



Visual Supervision in Bootstrapped Information Extraction

Matthew Berger^{1,2} Ajay Nagesh¹ Joshua A. Levine¹ Mihai Surdeanu¹ Hao Helen Zhang¹

¹University of Arizona

²Vanderbilt University

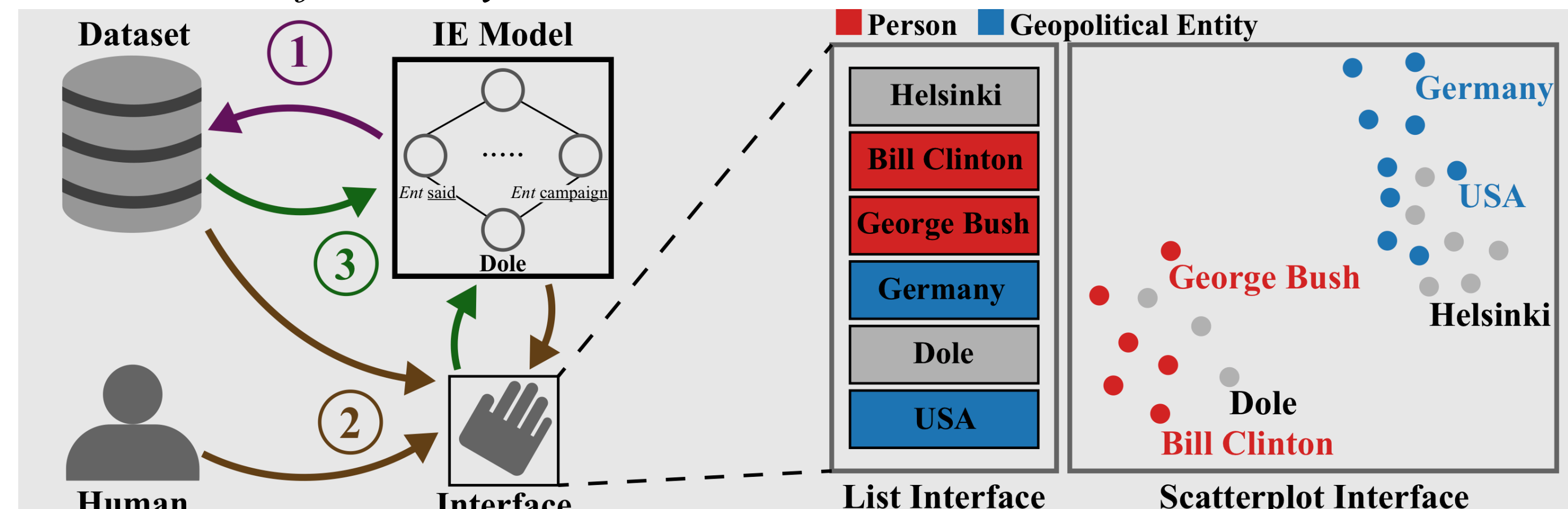


Motivation

Data Annotation in Information Extraction: list-based interface of entities ranked by uncertainty

Is this the best we can do?

We study different visual interfaces, sampling criteria, and interactions for entity annotation.



- **1:** Bootstrapping model automatically labels entities, performs update
- **2:** Human labels entities in visual interface
- **3:** Model updates based on human labels

Embedding-Based Bootstrapping

Model: semi-supervised word embeddings, distributional similarity of *entities* with *patterns*.

Entity $\mathbf{x}_e \leftarrow \boxed{\text{John said on Saturday}} \rightarrow \mathbf{x}_p$ Pattern

