Reducing Label Cost by Combining Feature Labels and Crowdsourcing

Combining Learning Strategies to Reduce Label Cost
7/2/2011

Jay Pujara  jay@cs.umd.edu
Ben London  blondon@cs.umd.edu
Lise Getoor  getoor@cs.umd.edu

University of Maryland, College Park
Labels are expensive

- Immense amount of data in the real world
- Often, no corresponding glut of labels
  - Precise labels may require expertise
  - Must ensure training labels have good coverage
Two strategies to mitigate cost

- Leverage unlabeled data in learning
- Find a cheaper way to annotate
Two strategies to mitigate cost

• Leverage unlabeled data in learning
  ◦ **Bootstrapping:** Use your labeled data to generate labels for unlabeled data
  ◦ **Active Learning:** Choose the most useful unlabeled data to label

• Find a cheaper way to annotate
  ◦ **Feature Labels:** Use a heuristic to generate labels
  ◦ **Crowdsourcing:** Get non-experts to provide labels
Feature Labels + Bootstrapping

- Feature Labels
  - Choose features that are highly correlated with labels
  - Remove features from input and use as labels
  - Possibly introduces bias into training data

- Bootstrapping
  - Train a classifier on labeled data
  - Predict labels on unlabeled data
  - Use the most confident predictions as labels

McCallum, Andrew and Nigam, Kamal. Text classification by bootstrapping with keywords, EM, and shrinkage. ACL99
Active Learning + Crowdsourcing

- **Active Learning**
  - Train a classifier
  - Predict labels on unlabeled data
  - Choose least confident predictions for label acquisition

- **Crowdsourcing**
  - Provide data to non-experts, reward for labels
  - Few requirements/guarantees about labelers
  - Resulting labels may be noisy, gamed

Ambati, V., Vogel, S., and Carbonell, J. Active learning and crowd-sourcing for machine translation. LREC10
Comparing **Learning/Annotation Strategies**

- **Active Learning**
  - Find labels for uncertain instances

- **Bootstrapping**
  - Find labels for certain instances

- **Feature Labels**
  - High precision, Low coverage

- **Crowdsourcing**
  - Low precision, High coverage
**Active Bootstrapping**

- Input: Feature label rules $F$, unlabeled data, $U$ and constants $T, k$ and $\alpha$
- Initialize $S$ by applying feature labels $F$ to data $U$
- For $t = 1, \ldots, T$:
  - Train a classifier on $S$
  - Predict labels on $U$
  - Add top-$k$ most certain positive predictions to $S$
  - Add top-$k$ most certain negative predictions to $S$
  - Add crowdsourced responses to top-$\alpha k$ uncertain predictions to $S$
  - $U = U - S$
- Output: Classifier trained on $S$
Evaluation on Twitter dataset

- **Task:** Sentiment Analysis (happy/sad tweets)
- **Data:** 77920 normalized* tweets originally containing emoticons (6/2009-12/2009)
- **Evaluation Set:** 500 hand-labeled tweets
- **Feature labels:** happy and sad emoticons from Wikipedia
- **Crowdsourcing:** HIT on Amazon’s Mechanical Turk platform. Use known evaluation set labels to validate results
- **Active Learning/Bootstrapping:** Use MEGAM maximum entropy classifier label probabilities

---

Yang, Jaewon and Leskovec, Jure. Patterns of temporal variation in online media. WSDM11

Wikipedia: List of Emoticons

Experiments on Twitter dataset

• Compare different approaches:
  ◦ Feature Labels + Bootstrapping
    • Start with seed set of 1K, 2K, 10K feature labels
    • Add 10% of seed set in each iteration
  ◦ Crowdsourcing + Bootstrapping
    • Start with 2000 crowdsourced labels (1000 instances)
    • After validation, 670 labels
    • Add 200 new labels in each iteration
  ◦ Active Bootstrapping (k=50, \(\alpha=2\))
    • Start with 1000 labels, add 100* crowdsourced and 100 bootstrapped labels in each iteration
Results:

Active Bootstrapping vs. Feature Labels + Bootstrapping

- Same amount of data per iteration
- Active Bootstrapping outperforms Feature Labels + Bootstrapping, at minimal cost ($16)
Results:

Active Bootstrapping vs. Feature Labels + Bootstrapping

- Even with additional starting data, Feature Labels + Bootstrapping starts well but is eventually overcome by Active Bootstrapping
Results:

**Active Bootstrapping vs. Crowdsourcing + Bootstrapping**

- Both methods cost about the same ($16), but **Active Bootstrapping** clearly outperforms.
Active Bootstrapping combines the best of both worlds:
- Minimal time/expense from domain expert (to create feature labels)
- Crowdsource the rest
Results:
Summary

<table>
<thead>
<tr>
<th>Method</th>
<th>Err, 10</th>
<th>Err, 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Lables, 1K</td>
<td>.332</td>
<td>.367</td>
</tr>
<tr>
<td>Feature Lables, 2K</td>
<td>.302</td>
<td>.353</td>
</tr>
<tr>
<td>Feature Lables, 10K</td>
<td>.295</td>
<td>.348</td>
</tr>
<tr>
<td>Crowdsource, 2K</td>
<td>.374</td>
<td>.478</td>
</tr>
<tr>
<td>Active Bootstrapping</td>
<td>.332</td>
<td>.292</td>
</tr>
</tbody>
</table>
Thank You!

- Reduce label cost by combining strategies
- Introduce algorithm, **Active Bootstrapping**:
  - Combines complementary annotation strategies (feature labels and crowdsourcing)
  - Combines complementary learning strategies (bootstrapping and active learning)
- Evaluate on a real-world dataset/task (sentiment analysis on Twitter), show superior results


Questions?