Speeding Up Syntactic Learning Using Contextual Information*

Leonor Becerra-Bonache leonor.becerra@univ-st-etienne.fr
Elisa Fromont elisa.fromont@univ-st-etienne.fr
Amaury Habrard amaury.habrard@univ-st-etienne.fr
Michaël Perrot michael.perrot@etu.univ-st-etienne.fr
Marc Sebban marc.sebban@univ-st-etienne.fr
Laboratoire Hubert Curien, Université Jean Monnet, Saint-Etienne, France

Editors: Jeffrey Heinz, Colin de la Higuera and Tim Oates

Abstract

It has been shown in (Angluin and Becerra-Bonache, 2010, 2011) that interactions between a learner and a teacher can help language learning. In this paper, we make use of additional contextual information in a pairwise-based generative approach aiming at learning (situation,sentence)-pair-hidden markov models. We show that this allows a significant speed-up of the convergence of the syntactic learning. We apply our model on a toy natural language task in Spanish dealing with geometric objects.
Keywords: Language learning, Pair-HMM, Context.

1. Introduction

The correspondence between sentences and the situations (or contexts) in which they are made seems to play an important role in the early stages of children's language acquisition. Thanks to this correspondence, the child tries to figure out how to use the language and the adult tries to understand the imperfect sentences produced by the child.

Angluin and Becerra-Bonache (2010, 2011) have proposed a model that takes into account aspects of this correspondence. The learner and the teacher interact in a sequence of situations by producing sentences that intend to denote one object in each situation. The learner uses cross-situational correspondences to learn to comprehend and produce denoting sentences in a given situation (there is no explicit semantic annotation of the sentences). The goal of the learner is to be able to produce every denoting sentence in any given situation. One of their main results is that the access to the semantics of shared situations facilitates language learning. However, the proposed model can only learn in the presence of a situation and can only generate a finite language.

Our work, mainly inspired by this model, is also related to (Wong and Mooney, 2007), except that they aim to learn logical forms from natural language sentences. However, our problem is the inverse, we model the situations by using a logical formulation and we aim at learning the natural language rather than the logical representations.

* This work was supported by Pascal 2 Network of Excellence.

Our contribution in this paper is twofold: (i) we learn a joint Pair-Hidden Markov model from a training set composed of (situation,sentence)-pairs; (ii) we study the impact of this generative joint model to increase the convergence of the language learning compared with a marginal model only trained from syntactic information.

2. A (situation,sentence)-pair Hidden Markov Model

2.1. Situations and sentences

The input given to the learner are pairs consisting of a *situation* and a *sentence* denoting something in that situation. We exploit these notions of *situation* and *sentence* in the context of geometric objects (inspired from Angluin and Becerra-Bonache (2010, 2011)).

A **situation** is a sequential description of two objects in the space. Each one has 3 possible shapes (*square, triangle, circle*) and 3 possible colors (*red, blue, green*). There could be 3 different relations between the 2 objects (*to the left of, to the right of, above*). For example, a situation can take the form of the sequence *The red square to the left of the blue triangle* and can be represented by $\{re1(x_1), sq1(x_1), le2(x_1, x_2), bl1(x_2), tr1(x_2)\}$.

A **sentence** is a finite sequence of words. We say that a sentence is *denoting* in a situation if it is a (possibly incomplete) description of the current situation. On this toy example, we arbitrarily say that a sentence is denoting if it describes at least the **first object** alone or **the first object and its relation with the second one**. For example, the sentences *The blue triangle* or *The red square to the left of* are not a description of the previous situation and, thus, are not denoting sentences. However, the sentence *The square to the left of the blue triangle* or only *The square* are both denoting in that situation.

