Heterogeneous Supervisory Data Sources for Semantic Parsing

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Abstract

A number of semantic annotation efforts have produced a variety of annotated corpora, capturing various aspects of semantic knowledge in different formalisms. Due to to the cost of these annotation efforts and the relatively small amount of semantically annotated corpora, we argue it is advantageous to be able to leverage as much annotated data as possible. This work presents a preliminary exploration of the opportunities and challenges of learning semantic parsers from heterogeneous semantic annotation sources.

1 Introduction

Multiple annotated resources capture relational semantic information, including FrameNet (Fillmore et al., 2002), PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2008), and the AMR Bank (Banarescu et al., 2013) for English. Each of these is the result of painstaking lexicography and annotation efforts by teams of linguists over several years. However, they are the result of largely independent annotation efforts which make different theoretical commitments. Here, we consider how we might best leverage such heterogenous resources to improve semantic parsing, so as to let semantic resources created at considerable cost not go to waste.

Supervised semantic parsers are usually trained with the representation and annotations from a single resource. For instance, SEMAFOR (Das et al., 2010, 2013) is a frame-semantic parser trained on the full-text annotated corpus of FrameNet. PropBank has likewise served as the training resource for semantic role labeling (SRL) systems (e.g., Punyakanok et al., 2008; see Palmer et al., 2010 for a review). FrameNet and PropBank seek to encode very similar aspects of meaning, including event predicates and their labeled arguments; yet the two projects have developed different representations, lexicons, and corpus annotation approaches due to different design considerations. Broadly speaking, FrameNet contains richer forms of linguistic detail (e.g., predicates are semantically organized with respect to one other in FrameNet), but the PropBank representation is much more conducive to large-scale annotation, and PropBank therefore has much greater type and token coverage of verbs. Moreover, the annotated corpora of FrameNet and PropBank capture different domains. As such, the two resources can be seen as partially overlapping but partially complementary. We hypothesize that systems could exploit this complementarity to gain robustness at semantic analysis. Encouragingly, some studies of dependency parsing has demonstrated that divergent treebanks can be leveraged to improve parsing quality (Zhou and Zhao, 2013; Johansson, 2013).

From a supervised learning perspective, the problem of combining the knowledge across these resources can be viewed as a joint learning problem. If we treat the annotation schema of each resource as one label-space, then we can formulate this as a multitask learning problem across tasks with different label spaces. The goal of a system should then be to learn a model over the different label spaces. This problem also manifests in the information extraction community, where there are multiple knowledge-bases confirming to different ontologies.

To see an example illustrating the need for models that integrate resources, consider the following sentence:

> *Taxing* wealthy individuals *appeals* to liberals; Brown *urges abolishing* regulations.

The four verbal predicates in this sentence are italicized. Figure 1 shows the frame-semantic parse predicted by SEMAFOR. We observe:

1. *urges* is mapped correctly to a frame, and its arguments to frame elements.

2. *individuals* receives the correct frame label, PEOPLE, but is missing an argument ("wealthy" should fill the **Descriptor** frame element). In other wordings of this sentence, "wealthy" was correctly identified as an argument to PEOPLE, but incorrectly labeled with the **Ethnicity** frame element.

3. *appeals* is correctly identified as a target, but the wrong frame label is chosen (a sense disambiguation error). The parse suggests that somebody is making an appeal to liberals, whereas the correct analysis would use the EXPERIENCER_OBJ frame to represent that an idea (taxing wealthy individuals) provokes an emotional response on the part of liberals. Because the inventory of valid arguments is different for each frame, and SEMAFOR predicts frame labels first, an error in frame identification will likely produce incorrect argument analyses.

4. *abolishing* is not identified as a target because it is absent from the FrameNet lexicon. In principle, however, it should be recognized as evoking the PROHIBITING frame, which contains nearly synonymous verbs. SEMAFOR[$^{NS}_{S}$ the demo version used for the figure?] attempts to disambiguate such OOV lexical items to existing frames (Das and Smith, 2011).

