The Illinois-Columbia System in the CoNLL-2014 Shared Task

Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth, and Nizar Habash
The CoNLL-2014 Shared Task

- Extends last year’s shared task
- CoNLL-2013 competition – five error types (account for about 50% of errors in the CoNLL data)
  - Articles
  - Prepositions
  - Noun number
  - Verb agreement
  - Verb form
- CoNLL-2014 evaluates with respect to all errors (28 error types)
- Our system ranked first on after-revisions data and second on before-revisions data
System Design and Goals

- Build a **robust** system that can make use of
  - Machine-learning methods
  - Cheap linguistic resources
    - **Native English data (the Google Web 5-gram corpus)**
  - Knowledge of the error patterns of the learners
    - **Annotated learner data (training data of the shared task)**
  - Inexpensive but reliable linguistic knowledge
The Illinois-Columbia System: Overview

- Based on the Illinois system that ranked first in the CoNLL-2013 shared task
- Extends the Illinois system in several respects:
  - Targets additional error types
  - Uses model combination for robustness
  - Uses joint inference to eliminate inconsistent predictions
The Illinois-Columbia System

- Implements ideas proposed in our prior work in this area:
  - **Adaptation**, i.e. developing models that are aware of error patterns, using scarce annotation): NAACL’10, ACL’11, BEA’12
    - Especially important when training on native English data
    - Can also be used when training on learner data
  - **Algorithmic perspective**: ACL’11
  - **Linguistically-inspired approach** to correcting open-class errors: EACL’2014
  - **Joint inference**: EMNLP’13
    - To eliminate inconsistent predictions made by individual models
Overview

- The baseline Illinois system
  - Learning algorithm
  - Training data
  - Adaptation
  - Linguistic knowledge

- New features in the Illinois-Columbia system
  - Additional error types
  - Model combination
  - Joint inference
Key Dimensions of the Illinois System

- Algorithmic perspective
- Data source: native vs. annotated
- Model adaptation to learner errors
- Linguistic knowledge
Overview of the Illinois System

- Basic pre-processing: POS tagging and shallow parsing using the *Illinois POS tagger and chunker*
- Five machine-learning modules are implemented:

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Training data</th>
<th>Learning arg.</th>
<th>Adaptation</th>
<th>Ling. knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article Prep.</td>
<td>Learner</td>
<td>AP</td>
<td>Error inflation (NAACL’10, BEA’12)</td>
<td>Features</td>
</tr>
<tr>
<td>Noun</td>
<td>Native</td>
<td>NB</td>
<td>-</td>
<td>Candidate generation</td>
</tr>
<tr>
<td>Verb agr.</td>
<td>Native</td>
<td>NB</td>
<td>-</td>
<td>Candidate generation, separate learning (EACL’14)</td>
</tr>
<tr>
<td>Verb form</td>
<td>Native</td>
<td>NB</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1**: Overview of the Illinois system along the key dimensions.

- AP – Averaged Perceptron
- NB – Naïve Bayes
Novel Components in the CoNLL-2014 System

- Extends last year’s system along several dimensions:
  - Expanded set of errors
    - Word form
    - Mec (punctuation, capitalization)
    - Style
  - Model combination
  - Joint inference
Targeting Additional Errors

- Word form
- Mec (punctuation and capitalization)
- Style
Word Form Errors

- Example:
  - “The application of surveillance technology serves as a warning to the "murders/murderers and they might not commit more murder”

- Candidates: which words should be corrected?
  - Consider those that occur in the training data as word form errors

- Confusion sets: what are the possible alternatives for a given word?
  - 45% of corrections in the development data also occur in training
  - In addition, we generate inflected verb forms and noun forms for words tagged as verbs and nouns

- Learning: NB with adaptation trained on the Google corpus
Mec Error Category

- Errors in comma usage
  - Two classifiers:
    - A learned module for missing and extraneous commas (AP classifier on learner data with adaptation)
    - A pattern-based module (patterns are extracted from the training data)

- Capitalization
  - Pattern-based module (patterns are extracted from the training data)
Style Errors

- Example:
  - don’t → do not
  - [clause], however [clause] → [clause]; however [clause]
Model Combination

- The Illinois system (2013) trains individual error-specific components on either learner or native data
  - Learner data
    - Similar genre and word usage
    - Linguistic annotation (POS, parsing, etc.)
  - Native data
    - Large amounts of cheap data
    - May provide more coverage

- This year, we use model combination:
  - An AP classifier with rich features trained on learner data
  - A NB classifier with word n-gram features trained on native data
Individual modules make **inconsistent predictions**: Both the **noun** and the **article** classifier identify the problem because the other word is used as part of context features:

They believe that **such situation** must be avoided.

such situation → such a situations

**We use joint inference** implemented on top of individually-learned models using the ILP formulation (Roth&Yih’04)
Performance of the Illinois-Columbia System on the Development Data

<table>
<thead>
<tr>
<th>Model</th>
<th>F0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The (baseline) Illinois system</td>
<td>33.17</td>
</tr>
<tr>
<td>+Model combination</td>
<td>34.92</td>
</tr>
<tr>
<td>+Additional errors</td>
<td></td>
</tr>
<tr>
<td>Word form</td>
<td>36.07</td>
</tr>
<tr>
<td>Mec (punc. and cap.)</td>
<td>36.52</td>
</tr>
<tr>
<td>Style</td>
<td>37.09</td>
</tr>
<tr>
<td>+Joint inference</td>
<td>37.13</td>
</tr>
</tbody>
</table>

Table 7: Modules marked with a “*” helped on the test data, while those marked with a “-” hurt the performance.
Conclusion

- We have presented the Illinois-Columbia system that participated in the shared task.
- We have described the key design principles of the Illinois-Columbia system that were also used in the Illinois system and presented and evaluated the new components.