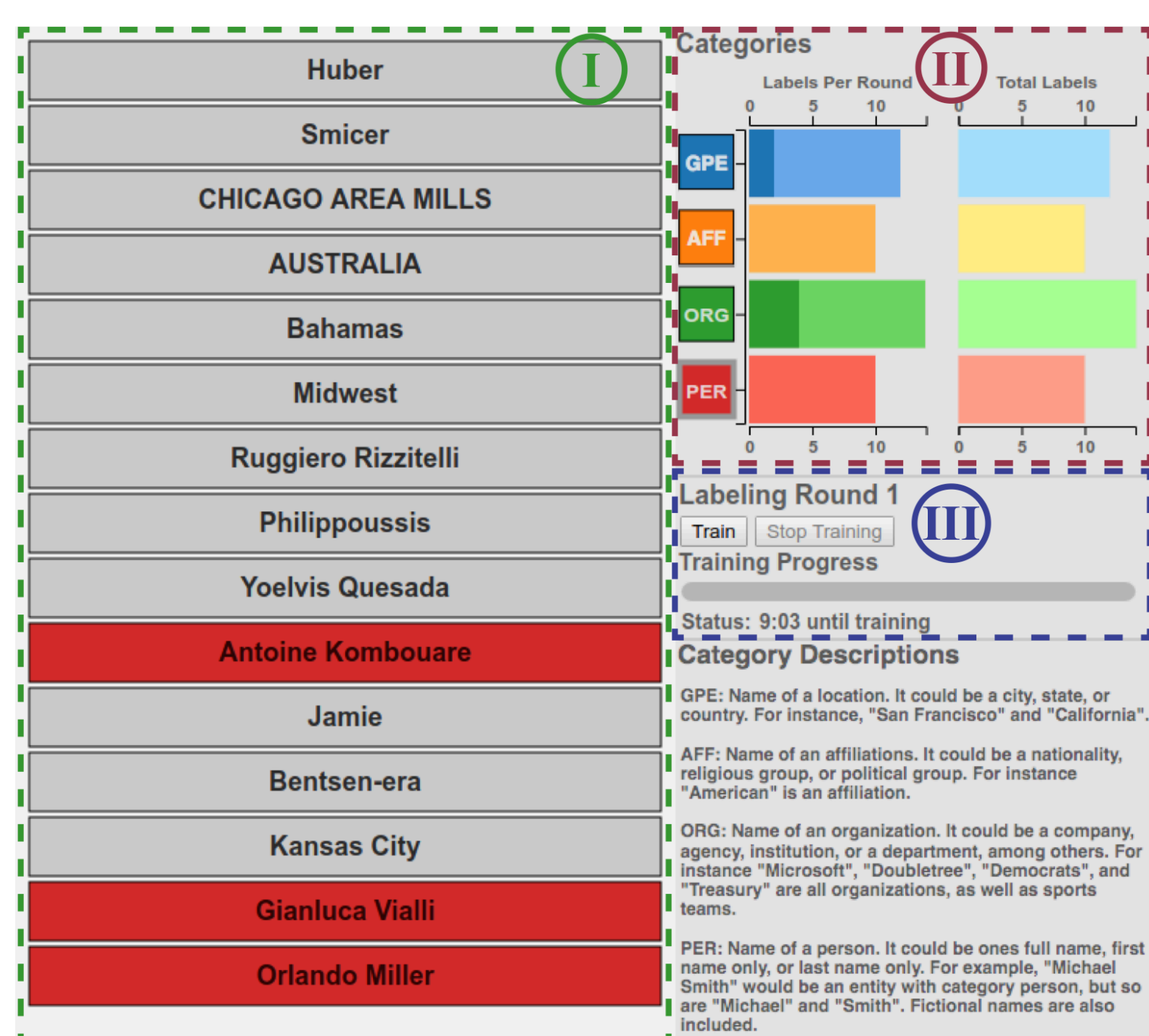
Unsupervised: Skip-Gram

$$SG = \sum_{(e,p) \in C_p} [\log(\sigma(\mathbf{x}_e^\top \mathbf{x}_p)) + \sum_{n \in N} \log(\sigma(-\mathbf{x}_e^\top \mathbf{x}_n))]$$

