Learning Multilingual Meta-Embeddings for Code-Switching Named Entity Recognition

Genta Indra Winata, Zhaojiang Lin, Pascale Fung

RepL4NLP, ACL 2019

CAiRE
Centre for Artificial Intelligence Research

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

“**walking dead** le quita el apetito a cualquiera”

English TV series title Spanish phrase

(Translation: **walking dead** takes away the appetite of anyone)

1. How do we represent a mixed-language sentence with pre-trained word embeddings?

2. For each token, do we need a language label for choosing pre-trained word embeddings to solve lexical ambiguity?

Multilingual Meta-Embeddings (MME)

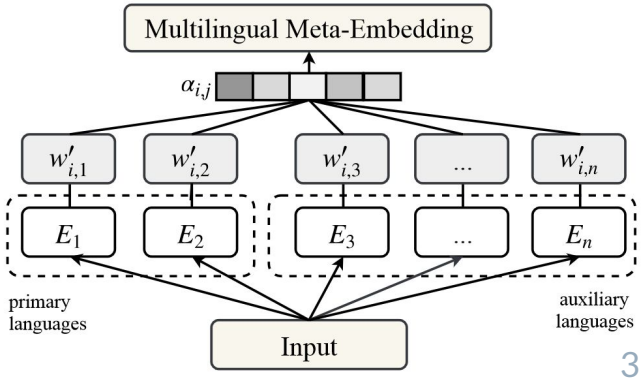
an effective method to learn:

1. language-agnostic meta-representations

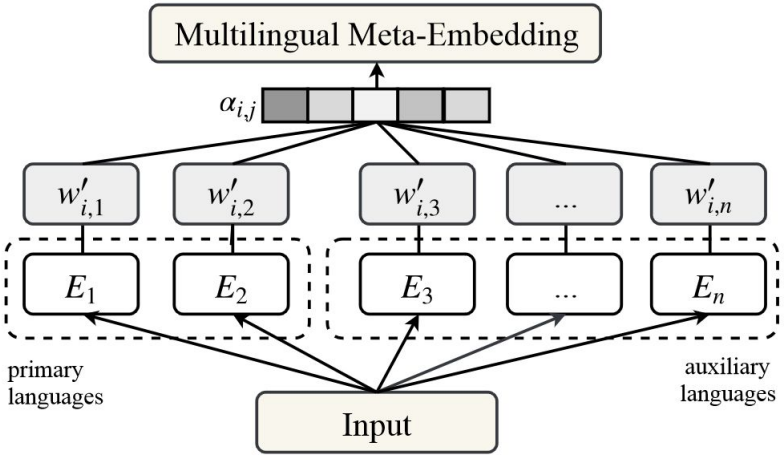
   by leveraging monolingual pre-trained embeddings **without any language labels**

2. how to combine different embeddings

   utilizes information from **semantically similar embeddings** via self-attention mechanism



How we build **MME**?



$$\mathbf{w}'_{i,j} = \mathbf{a}_j \cdot \mathbf{w}_{i,j} + b_j$$

$$\mathbf{w}_i^{MME} = \sum_{j=1}^{n} \alpha_{i,j} \mathbf{w}'_{i,j},$$

$$\alpha_{i,j} = \frac{e^{\phi(\mathbf{w}'_{i,j})}}{\sum_{j=1}^{n} e^{\phi(\mathbf{w}'_{i,j}))}}.$$

1. Generate **`n`** word representations from pre-trained word embeddings

2. Apply non-linear projection layer

   Map word embeddings from the **language-specific space to a multilingual shared space**.

3. Apply the self-attention mechanism

   Calculate the importance of each language embeddings. We weighted sum all embeddings.

“Our model improved after we added more languages”

We evaluate our model in two settings on English-Spanish Dataset

Multilingual

	Languages	F1 %
MONO	EN ES	62.75 62.91
CONCAT LINEAR	EN + ES + PT + CA EN + ES + PT + CA	64.99 65.33
Ours MME	 EN + ES EN + ES + PT + CA	 65.30 **66.63**

Cross-lingual

	Languages	F1 %
MONO	PT CA	54.86 53.96
CONCAT LINEAR	PT + CA PT + CA	58.28 60.72
Ours MME	 PT + CA	 **61.75**

