Syntax-aware Neural Semantic Role Labeling with Supertags

Dan Friedman*, Jungo Kasai*, Robert Frank, Owen Rambow, and Dragomir Radev

Yale University. * Equal Contribution.

LILY Lab

Introduction

Semantic role labeling (SRL) is the task of identifying the semantic relationships between each predicate in a sentence and its arguments (Gildea and Jurafsky, 2002). SRL information has been shown to be useful for such downstream tasks as machine translation, multi-document summarization, information extraction, and question answering.

Until recently it was believed that SRL systems require syntactic information to perform well. But recent work has demonstrated that deep neural networks can achieve competitive and even state-of-the-art performance without any syntactic information at all. These systems have the benefits, for example, of being simpler to implement and performing more robustly on foreign languages and out-of-domain data, cases where syntactic parsing is more difficult.

We show that using supertags is an effective middle ground between using full syntactic parses and using no syntactic information at all.

Supertag Design

We use dependency grammar-based supertags (Ouchi et al. (2014)). For a word w:

Model 0

Encodes the dependency relation and the relative position (direction) between w and its head, i.e. left (L), right (R), or no direction (ROOT).

Model 1

Adds to the "parent information" from Model 0 the information of whether w possesses dependents to its left (L) or right (R).

Model 2

Extends Model 1 by adding information about w's obligatory dependents, when w is a verb. When w lacks such obligatory children, we encode whether it possesses non-obligatory dependents to the left (L) or right (R) as in Model 1.

Model TAG

Encodes the dependency relation and the relative position (direction) of the head of a word similarly to Model 0 if the dependency relation is non-obligatory, and the information about obligatory dependents of verbs if any similarly to Model 2.

