Intelligent
Embedded Systems

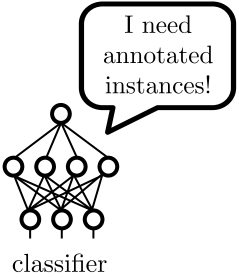
Multi-annotator Probabilistic Active Learning

Marek Herde, Daniel Kottke, Denis Huseljic, Bernhard Sick
{marek.herde | daniel.kottke | dhuseljic | bsick}@uni-kassel.de

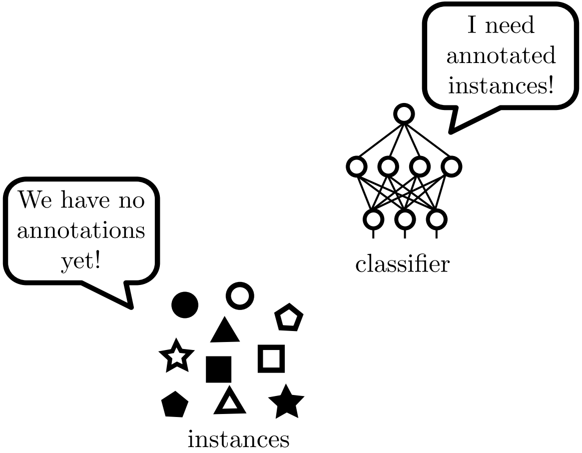
University of Kassel, Germany
Electrical Engineering / Computer Science

December 9, 2020

Motivation
Annotation Costs

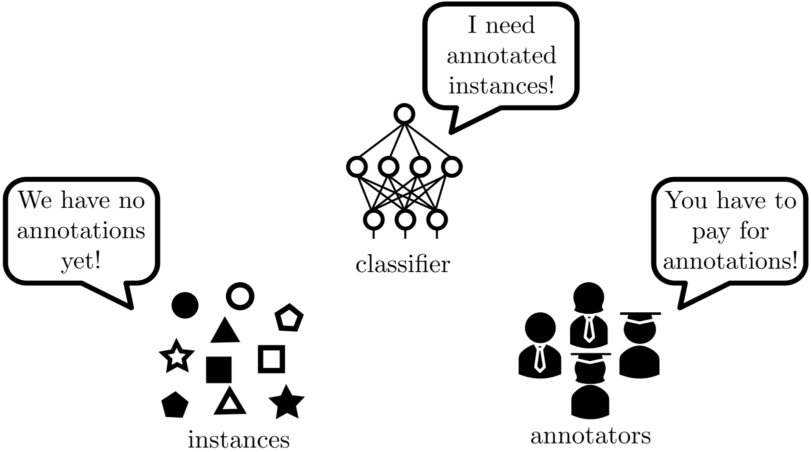


Motivation
Annotation Costs

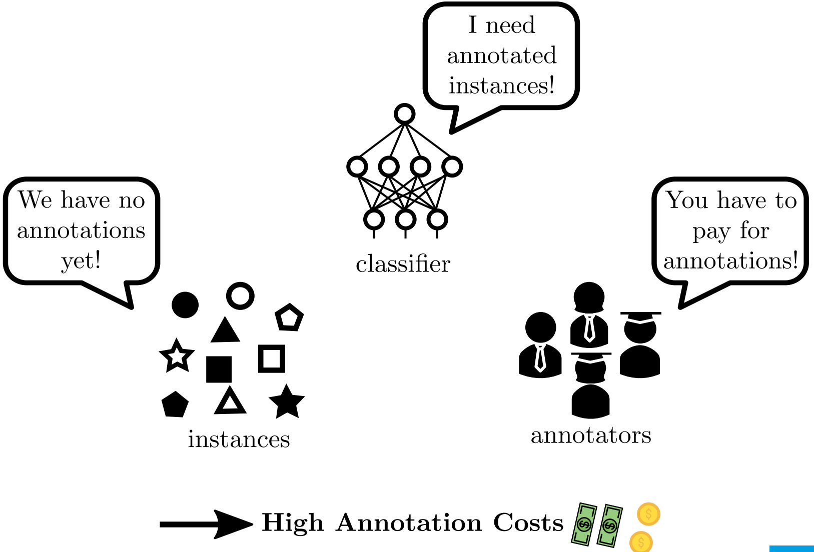


Motivation

Annotation Costs



Motivation
Annotation Costs

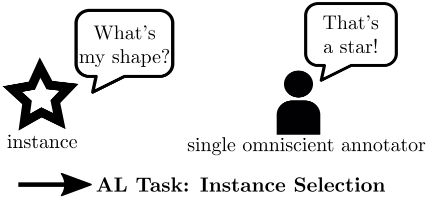


Active Learning (AL)

Through an intelligent selection of training data, AL aims at minimizing the annotation costs and the classifier's misclassification risk.

