

Cross-Lingual Bootstrapping for Semantic Role Labeling

Mikhail Kozhevnikov, [Ivan Titov](#)

Semantic role labeling

- ▶ The task is given a sentence to identify underlying predicate-argument structure:

Jack opened the lock with a paper clip

Closely related to relation extraction

Semantic roles (from the PropBank guidelines):

Arg0: opener – an initiator/doer in the event [**Who?**]

Arg1: the thing opening [**What?**]

Arg2: instrument [**Using what?**]

- ▶ Challenges for standard supervised learning:

- ▶ Very tied to lexical information (e.g, roles are, to some degree, predicate-specific)
- ▶ Requires a lot of expensive annotation to get an accurate model

We do not have much or any for many languages

The resource-poor setting

Little labeled data is not hard to obtain

▶ Our setting

- ▶ We have some annotation for a pair languages: either little annotation for both or little annotation for one of them
- ▶ We have parallel data (sentences and their translations) for this pair of languages

▶ Idea:

- ▶ Learn models for both and enforce agreement on the parallel data

$$\max L_1(\theta_1) + L_2(\theta_2) + R(\theta_1, \theta_2, \Sigma)$$

Objective on
language 1

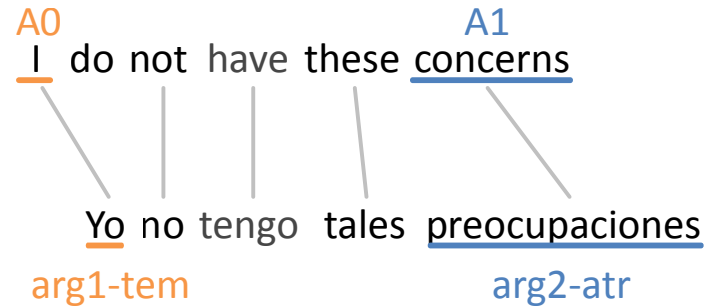
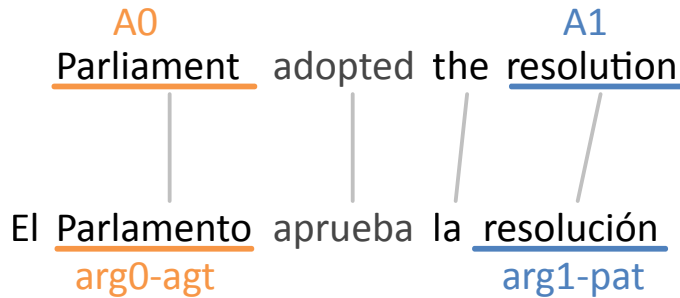
Objective on
language 2

Regularizing:
agreement on
parallel data

Challenge: enforcing agreement for SRL is not easy

Why enforcing agreement is not easy?

- ▶ Different formalisms are used for different languages
- ▶ The underlying mapping (if exists) is different for each predicate pair

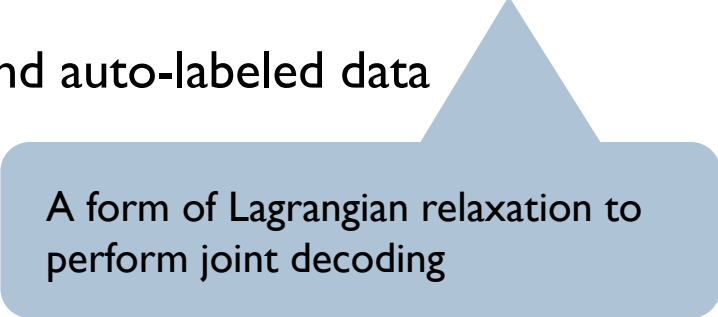


- ▶ Shifts in translation:
 - ▶ "John sold a car to Mary" can be translated as "Mary bought a car from John"
- ▶ Our goal is to estimate an agreement model $R(\theta_1, \theta_2, \Sigma)$

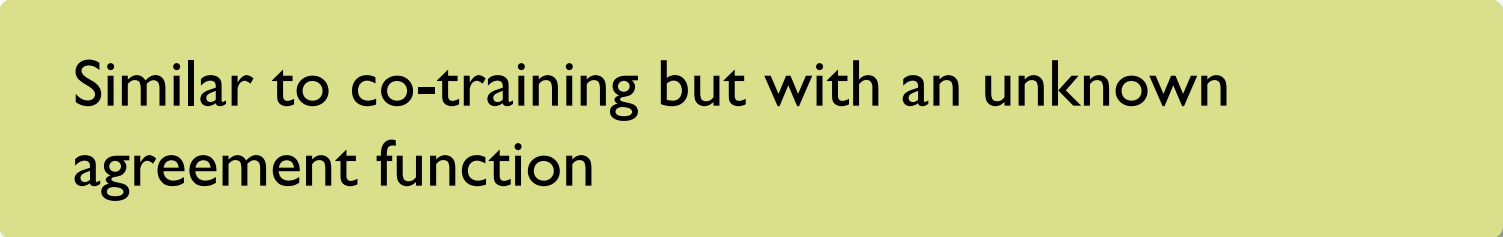
Σ - the agreement model parameters

Crosslingual co-training algorithm

1. Train SRL models on initial labeled datasets and label the parallel data
2. Train an agreement model on the parallel data
3. Run the joint inference step on parallel data to refine the predictions
4. Retrain the SRL models using both initial and auto-labeled data



A form of Lagrangian relaxation to perform joint decoding



Similar to co-training but with an unknown agreement function

Results (500 labeled sentences)

	Superv. baseline	Self- training	Asymmetr set-up	Symmetric set-up	'Oracle' Agreement
En-Cn	76.5	75.1	76.5 (+0.0)	76.4 (-0.0)	76.5 (+0.0)
En-Cz	55.8	55.7	56.1 (+0.4)	56.3 (+0.5)	56.7 (+1.0)
En-Cz, ood	57.0	57.0	57.5 (+0.5)	56.6 (-0.4)	57.5 (+0.5)
En-De	61.0	58.6	60.5 (-0.6)	60.6 (-0.5)	61.1 (+0.1)
En-De, ood	64.3	59.7	68.0 (+3.7)	67.8 (+3.5)	69.1 (+4.8)
En-Es	62.3	61.8	63.9 (+1.5)	64.4 (+2.1)	65.2 (+2.9)

Out-of-domain
evaluation

Only moderate improvements in some contexts:
better agreement models are needed