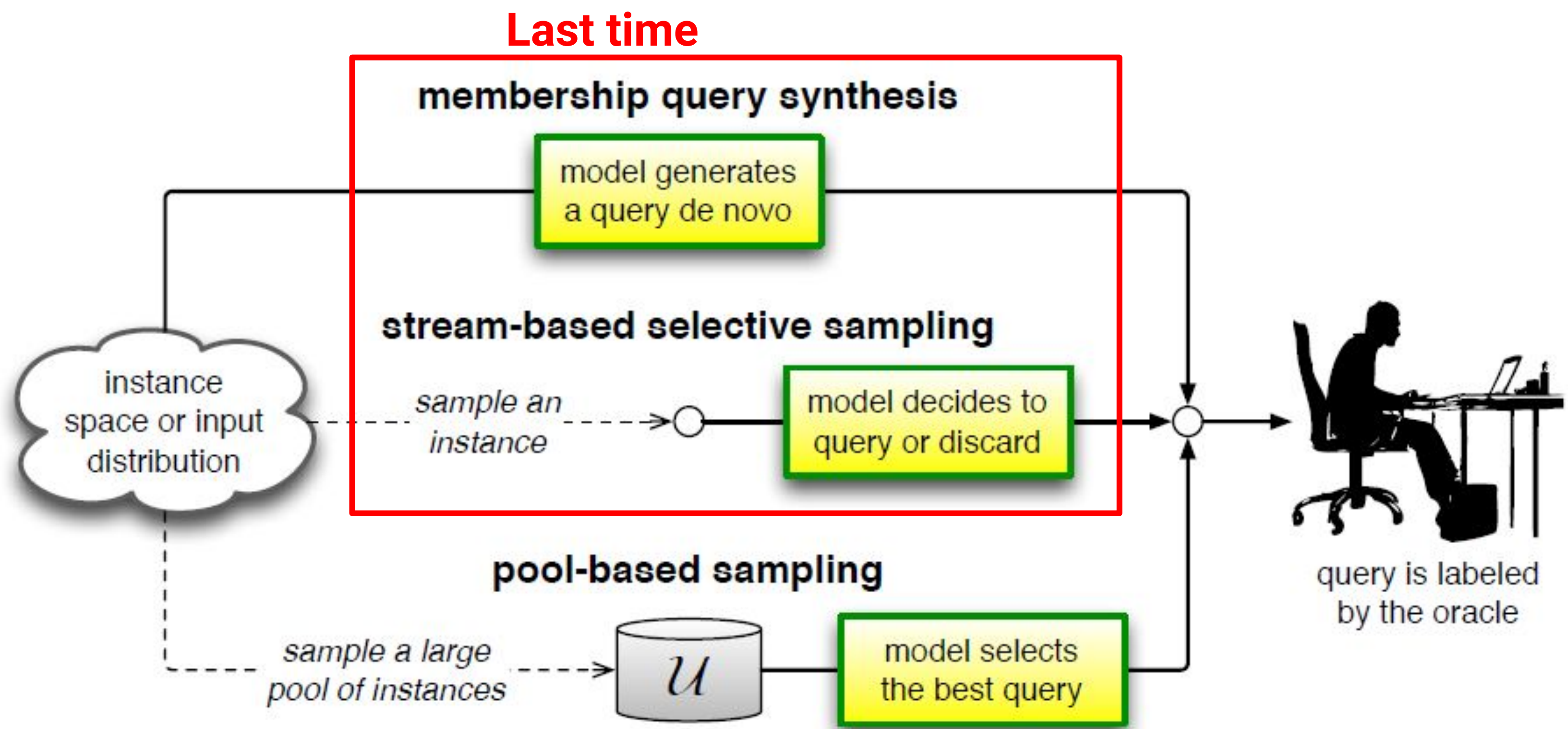
Active Learning - the pool-based selective sampling (part 1)

