Promoting Fairness in Learned Models by Learning to Active Learn under Parity Constraints

Amr Sharaf
University of Maryland
amr@cs.umd.edu

Hal Daumé III
University of Maryland
Microsoft Research
me@hal3.name

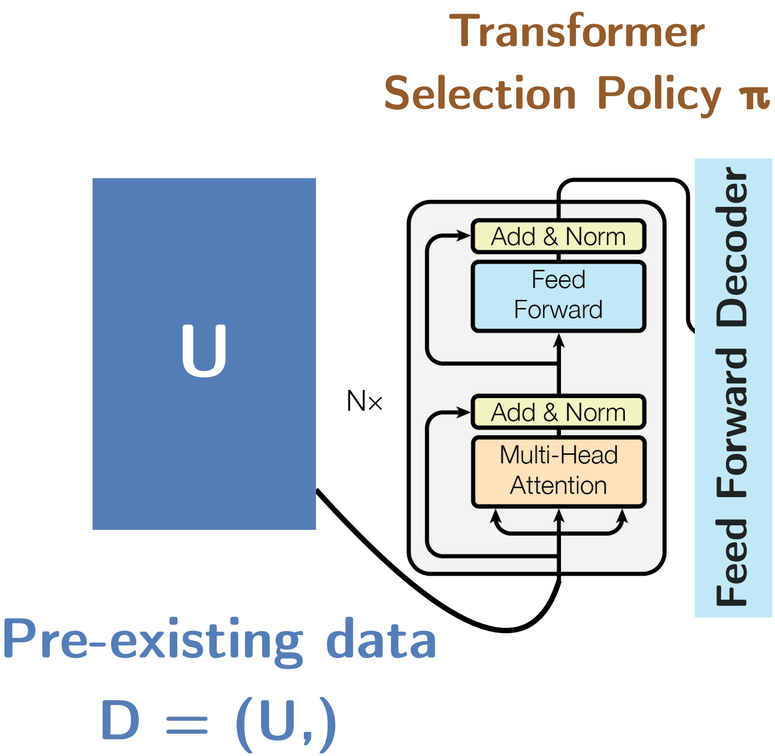
Can we learn to active learn under fairness parity constraints?

