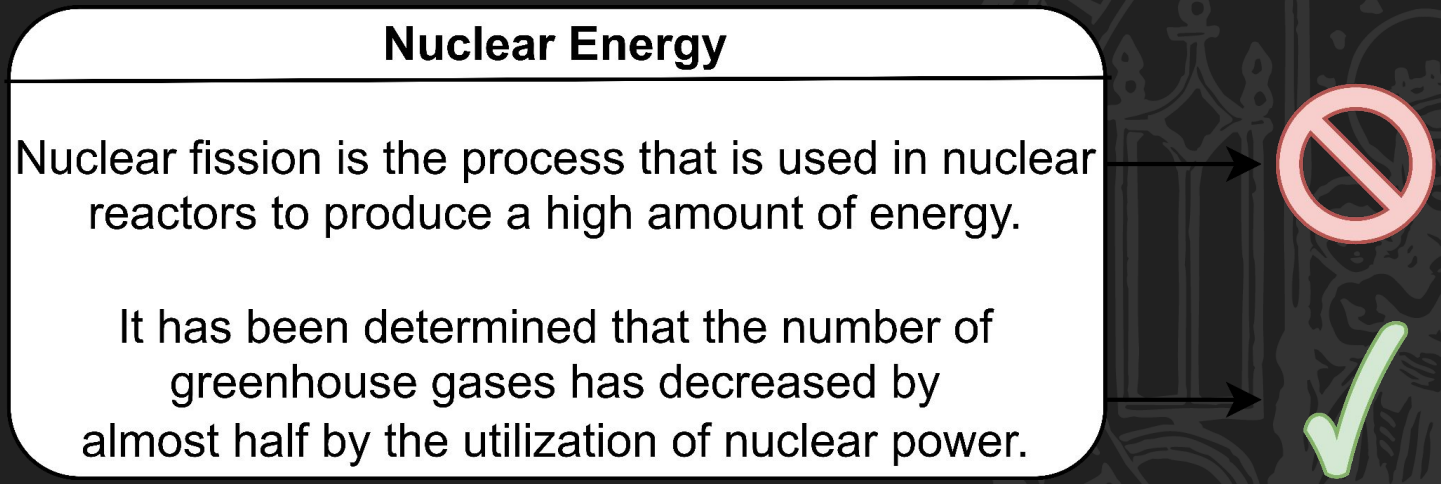
How to process Long Videos?

Argument Mining

- *An argument is a claim supported by reasons*
- Real world arguments are **diverse**
- Context specifies whether a text span is an **argument** or **not**
  - Context is often captured differently across domains
  - => Lack of consistently annotated data across text domains
- Recent work concentrated on specific text domains, e.g. tweets
  - annotated data often consists of specific annotation schemes

Argument Identification

span of text that expresses ***evidence*** or ***reasoning*** to either ***support*** or ***oppose*** a given topic

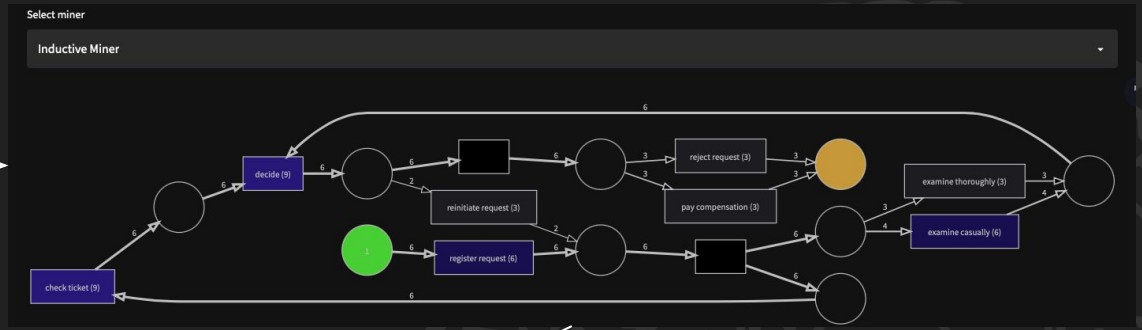


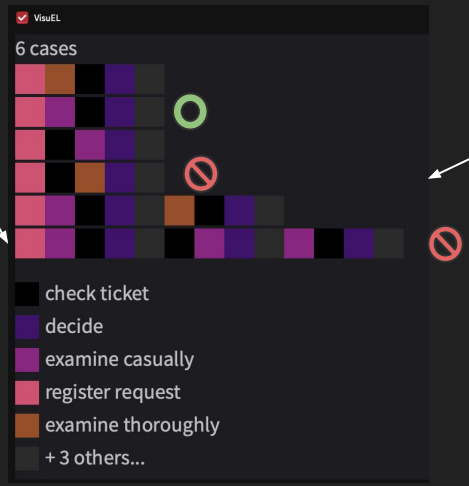
BA/MA Thesis: SotA Survey for Argument Identification

