To Re(label), or Not To Re(label)

Christopher H. Lin  
University of Washington

Mausam  
IIT Delhi

Daniel S. Weld  
University of Washington
Penguin

Bear

Giraffe
Triceratops

Bear

Giraffe
Penguin

Triceratops

Penguin

Penguin
\[
U(p(\beta_r | y_{i'}, \nu)) = \left\| E_p(\beta_r) (\beta_r) - E_{p(\beta_r|y_{i'}, \nu)} (\beta_r) \right\|_2 \\
\approx \left\| E \left( \frac{1}{S-1} \sum_{s=2}^{S} Z_r^{s-1} \left[ (\gamma|\gamma^{s-1}, Z^{s-1}) - (\gamma|\gamma^{s-1}, Z^{s-1}) \right] \right) \right\|_2.
\]

\[
Q(\alpha, \beta) = E \left[ \ln p(l, z|\alpha, \beta) \right] \\
= E \left[ \ln \prod_j \left( p(z_j) \prod_i p(l_{ij}|z_j, \alpha_i, \beta_j) \right) \right] \\
since l_{ij} are cond. indep. given z, \alpha, \beta \\
= \sum_j E \left[ \ln p(z_j) \right] + \sum_{ij} E \left[ \ln p(l_{ij}|z_j, \alpha_i, \beta_j) \right]
\]

\[
p(z|L, \theta) = \int p(z, q|L, \theta) dq = \prod_{j \in [M]} \int_0^1 p(q_j|\theta) q_j^{\gamma_j} (1 - q_j)^{\gamma_j - c_j} dq_j \overset{def}{=} \prod_{j \in [M]} \psi_j(z_{\mathcal{N}_j}),
\]

[Dawid et al 79, Whitehill et al 09, Welinder et al 10, Raykar et al 10, Wauthier et al 11, Karger et al 11, Kajino et al 12, Baba et al 13, Liu et al 12, Lin et al 12, etc, etc...]
How should we best spend a fixed budget $b$
when training a classifier?
budget $b = 9$
Unilabel?

9 examples with labels that are 75% accurate?
3 examples with labels that are 85% accurate?
1 example with a label that is 99% accurate?
How should we best spend a fixed budget \( b \) when training a classifier?
How to relabel

Yan et al 2011,
Dekel et al 2010,
Sheng et al 2008, Ipeirotis et al 2013,
Zhao et al 2011]
<table>
<thead>
<tr>
<th>Dataset</th>
<th># Features</th>
<th># Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Breast Cancer</td>
<td>9</td>
<td>699</td>
</tr>
<tr>
<td>(b) Bank Note Authentication</td>
<td>4</td>
<td>1372</td>
</tr>
<tr>
<td>(c) Seismic Bumps</td>
<td>18</td>
<td>2584</td>
</tr>
<tr>
<td>(d) EEG Eye State</td>
<td>14</td>
<td>14980</td>
</tr>
<tr>
<td>(e) Sonar</td>
<td>60</td>
<td>208</td>
</tr>
<tr>
<td>(f) Breast Cancer Diagnostic</td>
<td>30</td>
<td>569</td>
</tr>
<tr>
<td>(g) Hill-Valley</td>
<td>100</td>
<td>606</td>
</tr>
<tr>
<td>(h) Hill-Valley with Noise</td>
<td>100</td>
<td>606</td>
</tr>
<tr>
<td>(i) Internet Ads</td>
<td>1558</td>
<td>2359</td>
</tr>
<tr>
<td>(j) Gisette</td>
<td>5000</td>
<td>6000</td>
</tr>
<tr>
<td>(k) Farm Ads</td>
<td>54877</td>
<td>4143</td>
</tr>
<tr>
<td>(l) Spambase</td>
<td>57</td>
<td>4601</td>
</tr>
</tbody>
</table>
Unilabeling better in over half the datasets!!
Factors that Affect Relabeling Efficacy

Inductive Bias
Worker Accuracy
Budget
Assumptions

Passive Learning
Binary Classification
Identical Workers
Constant Cost
Majority Vote
j/k Relabeling

Penguin

Penguin

Triceratops
2/3 Relabeling

Penguin

Penguin
Factors that Affect Relabeling Efficacy

Inductive Bias
Worker Accuracy
Budget
strong inductive bias $\rightarrow$ low expressiveness

weak inductive bias $\rightarrow$ high expressiveness
True Concept: 65 and older -> “Senior Citizen”
True Concept: 65 and older -> “Senior Citizen”
Strong Inductive Bias Classifier

(low expressiveness)

True Concept: 65 and older -> “Senior Citizen”
True Concept: 65 and older -> “Senior Citizen”
Weaker Inductive Bias

Increases Relabeling Power
budget $b = 500$
workers 75% accurate
Controlling Inductive Bias via

<table>
<thead>
<tr>
<th># Features</th>
<th>Type of Classifier</th>
<th>Regularization</th>
</tr>
</thead>
</table>


Relabeling Accuracy / Unilabeling Accuracy

Number of Features (≈VC Dimension)

2/3 Relabeling
Relabeling Accuracy / Unilabeling Accuracy

Number of Features (≈VC Dimension)

Weaker Inductive Bias

2/3 Relabeling
Weaker Inductive Bias

Relabeling Better

Number of Features (≈VC Dimension)

