

Detecting negation scope is easy, except when it isn't

Federico Fancellu¹ Adam Lopez¹ Bonnie Webber¹ Hangfeng He²

¹ILCC, School of Informatics, University of Edinburgh

²School of Electronics Engineering and Computer Science, Peking University

f.fancellu@sms.ed.ac.uk, {alopez, bonnie}@inf.ed.ac.uk,
hangfenghe@pku.edu.cn

Negation Scope Detection (at the string level)

- ▶ **Input:** a sentence containing at least one negation marker (or **cue**)
- ▶ **Task:** classify a token as part of the **scope** of the **cue** or not (binary classification)

I am Italian but I do n't eat pizza

Negation Scope Detection (at the string level)

- ▶ **Input:** a sentence containing at least one negation marker (or **cue**)
- ▶ **Task:** classify a token as part of the **scope** of the **cue** or not (binary classification)

I am Italian but I do n't eat pizza

It is not the case that I eat pizza

Negation Scope Detection (at the string level)

- ▶ **Input:** a sentence containing at least one negation marker (or **cue**)
- ▶ **Task:** classify a token as part of the **scope** of the **cue** or not (binary classification)

I am Italian but I do n't eat pizza

It is not the case that I eat pizza

It is the case that I am Italian

Neural Networks for Negation Scope Detection [Fancellu et al., 2016]

- ▶ Bi-LSTM for negation scope detection
- ▶ Performance on par or better than previous heavily-engineered or heuristics-based approaches
- ▶ Tested on Conan-Doyle neg. [Morante et Daelemans, 2012]

This work

- ▶ Several corpora annotated with negation scope
 - ▶ Different annotation decisions
 - ▶ Different domains
- ▶ Our question: **Does it work on these corpora?**
 - ▶ BioScope (EN) [Vincze et al., 2009]
 - ▶ 3 sub-corpora (Abstract, Full, Clinical)
 - ▶ SFUProductReview (EN) [Konstantinova et al., 2012]
 - ▶ CNeSp (ZH) [Zou et al., 2015]
 - ▶ 3 sub-corpora (Product, Financial, Scientific)

Joint model

- ▶ Same bi-LSTM architecture, same features
- ▶ Add a 4-parameter transition matrix to create the dependency on the previous output

$$p(s|w, c) = \prod_{i=1}^n p(s_i | s_{i-1}, w, c)$$

Evaluation

- ▶ Evaluation
 - ▶ Token-level: F_1 on tokens correctly classified
 - ▶ Scope-level: Accuracy of full scopes we correctly match
- ▶ Performance on par or better than previous work