- Goal: Create **Argument Identification Framework** that allows:
  - integration of new data sets
  - integration of different annotation schemes
  - integration of new language models
  - integration of further Argument Mining tasks such as Argument Quality
  - benchmarking of the above
- Benefits: get familiar with current SotA approaches in NLP (such as Bert, DeBERTa, ERNIE 3.0, T5), Data Science development in Python, first insights in the Data Science Tools (e.g. Weights & Biases, Pytorch Lightning, RayTune)
- Requirements: Familiar with Python and Pytorch or **willingness to learn it,** optional benefits: DL&AI lecture, CL lectures, coursera specializations

Active learning for process conformance checking

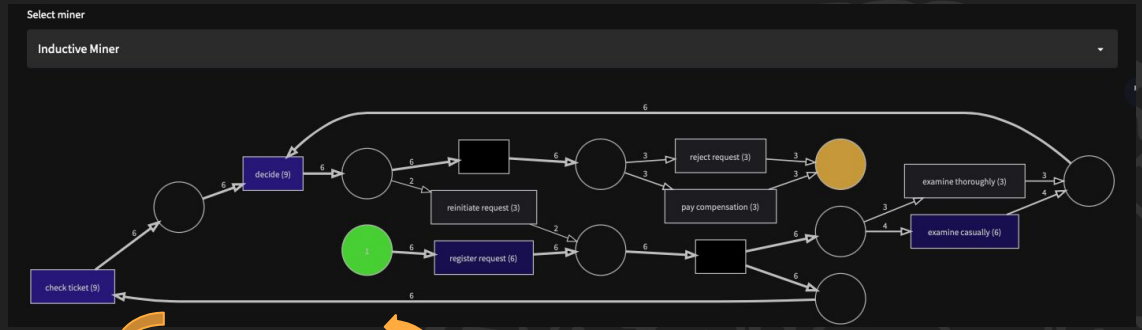
time:timestamp	Activity	case:concept:name
2010-12-31 10:06:00+01:00	examine thoroughly	1
2011-01-21 09:06:00+01:00	examine casually	5
2011-01-06 13:06:00+01:00	examine thoroughly	3
2010-12-30 14:16:00+01:00	examine casually	2
2011-01-08 14:43:00+01:00	examine thoroughly	4
2011-01-06 09:18:00+01:00	decide	3
2011-01-06 12:18:00+01:00	reinitiate request	3
2011-01-09 09:55:00+01:00	decide	3
2011-01-05 11:22:00+01:00	decide	2

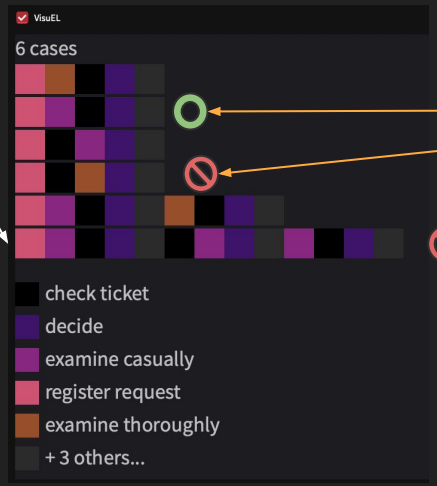




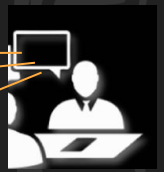
Active learning for process conformance checking

time:timestamp	Activity	case:concept:name
2010-12-31 10:06:00+01:00	examine thoroughly	1
2011-01-21 09:06:00+01:00	examine casually	5
2011-01-06 13:06:00+01:00	examine thoroughly	3
2010-12-30 14:16:00+01:00	examine casually	2
2011-01-08 14:43:00+01:00	examine thoroughly	4
2011-01-06 09:18:00+01:00	decide	3
2011-01-06 12:18:00+01:00	reinitiate request	3
2011-01-09 09:55:00+01:00	decide	3
2011-01-05 11:22:00+01:00	decide	2

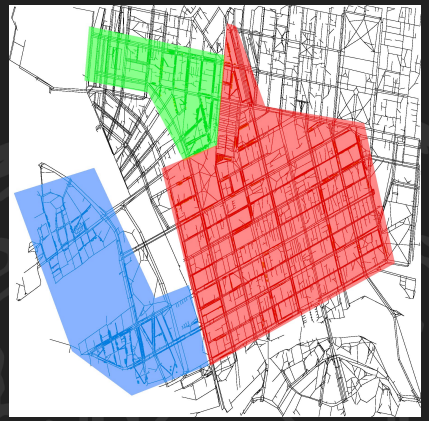




Conformant?



