

Semi-supervised SRL system with Bayesian inference

Alejandra Lorenzo, Christophe Cerisara

▶ **To cite this version:**

Alejandra Lorenzo, Christophe Cerisara. Semi-supervised SRL system with Bayesian inference. 15th International Conference on Intelligent Text Processing and Computational Linguistics, Apr 2014, Kathmandu, Nepal. pp.433. hal-01015414

HAL Id: hal-01015414
https://hal.science/hal-01015414

Submitted on 26 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-supervised SRL system with Bayesian inference

Alejandra Lorenzo and Christophe Cerisara

LORIA / UMR 7503, Vandoeuvre-les-Nancy, France
alelorenzo@gmail.com, cerisara@loria.fr

Abstract. We propose a new approach to perform semi-supervised training of Semantic Role Labeling models with very few amount of initial labeled data. The proposed approach combines in a novel way supervised and unsupervised training, by forcing the supervised classifier to overgenerate potential semantic candidates, and then letting unsupervised inference choose the best ones. Hence, the supervised classifier can be trained on a very small corpus and with coarse-grain features, because its precision does not need to be high: its role is mainly to constrain Bayesian inference to explore only a limited part of the full search space. This approach is evaluated on French and English. In both cases, it achieves very good performance and outperforms a strong supervised baseline when only a small number of annotated sentences is available and even without using any previously trained syntactic parser.

1 Introduction

1.1 Data scarcity in semantic role labeling

Semantic Role Labeling (SRL) is a major task in Natural Language Processing which provides a shallow semantic parsing of a text. Its primary goal is to identify and label the semantic relations that hold between predicates (typically verbs), and their associated arguments [1]. The analysis of semantic relations and predicate argument structures has many potential applications in Natural Language Processing (NLP). In particular, applications such as natural language understanding, machine translation, information extraction and question answering are shown to benefit from semantically annotated text. The extensive research carried out in this area resulted in a variety of annotated resources, which, in time, opened up new possibilities for supervised SRL systems. Although such systems show very good performance, they require large amounts of annotated data to be successful. This annotated data is not always available, very expensive to create and often language and domain specific [2].

To bypass these shortcomings, different solutions have been proposed. Unsupervised and semi-supervised learning techniques are two possible options to address the data scarcity problem. Unsupervised learning attempts to induce the annotations from large amounts of unlabeled data, while semi-supervised models are trained on both a limited quantity of labeled examples and a larger

unlabeled corpus. A first claass of semi-supervised systems for SRL exploit a bootstrapping approach, such as self-training and co-training [3, 4]. An alternative solution to combine labeled and unlabeled data is "semi-unsupervised" systems, which start from some unsupervised model and train a small number of this model's parameters on the limited labeled corpus available [5]. One example of the application of these semi-unsupervised approaches on the semantic role labeling task is described in [6].

1.2 Proposed semi-supervised approach

Our proposed approach to semantic role labeling is inspired by the work of [6]. However, one of the main differences is that, while they essentially use the labeled data to build an informed prior distribution over the unsupervised model parameters, we rather use the labeled data to train a supervised classifier which role is to generate a set of potential arcs, and then introduce these labels as virtual evidence [7] to constrain unsupervised Bayesian inference. The proposed system can thus either be viewed as a semi-supervised approach, where the output of the initial bootstrapped supervised system is filtered by the unsupervised model, or as a semi-unsupervised approach, where inference in the unsupervised model is constrained and guided by the supervised solutions. It is, to the best of our knowledge, the first SRL system that does not explicitly root itself in one or the other paradigm.

Another advantage of the proposed approach is that it produces labeled semantic roles and can thus be evaluated with supervised SRL metrics. Moreover, the proposed model automatically detects all candidate argument chunks and predicates and we do not make any assumption about predicate argument structures. Instead, we let the model infer a semantic structure by detecting not only the semantic role associated with each argument but also the predicate it shall be linked to. As shown in Section 3, the proposed approach shows competitive performance even under the assumption of no pre-existing syntactic parser. Hence, as opposed to many previous works that rely on either gold syntactic trees or trees obtained with supervised parsers trained on a large training corpus, we rather train our parser on the same small training corpus than for SRL. To summarize, the only inputs needed in the proposed approach are part-of-speech tags, and a very small initial corpus labeled with syntax and semantic arcs.