5. *taxing* is not identified as a target because it, too, is absent from the FrameNet lexicon. However, there is no apparent home for it in any existing frame; a TAXATION frame would have to be defined in the lexicon.[1]

All of these verbs are annotated in the PropBank corpus, which assigns a coarse sense number to each predicate and describes the predicate's core and non-core arguments. Even though PB and FN use different label spaces for predicates/frames and arguments/frame elements, we propose that PB could provide useful information for frame-semantic parsing with respect to (a) detecting and labeling arguments (#2 above), (b) disambiguating targets that have multiple senses in both the PB and FN lexicons (#3 above), and (c) labeling arguments where SEMAFOR is able to identify the correct frame for a target that is not in FN, provided that some other verbs in the same frame *are* present in both PB and FN (#4 above).

The following list indicates the scenarios where a joint model based on both FrameNet and PropBank data can improve frame-semantic parsing accuracies in a system like SEMAFOR. The examples and numbers presented below were obtained from the latest FrameNet release (1.5) and for PB we consider only the WSJ section of the annotations.

- Frames with none or very few annotations in FN. The frame MANIPULATION does not have any annotations for the verb *hold*, whereas PB-WSJ has 117 instances (that had an extractable argument mapping) for this frame-target combination. The frame EXPERIENCER_OBJ has several predicates without any annotations, for example *appeal*, *harass*, *worry*, and *boggle*. The first three of these have annotations in the PB-WSJ data. There are 984 such predicates in FN that have been assigned to frames, but with no sentence annotations available.
- Frames with few known targets or targets with no associated frames. For example, the frame GIVING has 19 known verbs as targets in FN. Many (near-)synonyms of *give* such as *allot*, *assign*, *designate*, and *allocate*, are not present in FN. Overall, there are 475 verbs in the PB WSJ data that are not targets for any frame in FrameNet. Some examples are: *involve*, *lurk*, *nominate*, *ladle*, *entice*, and *bank*.

Ideally, a joint learning system should also be able to suggest new targets for a frame based on the lexical similarity with the frame's existing targets.

Linguistic resources to map the different annotations have been built, but to a very limited extent. SemLink (Bonial et al., 2014) is one such database that maps and unifies different lexical resources of semantic information: such as PropBank (PB), FrameNet (FN), VerbNet (VN). The mappings that SemLink provides are available at two different levels: (a) sentence level parallel annotations - the WSJ section of the PB data has been annotated with the appropriate frames and frame-elements as well as VerbNet classes wherever possible, thereby giving detailed PB, VN and FN annotations for each sentence (b) concept/role-set level mappings—these are coarse-level mappings defined between a PB roleset and a VN role-set or a FN frame and a VN roleset. These mappings can be one-to-one, one-to-many or many-to-

[1] As part of the definition process the frame would be linked to related concepts that already have frames, such as GOVERNMENT_INSTITUTION, IMPOSING_OBLIGATION, and COMMERCE_COLLECT.

Figure 1: Frame semantic parse from SEMAFOR for a constructed example sentence.

many depending on the semantic generality of the involved rolesets. To go from a PB roleset to a FN frame, one has to go via the VN roleset first.

In this work, we will focus on SEMAFOR as the target system whose performance we want to improve. We are hence interested in using the SemLink mappings to get FN compatible data. The type (a) mappings can be directly used to augment the FN annotations; in the next section we present an analysis on the quantity and usability of the available mappings. Using the type (b) annotations first requires disambiguating the frame-to-roleset labels and then aligning the predicate arguments of the roleset with those expected in the frame. Note that the two argument sets might be of different cardinality.

2 Data analysis

We present some statistics of the data resources in Tables 1 and 2. The first table summarizes the SemLink mappings for the PB WSJ data. The total number of annotated PB instances are 74977, a majority of which do not have the corresponding FN labels due to various difficulties in mapping them. Around 31% of the predicates have the frame label `IN` meaning "indefinite". For these cases the mapping from VN to FN does not clearly indicate which Frame the instance should be. About 20% of the instances are labeled `NF` or "no frame" as FrameNet currently does not have an appropriate frame class for the semantic concept represented by the PB predicate. Of the remaining, 21% are incomplete as they only have frame labels and no argument annotations. Most of these are predicates with modifier arguments. This leaves about 29% usable mappings, 9% of which had argument pointer and other issues. Eventually, we were able to extract about 15323 instances.