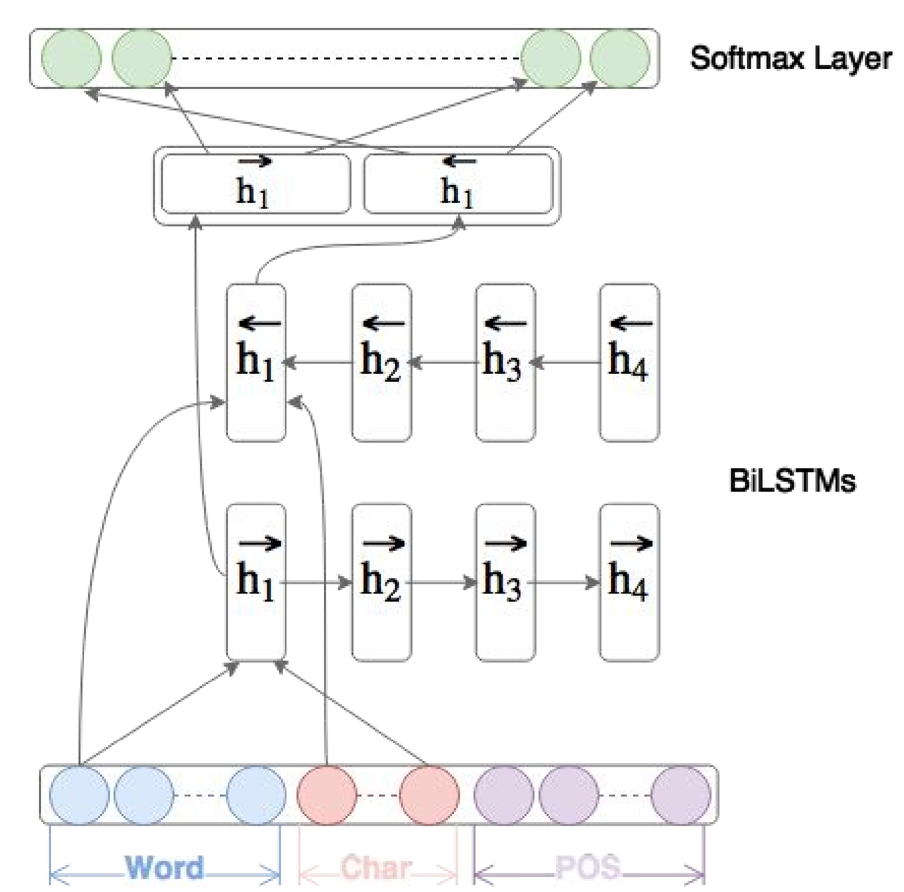
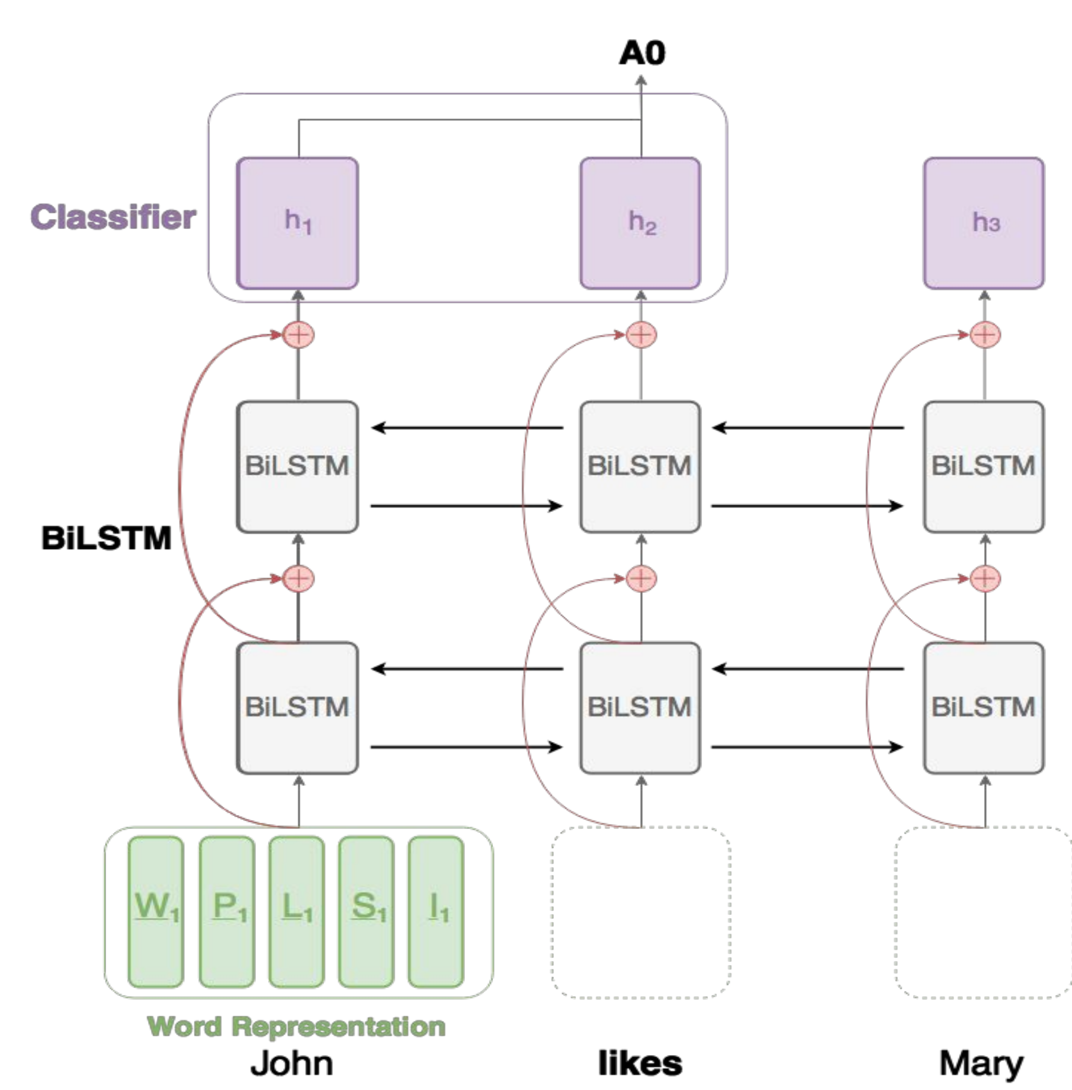


Figure 2. BiLSTM Supertagger



SRL architecture.

Input: Word embedding, POS embedding, lemma embedding, supertag embedding, andthe predicate indicator variable

Supertag	English # Stags	English Dev	English ID	English OOD	Spanish # Stags	Spanish Dev	Spanish ID
Model 0	99	92.93	94.17	88.71	88	92.97	92.67
Model 1	298	91.07	92.50	86.51	220	90.63	90.37
Model 2	692	90.60	92.05	85.40	503	90.08	89.84
Model TAG	430	92.60	94.17	87.46	317	92.33	92.18

Supertagging Results.

Non-ensemble System	P	R	F_1
FitzGerald et al. (2015) (global)	–	–	87.3
Roth and Lapata (2016) (global)	90.0	85.5	87.7
Marcheggiani et al. (2017)	88.7	86.8	87.7
Marcheggiani and Titov (2017)	89.1	86.8	88.0
BiLSTM	88.5	86.7	87.6
BiLSTM+DOut	88.0	87.6	87.8
BDH	88.3	87.8	88.1
BDH+Model 0	88.8	88.1	88.5
BDH+Model 1	89.0	88.2	**88.6**
BDH+Model 2	88.9	87.6	88.2
BDH+Model TAG	88.9	87.6	88.3