Andrzej Janusz
Daniel Kałuża

THE PLAN

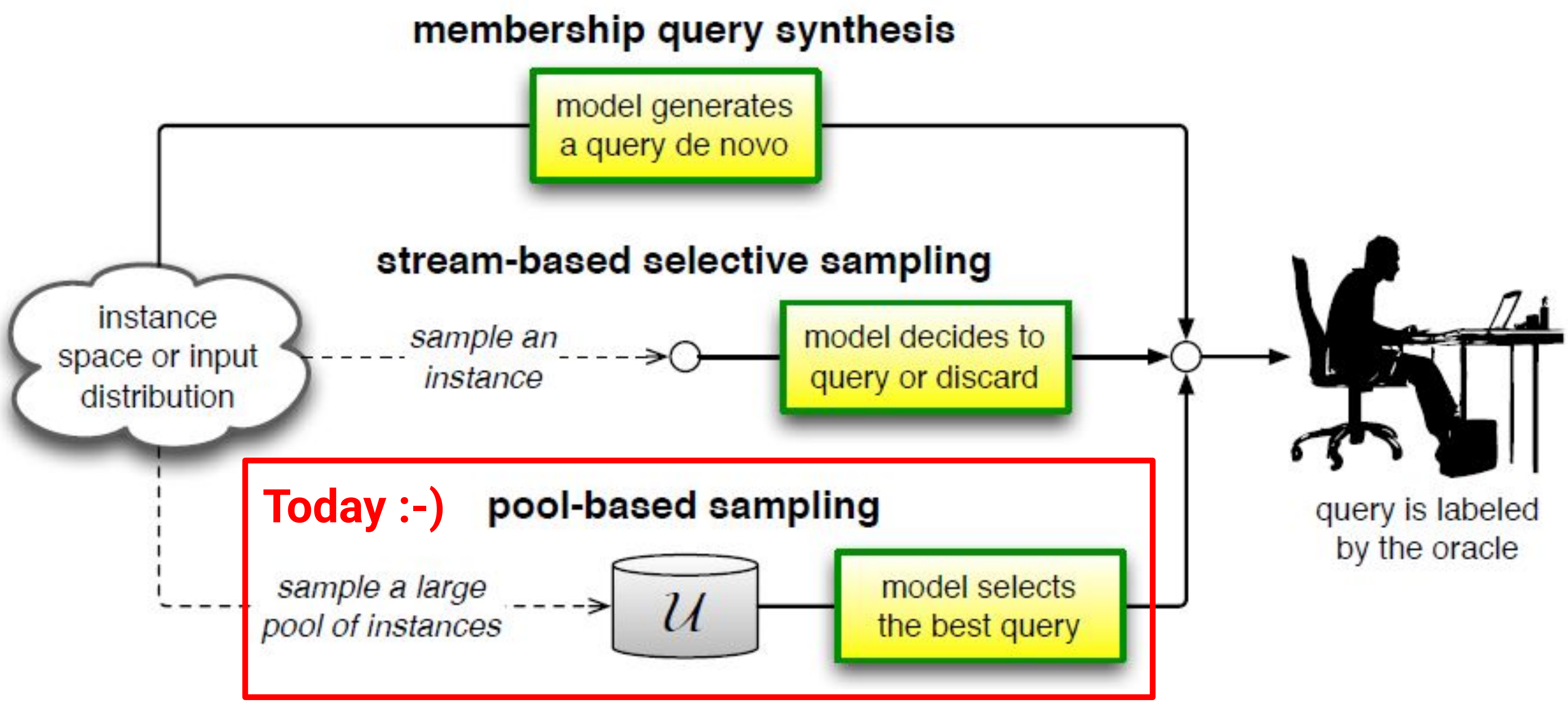
- A recap of the previous lecture.
- Uncertainty sampling.
- Exemplary measures.
- Use-cases.
- Evaluation in active learning experiments.
- Summary.

The three main Active Learning scenarios



Based on *Burr Settles: Active Learning Literature Survey (2010)*

The three main Active Learning scenarios



Based on *Burr Settles: Active Learning Literature Survey (2010)*

Active Learning as an optimization task

Formal task definition - we search for $U^* \subset DP$ such that:

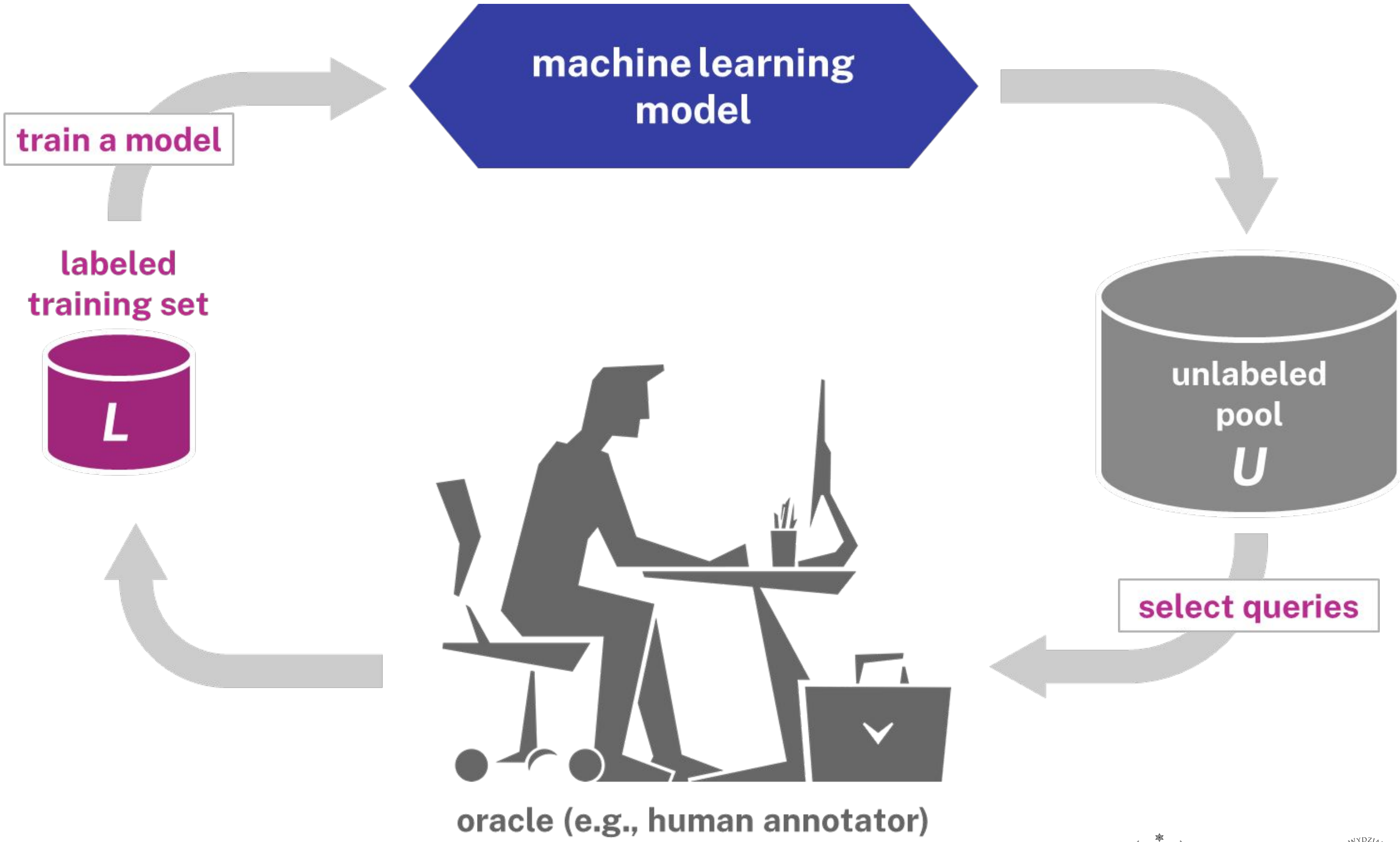
$$U^* = \underset{U:|U|=K}{\arg\max}\ \mathbb{E}_{(X,Y)}[q(Y, f^U(X))]$$

where f^U is a model trained on a subset $U \subset DP$ whose size is *K* and *q* is a predefined quality metric.

