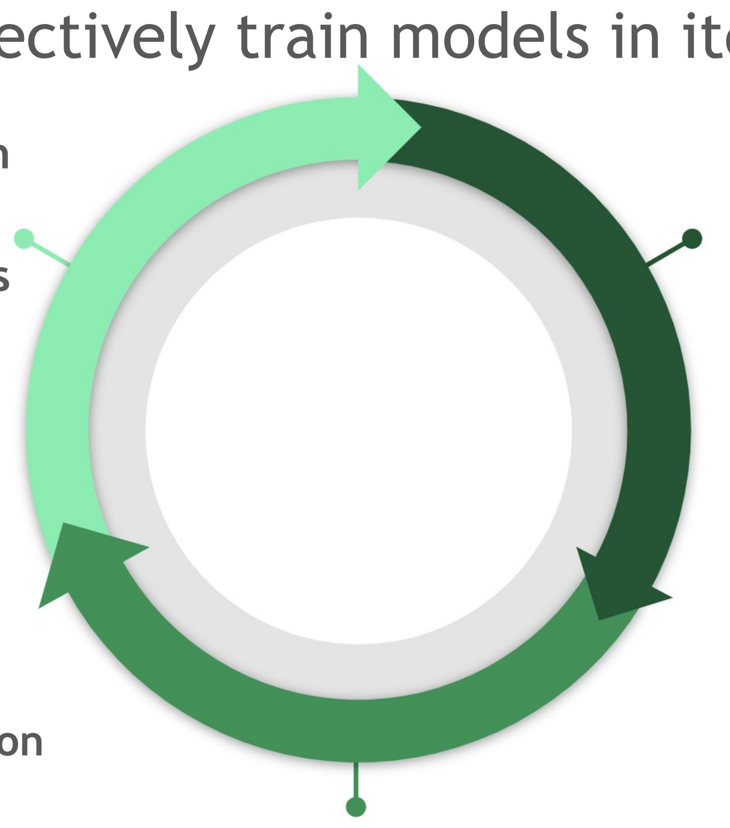
Low-resource Interactive Active Labeling for Fine-tuning Language Models

OSAKA UNIVERSITY
Megagon Labs

Seiji Maekawa, Dan Zhang, Hannah Kim, Sajjadur Rahman, Estevam Hruschka

1. Low-resource Active Learning (AL)

Active Learning: Acquires informative samples for human labeling to effectively train models in iterations.



Low-resource setting
(< 1000 samples acquired per iteration)

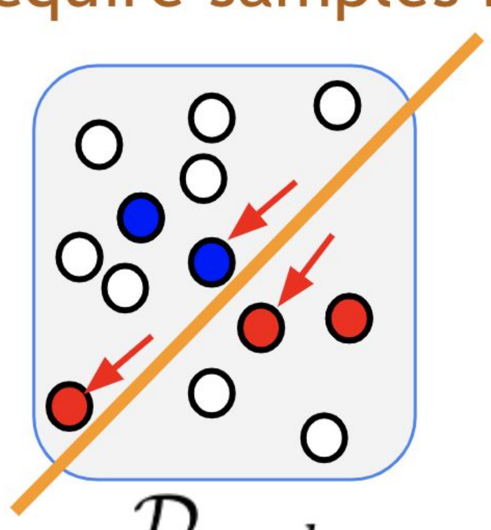
2. Problem Setting

- AL has been used to fine-tune LMs for NLP tasks
  - *sentiment analysis, document classification, …*
- Existing methods prioritize accuracy
  - *often overlooks labeling cost and iteration latency*
- Adoption of AL in practical settings such as labeling platforms can be challenging
  - *requires balancing all three objectives*
  - *adapt to different datasets and tasks*

How can we develop an AL method that balances *accuracy*, *cost*, and *latency* for diverse NLP tasks?

