A Correlated Worker Model for Grouped, Imbalanced and Multitask Data

An T. Nguyen [1] Byron C. Wallace Matthew Lease

University of Texas at Austin

UAI 2016

[1] Presenter

Overview

- A model of workers in crowdsourcing.

Overview

- A model of workers in crowdsourcing.
- Idea: Transfer knowledge of worker quality.

Overview

- A model of workers in crowdsourcing.
- Idea: Transfer knowledge of worker quality.
- Variational EM learning.

Overview

- A model of workers in crowdsourcing.
- Idea: Transfer knowledge of worker quality.
- Variational EM learning.

- Apply to two datasets:
  - Biomed Citation Screening: imbalanced, grouped.

Overview

- A model of workers in crowdsourcing.
- Idea: Transfer knowledge of worker quality.
- Variational EM learning.

- Apply to two datasets:
  - Biomed Citation Screening: imbalanced, grouped.
  - Galaxy Classification: multiple tasks.

Background

- Crowdsourcing: collect labels quickly at low cost.

Background

- Crowdsourcing: collect labels quickly at low cost.
- But (usually) lower quality.

Background

- Crowdsourcing: collect labels quickly at low cost.
- But (usually) lower quality.
- Common solution: collect 5 labels for each instance ...
- ... then aggregate them.

Background

- Crowdsourcing: collect labels quickly at low cost.
- But (usually) lower quality.
- Common solution: collect 5 labels for each instance ...
- ... then aggregate them.

- Most previous work: improve (the estimates of) **labels**.

Background

- Crowdsourcing: collect labels quickly at low cost.
- But (usually) lower quality.
- Common solution: collect 5 labels for each instance ...
- ... then aggregate them.

- Most previous work: improve (the estimates of) **labels**.
- Our work: improve (the estimates of) **worker qualities**.

Motivation

for estimating worker qualities

Motivation

for estimating worker qualities

Diagnostic insights.

Motivation
for estimating worker qualities

Diagnostic insights.

Help workers improve.

Motivation

for estimating worker qualities

Diagnostic insights.

Help workers improve.

Intelligent task routing (assign work to workers).

Worker Quality Measure

Accurary: simple but not enough.

Worker Quality Measure

Accurary: simple but not enough.

$\rightarrow$ Confusion matrix: $\Pr(\text{worker label}|\text{true label})$

Worker Quality Measure

Accurary: simple but not enough.

$\rightarrow$ Confusion matrix: $\Pr(\text{worker label}|\text{true label})$

Binary task (this work):

- Sensitivity: $\Pr(\text{positive}|\text{positive})$.
- Specificity: $\Pr(\text{negative}|\text{negative})$.

Setting

Input

- Crowd labels for each instance.
- No instance-level features (future work).

Setting

Input

- Crowd labels for each instance.
- No instance-level features (future work).

Output

- For each worker: sensitivity and specificity.

Setting

Input

- Crowd labels for each instance.
- No instance-level features (future work).

Output

- For each worker: sensitivity and specificity.

Eval. Metric

- RMSE on sen. and spe.

Setting

Input

- Crowd labels for each instance.
- No instance-level features (future work).

Output

- For each worker: sensitivity and specificity.

Eval. Metric

- RMSE on sen. and spe.
- gold sen. spe.: gold labels in whole dataset.

Challenges

Sparsity: many workers do only a few instances.

Challenges

Sparsity: many workers do only a few instances.

Data is imbalanced:

- A lot more negative than positive
- Difficult to estimate sensitivity

Idea

Transfer knowledge of worker quality

- Between classes.
- Within group.
- In multiple tasks.

Previous models

(Raykar et. al. 2010; Liu & Wang 2012; Kim & Ghahramani 2012)

Hidden vars:

- True label for each instance.
- Confusion mat. (sen. + spe.) for each worker.

Previous models

(Raykar et. al. 2010; Liu & Wang 2012; Kim & Ghahramani 2012)

Hidden vars:

- True label for each instance.
- Confusion mat. (sen. + spe.) for each worker.

Assumptions:

- Sen. & Spe. are independent params.
- A single group of workers.
- Multiple tasks: independent models.