Pool-based selective sampling

- We have a large pool of unlabeled instances *U*.
  - We evaluate the usefulness of the instances from *U* for the learner <u>at each iteration of the AL cycle</u>.
  - We may choose one or more instances to query the oracle.
  - The unlabeled data pool may grow in time but we assume that it is static in each iteration.
- An informativeness measure is used to evaluate all instances from the pool.
  - If the pool size is very large, some subsampling can be used…
- **How do we evaluate the informativeness of an instance?**

The active learning cycle - revisited



An example - Cancer tissue classification

- Rączkowski et al. (2019) describe an application of the pool-based active learning in the field of medical diagnostics.
  - Active learning framework chooses uncertain samples.
  - Instances - small tiles with tissues stained with hematoxylin and eosin (H&E).
  - Histopathologists annotate pixels of the image with tissue classes.
  - Deep neural network is used to learn from the annotated examples.
- The active learning approach resulted in 45% speed-up of the model learning process.

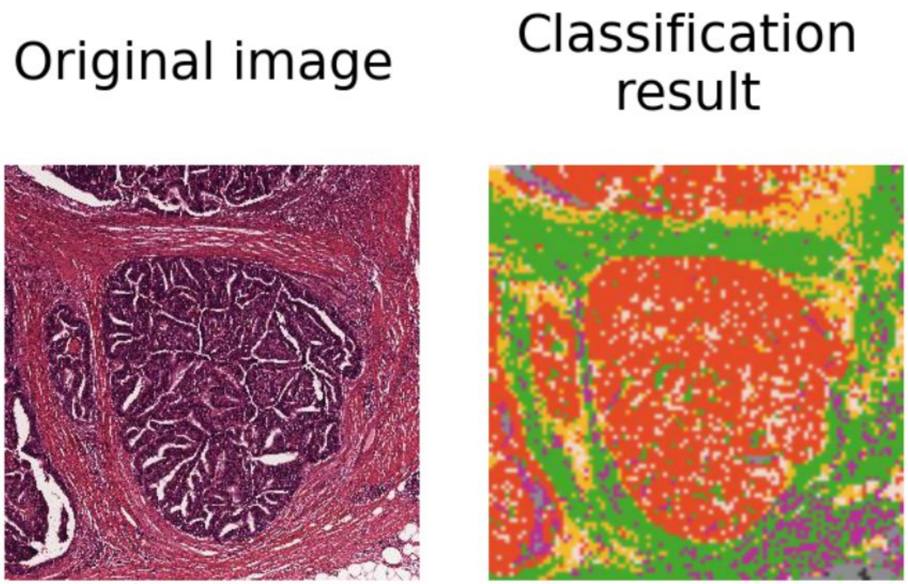
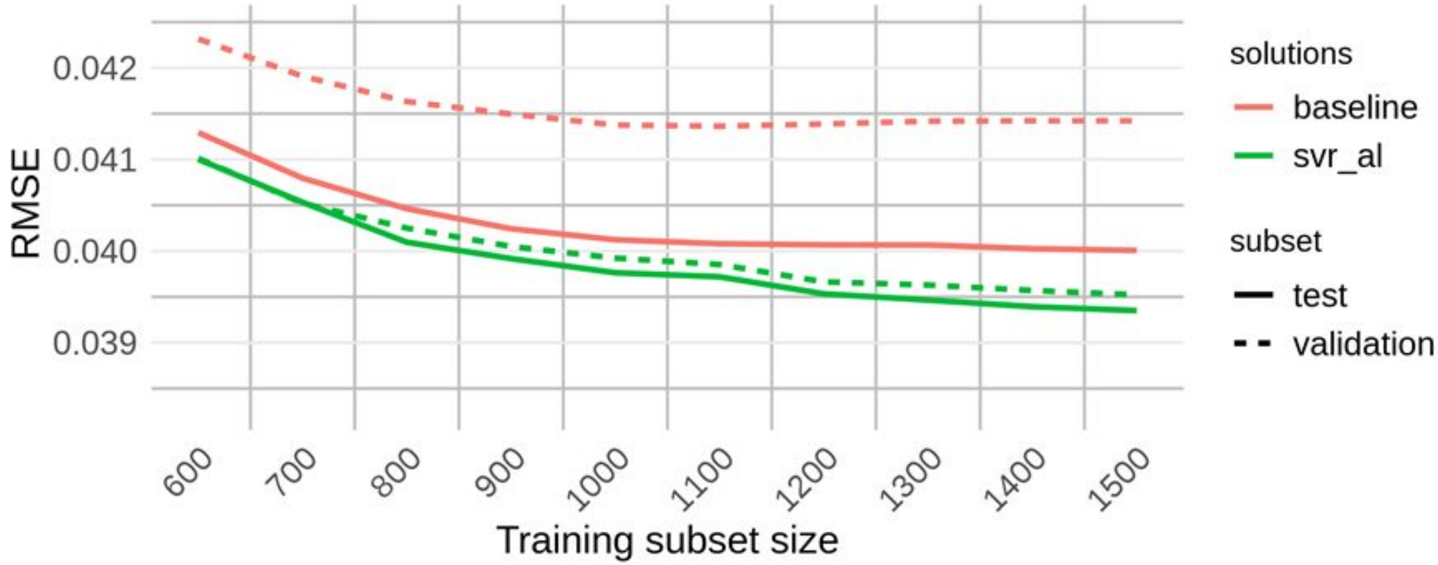


Image taken from Rączkowski et al. (2019): ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning.

