[$^{NS}_{S}$ provisional title:]Leveraging Additional Resources for Frame-Semantic Role Labeling

Abstract

The high cost of semantic structure annotation is a major obstacle to automating semantic analysis with broad coverage. The fully annotated datasets that exist are often small, hindering the robustness of models trained on them. However, low-resource tasks may benefit from exploiting *partially* annotated data, as well as data with *different* (but related) forms of annotation, for additional training data or features. This paper considers the argument identification and classification subtask of frame-semantic parsing, which to date has relied exclusively upon full-text annotations in the FrameNet resource.[$^{NS}_{S}$ is this true? I think Dipanjan used exemplars for the latent variable frame ID model, but he hasn't used them at all for arg ID, right?] We augment supervised learning with additional "indirect" training data and features so as to leverage additional resources internal and external to FrameNet (e.g., PropBank). [$^{NS}_{S}$ result]

1 Introduction

[$^{NS}_{S}$ sparseness is a challenge for many computational semantics tasks]

Frame-semantic parsing (Das et al., 2014) is a case in point. This is the task of automating the rich linguistic structure analyses of the FrameNet lexicon and corpus (Baker et al., 1998).[1] FrameNet represents kinds of events and other scenarios with an inventory of **frames** ([$^{NS}_{S}$ examples]). Each frame is associated with lexical **predicates** (verbs, nouns, adjectives, and adverbs) capable of evoking the scenario, and a set of **roles** (or **frame elements**) called to mind in order to understand the scenario. These roles may be implicit, but are frequently realized linguistically in the same sentence as the predicate. Given a sentence, frame-semantic parsing is the task of mapping tokens in the sentences to evoked frames, and for each evoked frame, finding and labeling its **argument** phrases with roles.[$^{NS}_{S}$ example]

FrameNet 1.5 defines a structured hierarchy of over 1,000 frames associated with [$^{NS}_{S}$ #] English lexical predicates, and also provides annotations for [$^{NS}_{S}$ # targets annotated total] attestations of these frames/predicates in corpora, annotated in context with their arguments. In FrameNet 1.5, a rather small number of sentences—[$^{NS}_{S}$ #], comprising [$^{NS}_{S}$ #] words—are provided with **full-text** annotations, i.e. the sentence has been analyzed for all available frames. But a full [$^{NS}_{S}$ #]% of sentences in FrameNet—the lexicographic **exemplars**—are annotated for only one frame per sentence, and have thus far not been exploited successfully for frame-semantic parsing. Here, we seek to leverage these exemplar sentences as well as the (type-level) hierarchical structure of the FrameNet lexicon.

In this paper, we address the **argument identification** subtask of finding and labeling arguments given a predicate in context and the frame it evokes. This is a form of semantic role labeling (SRL). Notably, another resource, **PropBank** (Kingsbury and Palmer, 2002), has been widely used for SRL (Palmer et al., 2010). PropBank annotations capture shallower lexical frames and arguments; additionally, PropBank provides [$^{NS}_{S}$ millions?] of words of fully annotated English sentences (annotation is much less expensive, but also potentially less valuable, because of the shallower representation). To get the best of both worlds, we aim to tap into PropBank's vast resources as indirect token-level supervision for FrameNet-style analysis. We hypothesize that PropBank analyses can serve as a weak signal for the FrameNet SRL task,

[1] `http://framenet.icsi.berkeley.edu/`

either by heuristically transforming PropBank annotations into FrameNet annotations to augment the training data, or by preprocessing sentences with a PropBank SRL system to obtain new features for FrameNet argument identification.

Our experiments expand the *training data* and/or the *feature space* of supervised argument identification in order to integrate evidence from all of these sources into SEMAFOR (Das et al., 2014), the leading open-source frame-semantic parser for English.[2] The results show that some of these sources of evidence succeed at boosting argument identification performance.[$^{\mathrm{NS}}_{\mathrm{S}}$ SOTA (without constraints)?]

2 Resources

[$^{\mathrm{NS}}_{\mathrm{S}}$ incl. data analysis of differences from full-text]

[$^{\mathrm{NS}}_{\mathrm{S}}$ w/in each mention: genre/overall vocabulary; coverage and distributions of predicates, FE labels; oracle coverage of FT test]

2.1 Exemplars

[$^{\mathrm{NS}}_{\mathrm{S}}$ how different from FT]

2.2 PropBank

[$^{\mathrm{NS}}_{\mathrm{S}}$ how different from FN]

2.3 SemLink

[$^{\mathrm{NS}}_{\mathrm{S}}$ limitations!]