Our Model

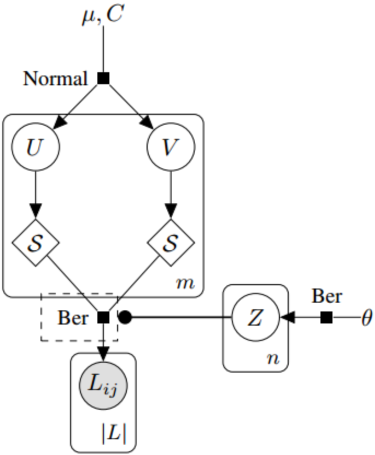
Assumptions:

- Sen. & Spe. are correlated.
- Multiple groups of workers (group membership is known).
- Sen. & Spe. in multiple tasks are correlated.

The Base Model

(i indexes instances, j indexes workers)

$$U_j, V_j \sim \mathcal{N}(\mu, C)$$

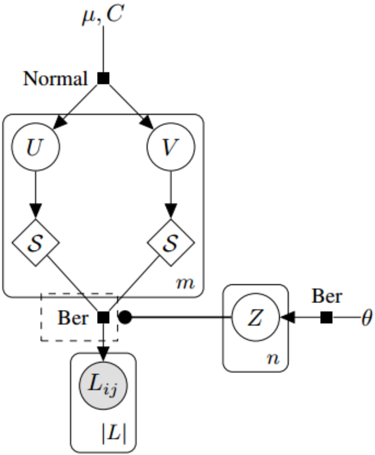


The Base Model

(i indexes instances, j indexes workers)

$$U_j, V_j \sim \mathcal{N}(\mu, C)$$

$$Z_i \sim \mathrm{Ber}(\theta)$$



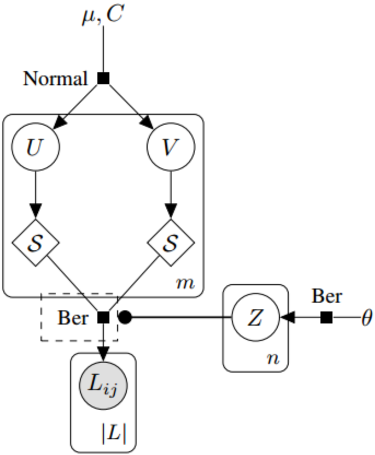
The Base Model

(i indexes instances, j indexes workers)

$$U_j, V_j \sim \mathcal{N}(\mu, C)$$

$$Z_i \sim \mathrm{Ber}(\theta)$$

$$L_{ij} | Z_i = 1 \sim \mathrm{Ber}(\mathcal{S}(U_j))$$



The Base Model

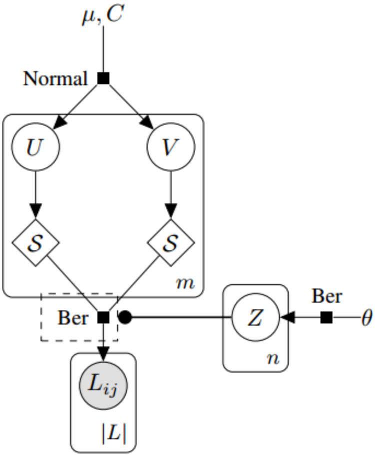
(i indexes instances, j indexes workers)

$$U_j, V_j \sim \mathcal{N}(\mu, C)$$

$$Z_i \sim \mathrm{Ber}(\theta)$$

$$L_{ij} | Z_i = 1 \sim \mathrm{Ber}(\mathcal{S}(U_j))$$

$$L_{ij} | Z_i = 0 \sim \mathrm{Ber}(\mathcal{S}(V_j))$$



Extensions

1. Worker Groups:
   - Know group membership.
   - Model each group k = a Normal dist (μ_k, C_k).

Extensions

1. Worker Groups:
   - Know group membership.
   - Model each group k = a Normal dist (μ_k, C_k).

2. Multiple tasks:
   - Assume two tasks.

Extensions

1. Worker Groups:
   - Know group membership.
   - Model each group k = a Normal dist (μ_k, C_k).
2. Multiple tasks:
   - Assume two tasks.
   - $(\mathrm{Sen}_1, \mathrm{Spe}_1)$ correlates with $(\mathrm{Sen}_2, \mathrm{Spe}_2)$.

Extensions

1. Worker Groups:
   - Know group membership.
   - Model each group k = a Normal dist (μ_k, C_k).
2. Multiple tasks:
   - Assume two tasks.
   - $(\mathrm{Sen}_1, \mathrm{Spe}_1)$ correlates with $(\mathrm{Sen}_2, \mathrm{Spe}_2)$.
   - $(U_1, V_1, U_2, V_2) \sim \mathcal{N}(\mu, C)$

Inference

For the Base Model

Approach: Variational EM

- E-step: infer $\Pr(U_{1..m}, V_{1..m}, Z_{1..n} | L)$.