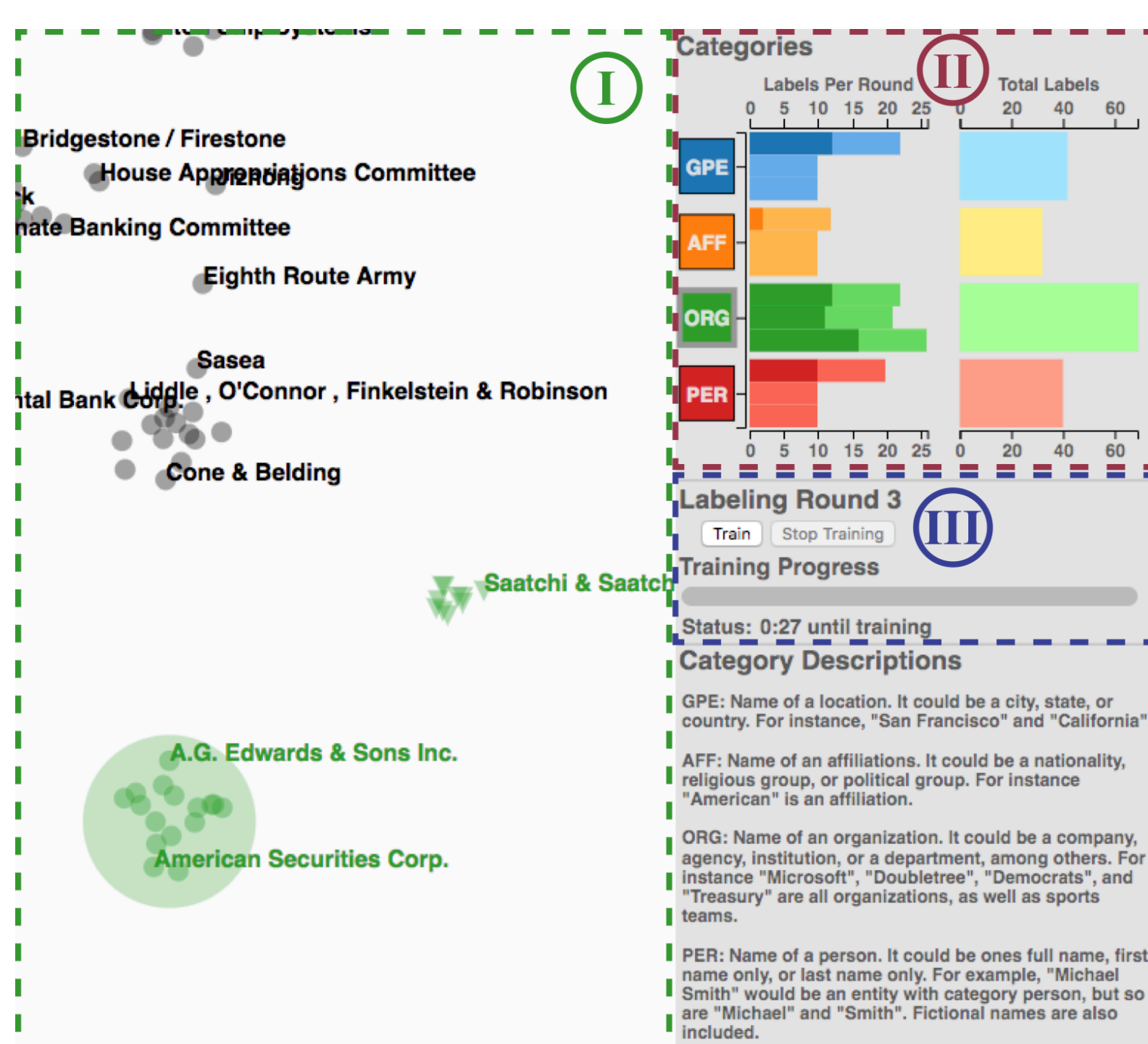
Supervised: Margin-based Hinge Loss

$$LM = \sum_{(a,b,c) \in E_l \rightarrow \text{labeled data (human/machine-promoted)}} [s(\mathbf{x}_a, \mathbf{x}_c) - s(\mathbf{x}_a, \mathbf{x}_b) + M]_+$$

Visual Interfaces



List



Scatterplot

- **I: labeling interface** – list with uncertainty sampling
- **II: entity labeling** – individual entity selection
- **III: bootstrapping** – round progression

