Phrase-Based & Neural Unsupervised Machine Translation

Presenter: Baosong Yang

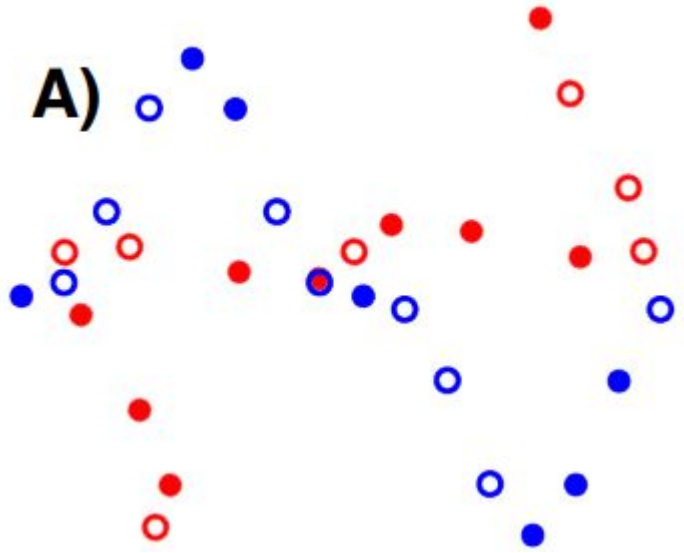
Motivation

- For low-resource language pairs.
- Using only monolingual corpora.

- Previous studies:
  - 1. Initialization
  - 2. Language model (Denoising)
  - 3. Back-translation

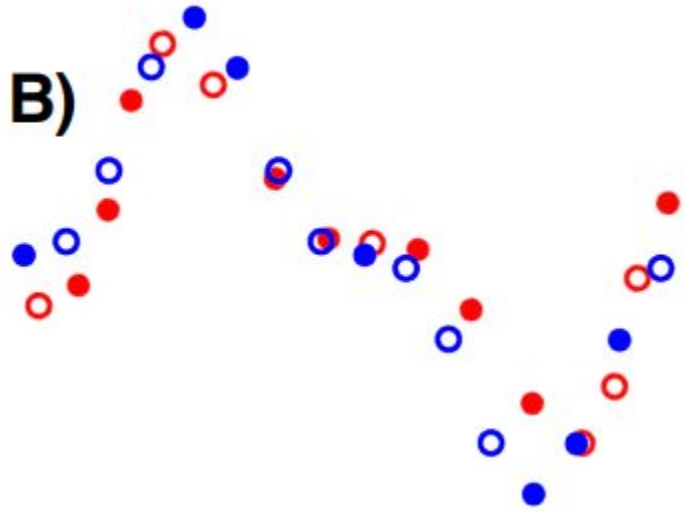
Unsupervised MT

- There are two monolingual datasets.



observed source sentence
unobserved translation of a target sentence
system translation of a target sentence

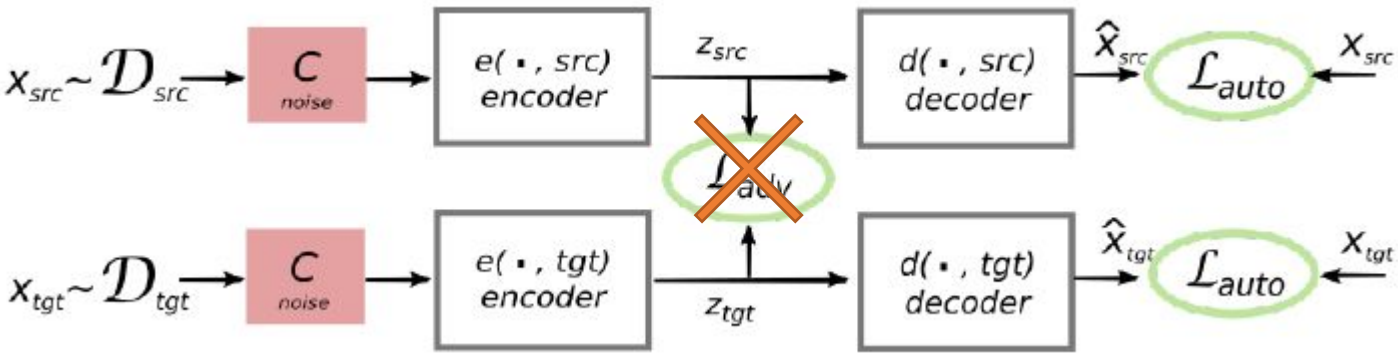
- Initialization
  - The two distributions are roughly aligned, e.g. by performing word-by-word translation with an inferred bilingual dictionary.