Image: Freepik.com

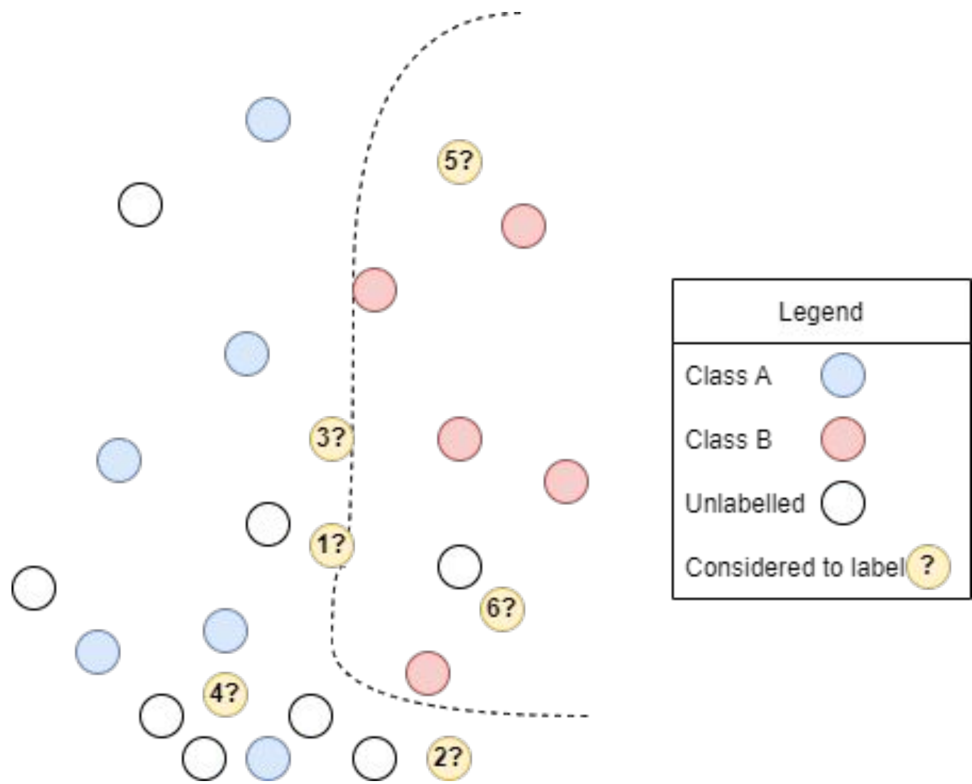
An exemplary application - AAIA’19 DMC



- Janusz et al. (2019) proposed a method based on a combination of informativeness density and diversity sampling for active learning of deck win-rates in a popular mobile video game Clash Royale.
- Historical win-rates were available for a large pool of decks. How will the win-rates change in a new season?
- Active learning outperformed random sampling and nu-SVR baselines. https://knowledgepit.ml/clash-royale-challenge/

Informativeness and uncertainty - again

- The informativeness can be considered from several perspectives:
  - Proximity to a decision boundary ≈ prediction uncertainty.
  - Representativeness.
  - Expected impact on the learner.
  - Expected influence on the lerner's generalization quality.



Uncertainty sampling

- The simplest and very popular approach:
  - We evaluate the prediction uncertainty for each instance.
    - For the classification task, it boils down to querying near the decision boundary region.
  - We want to minimize the epistemic uncertainty of the learner.
  - Measuring the epistemic uncertainty is not easy…
- All we need is a good measure.
  - How can we measure the proximity to the decision boundary?
  - What about the regression task?
  - What changes for the multi-label classification?
  - Other ML tasks?

Classification uncertainty sampling

- Popular classification uncertainty sampling methods:
  - Least confidence: $u^*_{LC} = \arg\max_u (1 - P_\theta(\hat{y}|u))$
  - Minimum margin: $u^*_M = \arg\min_u (P_\theta(\hat{y}_1|u) - P_\theta(\hat{y}_2|u))$
  - Shannon entropy: $u^*_H = \arg\max_u \left(-\sum_i P_\theta(y_i|u) \log P_\theta(y_i|u) \right)$
- Many other measures are vailable…
- **But most of the commonly used measures assume uniformly distributed decision thresholds!**

