Improving Morphology Induction with Spelling Rules

Jason Naradowsky
University of Massachusetts Amherst

Joint Work with Sharon Goldwater

narad@cs.umass.edu
Outline

- Morphology Induction
- Our Model
- Hyperparameters & Inference
- Experimental Results
- Conclusion
Morphology (Linguistics)

- The study of the internal structure of words:

  Antidisestablishmentarianism
Morphology (Linguistics)

The study of the internal structure of words:

Anti.dis.e stab lish.ment.arian.ism
Morphology (Linguistics)

- The study of the internal structure of words:
  - Anti.dis.establis.ment.arian.ism

Morphemes
Morphology (Linguistics)

- The study of the internal structure of words:

  Anti.dis.establish.ment.arian.ism

  stem
Morphology (Linguistics)

The study of the internal structure of words:

- **Anti.dis.establish.ment.arian.ism**
  - prefixes
  - stem
  - suffixes
Unsupervised Morphology Induction

- Observing just the words, find the best segmentation:
  - walking → walk.ing

- Applications:
  - Important component in many NLP tasks
  - Especially useful for morphologically-rich languages (Finnish, Arabic, Hebrew)
  - Cognitive Science: How do children learn this?
Underlying Assumption:

- User’s Goal: Find best (linguistic) solution.
- System Goal: Find most concise solution.

<table>
<thead>
<tr>
<th>Too Many Stems</th>
<th>Too Many Suffixes</th>
<th>Just Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>walk.</td>
<td>wa.lk</td>
<td>walk.</td>
</tr>
<tr>
<td>walks.</td>
<td>wa.lks</td>
<td>walk.s</td>
</tr>
<tr>
<td>walking.</td>
<td>wa.lking.</td>
<td>walk.ing</td>
</tr>
<tr>
<td>talk.</td>
<td>ta.lk</td>
<td>talk.</td>
</tr>
<tr>
<td>talking.</td>
<td>ta.lking</td>
<td>talk.ing</td>
</tr>
<tr>
<td>cat.</td>
<td>cat.</td>
<td>cat.</td>
</tr>
<tr>
<td>cat.s</td>
<td>cat.s</td>
<td>cat.s</td>
</tr>
</tbody>
</table>

Morphs: 6+2=8  3+5=8  3+3=6

Wednesday, July 15, 2009
**Underlying Assumption:**
- User’s Goal: Find best (linguistic) solution.
- System Goal: Find most concise solution.

<table>
<thead>
<tr>
<th>Too Many Stems</th>
<th>Too Many Suffixes</th>
<th>Just Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>walk.</td>
<td>wa.lk</td>
<td>walk.</td>
</tr>
<tr>
<td>walks.</td>
<td>wa.lks</td>
<td>walk.s</td>
</tr>
<tr>
<td>walking.</td>
<td>wa.lking.</td>
<td>walk.ing</td>
</tr>
<tr>
<td>talk.</td>
<td>ta.lk</td>
<td>talk.</td>
</tr>
<tr>
<td>talking.</td>
<td>ta.lking</td>
<td>talk.ing</td>
</tr>
<tr>
<td>cat.</td>
<td>cat.</td>
<td>cat.</td>
</tr>
<tr>
<td>cat.s</td>
<td>cat.s</td>
<td>cat.s</td>
</tr>
</tbody>
</table>

**Morphs:**
- Too Many Stems: $6+2=8$
- Too Many Suffixes: $3+5=8$
- Just Right: $3+3=6$
**Underlying Assumption:**

- **User’s Goal:** Find best (linguistic) solution.
- **System Goal:** Find most concise solution.