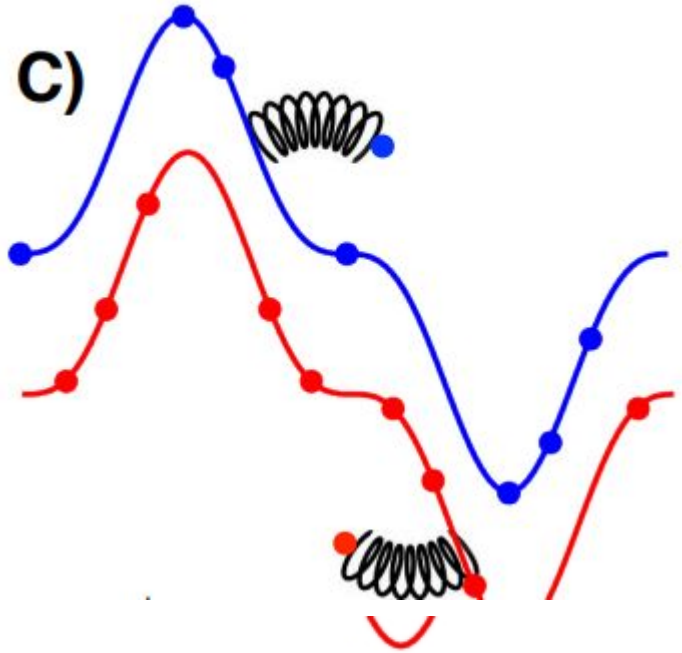


observed target sentence
unobserved translation of a source sentence
system translation of a source sentence

Unsupervised MT

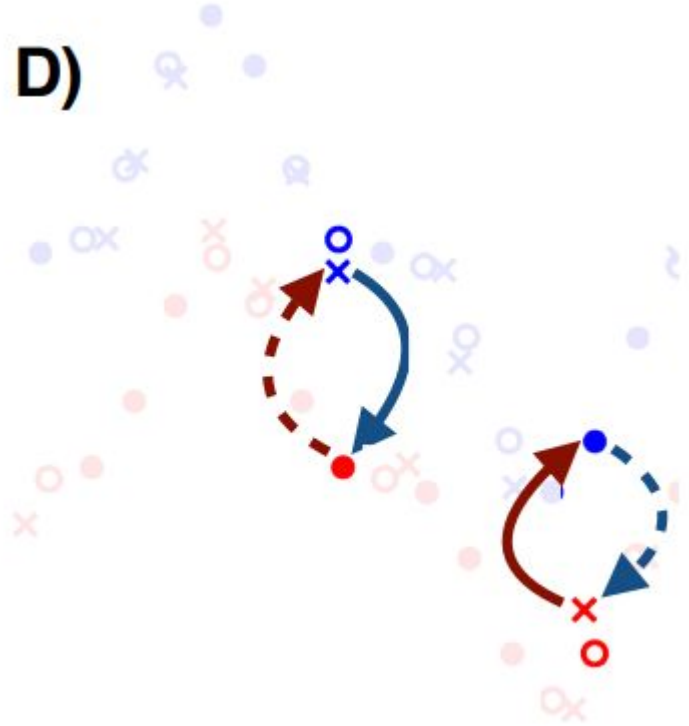
- Language modeling
  - A language model is learned independently in each domain to infer the structure in the data
  - Denoise/Correct sentences

