Crowdsourcing Multi-Label Classification

Jonathan Bragg Mausam Daniel S. Weld

University of Washington Computer Science & Engineering

Introduction

- AI can dramatically reduce crowdsourcing costs
- Binary / multiclass classification well-studied
- **Multi-label classification**
  - Common problem (e.g., tags on websites)
  - Previously unoptimized
- Our most sophisticated multi-label classifier:
  - Uses **learned model of label co-occurence** to infer true item-label relationships
  - Chooses questions that **maximize value of information** toward a joint classification

Problem Definition

- For each item, find the **subset of labels** that apply
  - n = # of items in dataset
  - m = # of labels
  - nm binary classification problems

Threshold Models

- Standard threshold voting
  - k votes per binary classification problem
  - Accept iff at least T out of k votes are positive
- Observation: k votes not always needed
- **Model #1: Lossless stopping**
  - Stop after T positive votes (or k-T+1 negatives)
  - No error compared to requesting all k votes
- **Model #2: One-away heuristic**
  - Stop after T-1 positive votes and no negative votes (or k-T negatives and no positives)
  - Small amount of error

Probabilistic Models

- **Model #3: Independent**
  - Assume labels are independent
  - m label parameters, 2 noisy worker parameters

$$P(label \mid votes) \propto P(label) \prod_{v \in votes} P(v \mid label)$$

- **Model #4: Multi-Label Naive Bayes (MLNB)**
  - Model pairwise label co-occurrence
  - $O(m^2)$ parameters

$$P(label \mid otherLabels) \propto P(label) \prod_{L \in otherLabels} P(L \mid label)$$

- **Submodularity** enables greedy selection of votes on labels that achieves (1-1/e) ≈ 63% of optimal

Threshold Results

- Optimize categorization step in Cascade algorithm for crowdsourcing taxonomy creation [2]
- n = 100 items, m = 33 labels, k = 15 votes

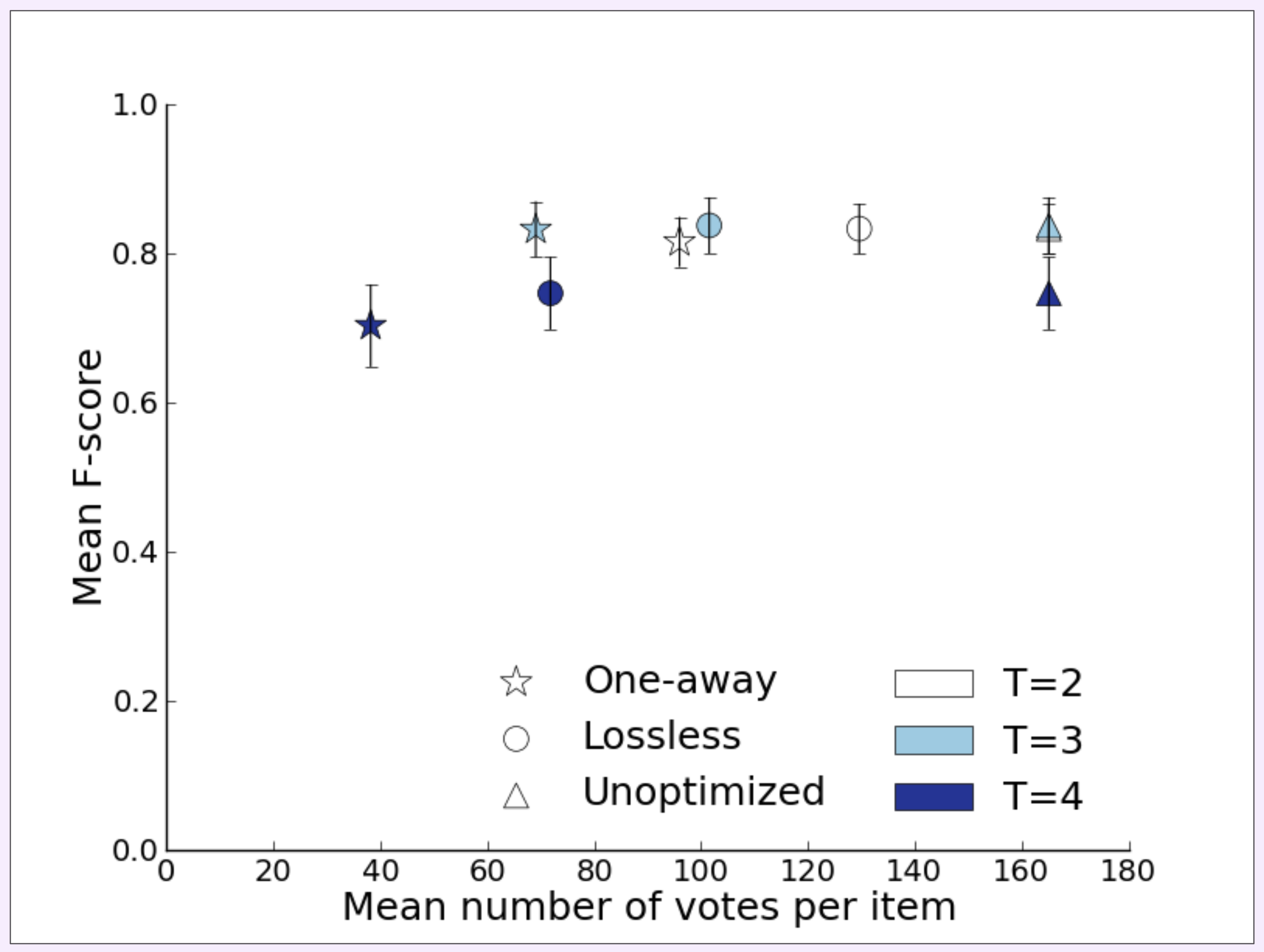

Figure 1: Lossless stopping reduces labor by 56% when T=4. One-away heuristic saves 58% when T=3.