2 Task definition

Semantic role labeling is the task of automatically finding the semantic roles for each predicate in a sentence. That is, finding out which constituents in a sentence are semantic arguments for a given predicate and then determining the appropriate role for each of these arguments. Different definitions for "semantic role" have been proposed. In this work, we use the definition provided by PropBank [8], since it is commonly used in the NLP community and there are available versions of this resource in the target languages. The semantic roles in

PropBank are defined with respect to individual verb senses or predicates. Thus, each predicate has a number of roles. In general, roles A0 and A1 attempt to capture Proto-Agent and Proto-Patient roles [9], and thus are more valid across verbs and verb instances than A2-A5 roles.

In this work, we focus on both determining which constituents in a sentence are semantic arguments for each predicate, and labeling these arguments with semantic roles. For this, we decompose the overall process into two main steps:

1. Candidate arcs generation: this process generates a set of possible candidate semantic arcs (see Section 3);
2. Bayesian inference: this process selects the most likely semantic arcs from the set of candidates (see Section 4).

Figure 1 illustrates the results produced by the proposed approach on an example sentence. The inferred dependency arcs and semantic relations are shown respectively above and below the sentence.

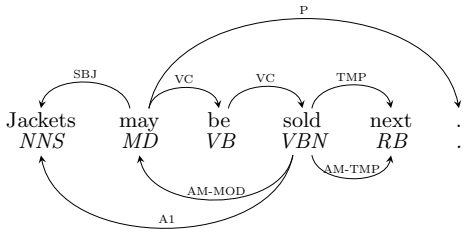


Fig. 1. Example of inferred sentence.

3 Candidate arcs generation

The first stage is summarized in Figure 2, with the following notation:

- $\mathcal{L}$ is the small initial manually labeled corpus used to train our supervised classifiers. It is typically composed of 50 sentences that are automatically tagged with an existing POS-tagger, and manually annotated with labeled dependencies and semantic relations.
- $\mathcal{U}$ is a large unlabeled corpus, only automatically annotated with POS tags.
- $\mathcal{T}$ is the test or gold manually annotated corpus. It is only used to evalute the performances.

This first stage only exploits an initial set of 50 manually labeled sentences to train both the MATE syntactic parser [10], and a Maximum Entropy semantic model M_s with L-BFGS optimization[1]. Both these supervised models are used to produce a set of candidate semantic arcs on the unlabeled corpus.

[1] We use the Stanford Classifier: `http://nlp.stanford.edu/software/classifier.shtml`

These candidate semantic arcs are then used in the next stage to constrain Bayesian inference to only use plausible semantic arcs, hence greatly reducing the size of the search space on the unlabeled corpus. It is thus very important to maximize the recall of M_s so that the set of candidate semantic arcs comprises most if not all gold arcs. In other words, M_s must miss as few semantic arcs as possible, which imply to maximize its recall while its precision shall be kept at a reasonable level but is not as important there.