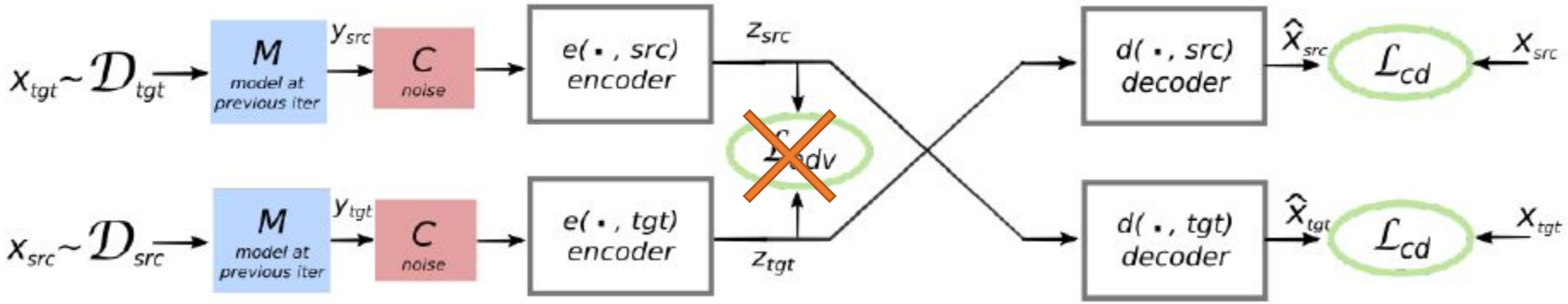




Unsupervised MT

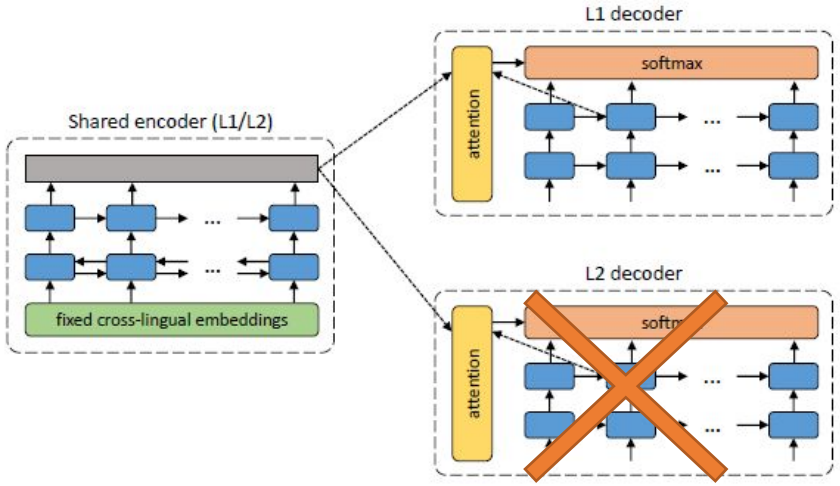
- Back-translation
  - Using source->target model to translate.
  - Using target->source model to reconstruct the sentence.
  - The discrepancy between the reconstruction and the initial sentence provides error signal.