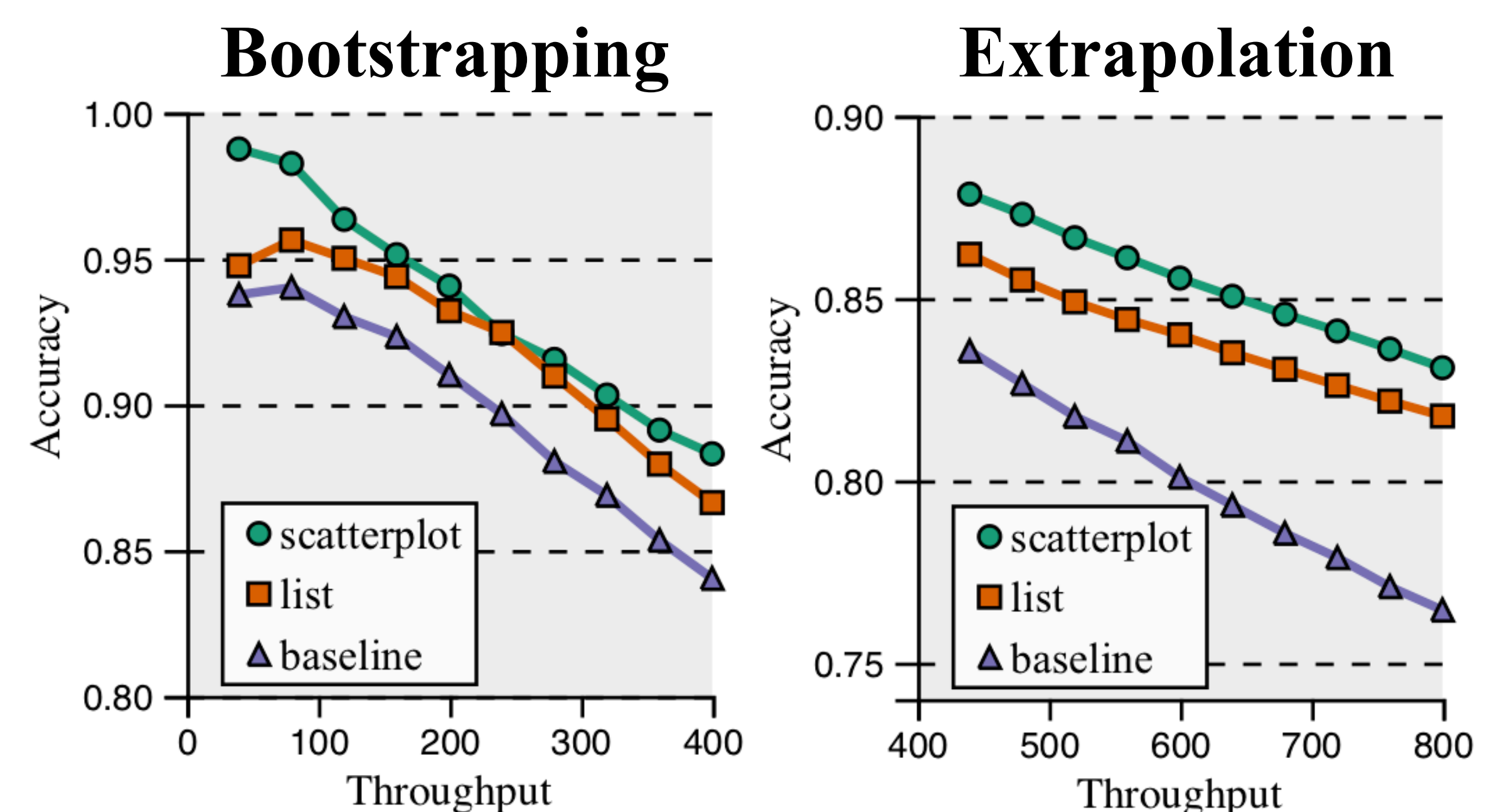
- **I: labeling interface** – 2D scatterplot with clustering & coverage sampling
- **II: entity labeling** – group-wise selection via area brushing
- **III: bootstrapping** – view projection of entities during training, determine advancement