Transferable RL Agents in MSRC

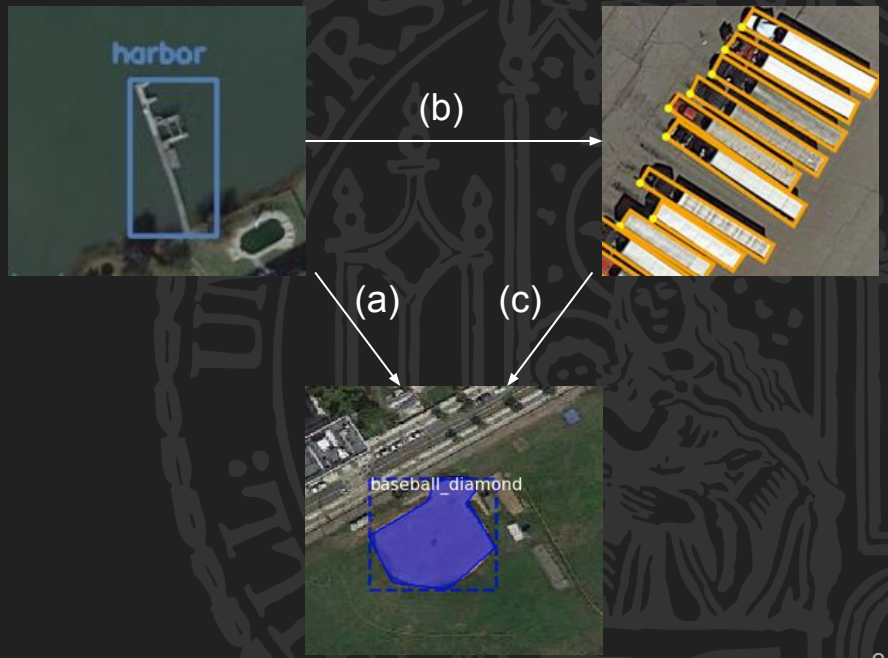


- Multi-Agent Stochastic Resource Collection
  - Multiple agents navigate through road network
  - Collect as many dynamic resources as possible
  - Example: Traveling Officer Problem
  - Solved with DRL
- Goals of thesis
  - Spatial transfer (train one area/road network transfer to different area)
  - Transfer wrt. number of agents
  - Scaling up
- Requirements: Familiar with Python and Pytorch or **willingness to learn it,** optional benefits: DL&AI lecture, RL lectures, coursera specializations

Computer Vision and Remote Sensing

Instance Segmentation with Oriented Bounding Boxes

(a) Tian, Zhi, et al. "Boxinst: High-performance instance segmentation with box annotations." *CVPR* 2021.
(b) ?
(c) ?



Computer Vision and Remote Sensing

Further topics:

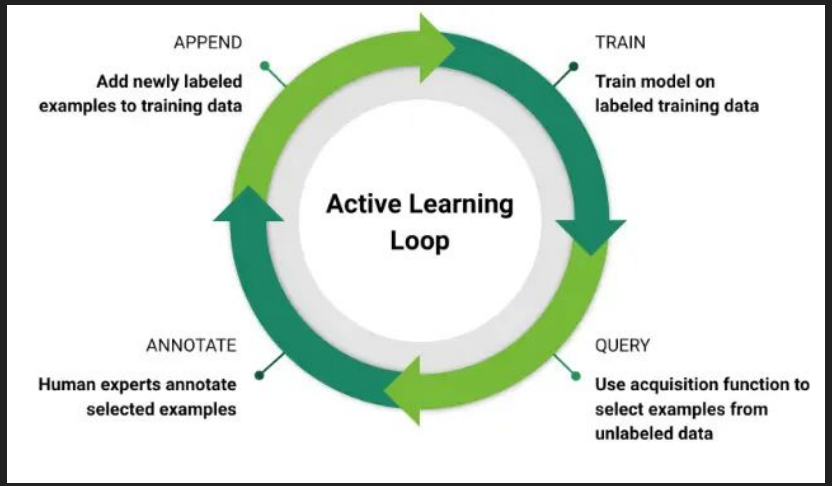
- Evaluation of segmentation models
  - Label efficiency
  - Detailed error categorization and analysis
- Weakly / semi- / unsupervised deep learning

