Interactive Refinement of Cross-Lingual Word Embeddings

Michelle Yuan[1*] Mozhi Zhang[1*] Benjamin Van Durme[2]
Leah Findlater[3] Jordan Boyd-Graber[1]

[1]University of Maryland

[2]John Hopkins University

[3]University of Washington

[*]Equal contribution

EMNLP 2020

NLP for Low-resource Languages

- Scarcity of both labeled and unlabeled data holds back applications in low-resource languages

NLP for Low-resource Languages

- Scarcity of both labeled and unlabeled data holds back applications in low-resource languages
- Cross-lingual word embeddings (CLWE) can bridge the gap by mapping words from different languages to a shared vector space

NLP for Low-resource Languages

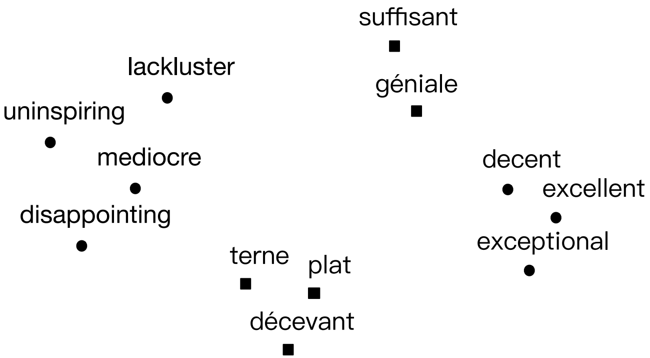
- Scarcity of both labeled and unlabeled data holds back applications in low-resource languages
- Cross-lingual word embeddings (CLWE) can bridge the gap by mapping words from different languages to a shared vector space
- CLWE not tailored for downstream tasks and need refinement (Zhang et al., 2020)

NLP for Low-resource Languages

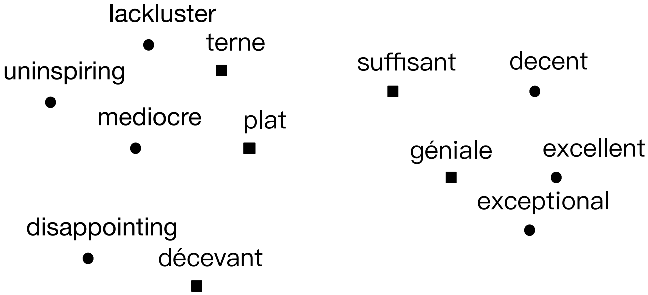
- Scarcity of both labeled and unlabeled data holds back applications in low-resource languages
- Cross-lingual word embeddings (CLWE) can bridge the gap by mapping words from different languages to a shared vector space
- CLWE not tailored for downstream tasks and need refinement (Zhang et al., 2020)

How can we quickly refine CLWE for low-resource NLP?

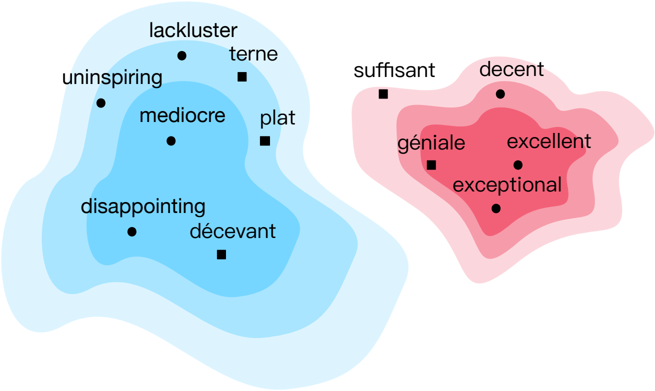
Refining CLWE



Refining CLWE



Refining CLWE



Classification clime: Areas in embedding space where words induce similar labels for a task

CLIME

CLassifying **I**nteractively with **M**ultilingual **E**mbeddings

CLIME

CLassifying **I**nteractively with **M**ultilingual **E**mbeddings

Input: pre-trained CLWE

1. Select keywords with *gradient-based salience* (Li et al., 2016)

CLIME

CLassifying **I**nteractively with **M**ultilingual **E**mbeddings

Input: pre-trained CLWE

1. Select keywords with *gradient-based salience* (Li et al., 2016)
2. Collect user feedback

CLIME

CLassifying **I**nteractively with **M**ultilingual **E**mbeddings

Input: pre-trained CLWE

1. Select keywords with *gradient-based salience* (Li et al., 2016)
2. Collect user feedback
3. Refine embeddings on user feedback through *retrofitting* (Mrkšić et al., 2017)

Keyword Selection

1. **Local salience** of word x_i in x:

$$S_{\mathbf{x}}(x_i) = \left\| \nabla_{\mathbf{E}_{x_i}} L(\mathbf{x}, y) \right\|_2 .$$