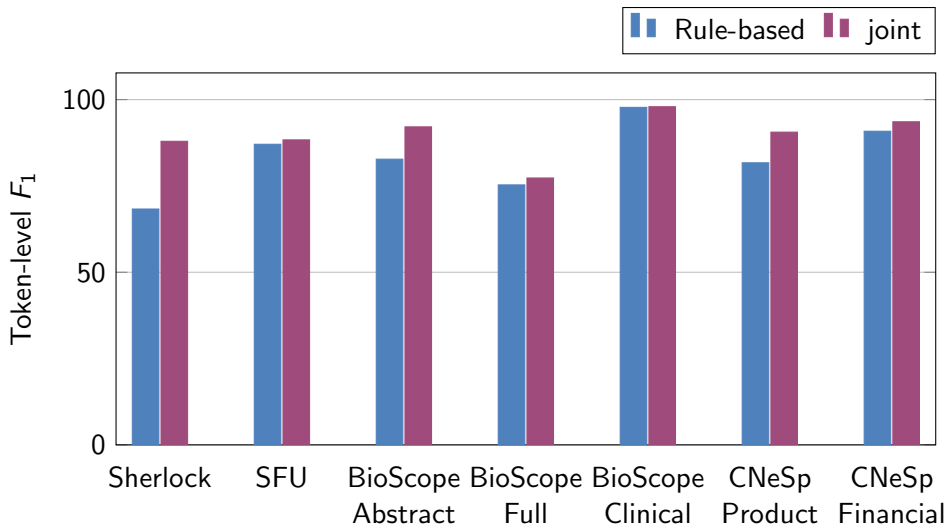
Rule-based scope detection

A lot of sentences where scope is delimited by punctuation

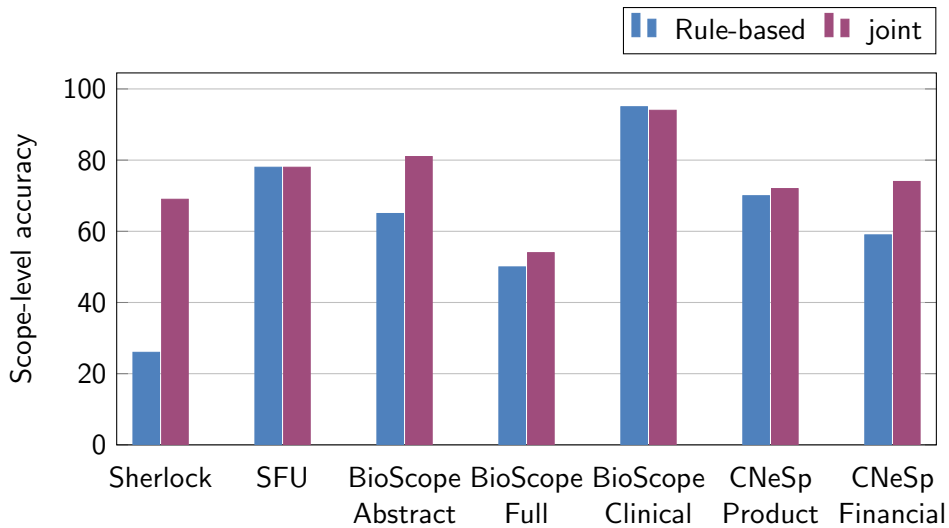
It helps activation , not inhibition of ibrf1 cells .



Results

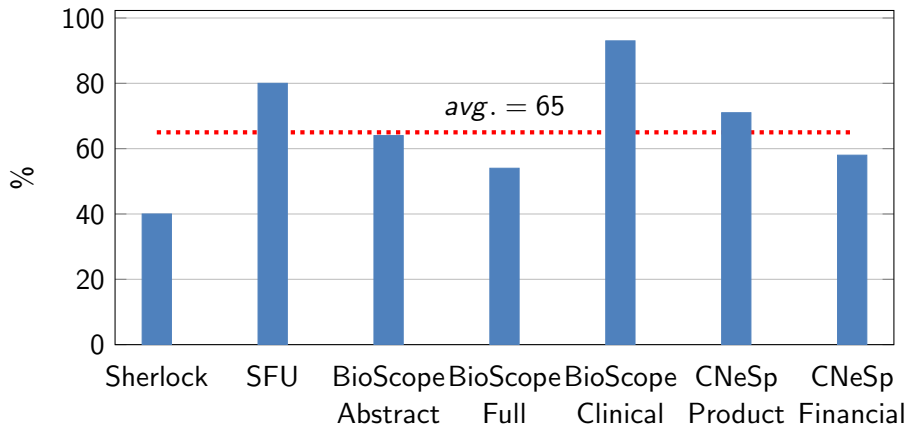


Results



Blame it on the training data

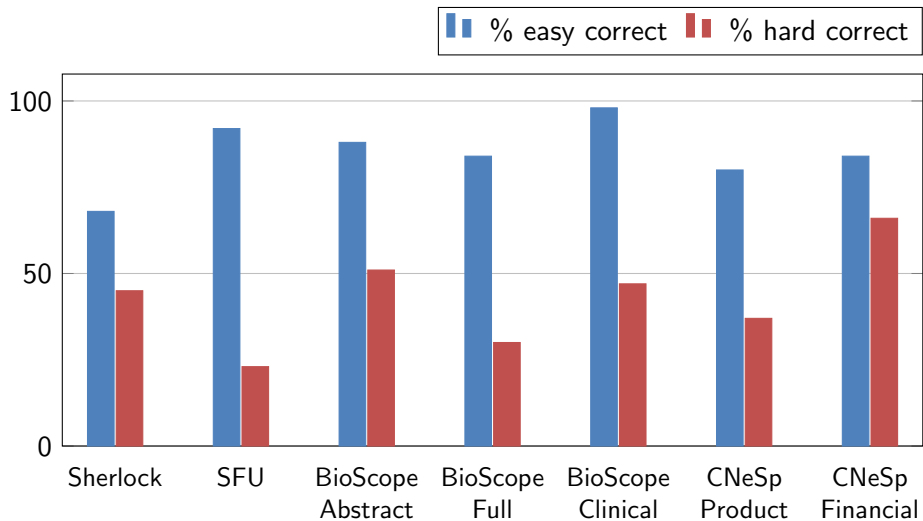
It helps activation , not inhibition of *ibrf1* cells .