3. Acquisition Strategies for Active Learning

Acquire samples near decision boundary

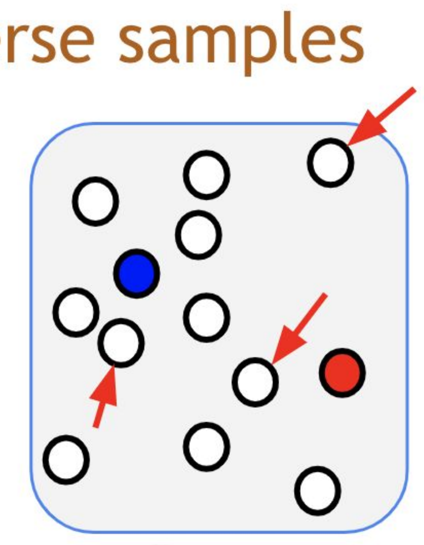


$\mathcal{D}_{pool}$: *Pool of unlabeled data*

: uncertainty-based acquisition

- Leverages model's predictive confidence
- Acquires low confidence samples

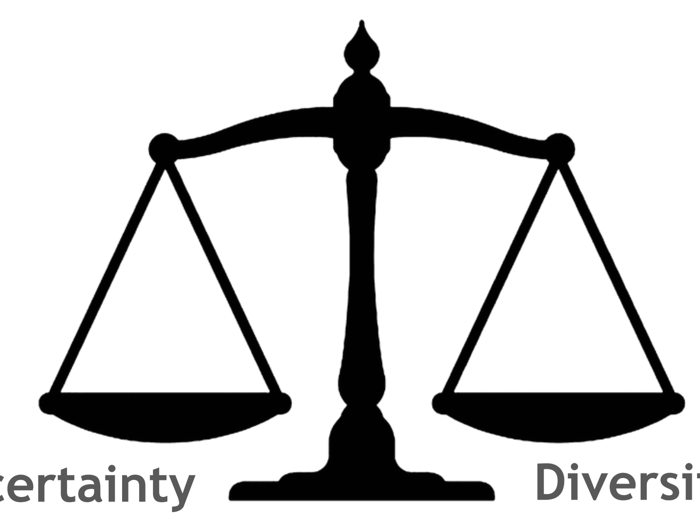
Acquire diverse samples



: diversity-based acquisition

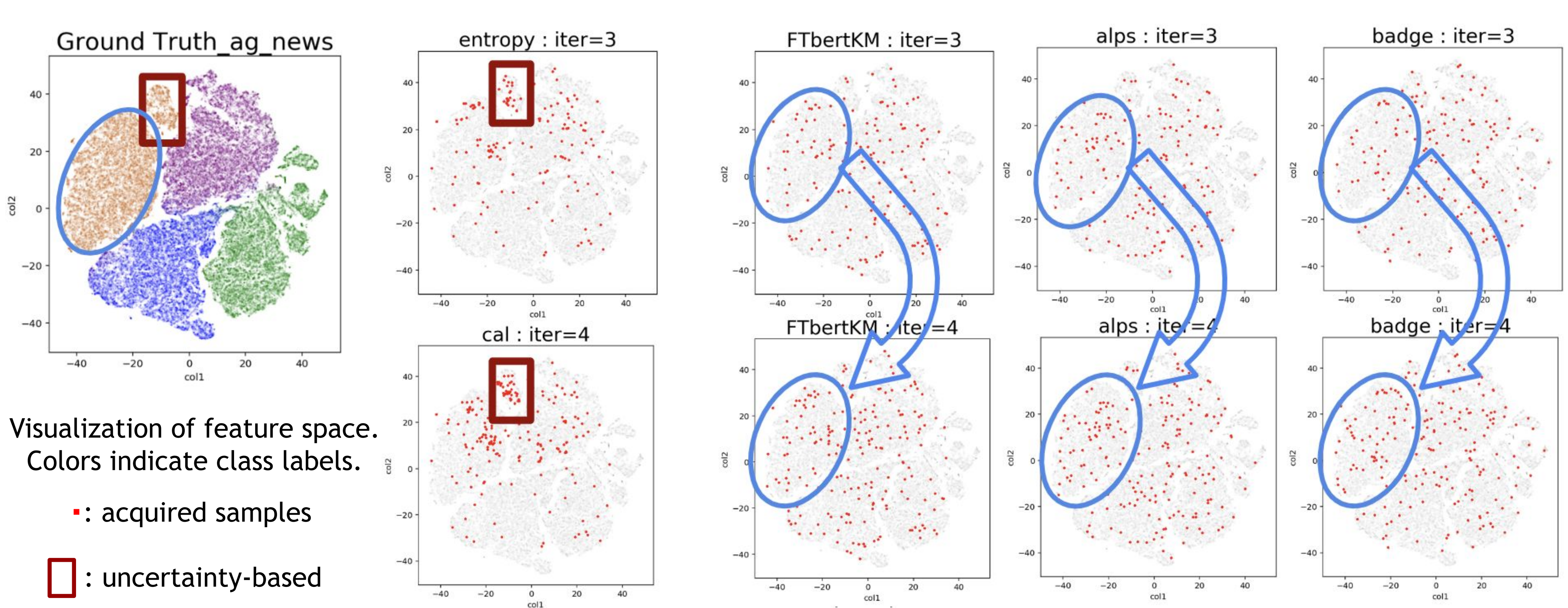
- Prioritizes coverage of classes
- Acquires a diverse group of samples