User Study

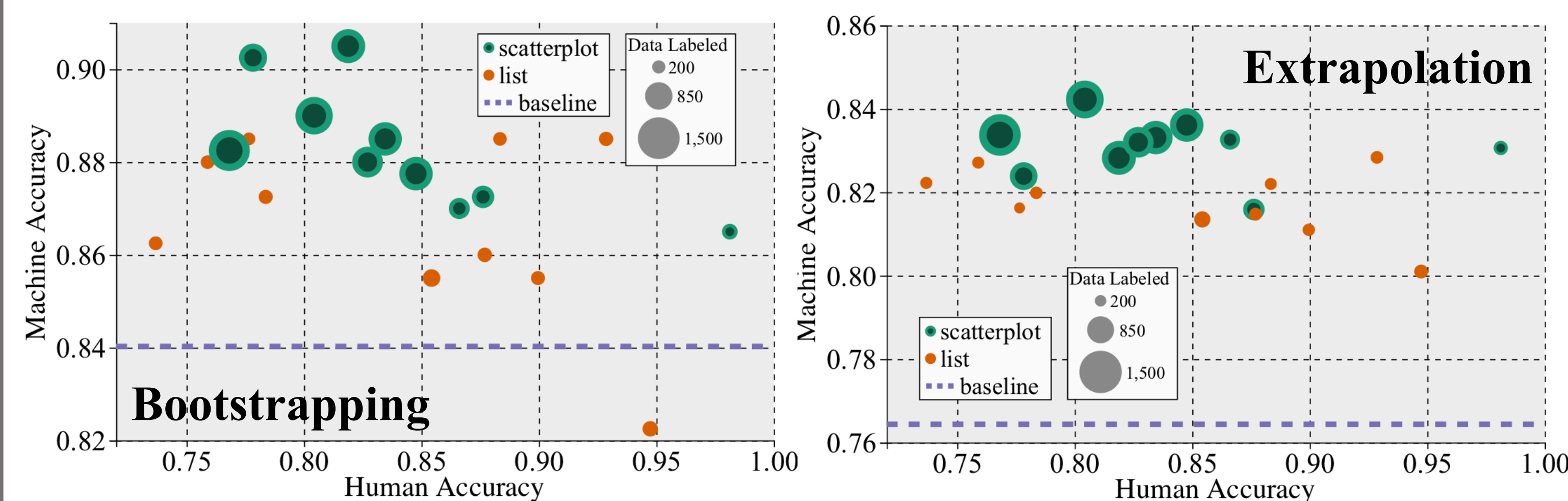
- User Base: 10 participants, within-subject design, e.g. presented with both interfaces.
- Dataset: Ontonotes (Weischedel et al. 2013), limited to 4 categories: people, organizations, geopolitical entities, religious/political affiliations
- Experimental Setup: participant performs 10 rounds of labeling, label entities for up to 1 minute in each round
- Evaluation: bootstrapping throughput of entities promoted in each round, as well as *extrapolation*, throughput of bootstrapper after obtaining all human labels.