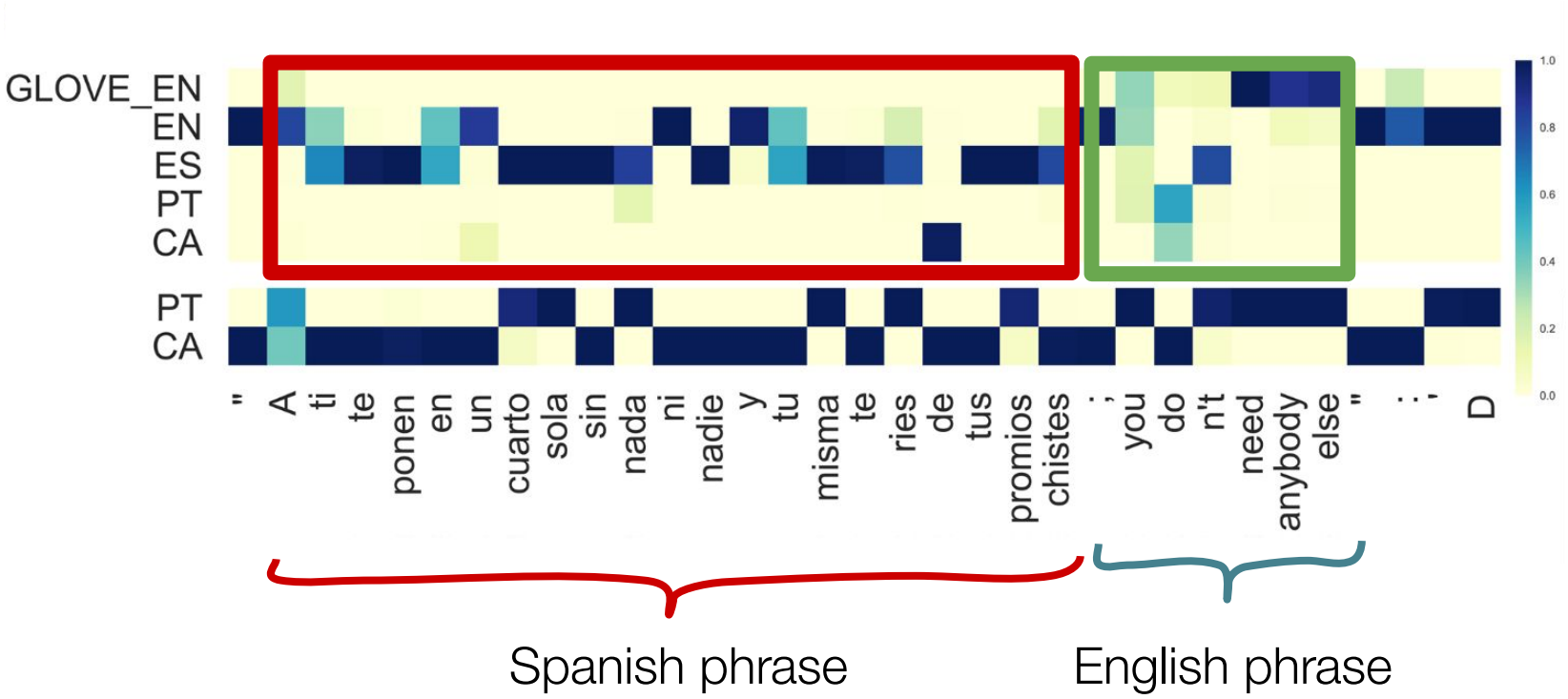
We run our experiments five times and compute the avg.

“MME achieves SOTA performance. Cross-lingual MME has the generalization ability to unseen languages.”

	Single Model	Ensemble
Trivedi, et al. 2018	61.89	63.76
Wang, et al. 2018	62.39	62.67
Winata, et al. 2018	62.76	N/A
Cross-lingual MME (ours)	**61.75**	**63.66**
Multilingual MME (ours)	**66.63**	**68.36**

Ensemble: We combine five models and use majority voting.

“Our model is able to recognize the language!”



Conclusion

Multilingual Meta-Embeddings:

1. is a language-agnostic approach
2. leverages multiple pre-trained word embeddings,
3. utilizes information from semantically similar embeddings,
4. achieves SOTA performance and has the generalization ability to unseen languages.

Thank you, grazie, terima kasih

Check our code :)

https://github.com/gentaiscool/meta-emb

Learning Multilingual Meta-Embeddings for
Code-Switching Named Entity Recognition