Hybrid strategies



- Integrate both diversity and uncertainty
- Acquire samples based on a dual-objective function

4. Limitations of Active Learning for Fine-tuning LMs



Visualization of feature space. Colors indicate class labels.

: acquired samples

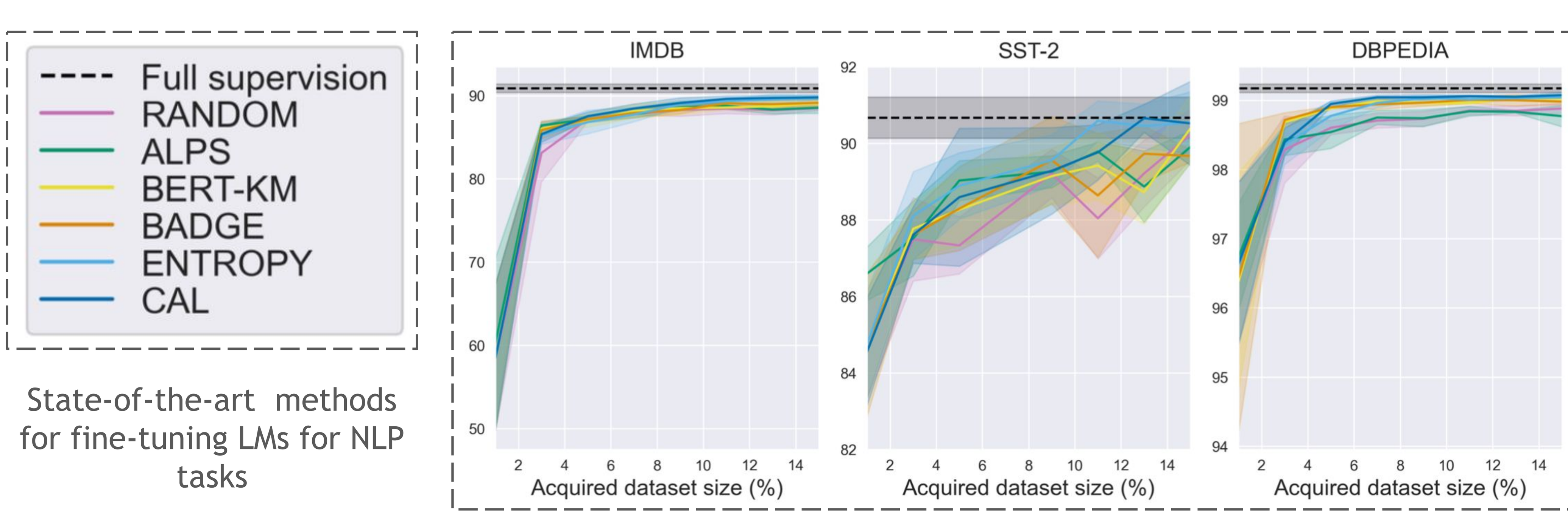
: uncertainty-based

: diversity-based

Intra-iteration redundancy:
- uncertainty-based methods prioritize samples near model's decision boundary
- tend to acquire similar samples in an iteration

Inter-iteration redundancy:
- diversity-based methods ignore model's confidence and prioritize coverage
- tend to acquire similar samples across iterations

Intra- and ***inter-***iteration sampling redundancies emerge in diversity- and uncertainty-based methods