Keyword Selection

1. **Local salience** of word x_i in x:

$$S_{\mathbf{x}}(x_i) = \left\| \nabla_{\mathbf{E}_{x_i}} L(\mathbf{x}, y) \right\|_2 .$$

2. **Global salience** of word w:

$$S(w) = \mathrm{IDF}(w, \mathbf{X}) \cdot \sum_{\mathbf{x} \in \mathbf{X} : w \in \mathbf{x}} S_{\mathbf{x}}(w).$$

Keyword Selection

1. **Local salience** of word x_i in x:

$$S_{\mathbf{x}}(x_i) = \left\| \nabla_{\mathbf{E}_{x_i}} L(\mathbf{x}, y) \right\|_2 .$$

2. **Global salience** of word w:

$$S(w) = \text{IDF}(w, \mathbf{X}) \cdot \sum_{\mathbf{x} \in \mathbf{X} : w \in \mathbf{x}} S_{\mathbf{x}}(w).$$

3. Select keywords that have highest global salience

Keyword Selection

Example sentences

`A disappointing dinner...meatballs were undercooked.` (negative)

`I was frustrated with the customer service.  Very disappointing.` (negative)

`An exceptional film.` (positive)

`I disliked waiting in line, but the ride was exceptional.` (positive)

Keyword Selection

Compute local salience

`A disappointing dinner...meatballs were undercooked.` (negative)

`I was frustrated with the customer service. Very disappointing.` (negative)

`An exceptional film.` (positive)

`I disliked waiting in line, but the ride was exceptional.` (positive)

Keyword Selection

Find words with highest global salience

`disappointing`

`exceptional`

Video Demo for User Interface

Retrofitting CLWE

1. **Feedback cost:** Pull positive neighbors p closer and negative neighbors n away

$$C_f(\mathbf{E}) = \sum_{k \in \mathcal{K}} \left(\sum_{n \in \mathcal{N}_k} \mathbf{E}_k^\top \mathbf{E}_n - \sum_{p \in \mathcal{P}_k} \mathbf{E}_k^\top \mathbf{E}_p \right).$$

Retrofitting CLWE

1. **Feedback cost:** Pull positive neighbors p closer and negative neighbors n away

$$C_f(\mathbf{E}) = \sum_{k \in \mathcal{K}} \left(\sum_{n \in \mathcal{N}_k} \mathbf{E}_k^\top \mathbf{E}_n - \sum_{p \in \mathcal{P}_k} \mathbf{E}_k^\top \mathbf{E}_p \right).$$

2. **Regularization:** Updated embeddings should not be too far away from original embeddings

$$R(\mathbf{E}) = \sum_{w \in \mathcal{V}} \left\| \hat{\mathbf{E}}_w - \mathbf{E}_w \right\|_2^2.$$

Retrofitting CLWE

1. **Feedback cost:** Pull positive neighbors p closer and negative neighbors n away

$$C_f(\mathbf{E}) = \sum_{k \in \mathcal{K}} \left(\sum_{n \in \mathcal{N}_k} \mathbf{E}_k^\top \mathbf{E}_n - \sum_{p \in \mathcal{P}_k} \mathbf{E}_k^\top \mathbf{E}_p \right).$$

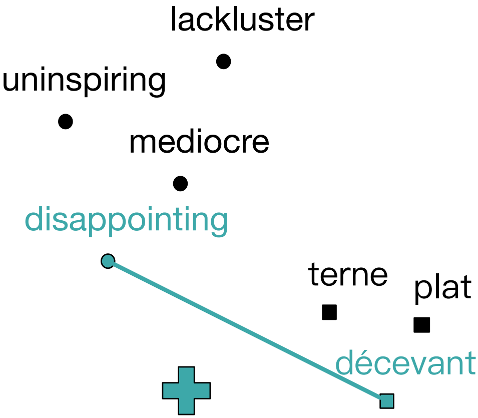
2. **Regularization:** Updated embeddings should not be too far away from original embeddings

$$R(\mathbf{E}) = \sum_{w \in \mathcal{V}} \left\| \hat{\mathbf{E}}_w - \mathbf{E}_w \right\|_2^2.$$