2.4 Illinois SRL system

3 Learning from multiple domains and representations

[$^{\mathrm{NS}}_{\mathrm{S}}$ Domain adaptation/multitask learning techniques]

3.1 Augmenting the Training Data

3.2 Frustratingly Easy

3.3 2-stage

3.4 Type-level hierarchy features

4 Experiments

[$^{\mathrm{NS}}_{\mathrm{S}}$ tuning regularizer for all experiments]

4.1 Baseline

[$^{\mathrm{NS}}_{\mathrm{S}}$ cost-augmented hinge rather than log loss, train with adadelta. runtime improvement, accuracies almost the same]

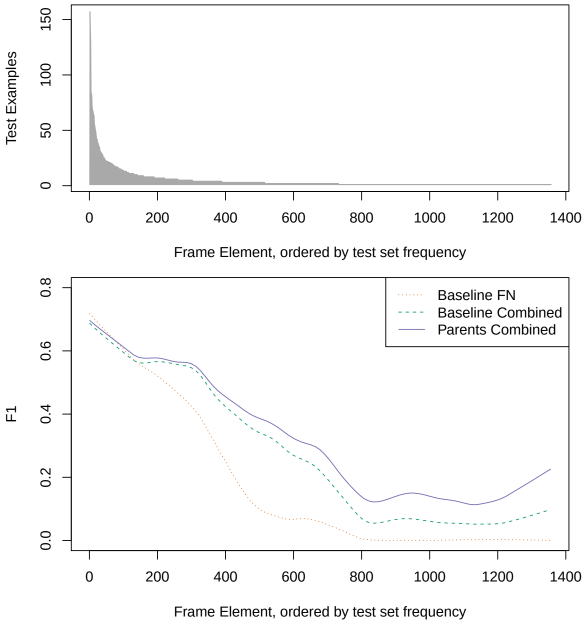

Figure 1: Count and F1 for each frame element appearing in the test set. F1 values have been smoothed with `loess`, with a smoothing parameter of 0.2.

[$^{\mathrm{NS}}_{\mathrm{S}}$ can mention recall-oriented experiments in a footnote]

[$^{\mathrm{NS}}_{\mathrm{S}}$ same FN 1.5 splits as Dipanjan]

[$^{\mathrm{NS}}_{\mathrm{S}}$ beam search decoding, not fancy hard constraints]

[$^{\mathrm{NS}}_{\mathrm{S}}$ preprocessing issues: removing duplicate sentences, merging adjacent split args in exemplars, OntoNotes PropBank preprocessing (NLTK), token-level SemLink details (such as filtering out sentences without mappable annotations; copy from WS paper)]

4.2 Evaluation

[$^{\mathrm{NS}}_{\mathrm{S}}$ main eval: FT test; new eval: exemplars]

4.3 Results

[$^{\mathrm{NS}}_{\mathrm{S}}$ results table without hierarchy features]

[$^{\mathrm{NS}}_{\mathrm{S}}$ hierarchy features: which ones work best (decided on baseline), how do they improve best result so far]

[$^{\mathrm{NS}}_{\mathrm{S}}$ Sam's curves on per-FE F1]

[$^{\mathrm{NS}}_{\mathrm{S}}$ comparison to prior work (baseline, best result). args+frames score vs. args only]

[$^{\mathrm{NS}}_{\mathrm{S}}$ discussion throughout]

5 Related Work

[$^{\mathrm{NS}}_{\mathrm{S}}$ Dipanjan's other papers; mention other PB SRL work?; anything using SemLink or combining

[2] `http://www.ark.cs.cmu.edu/SEMAFOR/`

	Test on FN			Test on Exemplars		
	P	R	F1	P	R	F1
Semafor baseline (trained on FN)	0.6603	0.5379	0.5929	0.649	0.336	0.4427
baseline trained on combined	0.66061	0.58234	0.61901	0.75639	0.65446	0.70174
trained only on exemplars	0.61084	0.49049	0.54409	0.77279	0.66228	0.71328
frust† on combined	0.65702	0.59043	0.62195	0.74018	0.61610	0.67247
siblings‡ on FN	0.67244	0.54763	0.60365	0.64922	0.39011	0.48737
siblings, trained on combined	0.65991	0.60406	0.63075	0.76369	0.68019	0.71952
parents⋆ on FN	0.67672	0.52790	0.59312	0.65440	0.38199	0.48239
parents, trained on combined	0.65920	0.60382	0.63029	0.76405	0.68616	0.72302
trained on FN+SemLink	0.655	0.3776	0.4791	-	-	-
with SRL augmented spans and features	0.70550	0.53178	0.60644	-	-	-

combined: FN + Exemplars training data

†: feature augmentation from the frustratingly easy DA paper

‡: for every feature f_i that fires for an argument a, fire an additional feature which is the conjunction: $(f_i \wedge \text{parent.frame} \wedge \text{parent.role} \wedge I_{hier})$ where parent.frame = parent(frame(a)). The parent's frame and role are obtained from the FN hierarchy. I_{hier} is an indicator to distinguish this feature from the regular conjunction features that use frame names and roles.

⋆: fire the siblings feature‡ and an additional feature: $(f_i \wedge \text{frame} \wedge \text{frame.role} \wedge I_{hier})$

resources for SRL?]

[$^{NS}_{S}$ multitask learning?]

6 Conclusion

[$^{NS}_{S}$ overall findings]

[$^{NS}_{S}$ future work: testing ground for improvements to PB and SemLink; automatic mappings between resources]

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proc. of COLING-ACL*, pages 86–90. Montreal, Quebec, Canada.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56.

Paul Kingsbury and Martha Palmer. 2002. From Tree-Bank to PropBank. In *Proc. of LREC*, pages 1989–1993. Las Palmas, Canary Islands.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. *Semantic Role Labeling*. Number 6 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool, San Rafael, CA.