COLING 2018

**SURVEY PAPER**

A survey paper provides a structured overview of the literature to date on a specific topic that helps the reader understand the kinds of questions being asked about the topic, the various approaches that have been applied, how they relate to each other, and what further research areas they open up. A conference-length survey paper should be about a sufficiently focused topic that it can do this successfully with in the page limits.

due March 16, -5 hrs (6:59 PM)

This paper acts as a survey of extant open-source parsers for Chinese Word Segmentation (CWS). We applied several of these systems on Chinese Treebank v. 8.0 and Universal Dependency Chinese-G. We then present an analysis of the types of errors persistent in the remaining percent margin and conclude that there exist certain types of context-resolvable errors that some kinds of post-processing systems might be well-suited to rectify.

survey paper (slash reproduction paper)

reviewing what's out there, including bake offs

testing those tools on CTB 8.0 and UD (v.?)

analyzing errors between news and wiki genres

concluding that there are context-resolvable errors in the n%

INTRO

* CWS is essential for downstream tasks
* ambiguity problem
  + the two types
* what’s missing from the literature ??

BACKGROUND

* bake-offs
* other systems put forth
* methodology
  + character-based vs word-based
  + into stat models (CRF+RNN, etc.)
* our hypothetical ideal methodology
  + publish the dataset when publish bake-off results

METHOD

* how we prepared the data, using the converter to simp/trad
* basic description of evaluator
  + what counts as a correct vs incorrect prediction

RESULTS

* discuss errors, contextualized into what's been discussed in previous bake-offs
  + that whole taxonomy
* show careful thought about what's going on in that remaining %
  + are they inevitable? are they resolvable?
  + [all??] state of the art systems probabilistic
    - some looking at context, but not linguistics
  + doesn't appear anyone is doing rule-based
  + >> context-resolvable EXAMPLE

DISCUSSION

* the kinds of problems these models are making makes room for some kind of hybrid system where rules can fill some of that gap
* you could imagine auto-detection
* or put it through a segmenter and then through a Constraint Grammar that says if this kind of mistake, fix it with this rule

CONCLUSION

* obviously the stats models are doing great, but why guess if you can know! write a rule and don't let your prob system guess wrong

12 MAR

we know that genres behave differently, cite Nivre, but the difference is we are looking at how parsers do on CHINESE as a whole (no genre distinctions)

* if we have time, evaluate on subgenres in ctb8

ZPar and Ensemble

heavy work in doing the background section

assume that the people reviewing my paper are the ones who wrote the background

[+respect]

NOTES:

don’t need to get too deep. quote the systems saying what they are.

Jieba hasn’t been reported on at all

Stanford hasn’t been reported on UD

Nivre BiRNN-CRF model <https://github.com/yanshao9798/tagger>

not working… pretend it doesn’t exist?

typo example (or any trad vs simp on same parser) is good example for SOME kind of post-processing (whether auto DECCA or CG or whatever)

kind of speculative (CG)

but in the future, take output of any parser >> CG and report on the upped percentages (ONE DAY)

CTB is NOT open!

ANOTHER CONTRIBUTION, at least one distinction, significantly worse on wiki data than news data  
(>>discussion we need future distinction of accuracy per genre, whereas that would be helpful information) especially if I feel like I could identify reasons why by analyzing types of errors we're making in CTB vs UD (try and find reasons one or two that wiki UD data is underperforming) Acronyms?  
  
(could still report ctb evaluation) in the context of saying that these models that have mostly been focused on NEWS?  
either way, performing better on news than expository (wiki) genre  
2014 write-up doesn't separate out genres and what is doing well in which genre

our main contribution  
analyzing errors, simply showing what is state-of-the-art on freely available systems (an argument toward more open science endeavors for this type of thing)  
matter of fact-ly say, previous bake-offs have been incredibly valuable and helpful, yet none of them are using open data and this hampers progress and new developments  
can't compare new tools  
  
(we wish we could be a reproduction paper, but have to do a survey because of closed data)  
  
very subtle and complementary, but need a baseline for any others who want to do comparisons  
  
what is Chinese NLP community? the establishment using closed, with lots of others trying to do open  
  
for reviewed purposes, don't want to ruffle feathers, but also want to encourage free-ware supporters, clearly make the point that it's more valuable  
  
open data for bake-offs, dilemma with potential training their models on the test data  
  
hypothetically, publish the dataset when publish bake-off results  
  
our paper, we aren't doing a bake-off (that exact thing where trained on CTB and test on CTB)