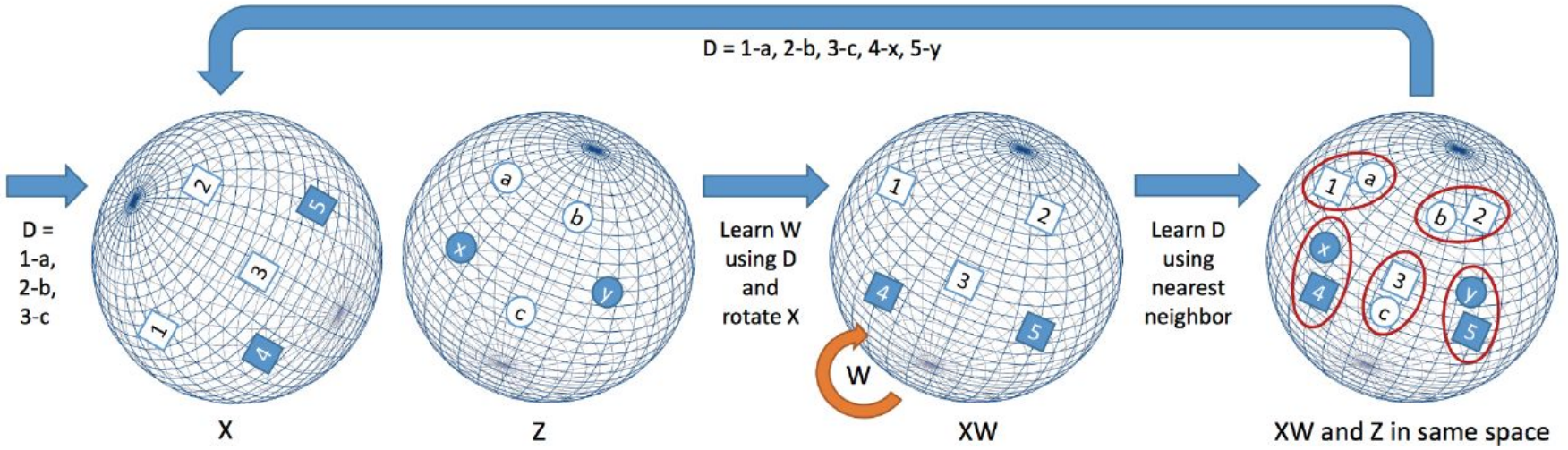
Unsupervised MT

- Other:
  - Shared encoder and decoder to learn a shared latent space
  - Initialization: Cross-lingual BPE embeddings using fastText



Unsupervised PB-SMT

- Initial phrase table: Align using MUSE, the 200 nearest neighbors.
  - Word Translation without Parallel Data



- Language model: KenLM
- Back-translation:
  - Back-translation to generate corpus.
  - Then training, then repeat...

Experiments

Model	en-fr	fr-en	de-en	en-de
(Artetxe et al., 2018)	15.1	15.6	-	-
(Lample et al., 2018)	15.0	14.3	13.3	9.6
NMT (LSTM)	24.5	23.7	19.6	14.7
NMT (Transformer)	25.1	24.2	21.0	17.2
PBSMT (Iter. 0)	16.1	15.4	14.5	10.3
PBSMT (Iter. n)	**27.1**	24.7	21.3	16.7
NMT + PBSMT	26.3	25.1	20.2	16.4
PBSMT + NMT	26.7	**27.1**	**23.6**	**19.2**

Table 2: **Comparison with previous approaches.** BLEU score for different models on the $en - fr$ and $en - de$ language pairs. Just using the unsupervised phrase table, and without back-translation (PBSMT (Iter. 0)), the PBSMT outperforms previous approaches. Combining PBSMT with NMT gives the best results.

