HLTDI: CL-WSD Using Markov Random Fields for SemEval-2013 Task 10

Alex Rudnick, Can Liu, Michael Gasser

HLTDI Lab
Indiana University
collaborators

Can Liu

Michael Gasser
Given an English sentence containing a polysemous target word...

we want to pick the right word in the output language.
We trained some classifiers to do CL-WSD. The ones that worked the best used evidence from multiple parallel corpora!
disambiguating letter

But a quick look at today's letters to the editor in the Times suggest that here at least is one department of the paper that could use a little more fact-checking.

All over the ice were little Cohens, little Levys, their names sewed in block letters on the backs of their jerseys.
But a quick look at today's letters to the editor in the Times suggest that here at least is one department of the paper that could use a little more fact-checking.

All over the ice were little Cohens, little Levys, their names sewed in block letters on the backs of their jerseys.
related work

"ParaSense or how to use Parallel Corpora for Word Sense Disambiguation"

Lefever et al. 2011 (and related papers)
three systems for CL-WSD
- just MaxEnt classifiers
- two layers of classifiers
- MRF: network of classifiers
preprocessing and extraction

- trained from the Europarl Intersection corpus

- about 880k sentences in all six languages

- find uses of the 20 different English words, along with their translations

- get (context, label) pairs!
tools

- NLTK tokenizer
- Stanford tagger
- Berkeley aligner
- TreeTagger (for lemmas)
el debate político en torno a la [central nuclear] de Temelín

the political debate on the Temelin [nuclear power plant]
el debate político en torno a la central nuclear de Temelín

the political debate on the Temelin nuclear power plant
el debate político en torno a la [central nuclear] de Temelín

the political debate on the Temelin nuclear power plant
features

- word form, tags, lemma
- three word window: word forms, POS tags, wordform/postag, lemma
- bigrams, tagged bigrams ...
one layer classifiers

- MaxEnt trained with megam
- one per word per language
- model $P(\text{label} \mid \text{features})$
- pickle for use later
two layer classifiers

Classifier stacking!

Use answers from four other languages as features during training.
two layer classifiers

\{ de=? , fr=? ,
   it=? , nl=? ,
   <features from before> \}
two layer classifiers

\{ de=\?, fr=\?, it=\?, nl=\?, \\
<features from before> \}
two layer classifiers

\{ \text{de=brief}, \text{fr=lettre,}, \text{it=lettera}, \text{nl=brief,}} \text{<features from before> } \}
MRF system

We need to make five decisions anyway.

Can making them all jointly help us out?
MRF system

Network of classifiers, inform each other about preferences at each time step

Pick labels for five variables jointly as an optimization problem
THE DISAMBIGUATION PENTAGRAM
loopy BP

At every time step, every node sends a message to all of its neighbors.

"Here's what my neighbors and I think about each of your options!"
loopy BP

Doing min-sum.

\[
\delta^t_{i \rightarrow j}(L_j) = \min_{l_i \in L_i} \left[ \phi_i(l_i) + \phi_{(i,j)}(l_i, l_j) + \sum_{k \in S \setminus \{i,j\}} \delta^t_{k \rightarrow i}(l_i) \right]
\]
loopy BP: messages

carta  20

letra  1000

mensaje  57

documento  70

comunicación  40
difficulties... or opportunities?!

To get the weights for the MRF edges, we need bitext for all languages, pairwise...

But we don't need that for the two-layer stacking approach...
future work

Come see us at HyTra in Bulgaria!

Working on sequence models for whole-sentence lexical selection
more future work

Lexical selection for low-resource hybrid RBMT

Spanish<->Guarani in progress!
Comments? Questions!

Thanks!