Probabilistic Results

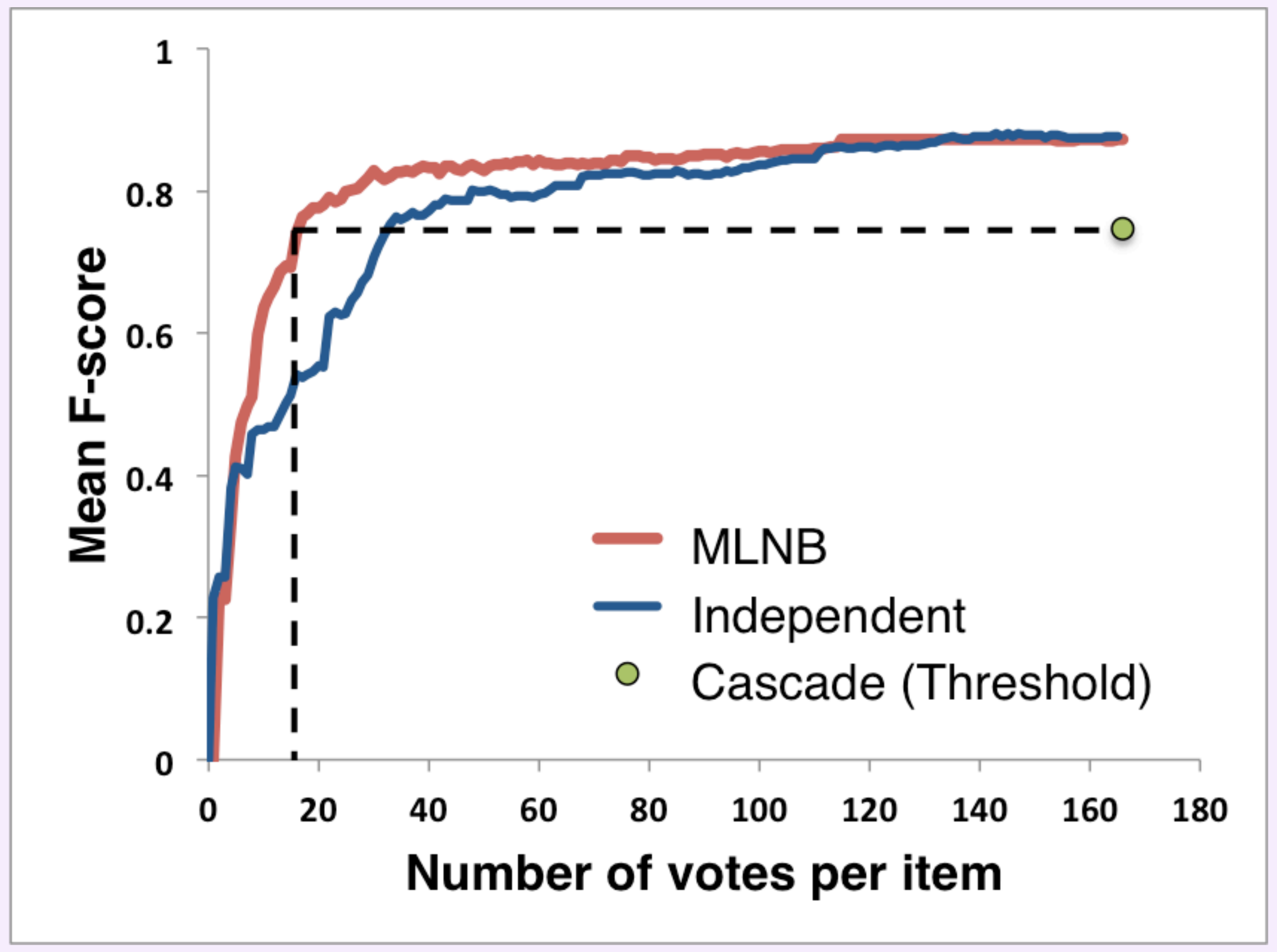

Figure 2: MLNB matches performance of Cascade (standard threshold voting, T = 4 out of 5) after 16 votes per item, compared to 165 used by Cascade.

Conclusions

- **MLNB uses less than 10% as much labor** as Cascade (standard threshold voting)
- Extension: selecting batches of questions (labels)
  - Useful in online labor marketplaces
  - Little reduction in accuracy
- See [1] for full results

References

[1] Bragg, J.; Mausam; and Weld, D. S. Crowdsourcing multi-label classification for taxonomy creation. In *Proceedings HCOMP '13*. To appear.

[2] Chilton, L. B.; Little, G.; Edge, D.; Weld, D. S.; and Landay, J. A. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings CHI '13*.

For more details, visit jonathanbragg.com/hcomp13

Acknowledgements: This work was supported by the WRF / TJ Cable Professorship, Office of Naval Research grant N00014-12-1-0211, and National Science Foundation grants IIS 1016713 and IIS 1016465.