State-of-the-art methods for fine-tuning LMs for NLP tasks

Unintended increase in the overall labeling budget
- *acquires* **redundant** samples in each iteration

Latency of sample acquisition hampering interactivity
- leverages the **entire unlabeled data** for acquisition decision

Marginal gain in accuracy compared to cost and latency
- partly due to **redundant** sample acquisition

Lack of adaptability to diverse datasets
- due to **one-size-fits-all** acquisition strategies

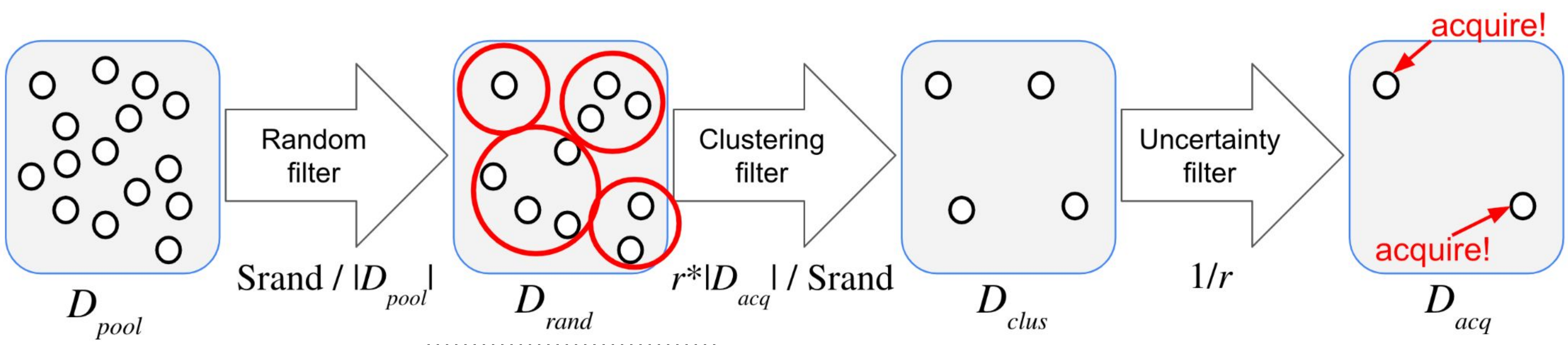
Low-resource Interactive Active Labeling for Fine-tuning Language Models

OSAKA UNIVERSITY

Megagon Labs

Seiji Maekawa, Dan Zhang, Hannah Kim, Sajjadur Rahman, Estevam Hruschka

5. Tyrogue: Interactive, Adaptive, and Hybrid Active Learning



D_{pool} : *unlabeled data pool*

D_{rand} : randomly sampled data. Srand = |D_{rand}|

r : control parameter for uncertainty-based acquisition

D_{acq} : acquired labeling candidates

Tyrogue overview

Three step filtering approach to balance accuracy, latency, and cost: ***random*** → ***diversity-based*** → ***uncertainty-based***

- random filtering lowers acquisition latency by reducing the candidate pool (a reasonable D_{rand} ensures comparable accuracy)
- clustering filter ensures acquisition of diverse samples leading to better coverage and lesser redundancy
- uncertainty filter acquires samples that the model is least certain about to improve predictive confidence

Adaptive acquisition by balancing diversity and uncertainty

- by varying r users can steer the acquisition strategy of Tyrogue to adapt for diverse NLP tasks and dataset types