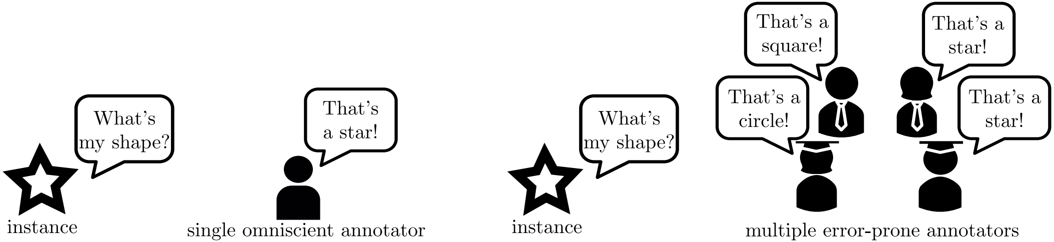
Active Learning (AL)

Through an intelligent selection of training data, AL aims at minimizing the annotation costs and the classifier's misclassification risk.



Active Learning (AL)

Through an intelligent selection of training data, AL aims at minimizing the annotation costs and the classifier's misclassification risk.

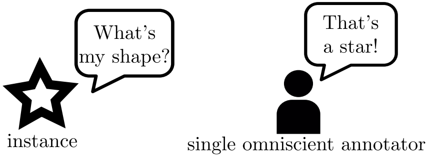


→ **AL Task: Instance Selection**

→ **AL Task: Instance + Annotator Selection**

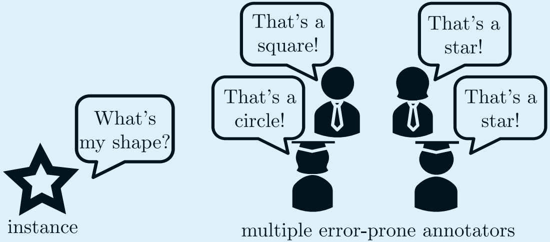
Active Learning (AL)

Through an intelligent selection of training data, AL aims at minimizing the annotation costs and the classifier's misclassification risk.



→ **AL Task: Instance Selection**

Scope: AL with Multiple Error-prone Annotators



→ **AL Task: Instance + Annotator Selection**

We propose **multi-annotator probabilistic active learning (MaPAL)**, which

We propose **multi-annotator probabilistic active learning (MaPAL)**, which

- estimates the performance of annotators as a function of instances.

We propose **multi-annotator probabilistic active learning (MaPAL)**, which

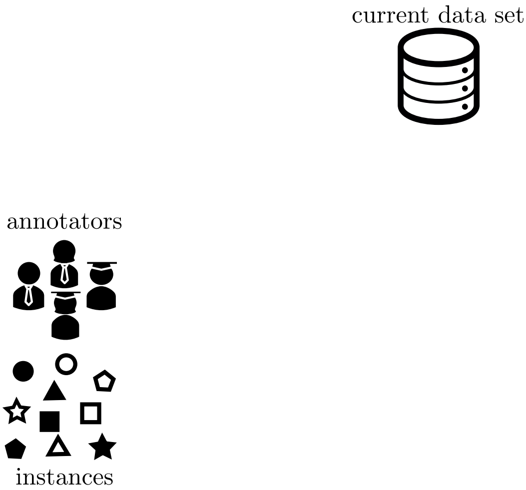
- estimates the performance of annotators as a function of instances.
- jointly selects an instance-annotator pair maximizing the classifier's expected performance.

