Multilingual model using cross-task embedding projection

Jin Sakuma
The University of Tokyo
jsakuma@tkl.iis.u-Tokyo.ac.jp

Naoki Yoshinaga
Institute of Industrial Science, the University of Tokyo
ynaga@iis.u-Tokyo.ac.jp

Summary

- Obtain **fully task-specific multilingual models** by learning **cross-task embedding projection**
- Propose **locally linear mapping** for **cross-task projection which has analytical solution**
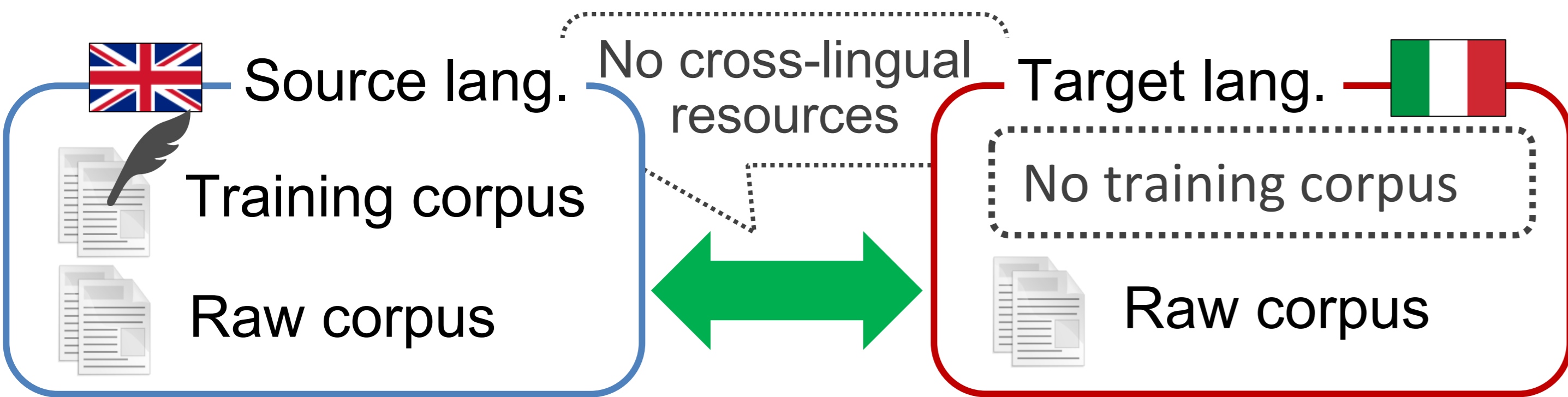- Empirically confirmed the effectiveness of our method in topic classification and sentiment analysis

Background

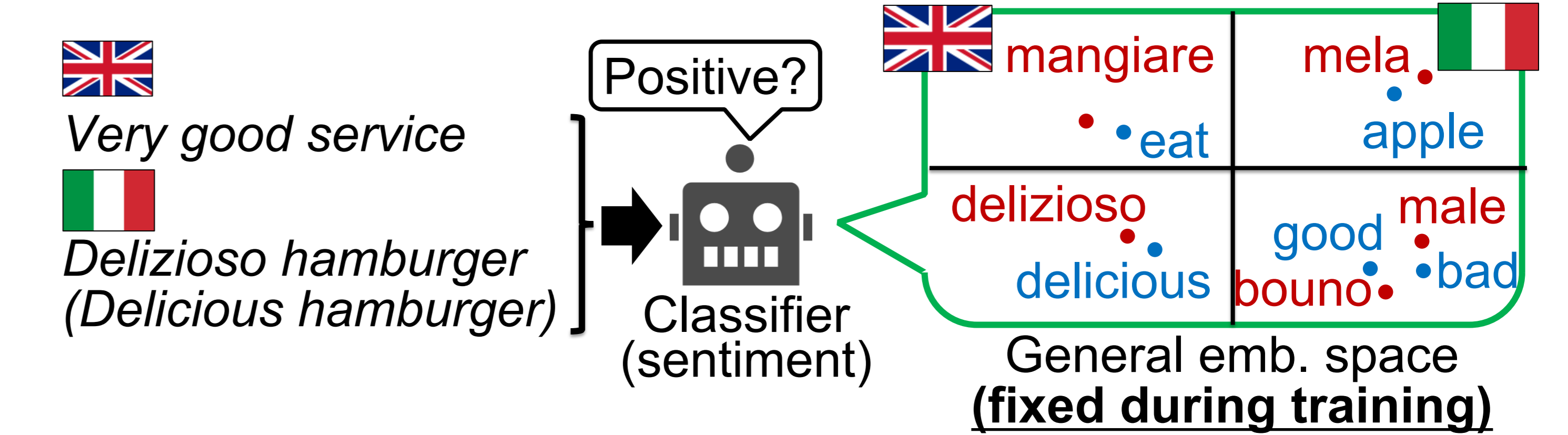Only a few languages have sufficient annotated resources for supervised learning (esp., deep learning)