The second table compares the extracted PB-WSJ annotations with the original FN annotations (version 1.4) used to train the SEMAFOR model. The number of annotations at the sentence-level

Table 1: Statistics of PB-WSJ data from SemLink

FN frame annotation	Verb tokens from PB	% of all
Frame label = `NF`	14624	20%
Frame label = `IN`	22982	31%
Frame with no arguments	15533	21%
Frame with at least 1 argument mappable	15323	20%
Instances not mapped due to other issues	6516	9%
Total PB predicate tokens (instances)	74977	100%

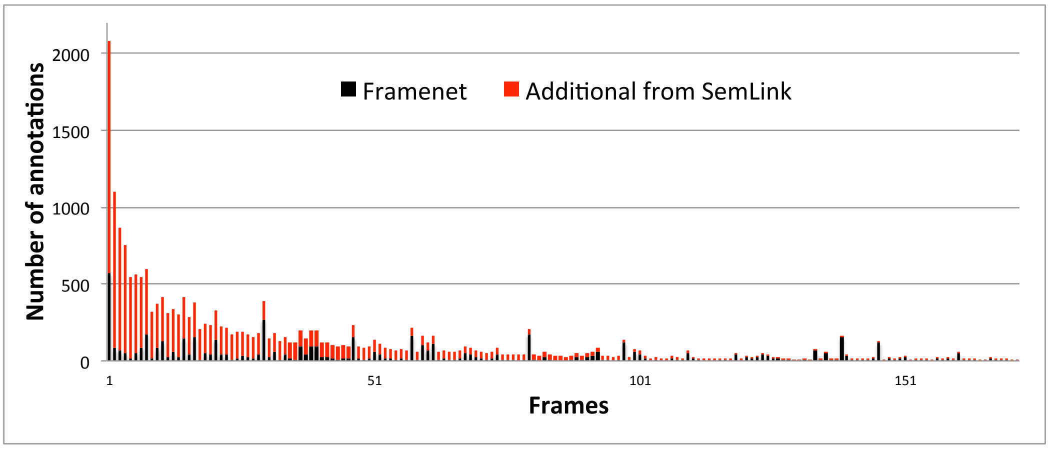
Table 2: Comparison of annotation density

Annotation unit	Count
Sentences in FN 1.4	2780
Annotation sets in FN 1.4 (i.e frame instances)	23940
Annotation density	8.6
Sentences in PB WSJ data	38594
Verb tokens in PB WSJ	74977
Annotation density	1.9
Sentences extracted from SemLink PB-WSJ	12382
Annotation sets extracted	15323
Annotation density	1.2

and frame-level are shown. It is evident that FN has a much higher annotation density - around 9 frames per sentence, as compared to around 1.2 for the PB derived data .

The mappings we obtained increase the number of annotations for around 170 frames. Figure 2 shows a stacked bar-chart plotting the annotations for every frame, sorted in decreasing order of additional frames obtained. The highest is around 1500 new annotations for the frame STATEMENT.

Figure 2: The number of new annotations per frame obtained upon extracting the WSJ mapping from SemLink. The bars are coloured (red) to indicate the contribution from the new annotations. Only frames for which new annotations were found are shown.



2.1 Noisy SemLink annotations

Since SemLink was built, FrameNet has updated frame definitions, some new frames were introduced and old ones deleted or divided into finer categories, irrelevant predicates moved to proper frames etc. Hence some of the SemLink annotations are obsolete and sub-optimal. For example, the frame STATEMENT does not contain the target `complain.v` anymore as a new frame COMPLAIN was introduced. There are around 3000 such instances with obsolete annotations. These can be detected easily and they are also not semantically absurd, but correcting these will require re-annotation on a case-by-case basis.