Relabeling Accuracy / Unilabeling Accuracy

2/3 Relabeling
Weaker Inductive Bias

Relabeling Better

Relabeling Accuracy / Unilabeling Accuracy

Number of Features (≈VC Dimension)

2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

2 4 8 16 32 64 128 256
Weaker Inductive Bias Increases Relabeling Power
Relabeling Accuracy / Unilabeling Accuracy

# Features = 50

Classifier

Support Vector Machine
Logistic Regression
Random Forest
Decision Tree
Nearest Neighbor

2/3 Relabeling
3/5 Relabeling
4/7 Relabeling
Relabeling Accuracy / Unilabeling Accuracy

Classifier

2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

Support Vector Machine
Logistic Regression
Random Forest
Decision Tree
Nearest Neighbor

Relabeling Accuracy / Unilabeling Accuracy

0.97
0.99
1.01
1.03
1.05
1.07
1.09

Classifier
Factors that Affect Relabeling Efficacy

**Inductive Bias**

- Weaker Inductive Bias Increases Relabeling Power

**Worker Accuracy**

**Budget**
Factors that Affect Relabeling Efficacy

Inductive Bias

Worker Accuracy

Budget
Training Example Accuracy Gain

Worker Accuracy

[Sheng et al 08, Ipeirotis et al 13]
Relabeling Accuracy / Unilabeling Accuracy

- 2/3 Relabeling
- 3/5 Relabeling
- 4/7 Relabeling

Worker Accuracy (p)
Moderately Accurate Workers Maximize Relabeling Power

Worker Accuracy (p)

Relabeling Accuracy / Unilabeling Accuracy

2/3 Relabeling
3/5 Relabeling
4/7 Relabeling
Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Moderately Accurate Workers Maximize Relabeling Power

Budget
Factors that Affect Relabeling Efficacy

Inductive Bias
Worker Accuracy
Budget
Unilabeling Beats Relabeling with Higher Budgets

Classifier Accuracy vs. Budget

- Unilabeling
- 2/3 Relabeling
- 3/5 Relabeling
- 4/7 Relabeling
Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Moderately Accurate Workers Maximize Relabeling Power

Budget

Smaller Budgets Increase Relabeling Power
<table>
<thead>
<tr>
<th>Dataset</th>
<th># Features</th>
<th># Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Breast Cancer</td>
<td>9</td>
<td>699</td>
</tr>
<tr>
<td>(b) Bank Note Authentication</td>
<td>4</td>
<td>1372</td>
</tr>
<tr>
<td>(c) Seismic Bumps</td>
<td>18</td>
<td>2584</td>
</tr>
<tr>
<td>(d) EEG Eye State</td>
<td>14</td>
<td>14980</td>
</tr>
<tr>
<td>(e) Sonar</td>
<td>60</td>
<td>208</td>
</tr>
<tr>
<td>(f) Breast Cancer Diagnostic</td>
<td>30</td>
<td>569</td>
</tr>
<tr>
<td>(g) Hill-Valley</td>
<td>100</td>
<td>606</td>
</tr>
<tr>
<td>(h) Hill-Valley with Noise</td>
<td>100</td>
<td>606</td>
</tr>
<tr>
<td>(i) Internet Ads</td>
<td>1558</td>
<td>2359</td>
</tr>
<tr>
<td>(j) Gisette</td>
<td>5000</td>
<td>6000</td>
</tr>
<tr>
<td>(k) Farm Ads</td>
<td>54877</td>
<td>4143</td>
</tr>
<tr>
<td>(l) Spambase</td>
<td>57</td>
<td>4601</td>
</tr>
</tbody>
</table>
Relabeling Accuracy / Unilabeling Accuracy

< 100 features

> 100 features

Dataset

Relabeling Accuracy / Unilabeling Accuracy

- 2/3 Relabeling
- 3/5 Relabeling
- 4/7 Relabeling

b a d c f l e g h i j k
Relabeling Accuracy / Unilabeling Accuracy

55% accurate workers

Dataset

2/3 Relabeling
3/5 Relabeling
4/7 Relabeling
75% accurate workers

Relabeling Accuracy / Unilabeling Accuracy vs. Dataset

- 2/3 Relabeling
- 3/5 Relabeling
- 4/7 Relabeling
Factors that Affect Relabeling Efficacy

**Inductive Bias**
Weaker Inductive Bias Increases Relabeling Power

**Worker Accuracy**
Moderately Accurate Workers Maximize Relabeling Power

**Budget**
Smaller Budgets Increase Relabeling Power
Future Work
Assumptions

Passive Learning
Identical Workers
Majority Vote
Binary Classification
Constant Cost
Assumptions

- Passive Learning → Active Learning
- Identical Workers → Real Workers
- Majority Vote → Sophisticated AI
- Binary Classification → General Classification
- Constant Cost → Varying Costs
Assumptions

- Passive Learning → Active Learning
- Identical Workers → Real Workers
- Majority Vote → Sophisticated AI
- Binary Classification → General Classification
- Constant Cost → Varying Costs

End-to-end Decision Theoretic System
Factors that Affect Relabeling Efficacy

**Inductive Bias**

Weaker Inductive Bias Increases Relabeling Power

**Worker Accuracy**

Moderately Accurate Workers Maximize Relabeling Power

**Budget**

Smaller Budgets Increase Relabeling Power