Classification uncertainty - analysis

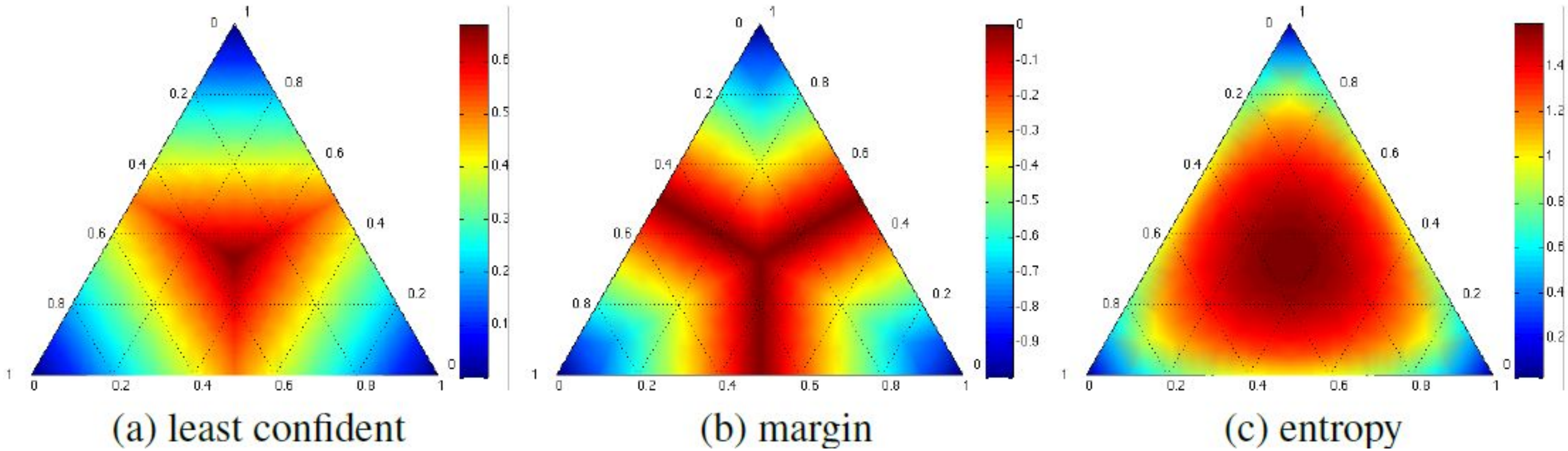
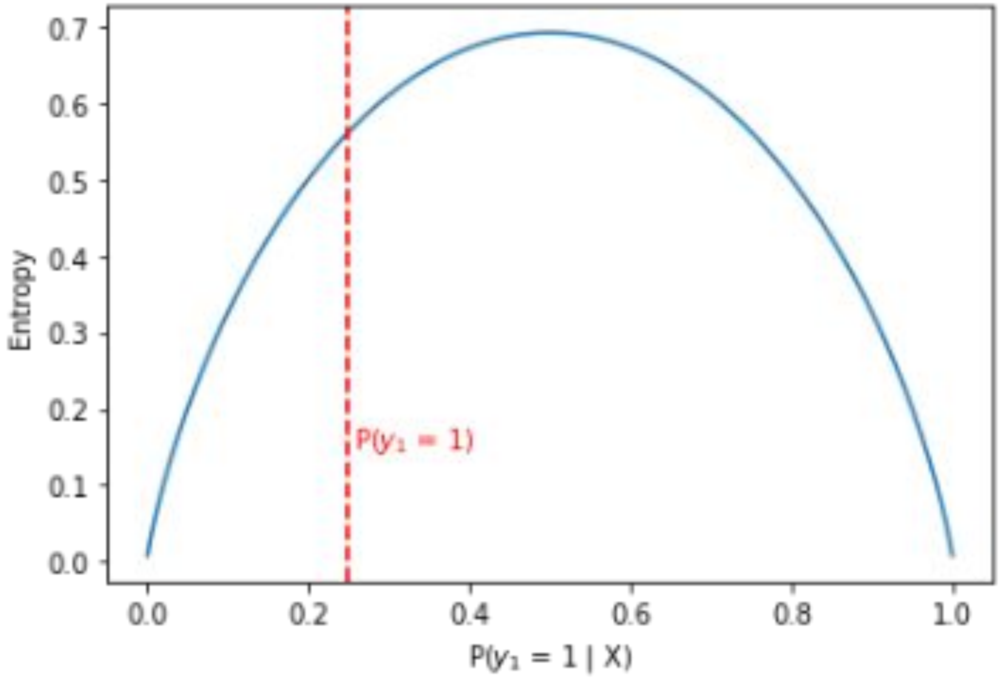


Image taken from Burr Settles: Active Learning Literature Survey (2010)

Imbalanced classification problems

- What if the class distribution is imbalanced and the quality measure gives different weights to classes?
  - We may want to shift the decision thresholds!



Informativeness for imbalanced classification

- Standard uncertainty measures can be adjusted so that they take their maximum at any given class distribution.
  - A simple rescaling trick!

$$< p_1, \ldots, p_i, \ldots, p_k > \ \text{-} \ \textit{predicted distribution}$$