We propose **multi-annotator probabilistic active learning (MaPAL)**, which

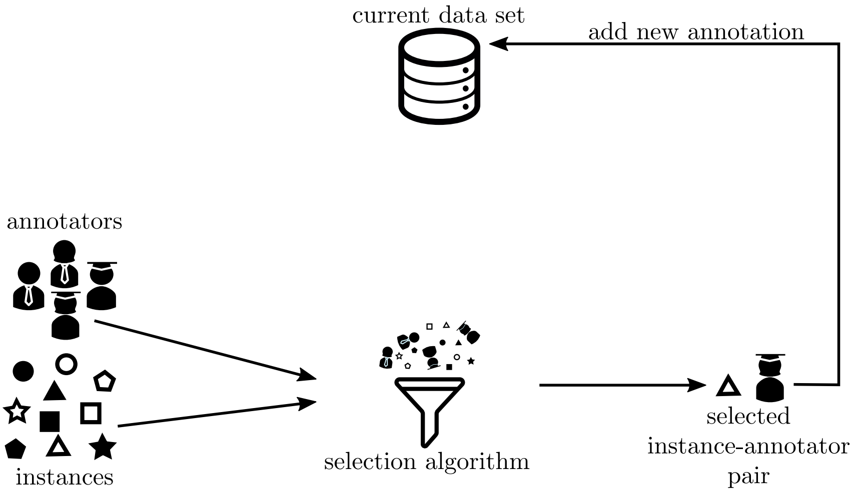
- estimates the performance of annotators as a function of instances.
- jointly selects an instance-annotator pair maximizing the classifier's expected performance.
- showed superior and robust performance compared to five related AL strategies in an experimental evaluation over various data sets.

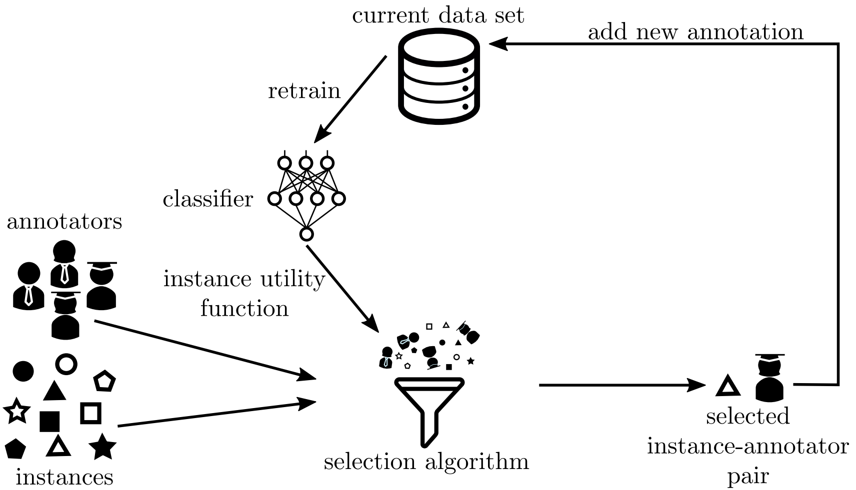
We propose **multi-annotator probabilistic active learning (MaPAL)**, which

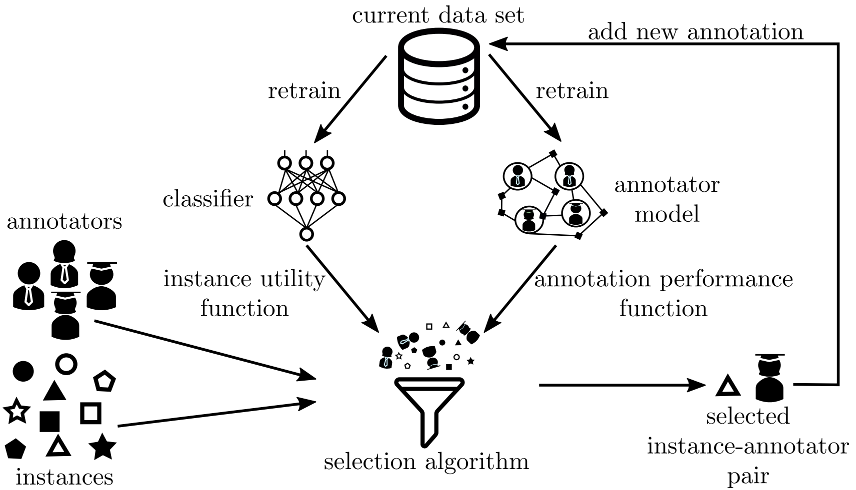
- estimates the performance of annotators as a function of instances.
- jointly selects an instance-annotator pair maximizing the classifier's expected performance.
- showed superior and robust performance compared to five related AL strategies in an experimental evaluation over various data sets.