Easy vs. hard instances

- ▶ **Easy:** predictable by punctuation
It helps activation , **not** inhibition of **ibrf1** cells .
- ▶ **Hard:** not predictable by punctuation
I do **not** use the **56k conextant winmodem** since I have cable access for the internet and he does not either .

Error analysis: *dev* set



Error analysis: *dev* set

- ▶ Most of the errors are due to the model trying to match punctuation boundaries

surprisingly , expression of **neither bhrf1 nor bcl-2** in a b-cell line **bjab** , protected by the cells from anti-fas-mediated apoptosis

Error analysis: *dev* set

- ▶ Most of the errors are due to the model trying to match punctuation boundaries

surprisingly , expression of neither bhrf1 nor blc-2 in a b-cell line bjab , protected by the cells from anti-fas-mediated apoptosis

Error analysis: *dev* set

- ▶ Most of the errors are due to the model trying to match punctuation boundaries

surprisingly , expression of neither bhrf1 nor blc-2 in a b-cell line bjab , protected by the cells from anti-fas-mediated apoptosis

I do **not** use the 56k conextant winmodem since I have cable access for the internet .

Error analysis: *dev* set

- ▶ Most of the errors are due to the model trying to match punctuation boundaries

surprisingly , expression of neither bhrf1 nor blc-2 in a b-cell line bjab , protected by the cells from anti-fas-mediated apoptosis

I do **not** use the 56k conextant winmodem since I have cable access for the internet .

Why does it happen?

Different corpora, different annotation styles

BioScope & SFU	
CNeSp	
Sherlock	

Why does it happen?

Different corpora, different annotation styles

BioScope & SFU	It helps activation , not inhibition of ibrf1 cells .
CNeSp	It helps activation , not inhibition of ibrf1 cells .
Sherlock	

Subject is seldom annotated

Why does it happen?

Different corpora, different annotation styles

BioScope & SFU	It helps activation , not inhibition of ibrf1 cells .
CNeSp	It helps activation , not inhibition of ibrf1 cells .
Sherlock	It <u>helps</u> activation , not inhibition of ibrf1 cells .

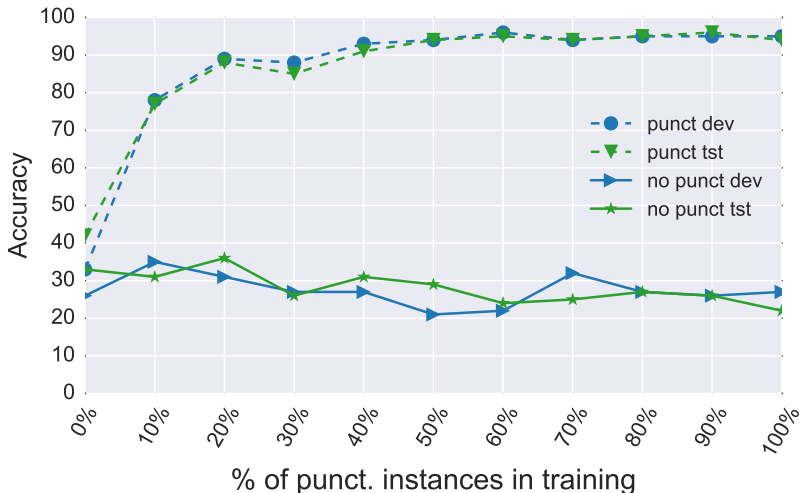
Subject is always annotated, omitted verb is retrieved

Is this problem caused by the annotation guidelines?

- ▶ We re-annotated 100 randomly selected sentences of 3 corpora using the Sherlock guidelines

Data	Easy original	Easy Sherlock
SFU	87%	42%
BioScope Abstract	84%	34%
CNeSp Financial	68%	45%

Undersampling is not enough



Conclusions

- ▶ GOOD PERFORMANCE FEELS GREAT BUT UNDERSTANDING YOUR MODEL FEELS EVEN BETTER!
- ▶ **Detecting negation scope is easy, except when it isn't:**
 - ▶ focus detection on those more difficult cases?