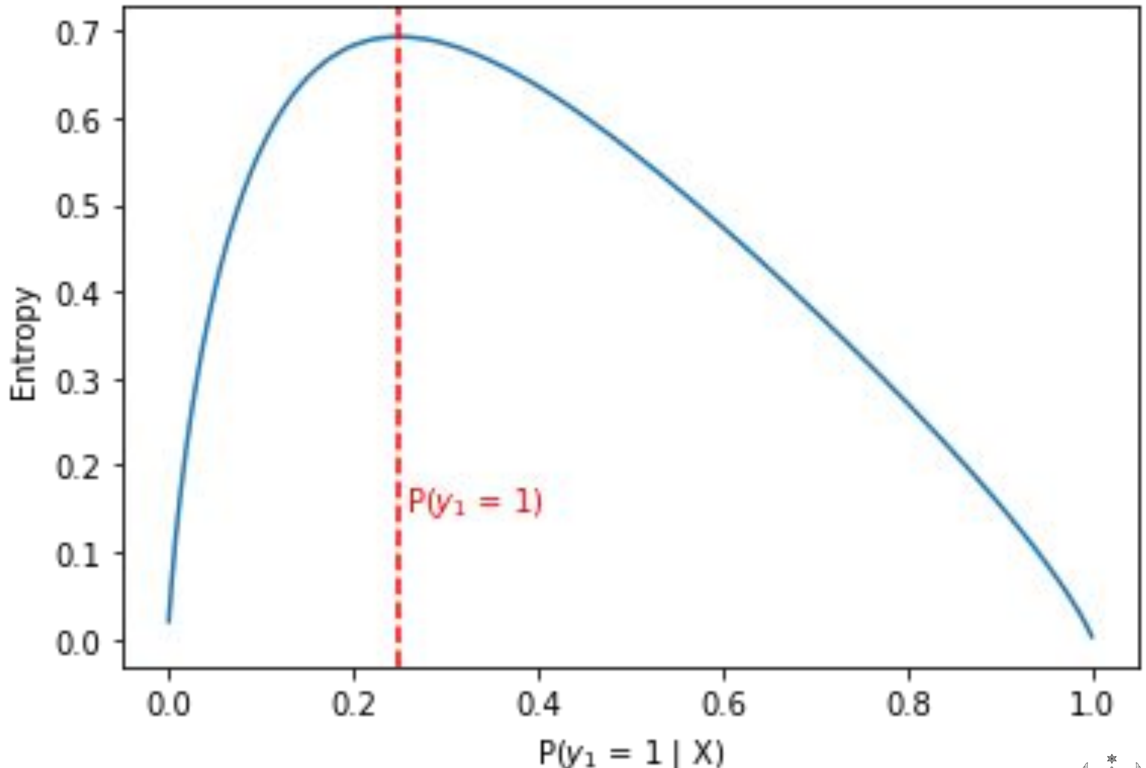
$$< r_1, \ldots, r_i, \ldots, r_k > \ \text{-} \ \textit{target distribution}$$

$$Let\ c = \sum_{i=1,\ldots,k} \frac{p_i}{r_i},\ then\ p_i \longrightarrow \frac{1}{c} \cdot \frac{p_i}{r_i}$$

- **How to choose the right decision boundary?**

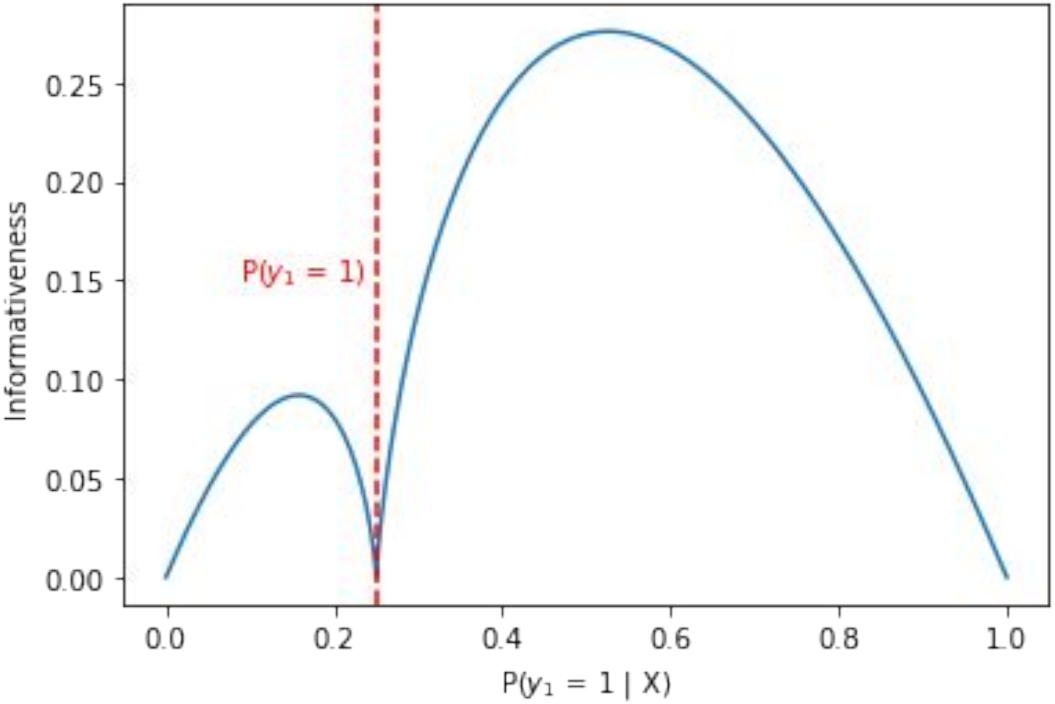
Rescaled decision boundaries

- If we consider a decision threshold at 0.25, we get a “decision threshold-centered entropy” :-)