Inference

For the Base Model

Approach: Variational EM

- E-step: infer $\Pr(U_{1..m}, V_{1..m}, Z_{1..n} | L)$.
- M-step: maximize parameters μ, C, θ.

Inference

For the Base Model

Approach: Variational EM

- E-step: infer $\Pr(U_{1..m}, V_{1..m}, Z_{1..n}|L)$.
- M-step: maximize parameters μ, C, θ.

Variational Inference:

- Approximate the (complex) posterior $\Pr(|)$...
- ... by a simpler function q.

Inference

For the Base Model

Approach: Variational EM

- E-step: infer $\Pr(U_{1..m}, V_{1..m}, Z_{1..n}|L)$.
- M-step: maximize parameters μ, C, θ.

Variational Inference:

- Approximate the (complex) posterior $\Pr(|)$...
- ... by a simpler function q.
- Minimize $\mathbb{KL}(q||p)$ ...
- ... equivalent to maximize a log-likelihood lower bound.

Inference

Meanfield Assumptions:

- q factorizes:

$$q(U_{1..m}, V_{1..m}, Z_{1..n}) = \prod_{j=1}^{m} q(U_j)q(V_j) \prod_{i=1}^{n} q(Z_i)$$

Inference

Meanfield Assumptions:

- q factorizes:

$$q(U_{1..m}, V_{1..m}, Z_{1..n}) = \prod_{j=1}^{m} q(U_j) q(V_j) \prod_{i=1}^{n} q(Z_i)$$

- Factors:

$$\begin{aligned} q(U_j) &= \mathcal{N}(\tilde{\mu}_{uj}, \tilde{\sigma}^2_{uj}) \\ q(V_j) &= \mathcal{N}(\tilde{\mu}_{vj}, \tilde{\sigma}^2_{vj}) \\ q(Z_i) &= \mathrm{Ber}(\tilde{\theta}_i) \end{aligned}$$

Inference

Meanfield Assumptions:

- q factorizes:

$$q(U_{1..m}, V_{1..m}, Z_{1..n}) = \prod_{j=1}^{m} q(U_j) q(V_j) \prod_{i=1}^{n} q(Z_i)$$

- Factors:

$$q(U_j) = \mathcal{N}(\tilde{\mu}_{uj}, \tilde{\sigma}^2_{uj})$$
$$q(V_j) = \mathcal{N}(\tilde{\mu}_{vj}, \tilde{\sigma}^2_{vj})$$
$$q(Z_i) = \mathrm{Ber}(\tilde{\theta}_i)$$

- Optimize with respect to

$$\{\tilde{\mu}_{uj}, \tilde{\sigma}^2_{uj}, \tilde{\mu}_{vj}, \tilde{\sigma}^2_{vj} | j = 1...m\} \text{ and } \{\tilde{\theta}_i | i = 1...n\}$$

Optimization

Coordinate Descent: update one var at a time.

Optimization

Coordinate Descent: update one var at a time.

Update Z_i:

$$q^*(Z_i = 1) \propto \exp\Big\{ \log \mathrm{Ber}(1|\theta) + \sum \mathbb{E}_{U_j \sim q(U_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j)) \Big\}$$

$$q^*(Z_i = 0) \propto \exp\Big\{ \log \mathrm{Ber}(0|\theta) + \sum \mathbb{E}_{V_j \sim q(V_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(V_j)) \Big\}$$

Optimization

Coordinate Descent: update one var at a time.

Update Z_i:

$$q^*(Z_i = 1) \propto \exp\Big\{ \log \mathrm{Ber}(1|\theta) + \sum \mathbb{E}_{U_j \sim q(U_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j)) \Big\}$$

$$q^*(Z_i = 0) \propto \exp\Big\{ \log \mathrm{Ber}(0|\theta) + \sum \mathbb{E}_{V_j \sim q(V_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(V_j)) \Big\}$$

Intuition:

- $Z_i \approx$ Prior $+ \sum \mathbb{E}$(Crowd labels for i)

Optimization

Coordinate Descent: update one var at a time.