❌ significant accuracy gap
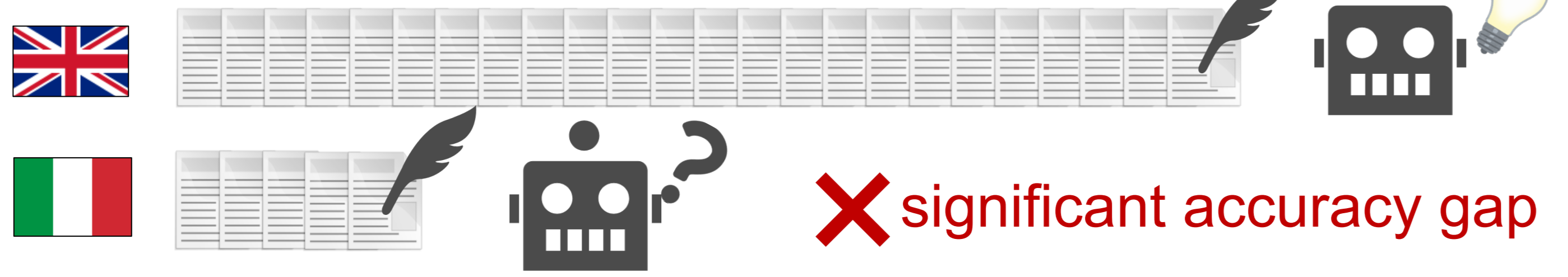
Multilingual models can be **trained in one language and applied to another** to utilize resources across languages.

[The target situation in this study]

Source lang. — No cross-lingual resources — Target lang.

Training corpus / Raw corpus ↔ No training corpus / Raw corpus

Related work

Multilingual models handle languages other than the one used in training by using **cross-lingual word embeddings (CLWE)**

Very good service

Delizioso hamburger (Delicious hamburger) → Positive? → Classifier (sentiment)

mangiare / mela / eat / apple / delizioso / delicious / good / buono / bad / male

General emb. space **(fixed during training)**

✔ No cross-lingual resources are required when using unsupervised CLWE [Artetxe+ 2018]

❌ General CLWE (Skip-gram etc.) is not optimized for the target task (the model performance is limited)

We want to have embeddings of multilingual models **optimized for the target task** to improve the accuracy
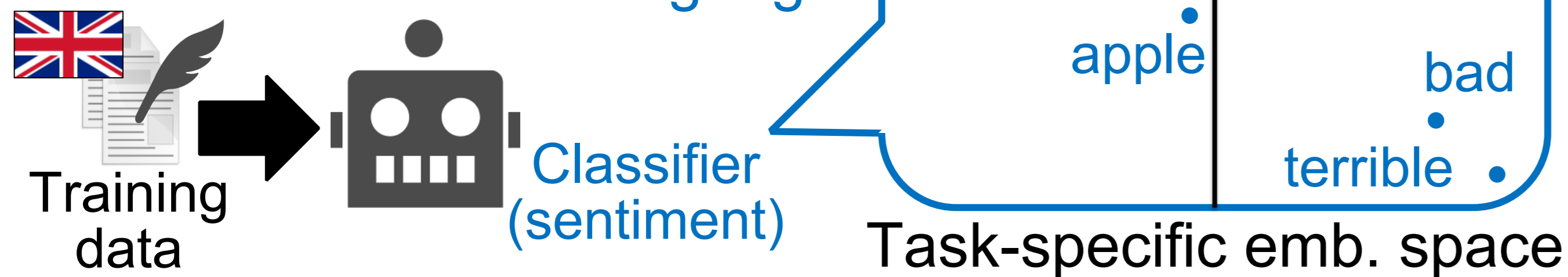
Proposal