PANDA Test Time Behavior

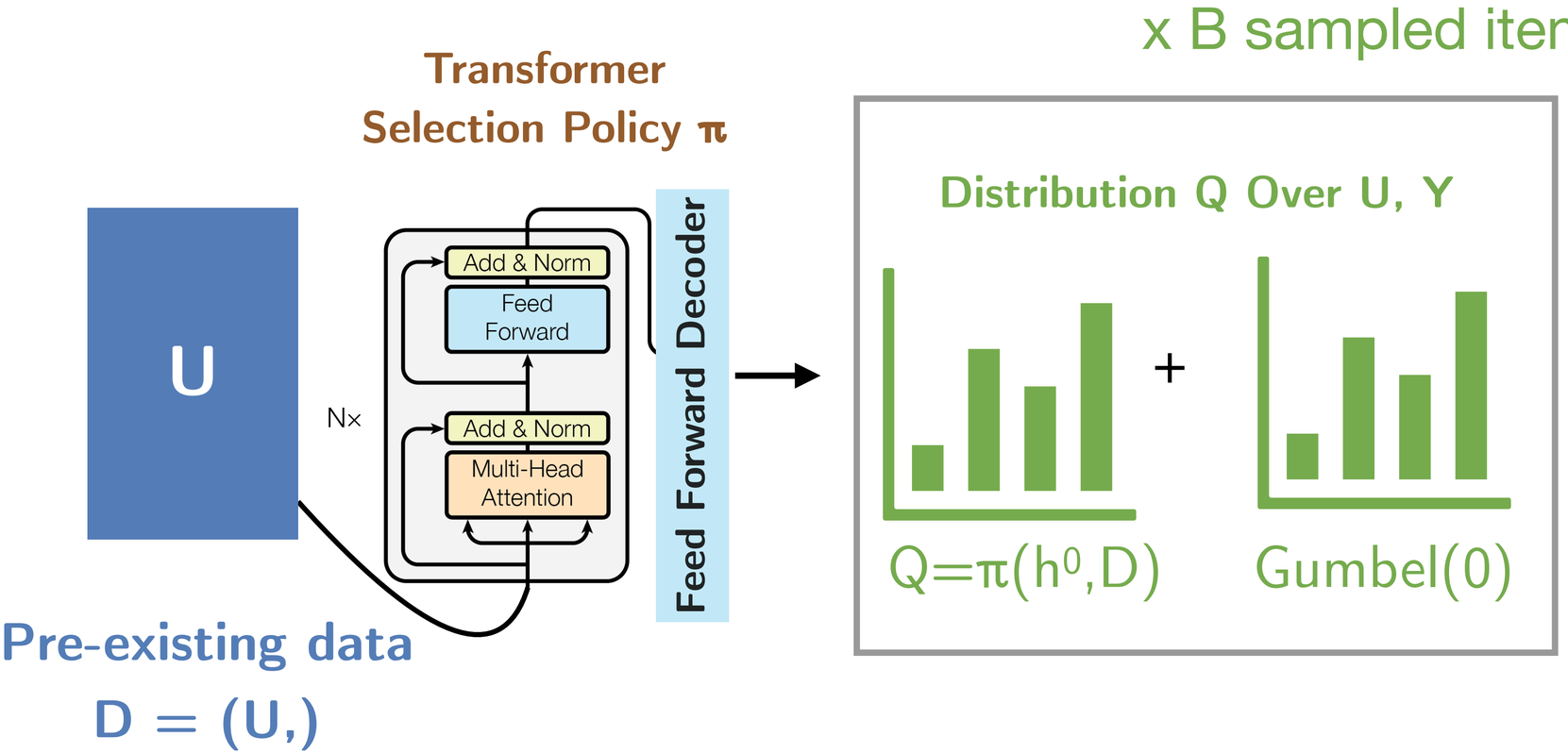
Pre-existing data

$D = (U,)$