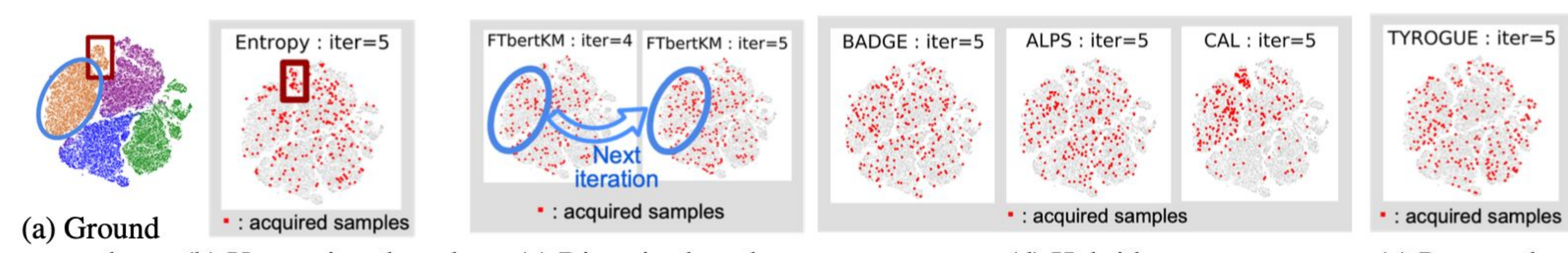
6. Experiment Setup

Methods
- ***Diversity-based acq.***: FTbertKM, Random
- ***Uncertainty-based acq.***: Entropy
- ***Hybrid acq. strategy*** : CAL, ALPS, BADGE

Scope
- ***Tasks***: Sentiment analysis, Topic classification, Natural language inference, Paraphrase detection

Metric
- **Average labeling cost** (given target accuracy)
- **Acquisition time** (per iteration)

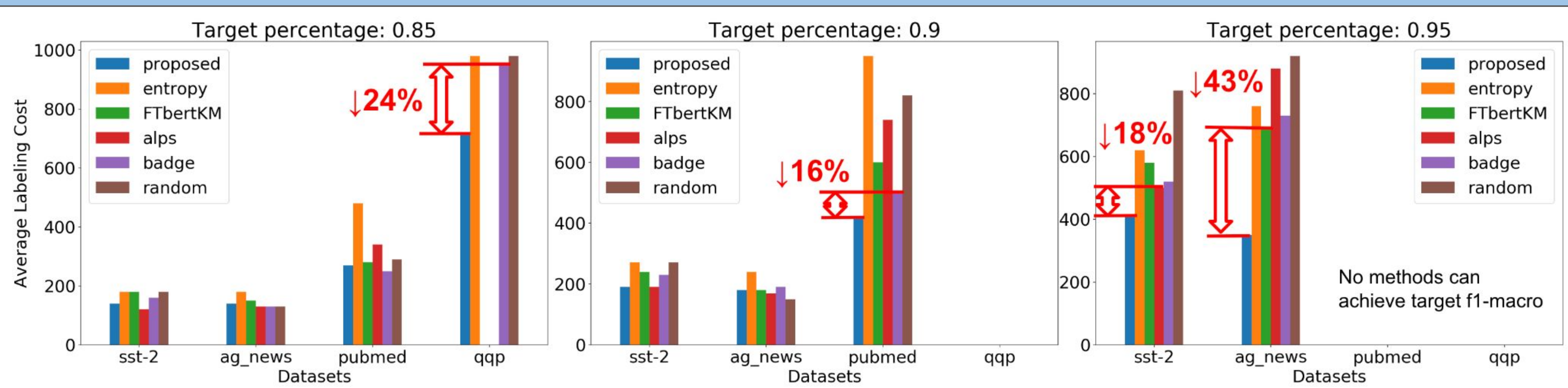
7. Effective Utilization of Labeling Budget



(a) Ground truth. (b) Uncertainty-based (c) Diversity-based (d) Hybrid. (e) Proposed.