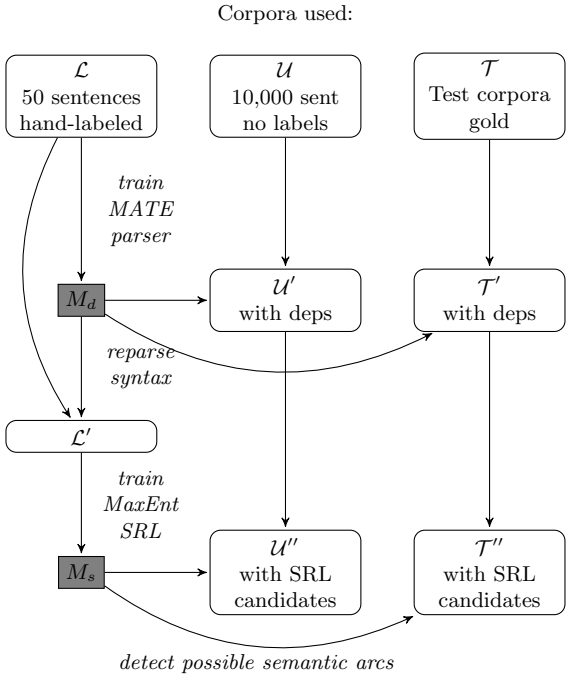


Fig. 2. Procedure to generate candidate semantic arcs, before inference

3.1 Supervised semantic model

Following common practice, all non-auxiliary verbs are selected as predicates[2]. Then the supervised Maximum Entropy semantic model M_s is trained on the small manually labeled corpus $\mathcal{L}$' as detailed in Figure 3. For this training, we first run an optional pre-processing step, which consists of a very simple rule based chunker that uses word forms and POS tags to segment noun and prepositional phrases, and thus reduce the number of possible arguments. The Maximum Entropy semantic model's features $\Phi(a, p)$ computed for each predicate-argument pair are shown in Table 1.

[2] In this work we do not disambiguate between predicate senses.

Table 1. Features used in M_s for each argument-predicate pair (a, p)

- Letter 4-gram, prefix and suffix 4-grams and lengths of a, p, (a, p), argument context bigram (a_{t-1}, a_{t+1}) and syntactic dependents of a
- POS tags of a and p
- Distance from a to p: -4 and less, -3, $\cdots$, $+3$, $+4$ and more
- indicator that a is the syntactic head of a NP or PP
- Dependency type from a
- Boolean true iff a and p are directly linked syntactically

1: **for** every predicate p (p is any non auxiliary verb as given by the POS tags) **do**
2: **for** every argument a (a is the estimated syntactic head of any NP or PP) **do**
3: Compute features $\Phi(a, p)$
4: **if** arc $(a, p) \in \mathcal{L}'$ **then**
5: Set class $c =$ label of $(a, p) \in \mathcal{L}'$
6: **else**
7: Set class $c = NOARC$
8: **end if**
9: Add the observation $(\Phi(a, p), c)$ to the training set T_r
10: **end for**
11: **end for**
12: Train the maximum entropy model on T_r

Fig. 3. Train MaxEnt semantic model

Once the semantic model is trained on the labeled corpus $\mathcal{L}$', it is applied on the unlabeled $\mathcal{U}$' and test $\mathcal{T}$' corpus as detailed in Figure 4.

4 Bayesian Model

The candidate semantic arcs proposed by the semantic model M_s are used as constraints during inference on $\mathcal{U}'' \cup \mathcal{T}''$ of the posterior of the Bayesian model. This Bayesian model is designed to encode standard linguistic features very similar to the ones used in most other unsupervised SRL models. The factors representing these features are described next and include lexical roles preferences $p(w|a)$, arguments position $p(pos|a)$ and syntactic roles preferences $p(d|a)$, which give the following joint on $\mathcal{U}'' \cup \mathcal{T}''$:

$$P(W, A, POS, D) = \prod_{u \in \mathcal{U}'' \cup \mathcal{T}''} \prod_{w_t \in A_u} P(w_t|a_t) P(pos_t|a_t) P(d_t|a_t)$$

where u is one sentence of the corpus, A_u is the observed set of argument candidates proposed by M_s for sentence u, d_t is the observed estimated dependency type from w_t, and (a_t, pos_t) are latent and chosen during inference from the set of candidate semantic arcs proposed by M_s for w_t. Exactly one semantic arc is chosen for each w_t during inference.

```
for every predicate p (p is any non auxiliary verb as given by the POS tags) do
    for every argument a (a is the estimated syntactic head of any NP or PP) do
        Compute features Φ(a, p)
        Use M_s to compute P(c|Φ(a, p)), where c spans all arc labels
        for every arc label c do
            if P(c|Φ(a, p)) > P(NOARC|Φ(a, p)) then
                Add the arc (a, p, c) to the set of candidate arcs
            end if
        end for
    end for
end for
```

Fig. 4. Compute candidate semantic arcs

4.1 Lexical roles preferences

$P(w|a)$ follows a Multinomial distribution smoothed with a symmetric Dirichlet with constant concentration hyper-parameter $\alpha = 0.001$. The same smoothing is applied to the two other factors described next. The α parameter has not been tuned at all but has been set beforehand to 0.001 in order to favor peaky distributions. w is the observed lexical form of the head of the candidate argument chunk, which is given by M_d (see Figure 2). This factor shall encode the fact that some words are more likely to play a given role than others. A typical example in French are personal pronouns "je, tu, il" (I, you, he), which are more likely to be A0, while "lui, leur" (him, them) are more likely to be A2. Another example is the preposition "de" (of), which is more likely to be AM than A0.