<table>
<thead>
<tr>
<th>Too Many Stems</th>
<th>Too Many Suffixes</th>
<th>Just Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>walk.</td>
<td>wa.lk</td>
<td>walk.</td>
</tr>
<tr>
<td>walks.</td>
<td>wa.lks</td>
<td>walk.s</td>
</tr>
<tr>
<td>walking.</td>
<td>wa.lking.</td>
<td>walk.ing</td>
</tr>
<tr>
<td>talk.</td>
<td>ta.lk</td>
<td>talk.</td>
</tr>
<tr>
<td>talking.</td>
<td>ta.lking</td>
<td>talk.ing</td>
</tr>
<tr>
<td>cat.</td>
<td>cat.</td>
<td>cat.</td>
</tr>
<tr>
<td>cat.s</td>
<td>cat.s</td>
<td>cat.s</td>
</tr>
</tbody>
</table>

**Morphs:** 6+2=8                  3+5=8                  3+3=6

Wednesday, July 15, 2009
Bayesian Morphology Induction
(Goldwater 2006)

\[ P(\text{word}) = P(\text{class}, \stem, \suffix) = \]
\[ P(\text{class}) \times P(\stem | \text{class}) \times P(\suffix | \text{class}) \]

- Each word consists of a stem and a suffix
  - (suffix can be the empty string)
- Multinomials with symmetric Dirichlet priors
  - No bias means most concise solution preferable
Generative Process: ‘walking’
Generative Process?: ‘napping’

- Class
  - Stem
  - Suffix
- ‘nap’
- ‘ping’
Generative Process?: ‘napping’
Spelling Rules

- Rules capture a one-character transformation in a particular context.
- 3 Types: Insertions, Deletions, and Null (no transformation)
- Left context more important in English
  - (we find 2 character left contexts most useful)
Outline

- Morphology Induction
- Our Model
- Hyperparameters & Inference
- Experimental Results
- Conclusion
A New Generative Process:

class

stem

suffix

‘nap’

‘ing’
A New Generative Process:

- **class**
- **suffix**
- **stem**
- **rule type**

- **‘nap’**
- **‘ing’**
- **INSERT**

Wednesday, July 15, 2009
A New Generative Process:

- class
- stem
- suffix
- rule type
- rule
- ‘nap’
- ‘ing’
- INSERT
- $\varepsilon \rightarrow p$
- ap_i
Our Model

\[ P(\text{class}, \text{stem}, \text{suffix}, \text{rule type}, \text{rule}) = \]
\[ P(\text{class}) \times \]
\[ P(\text{stem} | \text{class}) \times \]
\[ P(\text{suffix} | \text{class}) \times \]
\[ P(\text{rule type} | \text{context(\text{stem}, \text{suffix})}) \times \]
\[ P(\text{rule} | \text{rule type}, \text{context(\text{stem}, \text{suffix})}) \]

\[ \text{rule type} \in \{ \text{Insertion, Deletion, Null} \} \]

- Greatly increases search space:
  - About 28 times more possible solutions per word!
Outline

- Morphology Induction
- Our Model
- Hyperparameters & Inference
- Experimental Results
- Conclusion
Inference

- Alternate between:
  - Gibbs Sampling for the latent variables
    - (class, stems, suffix, etc)
  - Hyperparameter Updates
    - (update hyperparameters over priors on variables)
    - minimize free parameters!

- We run for 5 epochs of:
  - 10 Gibbs Sampling Iterations
  - 10 hyperparameter iterations

- Convergence much earlier
Hyperparameters

- Induced for class, stem, suffix, and rule variables
- Learn hyperparameters using Minka’s fixed-point method (Minka, 2003)
- Inducing all is principled, but also a computational burden
- Rule type prior set by linguistic intuition:
  - $\text{hyp(INSERTION)} = .001$
  - $\text{hyp(DELETION)} = .001$
  - $\text{hyp(NULL)} = .5$
Outline

- Morphology Induction
- Our Model
- Hyperparameters & Inference
- Experimental Results
- Conclusion
Data Sets & Evaluation

- 7487 different verbs from Wall Street Journal
- Gold Standard: CELEX lexical database
  - surface segmentation: walk.ing
  - abstract representation: 50655+pe

Evaluation Metrics:
- Underlying form accuracy
- Pairwise precision and recall
Underlying Form Accuracy