2.2. Teacher and learner as Pair-HMMs

Both our teacher and learner correspond to transducers that are represented in the form of Pair-HMMs. A Pair-HMM is a finite state generative model encoding distributions over pairs of input/output sequences (Durbin et al., 1998). More formally, a Pair-HMM is a 6-tuple $< Q, \Sigma, \Sigma', \Pi, A, B >$ with: Q the set of states, Σ the input alphabet, Σ' the output alphabet, Π the initial state matrix, A the transition matrix and B the edit operation matrices. An illustration of the Pair-HMM used in this study which models the linguistic competence of the teacher on a limited sub-language of Spanish is provided in Appendix A.

3. Evaluation of the model

The learning task we consider uses 243 different situations leading to $1,458$ possible (situation, sentence)-pairs. Indeed, for every situation, one can assign 6 possible denoting sentences: *shape, shape color, shape position shape, shape color position shape, shape position shape color, shape color position shape color.* The target language is composed by all the possible sentences that can be produced in any situation.

We consider the following experimental setup. Each experiment is repeated 5 times and we provide the averaged results. The learning set is constituted by all the possible (situations-sentences)-pairs, and we define a full order over the pairs according to a random draw. Then, we compare three learning methods: (i) A *joint learner* learning a pair-HMM from (situation,sentence)-pairs. (ii) A *marginal learner* learning a classical HMM from the sentences of the learning set only (this learner corresponds to learning the marginal distribution over the sentences). (iii) A learner using the algorithm Alergia (Carrasco and Oncina, 1994), to learn a probabilistic deterministic finite automata, from only the sentences of the learning set only. The parameter α of Alergia and the number of states of the Pair-HMM and HMM are assessed by cross-validation over a validation set constituted of 500 examples randomly drawn from the full learning sample, using the maximization of the log-likelihood criterion. The models are then learned from the full learning sample.

The quality of the learned models is evaluated according to two performance criteria: *correctness* and *completeness*. These criteria, similar to the well-known precision and recall criteria, allow us to assess if a learner has correctly learned the target language. The correctness is the ratio of correct sentences over all the sentences the learner can produce. The completeness corresponds to the rate of correct sentences, produced by the teacher, that the learner can read. The mean of correctness and completeness gives us a value between 0 and 1, which is called the *performance*. The goal of the learner is to achieve a level of performance as close to 1 as possible (0.99 is used as a threshold). We evaluate the correctness on $10,000$ sentences produced by the learner, and the completeness on the $1,458$ possible sentences produced by the teacher.

3.1. Performance results

In this section, we compare the evolution of the performance of the three approaches according to the number of training examples. The results are reported on Figure 1. We can see that the joint and marginal learners have the same rate of convergence. The expected level of performance (0.99) is reached quickly after about 180 examples for the two learners. This shows that the contextual information does not imply an acceleration of the convergence in terms of number of examples. However, these HMM-based models converge significantly faster than Alergia, that needs around 360 examples to converge. This behavior can be explained by the fact that Alergia is known to have poor results with small datasets. HMM-based models seem thus more appropriate in this context.

3.2. Evaluation of the impact of contextual information

If we have seen that contextual information does not improve the convergence according to the training set size, in comparison with a marginal learner, it actually accelerates the

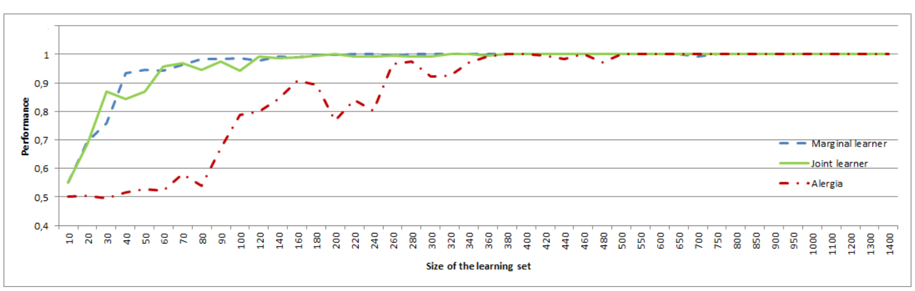

Figure 1: Performance according to the learning set size (mean over 5 trials).