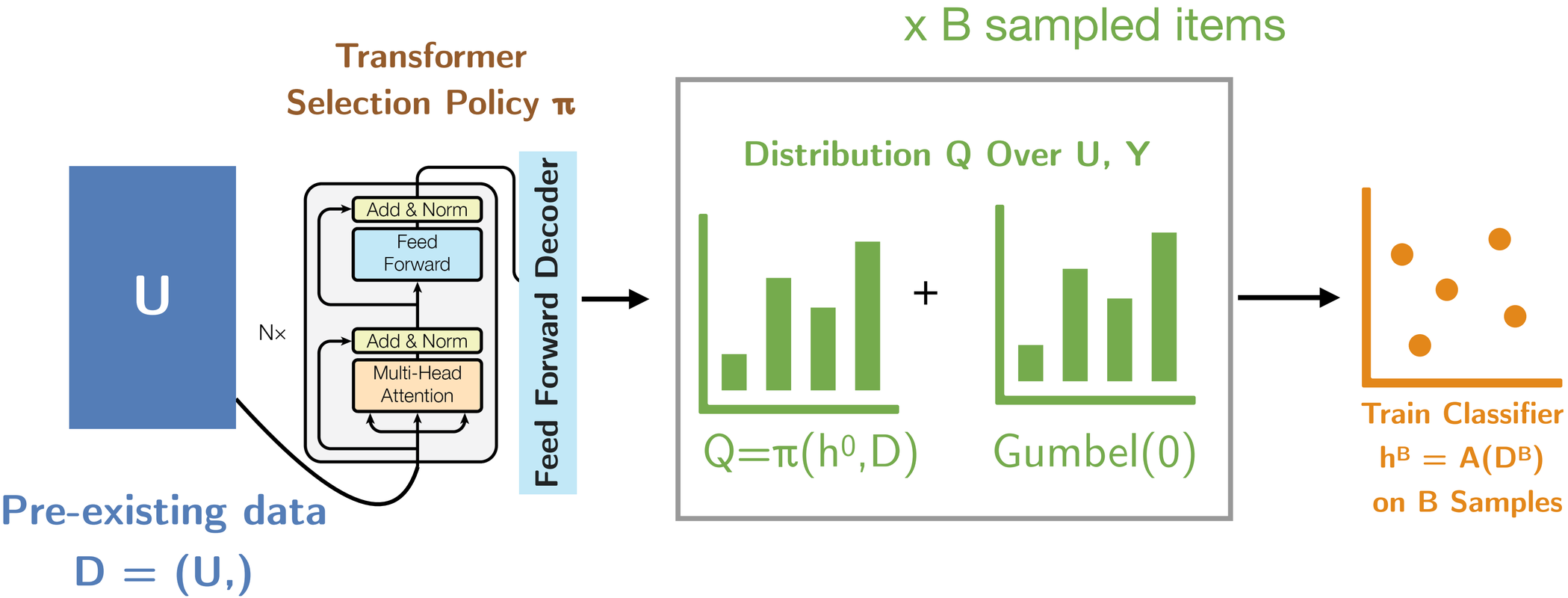
PANDA Test Time Behavior



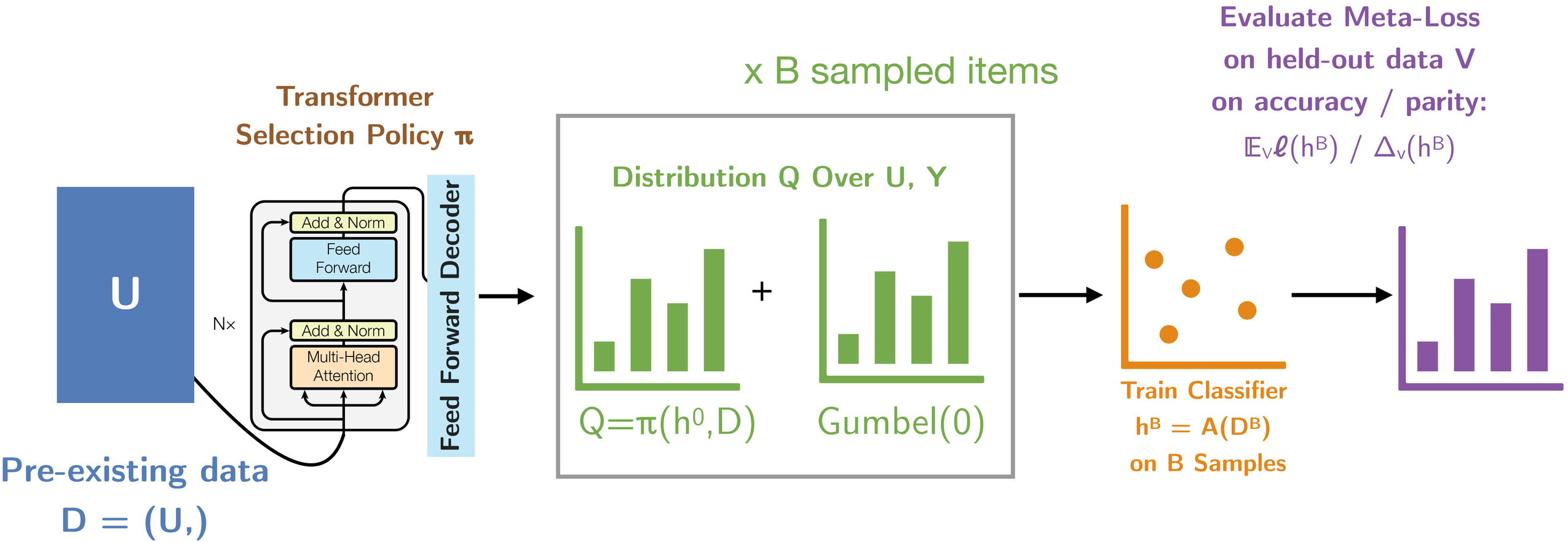
PANDA Test Time Behavior



