On the Applicability of Zero-Shot Cross-Lingual Transfer Learning for Sentiment Classification in Distant Language Pairs

ゼロショット転移学習を用いた多言語における感情分類の応用性

Andre Rusli, Makoto Shishido

言語処理学会第28回年次大会(NLP2022)
Hamamatsu, Japan – *Online*

TDU 東京電機大学

Agenda

- Backgrounds
  - Research objective
- Related works
- Experiments
  - Dataset
  - Scenarios
- Results
  - Fine-tuned supervised learning
  - Zero-shot transfer learning
- Conclusion and future works

Backgrounds

- Gap of resources for building a language model (LM) between languages
- Even though there is enough data, pre-training a language model from scratch requires a lot of computational resources
- Fine-tuning and cross-lingual transfer learning from a pre-trained LM
    - Some languages transfer better than the others
- XLM-R: massively multilingual Transformers pre-trained model (Conneau et al., 2020)

Research Objective

⇒ Experiment with and evaluate the performance of XLM-R for sentiment classification with zero-shot cross-lingual transfer learning between three languages; English, Japanese, and Indonesian.

Related Works

- XLM-R: a massively multi-lingual Transformers (MMT) model; a robustly trained RoBERTa; (Conneau et al., 2020)
  - CommonCrawl-100 data of 100 languages
  - 88/100 languages intersects with mBERT, but some of them are several orders of magnitude larger than mBERT
- SentencePiece: an unsupervised text tokenizer that does not depend on language-specific pre/postprocessing (Kudo & Richardson, 2018)
  - To handle multilingual texts in all three languages in our experiments. Also used inside of XLM-R.
- XLM-R for dependency parsing (Lauscher et al., 2020) and named entity recognition (Pfeiffer et al., 2020) in Japanese
- Cross-lingual transfer learning from English to Japanese using mBERT (Keung et al., 2020)
- Massive monolingual models for Indonesian based on BERT and XLM-R (Koto et al., 2020) (Willie et al., 2020)

Experiments: Datasets

AmazonEN: English Amazon product review sentiment dataset

- 160,000 data for fine- tuning. 4,000 data for evaluation.

AmazonJA: Japanese Amazon product review sentiment dataset

- 160,000 data for fine-tuning. 4,000 data for evaluation.

RakutenJA: Japanese Rakuten product review binary sentiment dataset

- 400,000 data for evaluation.

IndolemID: Indonesian Twitter and hotel review sentiment dataset

- 5,048 data for evaluation.

SmsaID: Indonesian multi-platform review sentiment dataset

- 1,129 data for evaluation.

Multilingual Amazon Reviews Corpus – Keung et al., 2020

Glyph – Zhang & LeCun, 2017

IndoLEM – Koto et al., 2020

SMSA – Purwarianti & Crisdayanti, 2019

Experiments: Datasets (examples)

English

- Positive ⇒ '"The best suspenders I've had so far. Sometimes the left side does not hold properly, but in general, they are still the best.'
- Negative ⇒ '"I bought 4 and NONE of them worked. Yes I used new batteries!'

Japanese

- Positive ⇒ '画像より色が暗く感じましたが、かわいいです。'
- Negative ⇒ 'やわらかいのはいいのですが しわがすごい。。'

Indonesian

- Positive ⇒ 'Cukup baik, pelayanan ramah. ruangan dan kamar mandi cukup bersih hanya saja tercium bau yg menggangu, sarapannya mantap over all memuaskan'
- Negative ⇒ 'Kamar bersih. Jadwal breakfast biasanya 07.00, ini belum ada apa2. Shower mati. Handuk tidak ada. Dan wifi ngadat. Kecewa.'

Experiments: Scenarios

Our experiments are divided into two main scenarios as follows.

1. **Fine-tuned supervised learning**

Fine-tune the XLM-RoBERTa$_{\mathrm{BASE}}$ pre-trained model using English-only, Japanese-only, and English+Japanese.

2. **Zero-shot transfer learning**

Use the fine-tuned model using AmazonEN to evaluate zero-shot cross-lingual transfer learning capability in AmazonJA, RakutenJA, SmsaID, and IndolemID datasets.

Results: Fine-tuned supervised learning

Table 1. Fine-tuning specifications and elapsed time

Fine-tuning Source Data	GPU	Ep-och (s)	Average elapsed time per epoch
AmazonEN	Tesla T4	4	33 minutes 5 seconds
AmazonJA	Tesla P100-PCIE-16GB	4	17 minutes 31 seconds
AmazonENJA	Tesla P100-PCIE-16GB	2	35 minutes 57 seconds

Table 2. Error percentage of the fully-supervised evaluation on the Multi-lingual Amazon Review Corpus. Results using mBERT are obtained from [15].

Model	EN-only	JA-only	EN&JA
mBERT	8.8	11.1	-
XLM-R_{BASE}	**7.35**	**7.25**	**7.19**

Results: Zero-shot transfer learning

Table 3. Error percentage of zero-shot cross-lingual transfer learning using XLM-R_{BASE} in comparison to a zero-shot mBERT from English data [15] and Japanese data [19]

Model	***AmazonJA***	***RakutenJA***
Zero-shot mBERT	19.04	-
Fully-supervised ULMFiT	-	4.45
XLM-R$_{BASE}$ w/ *AmazonEN*	**11.12**	**13.09**
XLM-R$_{BASE}$ w/ *AmazonENJA*	**7.05**	**8.51**

Table 4. Macro-averaged F1-score of zero-shot cross-lingual transfer learning using XLM-R$_{BASE}$ for Indonesian (IndoLEM [17], and SmSA [16]).

Model	***IndolemID***	***SmsaID***
Fully-supervised BERT	84.13	92.72
Fully-supervised mBERT	76.58	84.14
XLM-R$_{BASE}$ w/ *AmazonEN*	**72.19**	**86.77**
XLM-R$_{BASE}$ w/ *AmazonENJA*	**73.31**	**87.99**

Conclusion & Future Work

- This paper reports the results of experiments focusing on evaluating the applicability of cross-lingual transfer learning using the XLM-R pre-trained model.
  - Based on the results, zero-shot cross-lingual transfer learning between English, Japanese, and Indonesian yields promising results using XLM-R.
  - The models achieve the best result in one dataset and shows the applicability of cross-lingual transfer learning, considering that the models have not seen languages in the target dataset, it can outperform SOTA results in other datasets trained in a fully-supervised approach.
- All experiments are performed using the free version of Google Colab
- Future research steps include experimenting with more **hyperparameters**, evaluating other potential methods such as **few-shot transfer learning**, which has been proven to be useful to improve performance by adding just a few annotated data (Lauscher et al., 2020) and **meta-learning** (Nooralahzadeh et al., 2020).