current data set
annotators
instances

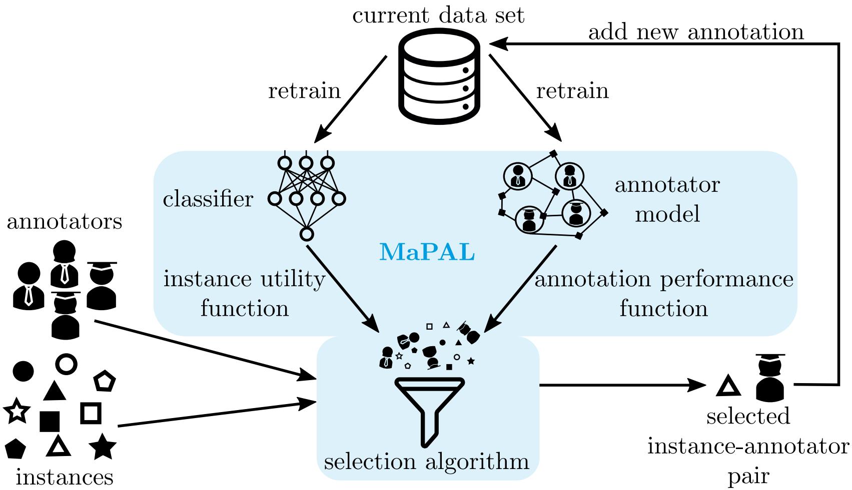
MaPAL
Structure



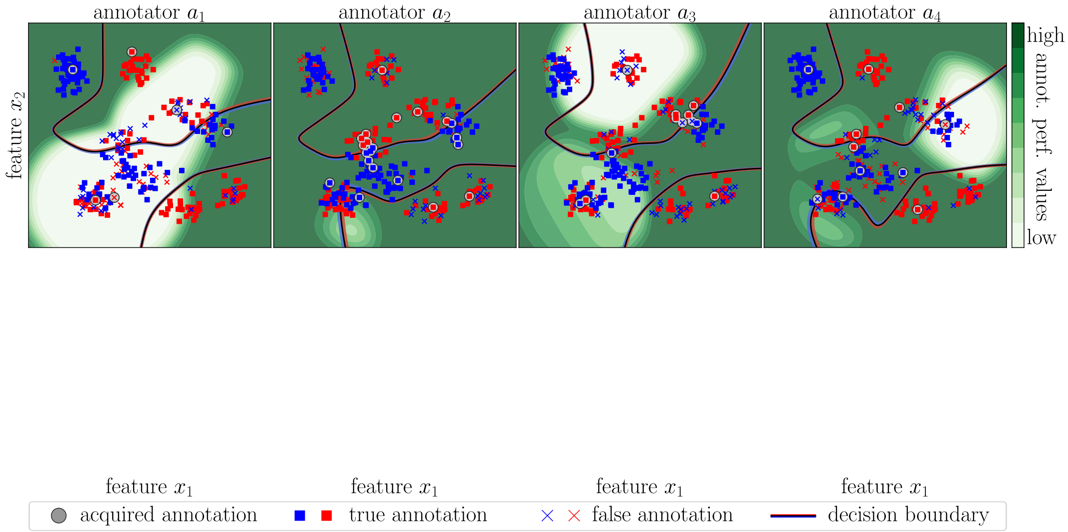

current data set
add new annotation
retrain
classifier
annotators
instance utility
function
instances
selection algorithm
selected
instance-annotator
pair


current data set
add new annotation
retrain
retrain
classifier
annotator
model
annotators
instance utility
function
annotation performance
function
selected
instance-annotator
pair
instances
selection algorithm