Requirements:

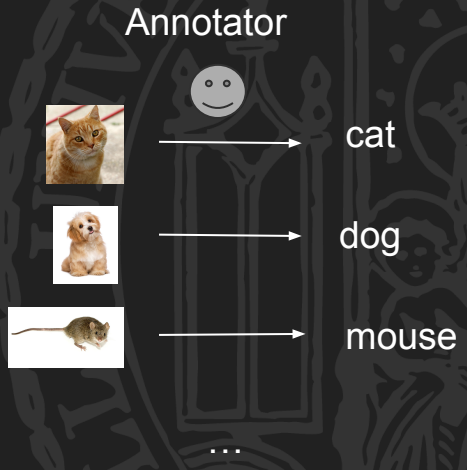
- Understanding of ML and DL
- Python and Pytorch or **willingness to learn it**
- Problem solving skills and little frustration tolerance :)

Active Learning

- Supervised learning requires annotations for each input
- Annotating data is expensive (especially when domain knowledge is required)
- Active learning is one way of dealing with limited or no labeled data

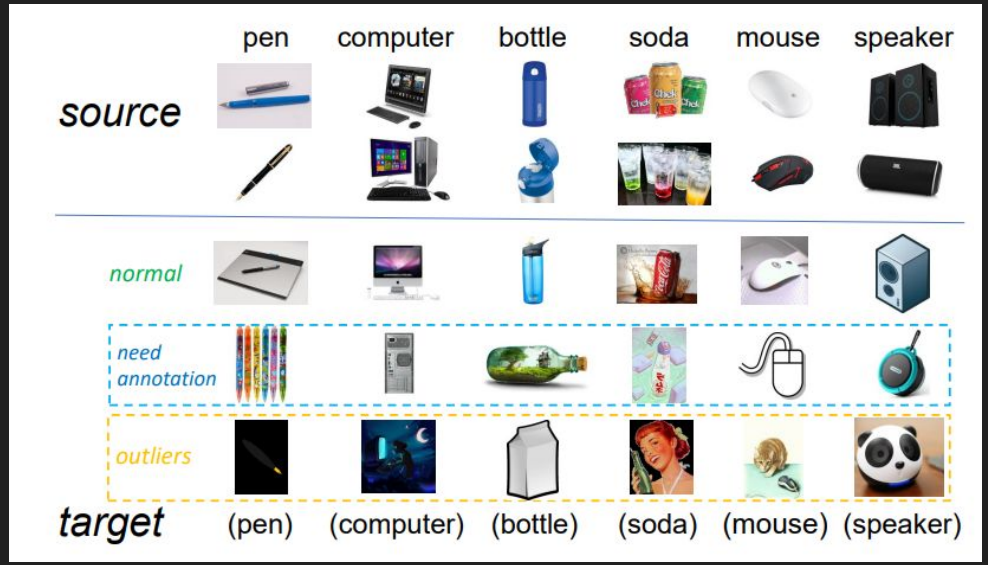


https://blogs.nvidia.com/blog/2020/01/16/what-is-active-learning/



Cost-Effective Active Learning for Domain Adaptation

Domain Adaptation: adapt model learned on source domain to target domain



Your Tasks:

- Read and understand scientific papers
- Implement a novel Single-Shot active domain adaptation approach
- Integrate existing methods
- Evaluation

Requirements:

- Familiar with Python
- optional benefits: Pytorch, DL&AI lecture

FU, Bo, et al. Transferable query selection for active domain adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. S. 7272-7281.

Useful References

1. Information on research areas and topics at our chair
   https://www.dbs.ifi.lmu.de/cms/forschung/index.html

2. Further information on bachelor and master theses at our chair
   https://www.dbs.ifi.lmu.de/cms/studium_lehre/bama_theses/index.html

3. Oberseminar Website
   https://www.dbs.ifi.lmu.de/cms/studium_lehre/lehre_master/oberseminar22/index.html

4. General Information by Course Coordination
   http://www2.tcs.ifi.lmu.de/~barths/stukoordination/abschlussarbeiten.html