- Construct the underlying stem from derivational data contained in the CELEX (using lemma ID number)
- Lookup suffix in dictionary:
  - e3S : -s
  - a1S : -ed
  - pe : -ing
- Match strings - UFA is % correct
## Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε ε → ε</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing ε → ε</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es ε → ε</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion ε → ε</td>
<td>44405+i</td>
</tr>
</tbody>
</table>
**Pairwise Precision and Recall**

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion</td>
<td>44405+i</td>
</tr>
</tbody>
</table>
## Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion</td>
<td>44405+i</td>
</tr>
</tbody>
</table>
## Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion</td>
<td>44405+i</td>
</tr>
</tbody>
</table>

1 match out of 1 arcs = 100% PP for this stem
## Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε ε → ε</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing ε → ε</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es ε → ε</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion ε → ε</td>
<td>44405+i</td>
</tr>
</tbody>
</table>
## Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state+ε</td>
<td>ε → ε</td>
</tr>
<tr>
<td>stating</td>
<td>state+ing</td>
<td>ε → ε</td>
</tr>
<tr>
<td>states</td>
<td>stat.es</td>
<td>ε → ε</td>
</tr>
<tr>
<td>station</td>
<td>stat+ion</td>
<td>ε → ε</td>
</tr>
</tbody>
</table>
### Pairwise Precision and Recall

<table>
<thead>
<tr>
<th>Word</th>
<th>Found</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state$^+\varepsilon$ $\varepsilon \rightarrow \varepsilon$</td>
<td>44380+i</td>
</tr>
<tr>
<td>stating</td>
<td>state$^+ing$ $\varepsilon \rightarrow \varepsilon$</td>
<td>44380+pe</td>
</tr>
<tr>
<td>states</td>
<td>stat.es $\varepsilon \rightarrow \varepsilon$</td>
<td>44380+a1S</td>
</tr>
<tr>
<td>station</td>
<td>stat$^+ion$ $\varepsilon \rightarrow \varepsilon$</td>
<td>44405+i</td>
</tr>
</tbody>
</table>

1 correct arc out of 2 arcs = %50 Recall for this stem
Results: Stems

- PP
- PR
- P F-Measure
- UFA

Baseline vs Our Model

Wednesday, July 15, 2009
Results: Suffixes

<table>
<thead>
<tr>
<th></th>
<th>baseline</th>
<th>our model</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>400</td>
<td>550</td>
</tr>
<tr>
<td>PR</td>
<td>700</td>
<td>850</td>
</tr>
<tr>
<td>P F-Measure</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>UFA</td>
<td></td>
<td>900</td>
</tr>
</tbody>
</table>

Wednesday, July 15, 2009
### Induced Rules:

<table>
<thead>
<tr>
<th>Freq</th>
<th>Rule</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>468</td>
<td>( e \rightarrow \varepsilon ) when before i</td>
<td>abate, abating</td>
</tr>
<tr>
<td>41</td>
<td>( \varepsilon \rightarrow e ) when after sh/ss/ch</td>
<td>match, matches</td>
</tr>
<tr>
<td>29</td>
<td>( \varepsilon \rightarrow p ) after p, before i or e</td>
<td>nap, napping</td>
</tr>
</tbody>
</table>

Of the top 20 types of induced rules, 568 of 623 correct = 91%

Incorrect rules: fated explained as fates.d with s-deletion rates explained as rat.s with an e-insertion
Conclusions

- Orthographic rules can help in morphology induction
- Greatly increases search space
- Joint inference over complimentary tasks can overcome the search burden and significantly improve performance in particular parts of task
- This may allow unsupervised generative models to compete more closely with unsupervised discriminative models (with contrastive estimation)
Future Work

- Extend to multiple suffixes
  - Test on more representative language samples
  - Test on more languages
- Leverage phonological information for asymmetric priors
  - Once we know ‘p’ is often doubled, and ‘t’ is similar to ‘p’, should imply ‘t’ may also often be doubled
  - May allow for character-to-character transformations
- Hierarchical Models
  - More like grammar induction than segmentation
  - Capture interaction between prefixes and suffixes