Update Z_i:

$$q^*(Z_i = 1) \propto \exp\Big\{ \log \mathrm{Ber}(1|\theta) + \sum \mathbb{E}_{U_j \sim q(U_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j)) \Big\}$$

$$q^*(Z_i = 0) \propto \exp\Big\{ \log \mathrm{Ber}(0|\theta) + \sum \mathbb{E}_{V_j \sim q(V_j)} \log \mathrm{Ber}(L_{ij}|\mathcal{S}(V_j)) \Big\}$$

Intuition:

- $Z_i \approx$ Prior $+ \sum \mathbb{E}$(Crowd labels for i)
- $\mathbb{E}$ wrt worker quality.

Optimization

Update U_j:

$$q^*(U_j) \propto \exp\Big\{\mathbb{E}_{V_j \sim q(V_j)} \log \mathcal{N}(U_j, V_j|\mu, C) + \\ \sum q(Z_i = 1) \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j))\Big\}$$

Optimization

Update U_j:

$$q^*(U_j) \propto \exp\Big\{\mathbb{E}_{V_j \sim q(V_j)} \log \mathcal{N}(U_j, V_j|\mu, C) +$$
$$\sum q(Z_i = 1) \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j))\Big\}$$

Intuition:

- $U_j =$ logit sensitivity of worker j.

Optimization

Update U_j:

$$q^*(U_j) \propto \exp\Big\{\mathbb{E}_{V_j \sim q(V_j)} \log \mathcal{N}(U_j, V_j | \mu, C) +$$
$$\sum q(Z_i = 1) \log \mathrm{Ber}(L_{ij} | \mathcal{S}(U_j))\Big\}$$

Intuition:

- U_j = logit sensitivity of worker j.
- $U_j \approx \mathbb{E}$(correlation with specificity) + ...

Optimization

Update U_j:

$$q^*(U_j) \propto \exp\Big\{\mathbb{E}_{V_j \sim q(V_j)} \log \mathcal{N}(U_j, V_j | \mu, C) + \\ \sum q(Z_i = 1) \log \mathrm{Ber}(L_{ij} | \mathcal{S}(U_j))\Big\}$$

Intuition:

- $U_j =$ logit sensitivity of worker j.
- $U_j \approx \mathbb{E}$(correlation with specificity) $+$ ...
- ... instances that worker j has labeled.

Optimization

Update U_j:

$$q^*(U_j) \propto \exp\Big\{\mathbb{E}_{V_j \sim q(V_j)} \log \mathcal{N}(U_j, V_j|\mu, C) + \\ \sum q(Z_i = 1) \log \mathrm{Ber}(L_{ij}|\mathcal{S}(U_j))\Big\}$$

Intuition:

- $U_j =$ logit sensitivity of worker j.
- $U_j \approx \mathbb{E}$(correlation with specificity) $+$ ...
- ... instances that worker j has labeled.

(Similar equation for V_j)

Optimization

Problem: $\mathbb{E}()$ difficult to compute.

Optimization

Problem: $\mathbb{E}()$ difficult to compute.

Solution: Laplace Variational Inference (Wang & Blei, 2013)

- Approximate these update equations...
- ... by Laplace approximation.

Optimization

Problem: $\mathbb{E}()$ difficult to compute.

Solution: Laplace Variational Inference (Wang & Blei, 2013)

- Approximate these update equations...
- ... by Laplace approximation.
- Details in the paper.

Learning

E-step: Infer posterior distribution over hidden vars.

Learning

E-step: Infer posterior distribution over hidden vars.

M-step: maximize μ, C, θ under posterior.

- μ, C: sample mean and Covariance.
- θ: average of $\{\tilde{\theta}_i | i = 1...n\}$.

Evaluation

Citizen Science:

Evaluation

Citizen Science:

- Workers **volunteer** ...
- ... to help science.
- Different from traditional crowdsourcing:
  - large scale.
  - (usually) higher quality.

Evaluation

Citizen Science:

- Workers **volunteer** ...
- ... to help science.
- Different from traditional crowdsourcing:
  - large scale.
  - (usually) higher quality.

Two real world scenarios:

- Biomedical Citation Screening.
- Galaxy Morphological Classification.

Scenario 1

Biomedical Citation Screening:

- Motivation: biomedical literature is huge.
- Need to find relevant citations.

Scenario 1

Biomedical Citation Screening:

- Motivation: biomedical literature is huge.
- Need to find relevant citations.

The RCT dataset:

- Identify Randomized Control Trials reports.

Scenario 1

Biomedical Citation Screening:

- Motivation: biomedical literature is huge.
- Need to find relevant citations.