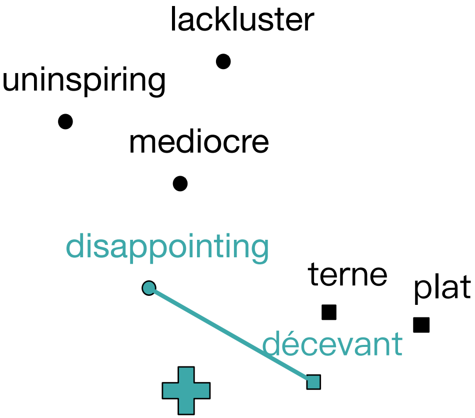
3. **Total cost function:**

$$C(\mathbf{E}) = C_f(\mathbf{E}) + \lambda R(\mathbf{E}).$$

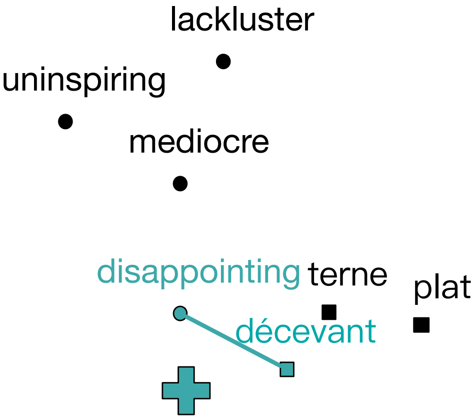
Retrofitting CLWE



Retrofitting CLWE



Retrofitting CLWE



Task

Document classification for detecting medical emergencies in low-resource languages

Task

Document classification for detecting medical emergencies in low-resource languages

Ilocano	... Nagtalinaed dagiti pito a balod ti Bureau of Jail Management and Penology (BJMP) ditoy ciudad ti Laoag iti isolation room gapo iti tuko ...

Experiments

- **Embeddings:** fastText (Bojanowski et al., 2017) aligned with RCSLS (Joulin et al., 2017)
- **Classifier:** CNN (Kim, 2014)
- **Source language:** English
- **Target language:** Ilocano

User Study

We want to classify documents into two categories:

1. Documents describing a medical emergency.
2. Documents **not** describing a medical emergency.

Which words are likely to appear in the same category as

ambulance

ENGLISH	ILOCANO
ambulances	ospital
medics	nars
hospital	tulong
emergency	pagbakuitan
nurses	pammaregta
Add word	Add word

Comparison

- **Base:** original CLWE and original training dataset

Comparison

- **Base:** original CLWE and original training dataset
- **Active:** original CLWE and training dataset augmented by active learning (entropy sampling)

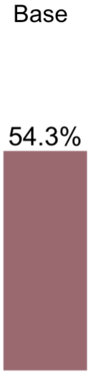
Comparison

- **Base:** original CLWE and original training dataset
- **Active:** original CLWE and training dataset augmented by active learning (entropy sampling)
- **CLIME:** refined CLWE and original training dataset

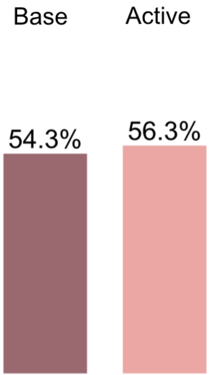
Comparison

- **Base:** original CLWE and original training dataset
- **Active:** original CLWE and training dataset augmented by active learning (entropy sampling)
- **CLIME:** refined CLWE and original training dataset
- **A+C:** refined CLWE and training dataset augmented by active learning

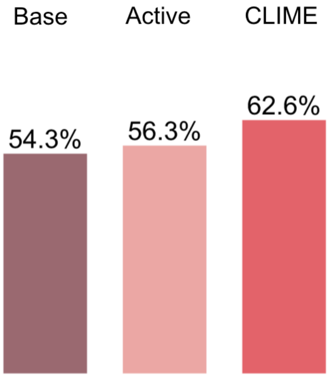
Results



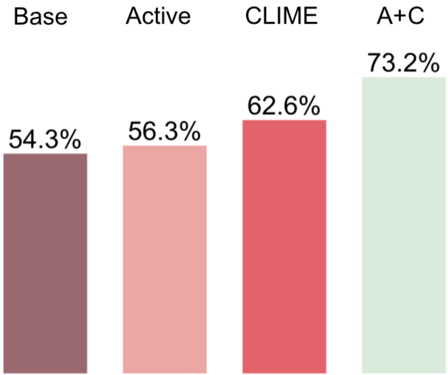
Results



Results



Results



References I

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *TACL*, 5:135–146.

Armand Joulin, Edouard Grave, Tomas Mikolov, and Piotr Bojanowski. 2017. Bag of tricks for efficient text classification. In *EACL*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *EMNLP*.

Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2016. Visualizing and understanding neural models in NLP. In *NAACL*.

Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Ira Leviant, Roi Reichart, Milica Gašić, Anna Korhonen, and Steve Young. 2017. Semantic specialisation of distributional word vector spaces using monolingual and cross-lingual constraints. *TACL*, 5:309–324.

Mozhi Zhang, Yoshinari Fujinuma, Michael J. Paul, and Jordan Boyd-Graber. 2020. Why overfitting isn't always bad: Retrofitting cross-lingual word embeddings to dictionaries. In *ACL*.