**ALWars: Combat-based Evaluation of Active Learning Strategies**

**Active Learning (AL) 101**
- Iterative, smart ordering of unlabeled samples for annotation
- Goal: prevent redundantly labeled samples

**Battle Mode**
- Replay of AL strategies with metrics for every AL iteration
- Allows deep understanding of AL strategies
- Reveals strategies time behaviour
- Based on AL framework ALiPy by Tang et al. with support for 25 AL strategies out-of-the-box ➔ easily extendable
- Support to upload new datasets
- Interactive (e.g. zoomable) plots

**Findings**
- Strategies strongly depend on dataset and experiment parameters
- One metric is not enough to compare AL strategies
- No clear superior AL strategy, often Random strategy performs comparably well

**Future Work**
- Large scale pre-computation of AL experiments
- Vision: AL benchmark

**Metrics**
- Standard ML metrics for test dataset
- AUC – area under learning curve, normalized by optimal curve

**Uncertainty Histogram**
- Majority of AL strategies use primarily some kind of uncertainty measure
- Uncertainty = 1 - classification probability

**Vector Space**
- PCA or configurable two features of dataset
- Labeled samples and current query highlighted

**Data Maps**
- Mean (confidence) and standard deviation (variability) of learner model
- Easy-to-learn, ambiguous, hard-to-learn samples can be identified

**Classification Boundaries**
- 2D plot with confidence displayed as classification boundary

**Uncertainty Histogram**
- Major part of AL strategies use primarily some kind of uncertainty measure
- Uncertainty = 1 - classification probability

**Classification Boundaries**
- Supported data types: HTML, text, images

**Vizualisation of AL query samples**
- Supported data types: HTML, text, images

**Learning Curve**
- ML metric in relation to AL iterations

**Vector Space**
- PCA or configurable two features of dataset
- Labeled samples and current query highlighted

**Future Work**
- Iterative, smart ordering of unlabeled samples for annotation
- Goal: prevent redundantly labeled samples

**Findings**
- Strategies strongly depend on dataset and experiment parameters
- One metric is not enough to compare AL strategies
- No clear superior AL strategy, often Random strategy performs comparably well

**GitHub Repository**
https://github.com/etikedi/etikedi/