PANDA Test Time Behavior

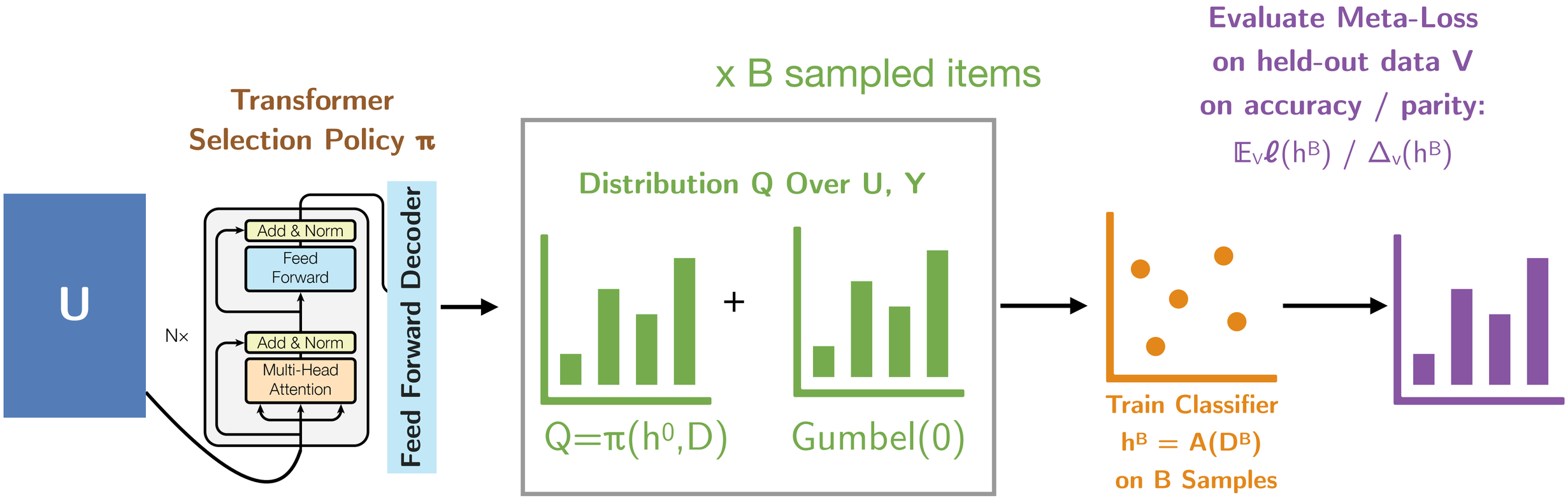


PANDA Test Time Behavior

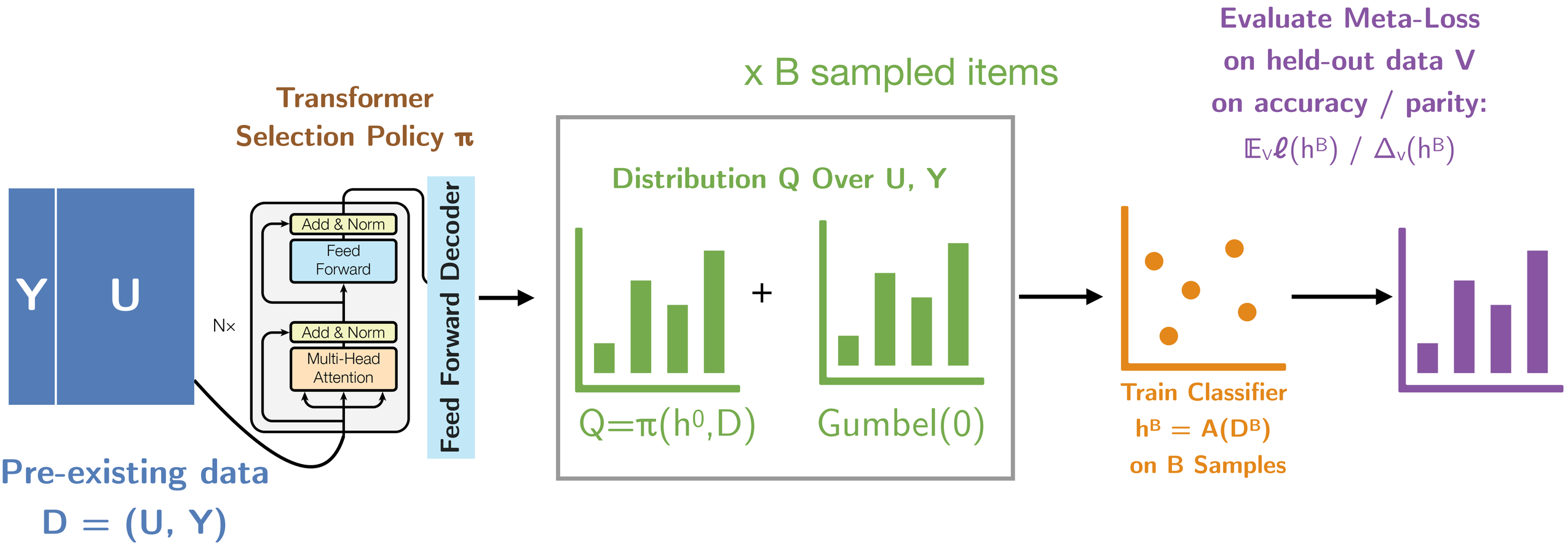


Goal: can we manage an efficacy vs annotation cost trade-off under a target parity constraint?