Additionally, there are some mistakes in some of the annotations due to the existence of multiple frame matches for a particular predicate. For example, in the sentence *McMoRan Energy Partners will be* `liquidated`, the frame for *liquidate* is KILLING - all 14 occurrences of liquidate have this error. The sentence *Speaker Jim Wright... attempting to* `direct` *the president* has the frame annotation BEHIND_THE_SCENES, which refers to film direction. There are 17 instances with this frame erroneously marked. These kind of errors are hard to detect. The SemLink mappings can thus not be used as gold-standard annotations to train models. One possibility is to use this data as low-confidence training data.

3 Models for joint learning

This problem of learning from multiple resources can be formulated in different ways. We present here a spectrum of possible approaches, each relying either on a different methodology or manipulating the available data differently. For simplicity, let us assume there are two resources D_1, D_2. Let $X_1, X_2 \in \mathcal{X}$ represent the set of data instances (i.e sentences) from the two sources and $Y_1 \in \mathcal{Y}$, $Y_2 \in \mathcal{Y}'$ be the labels.

- Learning a deterministic mapping $\phi : \mathcal{Y}' \rightarrow \mathcal{Y}$ between the label spaces from the two resources. Such a mapping can be used to transform the annotations from one resource into the labeling schema used by the other. We could then train a model using $\{X_1, Y_1\} \cup \{X_2, \phi(Y_2)\}$. The SemLink mappings we discussed earlier give us such a function ϕ that gives a limited and ambiguous mapping.
- A two-stage pipeline that trains a model on D_1 say θ_1 and apply it on X_2. The output labels can be used as features along with X_2 to train the model θ_2 on D_2.
- A bootstrapping approach that uses co-training similar to (Clark et al., 2003), where we iteratively improve one model using the output from the other. This is the same as the pipeline approach described above, performed for several iterations. At iteration t, we obtain θ_1^t and apply it on X_2. The output labels are used as features along with X_2 to train the model θ_2^t. In the next iteration $(t+1)$, the output labels of θ_2^t applied on X_1 can be used as features to train θ_1^{t+1}.
- A multi-task learning based objective function that maximizes the log likelihood over the two data sources D_1 and D_2. The parameters from the two tasks can be coupled together using a feature transformation (Argyriou et al., 2006), or based on feature correspondences learned using

unlabeled data (Blitzer et al., 2006).

4 Evaluation

As indicated in the introduction, a model encompassing multiple resources can fill in various types of gaps in the FN annotations. It is infeasible to measure improvements for predicates which are not already frame targets in FN. For existing frame-predicates with no argument annotations, the SemLink mappings can be used as test data.

Acknowledgments

References

A. Argyriou, T. Evgeniou, and M. Pontil. 2006. Multi-task feature learning. *NIPS*.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186. Association for Computational Linguistics, Sofia, Bulgaria.

John Blitzer, R. McDonald, and F. Pereira. 2006. Domain adaption with structural correspondence learning. *EMNLP*.

Claire Bonial, Kevin Stowe, and Martha Palmer. 2014. Renewing and revising semlink. *ACL*.

Stephen Clark, James R. Curran, and Miles Osborne. 2003. Bootstrapping pos taggers using unlabelled data. *CoNLL*.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2013. Frame-Semantic Parsing. *Computational Linguistics*, 40(1):9–56.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. *NAACL-HLT*.

Dipanjan Das and Noah A. Smith. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *ACL*.

Charles J. Fillmore, Christoper R. Johnson, and Miriam R. L. Petruck. 2002. Background to framenet. *International Journal of Lexicography*.

Richard Johansson. 2013. Training parsers on incompatible treebanks. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 127–137. Atlanta, United States.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation Journal*.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic Role Labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103.

V. Punyakanok, D. Roth, and W. Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*.

Guangyou Zhou and Jun Zhao. 2013. Joint inference for heterogeneous dependency parsing. In *ACL*.