4.2 Argument position

$P(pos|a)$, where pos is the position of the argument relative to its predicate, follows a similar Multinomial distribution than the previous factor. The position variable can take two values: left or right. This factor shall encode the fact that, in French and English, the relative position of the role is relevant. Hence, the active case as well as the declarative forms of sentences are largely dominant in the corpus, and thus the A0 role is more likely to occur before the verb.

4.3 Syntactic preferences

$P(d|a)$, where d is the dependency label that governs the argument, also follows a smoothed Multinomial distribution. This factor shall encode part of the well-known correlation between syntactic dependencies, such as subject, and semantic roles, such as A0.

4.4 Inference

Inference is realized with the Metropolis-Hastings algorithm. The chosen proposal distribution proposes, for one random sentence u and argument $w_t \in A_u$,

to replace its current semantic arc (a_t, pos_t) with a new one amongst the set of candidate arcs proposed by M_s for w_t, eventually attaching to a new predicate and/or with another label. Note that this proposal is based on the assumption that every argument is linked to exactly one predicate. This strong assumption is reasonable on the target French SRL corpus, in which less than 8% of the arguments are linked to more than one predicate.

For each of these possible moves, the proposal is non-uniform, in order to speed up convergence. We rather set the proposal distribution so that 80% of the time, the move that leads to the largest posterior is chosen, and we distribute the remaining 20% probability mass uniformly over the other possible moves. It is easy to check that the Bayesian model is identifiable with discrete and finite variables and that the detailed balance condition is verified. These conditions guarantee that inference converges towards a stationary posterior distribution. Before inference, the semantic arcs are initialized by choosing for each argument w_t, the semantic arc with the maximum score given by M_s.

5 Experimental validation

The proposed semi-supervised SRL approach is evaluated on French and English. In both cases we compare our approach with the MATE state-of-the-art supervised semantic parser [11] (called **MATEsrl**). For this comparison, both the **MATEsrl** system and the proposed approach are trained on the same 50 sentences. Furthermore, as explained in Section 3, instead of using gold syntactic dependencies, for all three corpora we use the dependencies obtained with the MATE syntactic parser (**MATEdep**), which is also trained only on the 50 sentences of the labeled corpus. The objective of this comparison is to prove that the proposed weakly-supervised approach outperforms a supervised approach when only a few number of annotated examples are available. The choice of this supervised **MATEsrl** system is motivated by its very good performances in general, and on the French corpus in particular. Indeed, we have also evaluated the performances of the **MATEsrl** system when trained on the full French corpus with 10-fold cross-validation, which then gives a labeled F1 of 98.7% and an unlabeled F1 of 98.8%. Note that cross-validation has only been used in this specific experiment, and not in any of the others, because in the other experiments, we only use 50 sentences for training and may thus use a large test set of 500 sentences.

In English, the proposed system has further been compared with the state-of-the-art semi-supervised SRL system presented in [12]. However, for this comparison we used slightly different settings, adapted to those described in [12]. For all the experiments, the scores are computed as in the CoNLL09 evaluation campaign, except for the labeled scores that assume that the gold labels of the predicates are known, because our system does not do sense detection. To simplify notations, we omit next the $'$ and $''$ when referring to the corpora derived from $\mathcal{L}$, $\mathcal{T}$ and $\mathcal{U}$.

5.1 Evaluations on French

Data. The data used in this evaluation is the French CLASSIK corpus [13]. The "gold" section of this corpus has been manually labeled and contains 1000 sentences in total. $\mathcal{L}$ is composed of the 50 first sentences, and $\mathcal{T}$ is composed of the 500 last sentences. We can thus make the size of $\mathcal{L}$ vary from 50 to 500 sentences when drawing Figure 5. $\mathcal{U}$ is composed of 10,000 sentences taken from the non-manually annotated part of the French CLASSIK corpus.