Last 2 rows: Tuning the one model on the data generated by the other

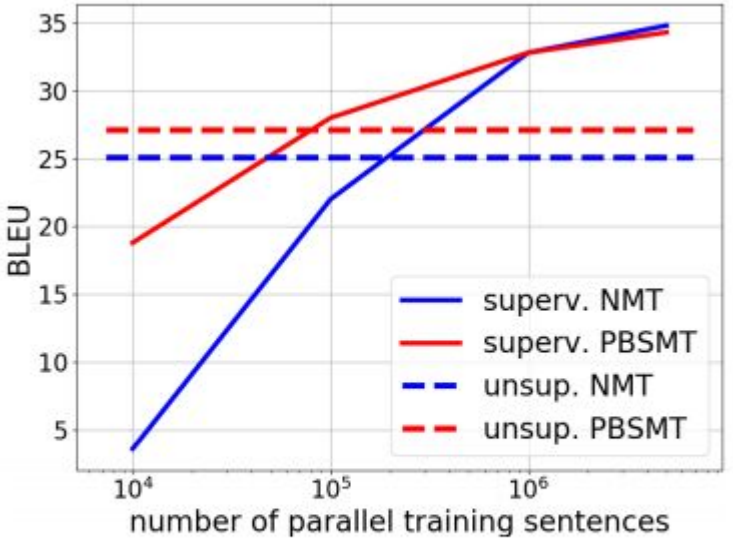


Figure 2: Comparison between supervised and unsupervised approaches on WMT'14 En-Fr, as we vary the number of parallel sentences for the supervised methods.

Compare to supervised

	en → fr	fr→ en	en→ de	de→ en	en→ ro	ro→ en	en→ ru	ru→ en
Unsupervised phrase table	-	15.42	-	14.50	12.99	-	-	7.68
Back-translation - Iter. 1	24.09	23.65	15.06	20.86	20.17	17.12	10.58	14.35
Back-translation - Iter. 2	26.05	24.44	16.66	21.28	20.80	19.25	12.22	15.20
Back-translation - Iter. 3	26.52	24.48	**16.74**	**21.30**	**21.04**	19.84	**12.46**	**15.35**
Back-translation - Iter. 4	26.85	**24.67**	16.71	-	-	**20.07**	12.40	-
Back-translation - Iter. 5	**27.09**	-	-	-	-	-	-	-

Experiments

Source	**Je rêve constamment d'eux, peut-être pas toutes les nuits mais plusieurs fois par semaine c'est certain.**
NMT Epoch 1	I constantly dream, but not all nights but by several times it is certain.
NMT Epoch 3	I continually dream them, perhaps not all but several times per week is certain.
NMT Epoch 45	I constantly dream of them, perhaps not all nights but several times a week it 's certain.
PBSMT Iter. 0	I dream of, but they constantly have all those nights but several times a week is too much. "
PBSMT Iter. 2	I had dreams constantly of them, probably not all nights but several times a week it is large.
PBSMT Iter. 8	I dream constantly of them, probably not all nights but several times a week it is certain.
Reference	**I constantly dream of them, perhaps not every night, but several times a week for sure.**

Source	**La protéine que nous utilisons dans la glace réagit avec la langue à pH neutre.**
NMT Epoch 1	The protein that we use in the ice with the language to pH.
NMT Epoch 8	The protein we use into the ice responds with language to pH neutral.
NMT Epoch 45	The protein we use in ice responds with the language from pH to neutral.
PBSMT Iter. 0	The protein that used in the ice responds with the language and pH neutral.
PBSMT Iter. 2	The protein that we use in the ice responds with the language to pH neutral.
PBSMT Iter. 8	The protein that we use in the ice reacts with the language to a neutral pH.
Reference	**The protein we are using in the ice cream reacts with your tongue at neutral pH.**

Source	**Selon Google, les déguisements les plus recherchés sont les zombies, Batman, les pirates et les sorcières.**
NMT Epoch 1	According to Google, there are more than zombies, Batman, and the pirates.
NMT Epoch 8	Google's most wanted outfits are the zombies, Batman, the pirates and the evil.
NMT Epoch 45	Google said the most wanted outfits are the zombies, Batman, the pirates and the witch.
PBSMT Iter. 0	According to Google, fancy dress and most wanted fugitives are the bad guys, Wolverine, the pirates and their minions.
PBSMT Iter. 2	According to Google, the outfits are the most wanted fugitives are zombies, Batman, pirates and witches.
PBSMT Iter. 8	According to Google, the outfits, the most wanted list are zombies, Batman, pirates and witches.
Reference	**According to Google, the highest searched costumes are zombies, Batman, pirates and witches.**

Conclusion

- The basic steps for unsupervised models
- Achieved the performance of supervised models
- Maybe a new opportunity.