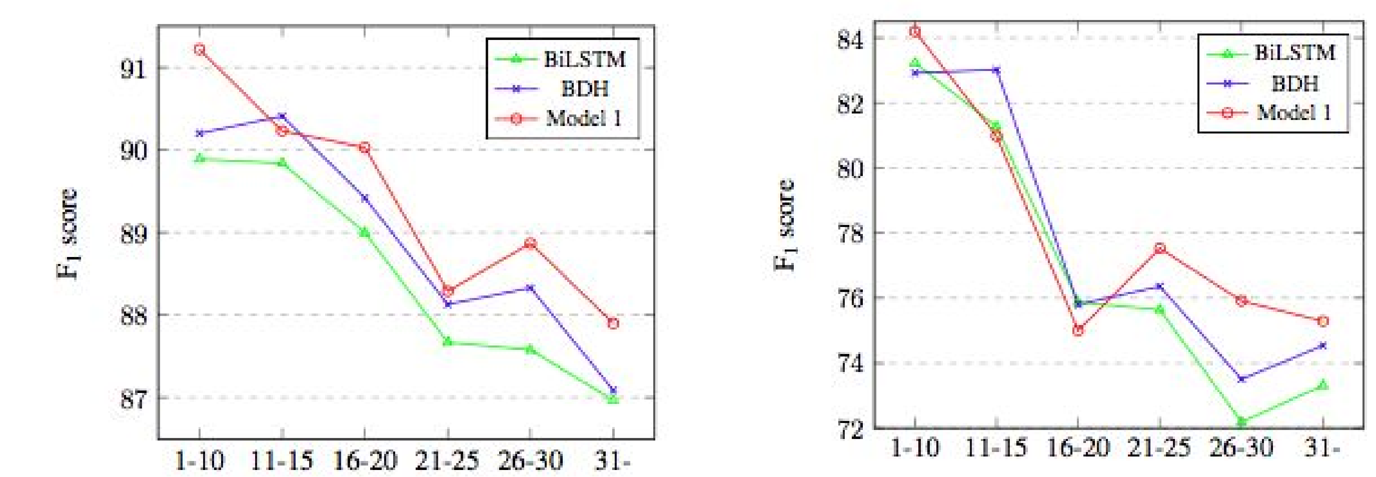
English in-domain Results.

Non-ensemble System	P	R	F_1
FitzGerald et al. (2015) (global)	–	–	75.2
Roth and Lapata (2016) (global)	76.9	73.8	75.3
Marcheggiani et al. (2017)	79.4	76.2	**77.7**
Marcheggiani and Titov (2017)	78.5	75.9	77.2
BiLSTM	77.2	75.6	76.4
BiLSTM+DOut	76.6	76.1	76.3
BDH	77.8	76.6	77.2
BDH+Model 0	77.4	76.3	76.8
BDH+Model 1	78.0	77.2	77.6
BDH+Model 2	77.6	76.3	76.9
BDH+Model TAG	78.6	76.8	**77.7**

English out-of-domain results.

Token	Model 0	Model 1	Model 2	Model TAG
No	DEP/R	DEP/R	DEP/R	DEP/R
,	P/R	P/R	P/R	P/R
it	SBJ/R	SBJ/R	SBJ/R	-
was	ROOT	ROOT+L_R	ROOT+SBJ/L_PRD/R	ROOT+SBJ/L_PRD/R
n't	ADV/L	ADV/L	ADV/L	ADV/L
black	NAME/R	NAME/R	NAME/R	NAME/R
Monday	PRD/L	PRD/L+L	PRD/L+L	-
.	P/L	P/L	P/L	P/L

Supertag Examples.



Results by sentence length (in-domain: left, ood: right)

Results and Discussion

Clear improvements in SRL performance, even outperforming systems that incorporate complete dependency parses. The gains in Model 1 come more from longer sentences, especially in the OOD test set. This implies that the supertag model is robust not only to the inter-domain difference but also to the sentence length, perhaps because supertags encode relations between words that are distant from each other in the sentence, information that a simple BiLSTM is unable to recover.

Conclusion and Future Work

We presented state-of-the-art non-ensemble SRL systems on the CoNLL-2009 English and Spanish data that make crucial use of dependency-based supertags. We showed that supertagging serves as an effective middle ground between syntax-agnostic approaches and full parse-based approaches for dependency-based semantic role labeling. Supertags give useful syntactic information for SRL and allow us to build an SRL system that does not depend on a complex architecture. We have also seen that the choice of the linguistic content of a supertag makes a significant difference in its utility for SRL.

We predicted supertags and semantic role labels independently for English and Spanish. However, sharing some part of a supertagger and role labeler across different languages could improve supertag- ging and SRL performance. In particular, if a pair of languages has morphological similarity such as Spanish and Catalan, we can jointly learn character- level representation from both languages. In future work, we will explore such multi-lingual transfer learning for supertagging and SRL.