Comparison with MATE. In this experiment, the **MATEsrl** parser is trained on the same 50 sentences from $\mathcal{L}$ than our proposed system. The results obtained by both the **MATEsrl** system and the proposed approach are shown, respectively, in the first and last rows of Table 2. Note that, although the **MATEsrl** performances are very high when trained on the full corpus, they drop down dramatically when trained on only 50 sentences. In this case, the proposed approach largely outperforms the supervised system.

The second and third rows in Table 2 show the performances obtained when using only the first stage of our system, that is the output of the Maximum Entropy classifier, without doing Bayesian inference. The first "all links" model simply includes all of the semantic arcs produced by our supervised classifier. On the average, this classifier produces about 5 candidate semantic arcs per argument. Obviously, the F-measure is quite low here, because this classifier has been designed to produce many more arcs than necessary, so that the subsequent inference step only selects a few of them. But the recall is more interesting than the F1 in this experiment, because it shows the best performance that can be reached with this first stage of classifiers.

The second "Optimum links" shows the best possible results that can be reached with the set of candidate links proposed by the classifier, given our restriction that every argument can be linked to at most one predicate. So in this line, we select for every argument the single arc that matches a corresponding arc in the gold semantic structure, or a random arc if none of the candidates is correct. This line gives the real upper bound of performances that can be reached by our system, given our restrictions and the current setup of the deterministic and supervised parts of the model, i.e., at the exclusion of the unsupervised model. This oracle does not consider the labels of the arcs. Note that its recall is slightly lower than the "All links" recall, because of the few arguments that are linked to several predicates. Its precision is also lower than 100% because of the false alarms from both predicate and argument detection.

Comparing the last two lines of Table 2, we can note that the proposed system only adds about 15% of errors more than the oracle system, in terms of F1, precision and recall. This suggests that the inference stage is doing correctly its job and that further improvements can be obtained mainly by working on the deterministic part and restrictions of our system.

Table 2. SRL experimental validation on French in terms of labeled and unlabeled F1-measure, Precision and Recall

System	F1 lab.	Prec. lab.	Rec. lab.	F1 unlab.	Prec. unlab.	Rec. unlab.
MATE	36.4	40.8	32.9	58.7	65.8	53.0
All links	31.4	20.6	**66.0**	33.3	21.8	**70.0**
Optimum links				76.7	86.2	69.1
Inference	**54.8**	62.7	48.7	**73.5**	84.1	65.3

5.2 Impact of quantity of labeled data

Figure 5 shows the evolution of the unlabeled F1 in function of the number of manually annotated sentences both for the supervised **MATEsrl** system and for the proposed weakly supervised system. As expected, the proposed system gives much better performances than the MATE system with a small amount of manually labeled data, but still remains better for up to 450 sentences, although the difference between both systems decreases when more labeled data is included. This suggests that the proposed model may still be improved by better tuning the complexity and number of features used to train our first stage classifier, depending on the size of the available corpus.

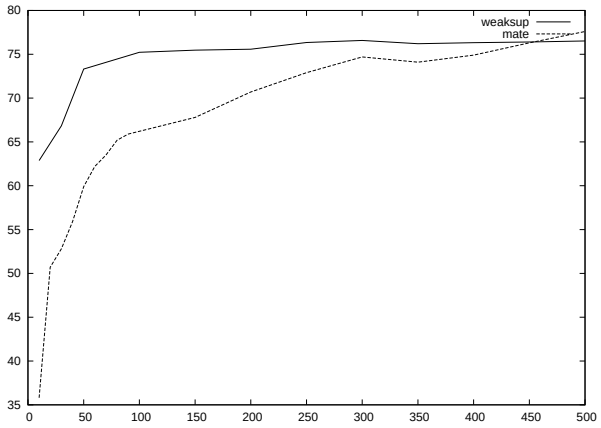


Fig. 5. Unlabeled F1 as a function of the number of manually annotated sentences

5.3 Evaluations on English

Two evaluation experiments are realized in English: first, we compare our approach with the supervised **MATEsrl** system, in a similar way as done in French. Second, we compare our system with the semi-supervised SRL presented in [12].