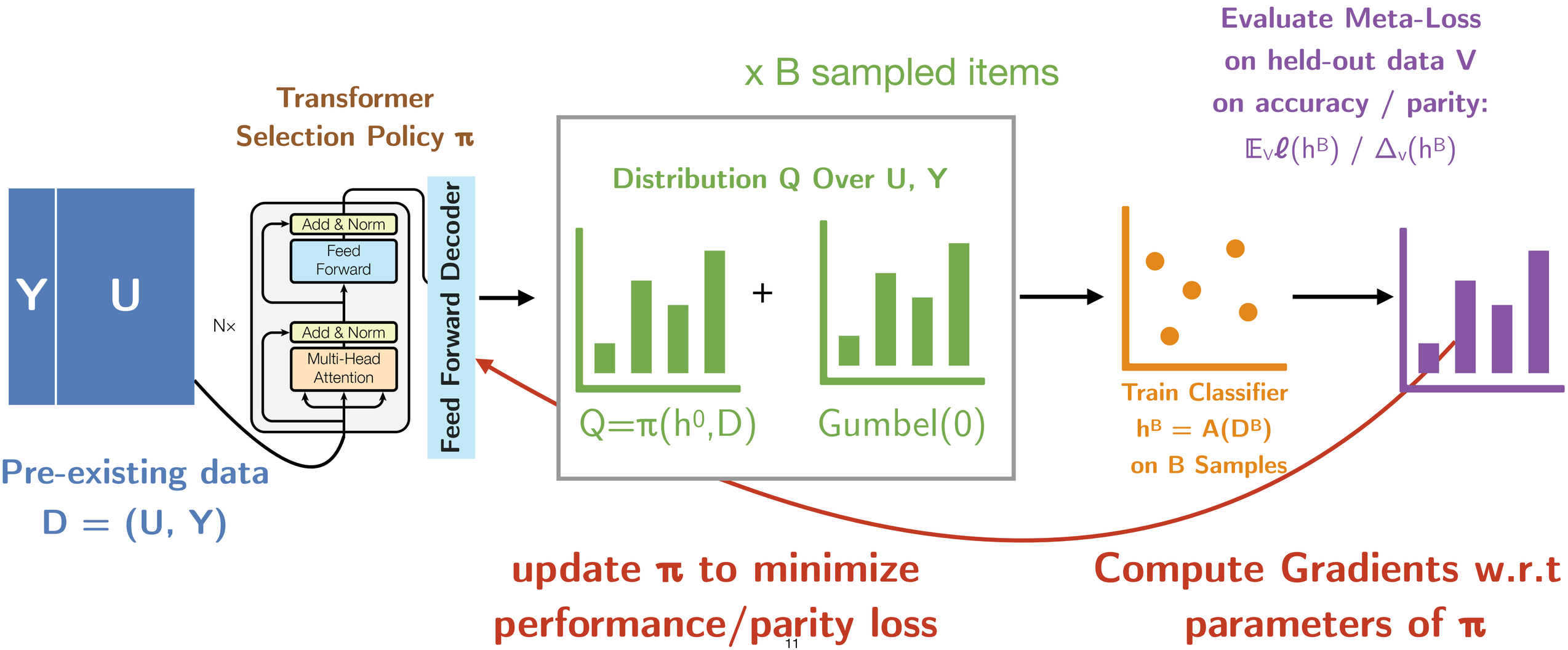
PANDA Train Time Behavior



PANDA Train Time Behavior

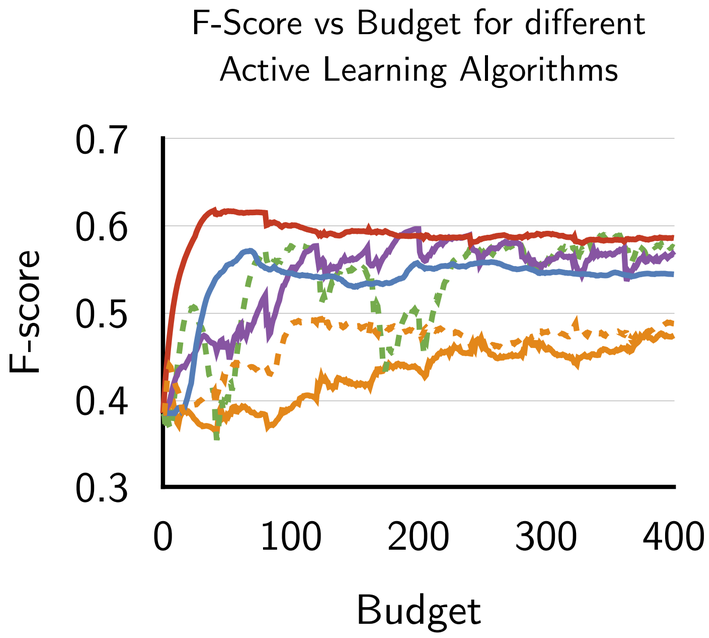


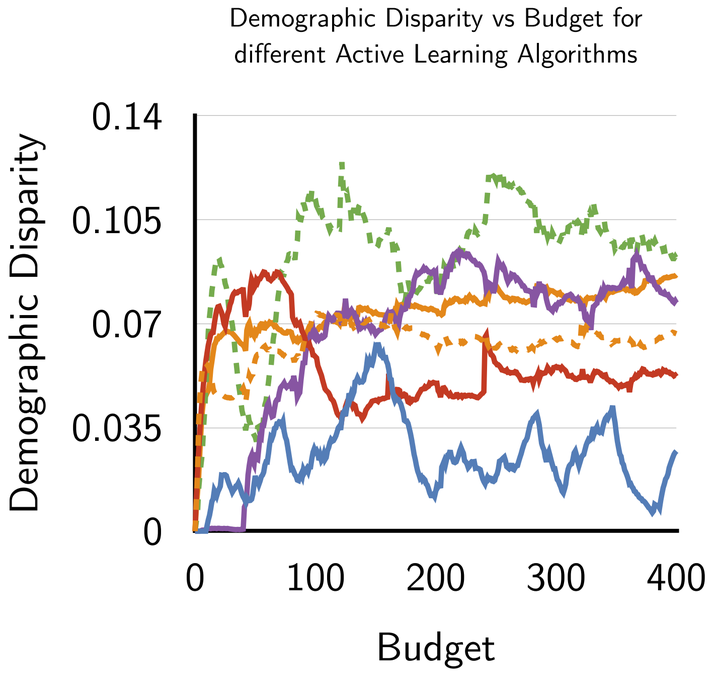
PANDA Train Time Behavior



Experimental Results

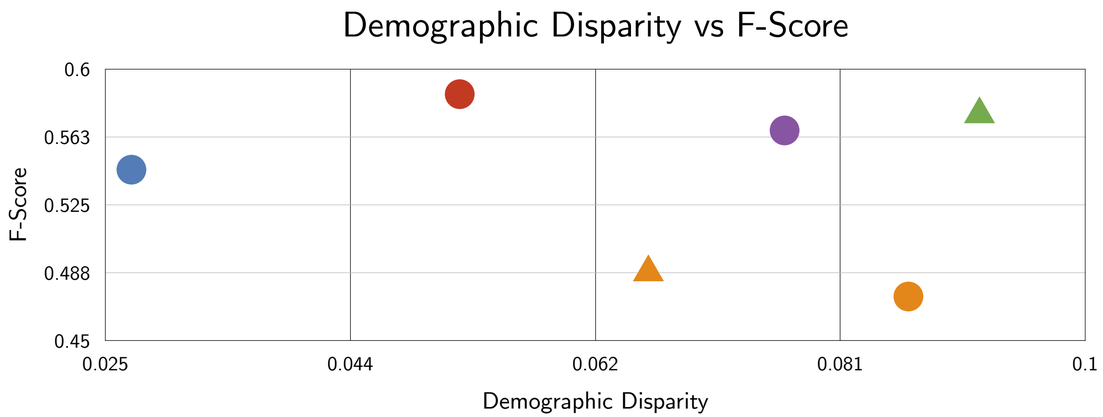
Experimental Results

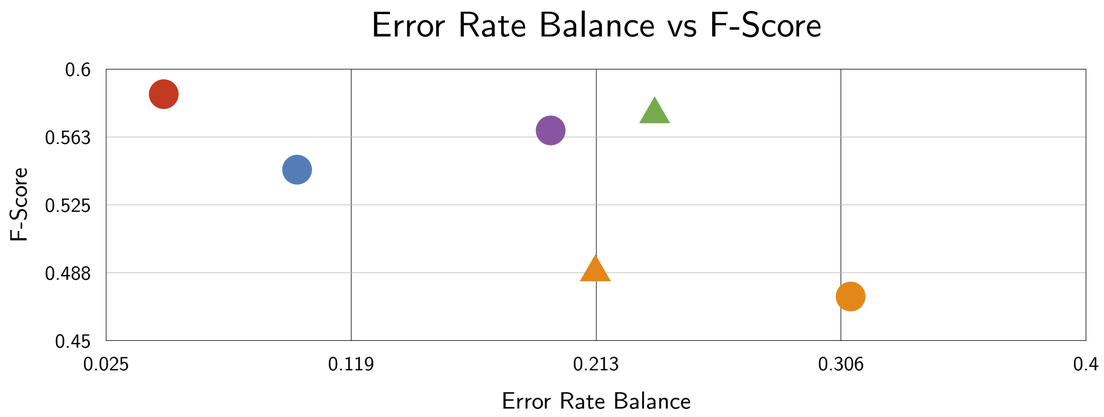




Random Sampling — Fairlearn — PANDA — Fair Active Learning — Entropy Sampling — Group Aware Random Sampling

Experimental Results





Random Sampling Fairlearn PANDA Fair Active Learning Entropy Sampling Group Aware

Conclusion

- Q: Can we learn to active learn under fairness parity constraints?
- A: Yes, using meta-learning + Forward Backward Splitting;
- We compare to alternative active learning strategies;
- PANDA outperforms alternative strategies in most setting.

Questions?

amr@cs.umd.edu