Results

Bootstrapping Throughput Over Each Round

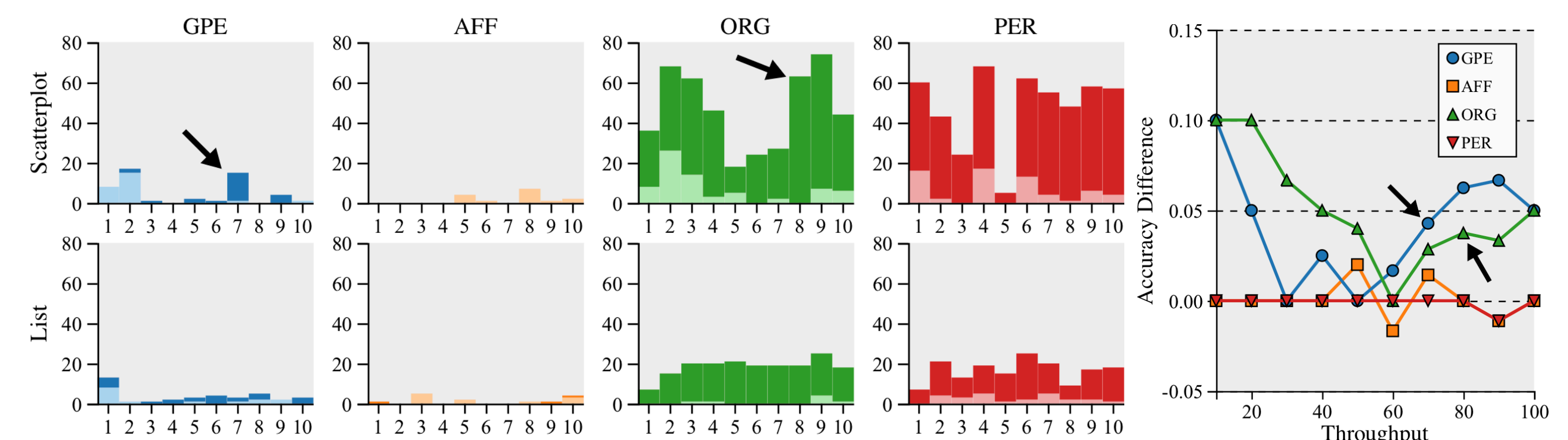


Participant performance at last round



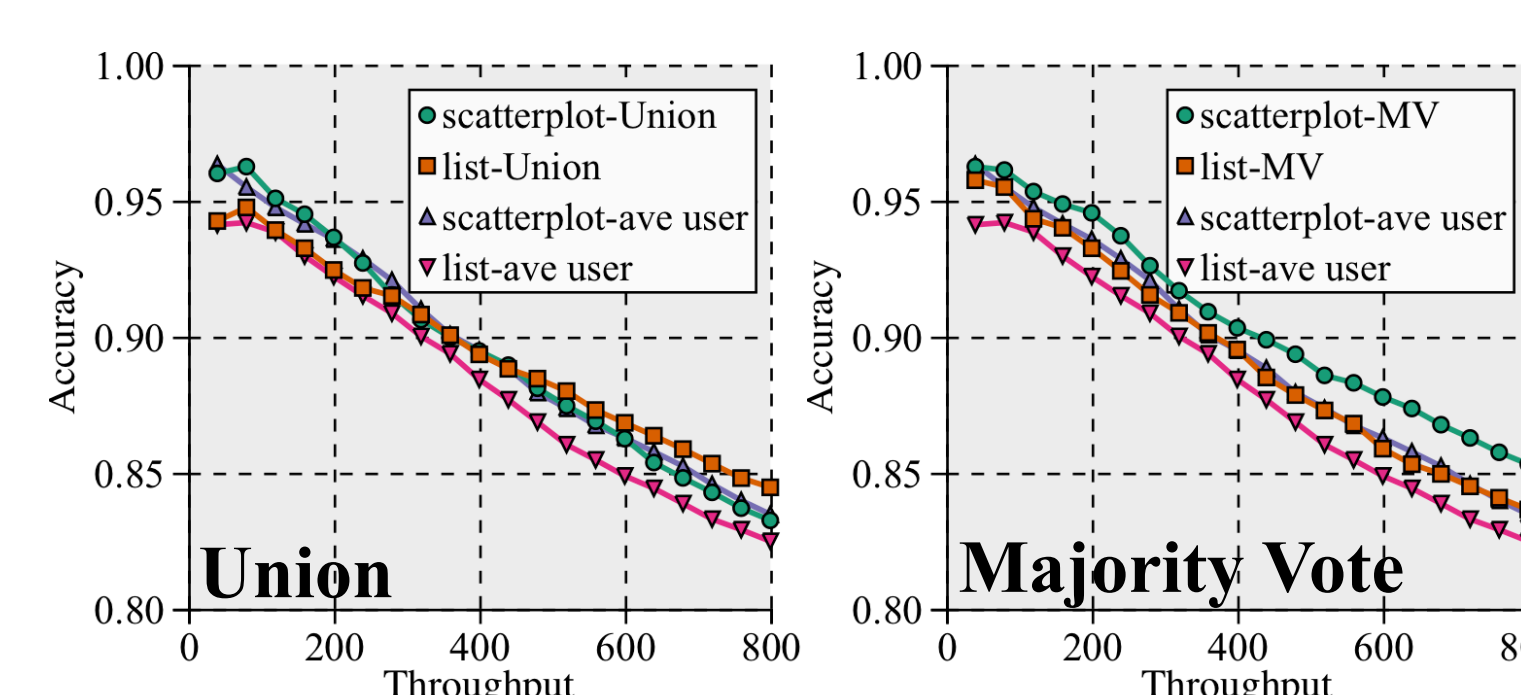
Takeaway: aim for an interface that yields *many annotations at reasonable accuracy*, rather than few, higher-quality ones

Example Participant Performance



Takeaway: labeling volume can counter noise in annotations

Labeling Consensus



Takeaway: can tolerate label noise, but not too much noise

Acknowledgements: NSF TRIPODS grant (NSF-1749858) and the Bill and Melinda Gates Foundation HBGDKi Initiative. Mihai Surdeanu discloses a financial interest in lum.ai. This interest has been disclosed to the University of Arizona Institutional Review Committee and is being managed in accordance with its conflict of interest policies.

Contact: matthew.berger@vanderbilt.edu