Data. The data used for the evaluations on English is the standard CoNLL 2008 shared task [14] version of Penn Treebank WSJ and PropBank. As done in [12], $\mathcal{T}$ corresponds to the test portion of the CoNLL 2008 corpus. $\mathcal{L}$ is composed of the first 50 (Table 3) or 400 (Table 4) sentences extracted from the training corpus of CoNLL 2008, while $\mathcal{U}$ is composed of the next 10,000 sentences.

Comparison with MATE. Both systems are trained on only 50 sentences. The results obtained by both the **MATEsrl** semantic parser and the proposed system are shown in Table 3. Like in French, the proposed approach largely outperforms the state-of-the-art supervised semantic parser when only a few number of annotated examples are available.

Table 3. SRL experimental validation on English in terms of labeled and unlabeled F1-measure, Precision and Recall

System	F1 lab.	Prec. lab.	Rec. lab.	F1 unlab.	Prec. unlab.	Rec. unlab.
MATE	26.0	31.7	22.1	53.9	65.5	45.8
Inference	**47.5**	**54.7**	**42.0**	**67.2**	**77.4**	**59.4**

Comparison with semi-supervised. We compare next our approach with the semi-supervised SRL system presented in [12], which also produces labeled semantic arcs. In [12] the authors present a Latent Words Language Model, which learns word similarities from unlabeled text and use them in different semi-supervised SRL systems as additional features or to automatically expand a small training set. They experiment with different sizes of the training corpus and show that for a small training corpus they outperform a state-of-the-art supervised baseline. We mimic their experimental setup by increasing the size of our training corpus from 50 to 400 sentences, which represents about 1% of the training corpus. The first two lines in Table 4 show the results presented in [12] when training on 5% of the training set. The first line is their baseline, a supervised SRL system, while the second line shows the results of the same supervised system, but using extra features given by the latent words language model. These features correspond to the estimated distribution of the latent words for every word for both the training and test set. The last two lines in the table show, respectively, the results achieved by our baseline, the **MATEsrl** parser, and our proposed system when training only on 1% of the training set. We can observe that with only 1% of the training data (i.e., about 400 sentences), the proposed system

Table 4. Comparison with **MATEsrl** and [12] for two training corpus sizes.

System	Training	F1 lab.
Supervised Baseline [12]	5%	40.49
LWFeatures [12]	5%	60.3
MATE SRL supervised	1%	54.4
Inference	**1%**	**60.6**

matches the performances of the semi-supervised approach trained on 5% of the training corpus (i.e. with about 1900 sentences).

6 Additional Related Works

A variety of algorithms have been proposed for semi-supervised learning[3]. And there are many more examples of applications of these semi-supervised approaches to SRL other than the ones described in the introduction and evaluation sections. For instance, [3] and [4] tested self-training and co-training on SRL; [17] used a graph-alignment method to SRL; and, more recently, [18] used a graph-based label propagation semi-supervised approach to improve the coverage of a frame-semantic parsing model and reported significant improvements over a state-of-the-art baseline, both in frame identification accuracy and full frame-semantic parsing F_1. Finally, another interesting approach, also related with our work, is "prototype-based" learning [19, 20]. In this approach, prior knowledge is specified declaratively, by providing "prototypes" (e.g., a list of representative words) for each label. Then they use distributional similarity between the words in the corpus and the prototypes as features in a generative model. Similarly, our proposed framework might support the inclusion of "prototypes" in the form of rules that generate candidate semantic arcs, as a replacement or in addition to the supervised Maximum Entropy model used in this work.

7 Conclusions and future work

We present in this work a new approach to SRL that is able to work competitively even when only a small amount of labeled data is available. The proposed approach exploits both supervised and unsupervised methods, without privileging one or the other by design. It is based on the combination of a supervised semantic role labeler that generates many more potential candidate arcs than traditional supervised systems, with a Bayesian unsupervised model that maximizes the joint posterior of several linguistic factors that are commonly used in the unsupervised SRL field. This combination is realized thanks to "virtual evidence" that acts as new types of constraints for Bayesian inference.

[3] We refer the reader to [15] or [16] for an overview on semi-supervised methods.