Tyrogue (e) minimizes redundant sampling compared to SOTA methods

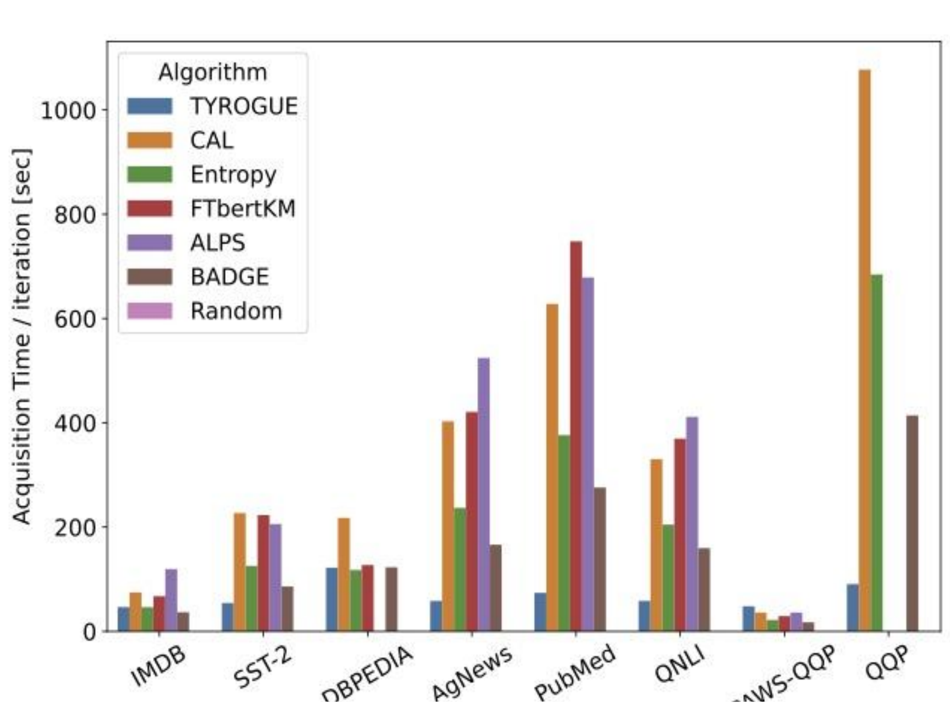
8. Labeling Cost Reduction with Comparable Accuracy



Minimization of redundant sampling reduces labeling cost

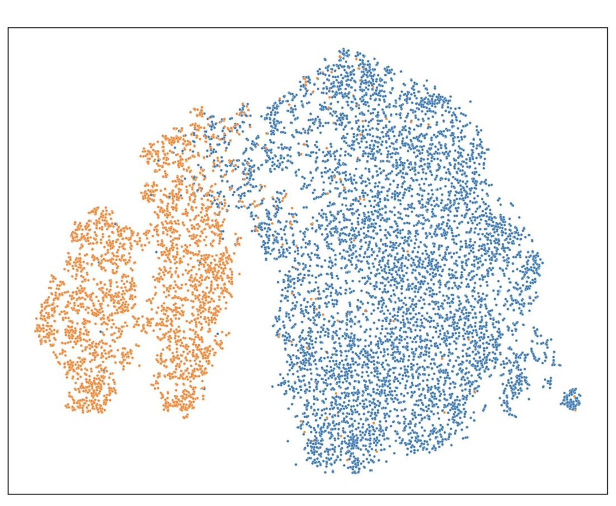
Despite reduced cost, **Tyrogue** exhibits comparable accuracy to SOTA

9. Low Latency Sampling



- SOTA methods exhibit **higher** latency
- Tyrogue is **indifferent** to dataset scale
- The impact of random filtering is more apparent for **larger datasets**

10. Impact of Datasets

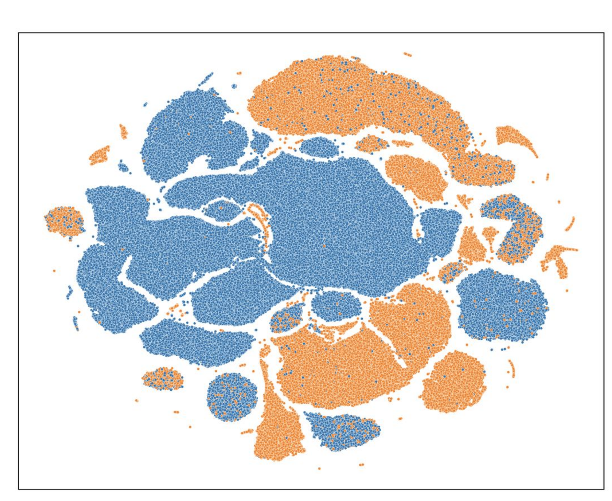


Uncertainty-based methods perform well for simple class distributions

PAWS-QQP

Hybrid methods such as Tyrogue perform well for complex class distributions

QQP



11. Concluding Remarks

TYROGUE

We develop a low-resource interactive AL method that minimizes labeling cost (by up to **43%**) and acquisition latency (by up to **11X**) while achieving comparable accuracy to SOTA methods

Future focus: **integrate** with labeling platforms, apply meta-learning to enable **automated** adaptive acquisition, build **transparent** methods to explain acquisition decisions