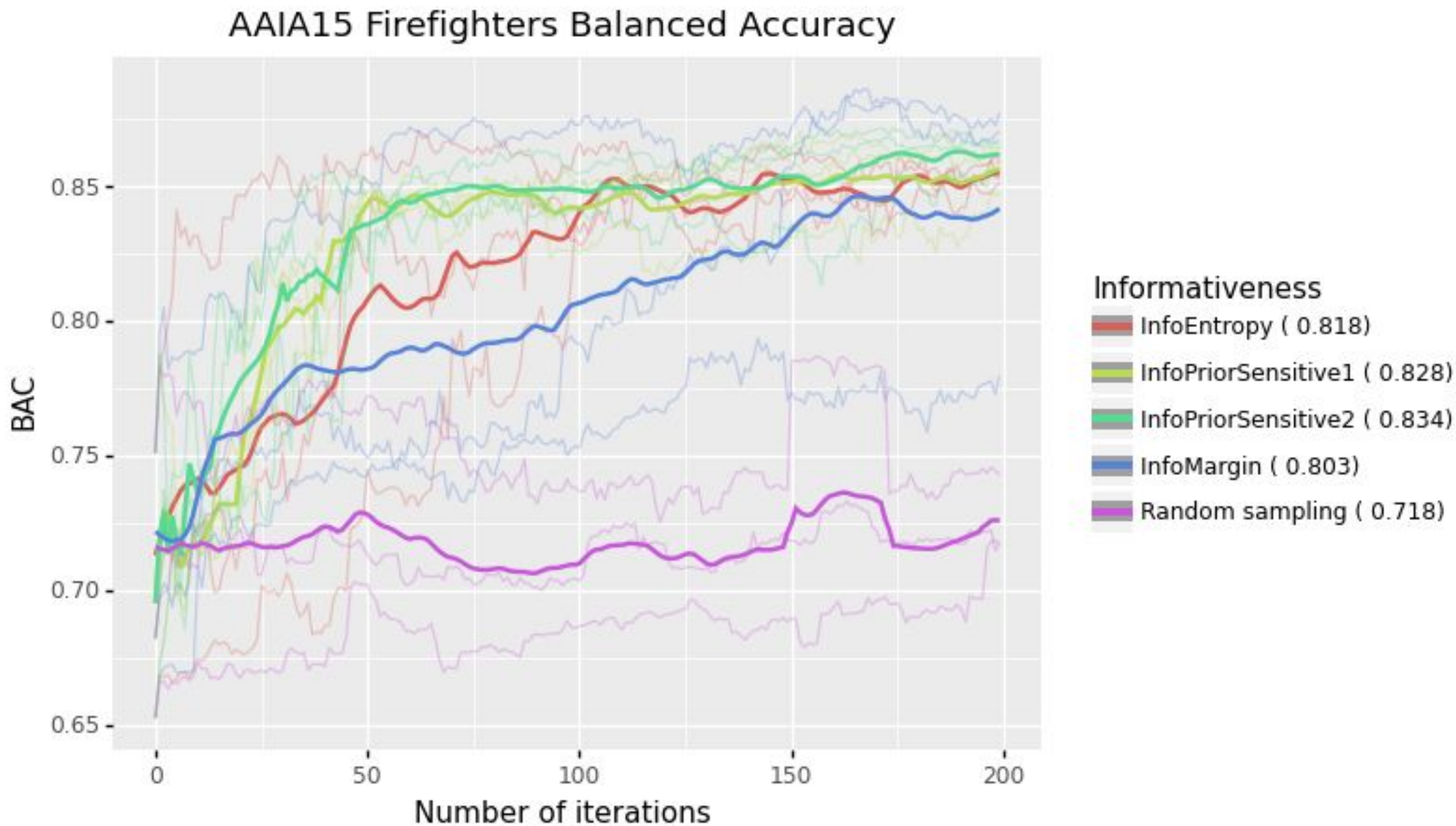


Other ideas

- What if we don't want to promote sampling from the decision boundary?
  - Cases very near to the boundary can be confusing to experts (i.e., our oracle).
  - Cases might be close to the boundary due to aleatoric uncertainty.