The RCT dataset:

- Identify Randomized Control Trials reports.
- Very imbalanced (3% positive).

Scenario 1

Biomedical Citation Screening:

- Motivation: biomedical literature is huge.
- Need to find relevant citations.

The RCT dataset:

- Identify Randomized Control Trials reports.
- Very imbalanced (3% positive).
- Workers: from in 2 groups...
- ... experts and novices

Scenario 1

Baselines:

- Majority Vote.
- Two Coin (Raykar et. al. 2010).

Scenario 1

Baselines:

- Majority Vote.
- Two Coin (Raykar et. al. 2010).

Our method: two versions

- Full-Cov: the full model.

Scenario 1

Baselines:

- Majority Vote.
- Two Coin (Raykar et. al. 2010).

Our method: two versions

- Full-Cov: the full model.
- Diag-Cov: constrain C to be diagonal.

Scenario 1

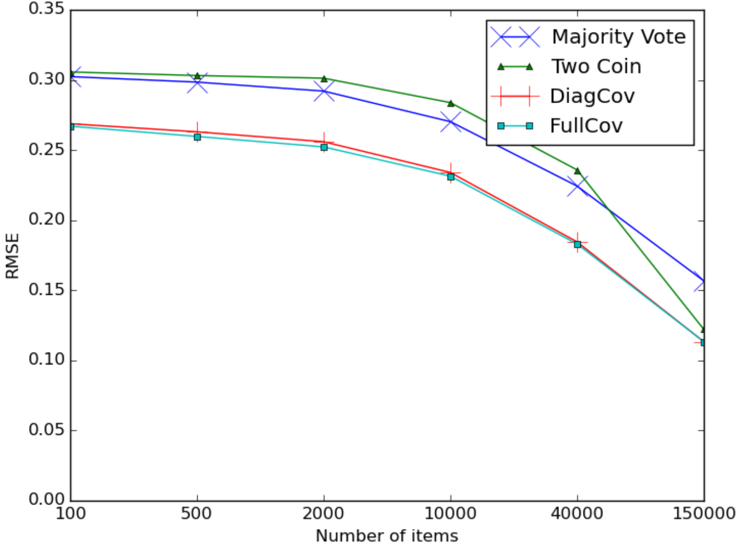
Baselines:

- Majority Vote.
- Two Coin (Raykar et. al. 2010).

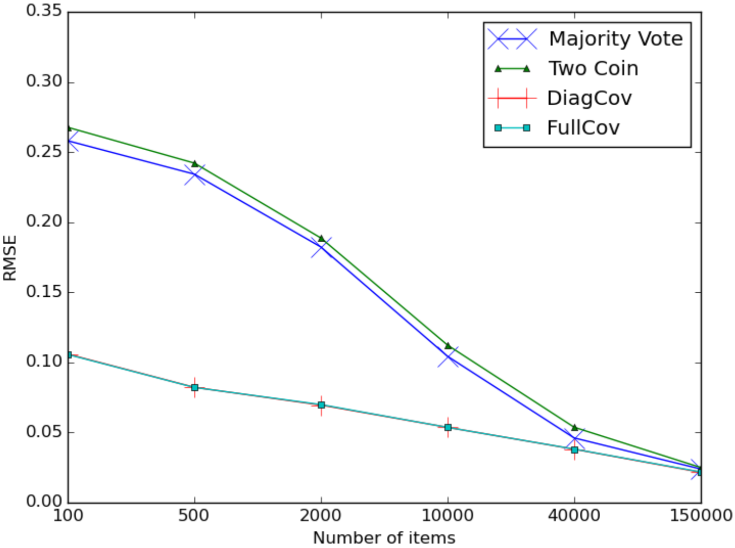
Our method: two versions

- Full-Cov: the full model.
- Diag-Cov: constrain C to be diagonal.
  - only model worker groups ...
  - ... but not model sen-spec correlation.

Results: Sensitivity



Results: Specificity



Discussion

Our method has two parts: group and correlation.

Discussion

Our method has two parts: group and correlation.

- Group provides most improvement.

Discussion

Our method has two parts: group and correlation.

- Group provides most improvement.
- Correlation gives additional boost for sen.

Scenario 2

Galaxy Morphological Classification:

- Motivation: Few astronomers, lot of galaxies.

Scenario 2

Galaxy Morphological Classification:

- Motivation: Few astronomers, lot of galaxies.