💡 Obtain **fully task-specific multilingual models** by **cross-task embedding projection**

Step1: Train general CLWE from raw monolingual corpora in the source and target languages [Artetxe+ 18]

mangiare / mela / eat / apple / delizioso / bad / buono / delicious / male / great / good

General emb. space

Step 2: Train a neural network in the source language

Training data → Classifier (sentiment)

Step 3: Induce task-specific CLWE for **words in the target language**

great / good / buono / delicious / eat / apple / bad / terrible

Task-specific emb. space

Locally linear mapping for cross-task emb. projection

Assumption:
Neighboring words in the general embedding space will be close in the task-specific embedding space

Step 3.1:
In the general emb. space, reconstruct the target word (bouno) as a weighted mean of neighboring source words (good, great)

bouno — good, great

$$\hat{\alpha}_{w*} = \arg\min_{\alpha_{w*}} \left| Y_w^{\text{gen}} - \sum_{i \in \mathcal{N}_w} \alpha_{wi} X_i^{\text{gen}} \right|$$

k nearest neighbors of w

- $\hat{\alpha}_{w*}$ learns relationship among neighboring words
- The **analytical solution** is obtainable by Lagrange multiplier method

Step 3.2:
Compute a task-specific embedding of the target word (bouno) to **preserve the local topology**

$$Y_w^{\text{spec}} = \sum_{i \in \mathcal{N}_w} \hat{\alpha}_{wi} X_i^{\text{spec}}$$

Experiments

❓ Does task-specific CLWE improve multilingual models?

[Tasks and datasets]
- **Topic classification (4 classes)**: RCV1/2 corpus in 8 languages
- **Sentiment analysis (3 classes)**:
  - Amazon reviews in 3 languages
  - Yelp reviews in 3 languages (evaluated on ABSA dataset)

[Models to compare]
- **CLWE fixed (baseline multilingual model)**:
  Bag-of-embeddings model with two-layer feed forward neural network with **its embedding layer fixed to general CLWE**
- **CLWE opt. (proposed task-specific multilingual model)**:
  Variant of CLWE fixed with **the embedding layer updated**; we made this model cross-lingual by our cross-task projection

Models are trained in English and tested in the target languages

[Results]

Models	Topic classification RCV 1/2								Sentiment analysis			Yelp - ABSA		
	es	fr	it	da	de	nl	pt	sv	de	fr	ja	es	nl	tr
CLWE fixed	0.36	0.77	0.54	0.62	**0.81**	0.79	0.52	0.82	0.80	**0.81**	**0.80**	0.73	**0.68**	0.57
CLWE opt.	**0.71**	**0.89**	**0.62**	**0.67**	0.81	**0.81**	**0.58**	**0.83**	**0.82**	0.81	0.79	**0.76**	0.68	**0.62**

✔ CLWE opt. outperforms CLWE fixed in most of the settings

Analysis

❓ Does cross-task projection correctly compute task-specific embeddings?

We observe the nearest neighbors to understand the obtained task-specific CLWE

	General	Topic classification
excellent (excellent)	excellente (excellent), excellents, bon (good), excellellentes, exceller (to excel)	execellents, excellente, excellentes, appreciable (appreciable), bons (good)
terrible (terrible)	terribles (terrible), horrible (horrible), terriblement (terribly), epouvantable, effroyable (terrifying)	terribles, horrible, meurtrie (wounded), **gwynplaine**, epouvantes (terrified)
economie (economy)	economie (economy), economique (economic), macroeconomie, geoeconomie, microeconomie	economique, economiques, conjuncture (conjuncture), **fmi (IMF)**, economique

- Unrelated words are close to words that will not contribute to classification (e.g., **gwynplaine**)
- Words that used often in the same topic are close to each other (e.g., **fmi**)

✔ Cross-task projection successfully obtains task-specific representations of words