Efficiently Maintaining Conditional Random Fields

Xixuan (Aaron) Feng
Guangde Chen
Statistical Information Extraction (IE) is important in both academic and industrial applications:
- To build knowledge bases
- To improve search, mining, etc.
- E.g.: isWiki, MSR Academic Search, Ali Baba, YAGO

Problem: Corpus is evolving!
- A single new training example forces us to train the model from scratch and so is too slow to apply to real applications

Goal: To incorporate new examples efficiently
- We focus on the most popular statistical model for IE – conditional random fields (CRFs)

Result: Evaluation over a text chunking dataset
Baselines

• **Two extremes**
  – Retrain-All: to train the model from scratch as every new example comes in
  – Retrain-New: only to use new examples

### Retrain-All

1. \( w \leftarrow 0 \)
2. \( w \leftarrow w - \sum \nabla l_i(w) \)
3. If \( w \) is not converged
   \[ \text{GOTO 2} \]

### Retrain-New

1. \( w \leftarrow 0 \)
2. \( w \leftarrow w - \nabla l_{new}(w) \)
3. If \( w \) is not converged
   \[ \text{GOTO 2} \]
Experiment Setup

• Dataset
  – CoNLL 2000 Shared Tasks
    http://www.cnts.ua.ac.be/conll2000/chunking/
  – Wall Street Journal
  – ~9K sentences = 5K (test set) + ~4K (training pool)

• Task: Text Chunking

• Features: 7.4 million
  – Word-based regular expressions
  – Part-of-speech

• 20 initial examples + 200 examples streaming in
  – From training pool
  – Evaluate quality using 5K-sentence test set
Baselines

• **Two extremes**
  – Retrain-All: to train the model from scratch as every new example comes in
  – Retrain-New: only to use new examples

Loss Value on 5K-sentence Test Set

Time to Incorporate New Examples
Increment Instead of Retrain

- Keep the model as initial guess when new examples come
  - Incr-All and Incr-New
  - We expected Incr-New gives bad quality results

Incr-All
- 1. \( w \leftarrow w_{previous} \)
- 2. \( w \leftarrow w - \sum \nabla l_i(w) \)
- 3. If \( w \) is not converged
  GOTO 2

Incr-New
- 1. \( w \leftarrow w_{previous} \)
- 2. \( w \leftarrow w - \nabla l_{new}(w) \)
- 3. If \( w \) is not converged
  GOTO 2
Increment Instead of Retrain

- Keep the model as initial guess when new examples come
  - We expected Incr-New gives bad quality results
  - But we were wrong!

![Loss Value on 5K-sentence Test Set](image1)

![Time to Incorporate New Examples](image2)
Analysis

• **Incr-New is not always good**
  – In general, gradient descent is robust to initial value
Analysis (Cont.)

- Cases that Incr-New is good
  - High dimensional model space and sparsity
Conclusion and Future Work

• We have evaluated and analyzed several strategies to incorporate new examples for CRF
  – Retrain-All, Retrain-New, Incr-All, Incr-New

• Next Steps
  – Evaluation with regularization
  – Formal proof of cases that Incr-New are good
  – Other data sets

• Thanks & Questions