Galaxy Zoo 2 dataset:

- Multiple questions: galaxy shape? number of spiral arms?...
- Have volunteers answering questions.

Scenario 2

Setting:

- Given all labels in source task ...
- ... and some labels in target task.
- Predict worker sen. and spe. in target task.

Scenario 2

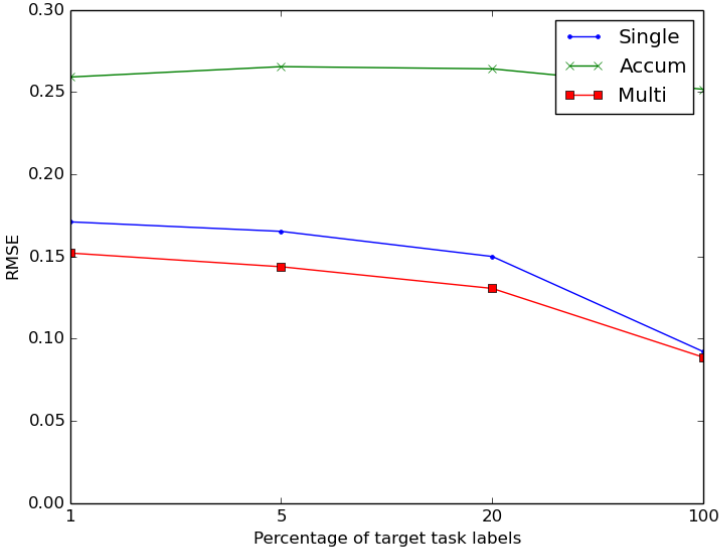
Setting:

- Given all labels in source task ...
- ... and some labels in target task.
- Predict worker sen. and spe. in target task.

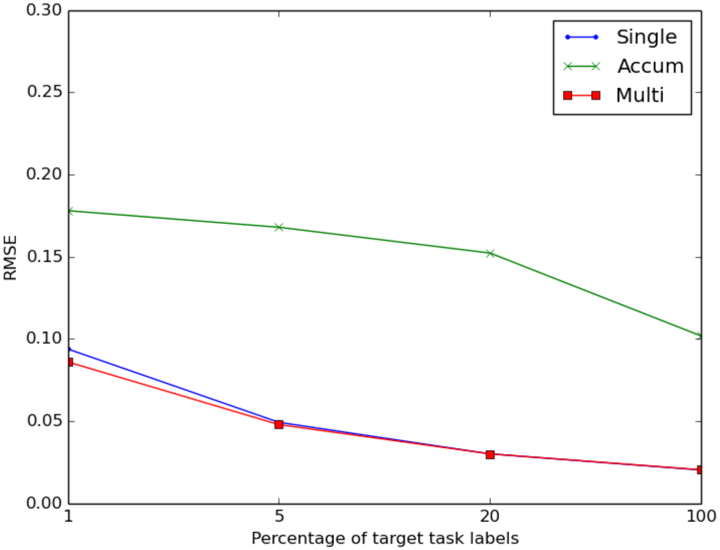
Compare:

- Single: only consider target labels.
- Accum: merge source labels to target.
- Multi: our multi-task model.

Result: Sensitivity



Result: Specificity



Discussion

Multi is surprisingly bad.

- Tasks are different, naive merge is bad.

Discussion

Multi is surprisingly bad.

- Tasks are different, naive merge is bad.

Our method

- has good improvement ...
- ... although sometimes modest.

Discussion

Multi is surprisingly bad.

- Tasks are different, naive merge is bad.

Our method

- has good improvement ...
- ... although sometimes modest.
- Again, tasks are different...
- Many workers better in source task ...
- ... but worse in target task.

Discussion

Multi is surprisingly bad.

- Tasks are different, naive merge is bad.

Our method

- has good improvement ...
- ... although sometimes modest.
- Again, tasks are different...
- Many workers better in source task ...
- ... but worse in target task.
- Our method still as good as the baseline.

Conclusion

Summary

- Model correlation to transfer knowledge.
- Empirically improve estimates of worker quality.

Conclusion

Summary

- Model correlation to transfer knowledge.
- Empirically improve estimates of worker quality.

Future work

- Extend: instance-level features.
- Application: tasks/instances routing.

Conclusion

Summary

- Model correlation to transfer knowledge.
- Empirically improve estimates of worker quality.

Future work

- Extend: instance-level features.
- Application: tasks/instances routing.

Question?