An example - a comparison on AAIA’15 data

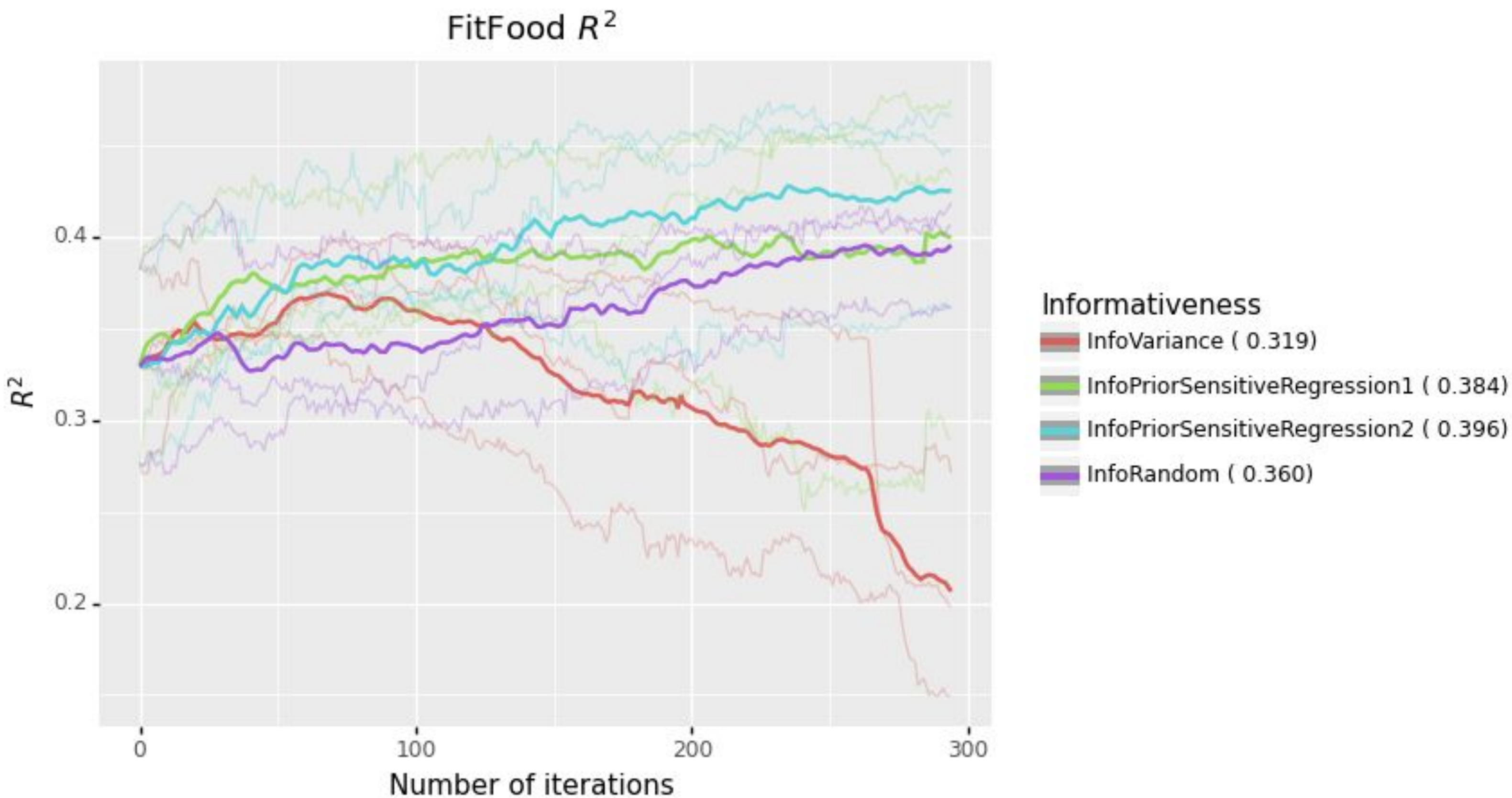


- Initial batch size: 200 (1% of the pool) and 200 iterations.
- XGBoost learner with default settings.

Regression uncertainty sampling

- Exemplary regression uncertainty sampling methods:
  - Variance-based: $u^*_{Var} = \arg\max_u Var(\Phi(u))$
  - Prior density-scaled: $u^*_{\mu} = \arg\max_u Var(\Phi(u)) \cdot \int_{E(\Phi(u))-\epsilon}^{E(\Phi(u))+\epsilon} \Phi(x)dx$
  - Differential entropy: $u^*_H = \arg\max_u \left(- \int \Phi(u)(x) \log \Phi(u)(x) dx \right)$
- Our model Φ needs to return distributions (not only the predictions)…
- **We may need to take into consideration the prior distribution of targets (*Φ*) - but how can we do that?!?**

An exemplary application - a regression task



- Initial batch size: 300 (1% of the pool), and 300 iterations.
- XGBoost trained using a natural gradient learner with a negative binomial prior.

Estimating the decision boundary

- Depending on the evaluation metric, it might be desirable to use a problem-specific decision boundary.
  - But we don’t have to many labels…
  - and we don’t want to do random sampling.
  - We may want to balance the predictions (e.g., to optimize the BAC metric).
- Instead of estimating the distribution of the target variable using known labels, **use the distribution of predictions**!
  - It works well for classification and regression problems.

Evaluation in active learning experiments

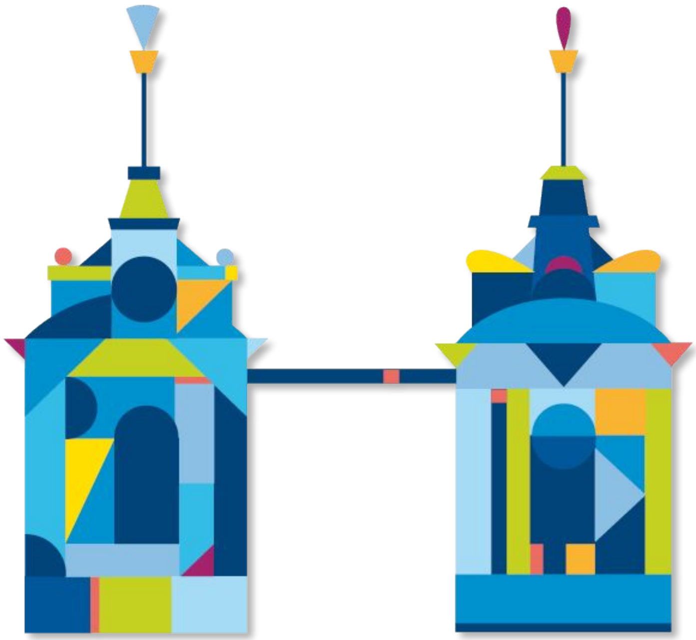
- We only simulate “real-life” problems.
- We don’t have to rely on real oracle - we have the labels.
  - We focus on testing the query selection methods…
  - or model updating techniques (more on this topic in future).
- The four main KPIs in active learning:
  - What performance level did we achieve after a fixed number of queries?
  - How many queries we needed to achieve the required performance level?
  - Area under the performance curve.
  - Stability of the model training process.

Summary

- We discussed the pool-based selective sampling approach to Active Learning.
- We focused on uncertainty sampling techniques.
- We discussed several uncertainty measures for classification and regression tasks which can be used to guide the AL process.
- We briefly talked about the performance evaluation in active learning experiments.
- We analyzed a few application examples for different ML tasks.

Literature:

1. B. Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, (2010).
2. R. Cassidy, E. S. Charles, J. D. Slotta, N. Lasry. Active Learning: Theoretical Perspectives, Empirical Studies and Design Profiles. Frontiers Media SA, (2019).
3. R. Monarch. Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-centered AI. Simon and Schuster, (2021).
4. Ł. Rączkowski, M. Możejko, J. Zambonelli, E. Szczurek. ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning. Scientific Reports 9, 14347 (2019).
5. D. Angluin. Queries revisited. In Proceedings of the International Conference on Algorithmic Learning Theory, pages 12–31. Springer-Verlag, (2001).
6. H. Wang, X. Chang, L. Shi, Y. Yang, Y.D. Shen. Uncertainty sampling for action recognition via maximizing expected average precision. IJCAI International Joint Conference on Artificial Intelligence, pages 964-970, (2018).
7. A. Janusz, Ł. Grad, M. Grzegorowski: Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction. FedCSIS 2019: 3-6, (2019).



QUESTIONS OR COMMENTS?

a.janusz@mimuw.edu.pl

or

d.kaluza@mimuw.edu.pl