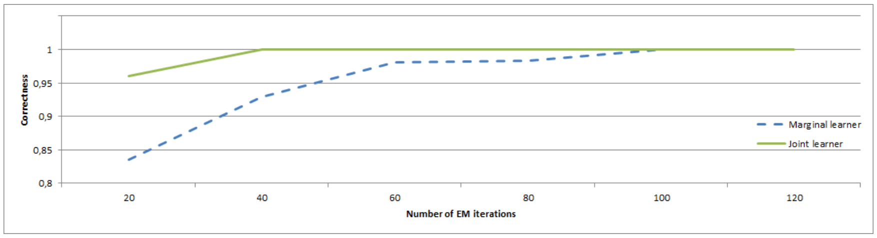

Figure 2: Evolution of the correctness of the joint and marginal learners according to the iterations of the EM algorithm.

convergence of the EM procedure. To illustrate this behavior, we consider an illustrative learning set of 180 elements from which the joint and marginal learners reach a level of completeness of 1. We then study the evolution of the correctness according to the iterations of the EM algorithm. The results are given on Figure 2. A level of correctness of 1 is achieved quickly after 40 iterations for the joint learner, while at least 100 iterations are needed for the marginal one. This experiment shows that the use of contextual information allows to significantly speed up the learning of the target language.

4. Discussion and future work

We have presented a new model able to take into account contextual information for language learning. The preliminary results obtained show that a joint learner is effectively able to accelerate language learning in comparison with purely syntactic learners. Indeed, learning with our model requires less examples than classical state merging based methods such as Alergia and less iterations than EM-based marginal learners. As a future work, we aim at doing larger experiments and considering different tasks with more complex languages. We also plan to work on theoretical frameworks allowing us to characterize when contextual information helps to learn well and on relationships with other approaches such as machine translation methods.

References

Dana Angluin and Leonor Becerra-Bonache. Effects of meaning-preserving corrections on language learning. In *CoNLL*, pages 97–105, 2011.

Dana Angluin and Leonor Becerra-Bonache. A model of semantics and corrections in language learning. Technical report, Yale University, April 2010.

R. C. Carrasco and J. Oncina. Learning stochastic regular grammars by means of a state merging method. In *ICGI*, pages 139–152, 1994.

R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

Y. W. Wong and R.J. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. In *ACL*, pages 960–967, 2007.

Appendix A. Teacher as a Pair-HMM

The linguistic competence of the teacher is modeled by a Pair-HMM. In our experiments, we consider a limited sub-language of Spanish. Figure 3 shows the pair-HMM used to model the teacher in this study.

Given a situation consisting of *a blue triangle to the left of a red square* and represented by $\{tr1(x_1), bl1(x_1), le2(x_1, x_2), sq1(x_2), re1(x_2)\}$, the teacher could produce the following sentences: *el triangulo, el triangulo azul, el triangulo a la izquierda del cuadrado, el triangulo azul a la izquierda del cuadrado, el triangulo a la izquierda del cuadrado rojo, el triangulo azul a la izquierda del cuadrado rojo.*

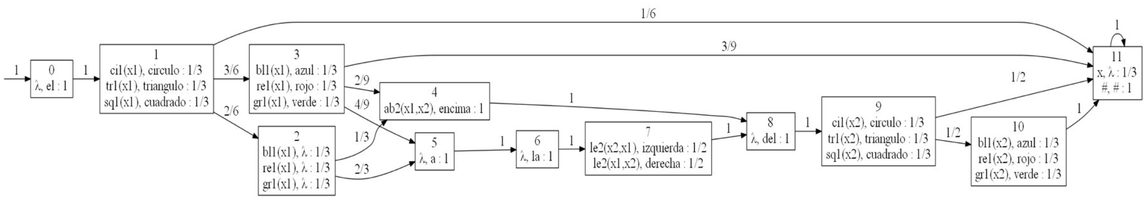

Figure 3: Teacher modeled by a Pair-HMM, where λ is the empty symbol. The last state, 11^{th} one, is the final state which allows the teacher to end his tasks of comprehension and production.