Because the proposed approach relies on a supervised SRL classifier, it produces labeled semantic roles and it does infer a semantic structure by detecting to which predicate each argument should be linked. This semi-supervised direction is very promising, specially for those domains and languages for which little or no annotated data is available.

We successfully evaluated the proposed model on two languages, French and English, showing, in both cases, consistent performances improvement over a state-of-the-art supervised SRL system on small amounts of labeled data. Furthermore, we showed for English that its accuracy reaches a level comparable to that of a state-of-the-art semi-supervised SRL systems even when the amount of labeled data is smaller.

The system could be improved in many ways, and in particular the proposed unsupervised model. We could, for instance, include some penalization term for sampling the same role for several arguments of a verb instance (at least for core roles).

Acknowledgments

This work has been partially funded by the French ANR project ContNomina.

References

1. Màrquez, L., Carreras, X., Litkowski, K.C., Stevenson, S.: Semantic role labeling: an introduction to the special issue. Comput. Linguist. **34** (2008) 145–159
2. Pradhan, S.S., Ward, W., Martin, J.H.: Towards robust semantic role labeling. Comput. Linguist. **34** (2008) 289–310
3. He, S., Gildea, H.: Self-training and Cotraining for Semantic Role Labeling: Primary Report. Technical report, TR 891, University of Colorado at Boulder (2006)
4. Lee, J.Y., Song, Y.I., Rim, H.C.: Investigation of weakly supervised learning for semantic role labeling. In: ALPIT. (2007) 165–170
5. Daumé III, H.: Semi-supervised or semi-unsupervised? In: Proc. NAACL Wokshop on Semi-supervised Learning for NLP. (2009)
6. Titov, I., Klementiev, A.: Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In: Proceedings of the International Conference on Computational Linguistics (COLING), Bombay, India (2012)
7. Jain, D., Beetz, M.: Soft evidential update via markov chain monte carlo inference. In: Proc. 33rd Annual German Conf. on Artificial Intelligence (KI 2010). (2010)
8. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: An annotated corpus of semantic roles. Comput. Linguist. **31** (2005) 71–106
9. Dowty, D.: Thematic proto-roles and argument selection. Language **67** (1991) 547–619
10. Bohnet, B.: Top accuracy and fast dependency parsing is not a contradiction. In: Proc. International Conference on Computational Linguistics, Beijing, China (2010)
11. Björkelund, A., Hafdell, L., Nugues, P.: Multilingual semantic role labeling. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task. CoNLL '09, Stroudsburg, PA, USA, Association for Computational Linguistics (2009) 43–48

12. Deschacht, K., Moens, M.F.: Semi-supervised semantic role labeling using the latent words language model. In: EMNLP. (2009) 21–29
13. van der Plas, L., Merlo, P., Henderson, J.: Scaling up automatic cross-lingual semantic role annotation. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2. HLT '11, Stroudsburg, PA, USA, Association for Computational Linguistics (2011) 299–304
14. Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., Nivre, J.: The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In: Proceedings of the Twelfth Conference on Computational Natural Language Learning. CoNLL '08, Stroudsburg, PA, USA, Association for Computational Linguistics (2008) 159–177
15. Zhu, X.: Semi-Supervised Learning Literature Survey. Technical report, Computer Sciences, University of Wisconsin-Madison (2005)
16. Pise, N.N., Kulkarni, P.: A survey of semi-supervised learning methods. In: Proceedings of the 2008 International Conference on Computational Intelligence and Security - Volume 02. CIS '08, Washington, DC, USA, IEEE Computer Society (2008) 30–34
17. Fürstenau, H., Lapata, M.: Graph alignment for semi-supervised semantic role labeling. In: EMNLP. (2009) 11–20
18. Das, D., Smith, N.A.: Semi-supervised frame-semantic parsing for unknown predicates. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1. HLT '11, Stroudsburg, PA, USA, Association for Computational Linguistics (2011) 1435–1444
19. Haghighi, A., Klein, D.: Prototype-driven learning for sequence models. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. HLT-NAACL '06, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 320–327
20. Haghighi, A., Klein, D.: Prototype-driven grammar induction. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. ACL-44, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 881–888