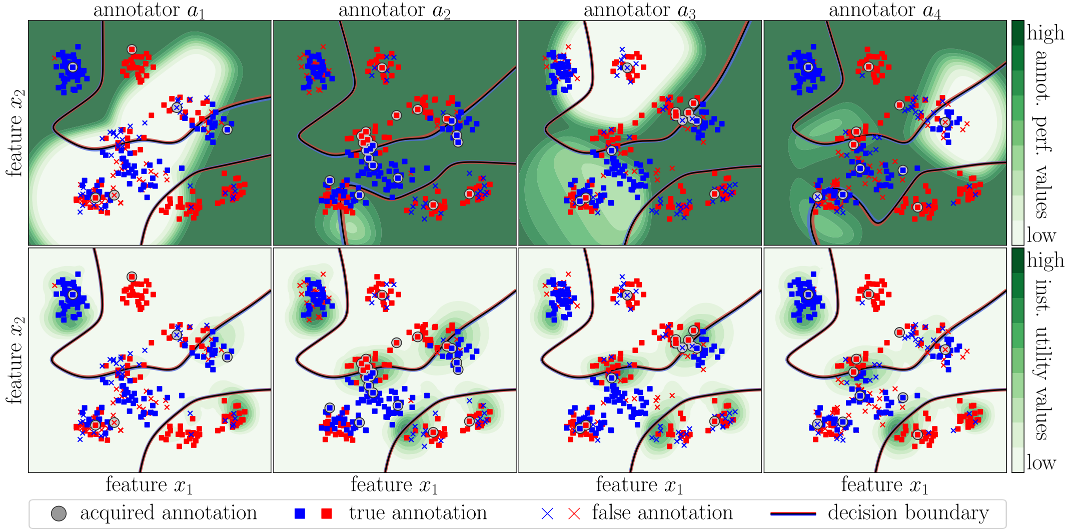
MaPAL
Structure



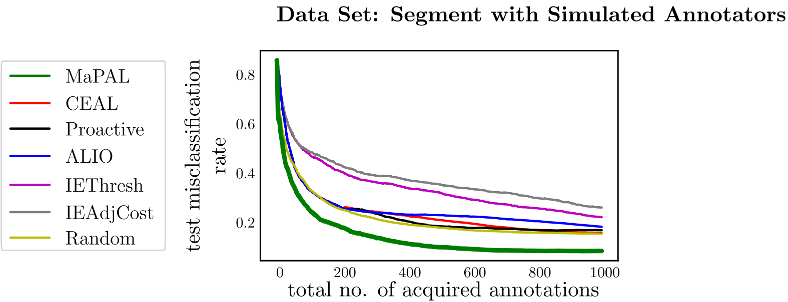
MaPAL
Visualization



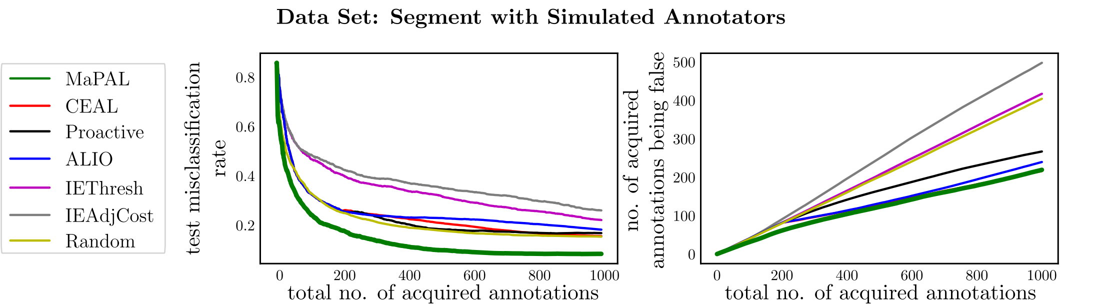
MaPAL
Visualization



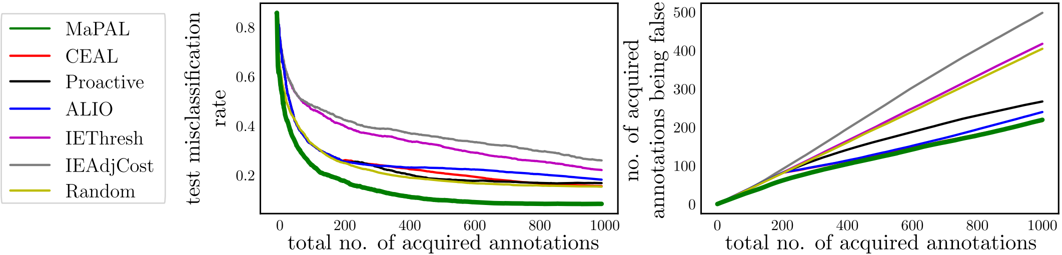
Evaluation
Learning Curves



Evaluation
Learning Curves



Data Set: Segment with Simulated Annotators



Further details on the experimental setup, additional results, and code are available at our repository

`https://github.com/mherde/mapal.`

Thank you for your attention!
Questions?