GL-CLeF: A Global-Local Contrastive Learning Framework for Cross-lingual Spoken Language Understanding


Libo Qin[1]


Qiguang Chen[1]


Tianbao Xie[1]


Qixin Li[1]


Jian-Guang Lou[2]


Wanxiang Che[1]


Min-Yen Kan[3]

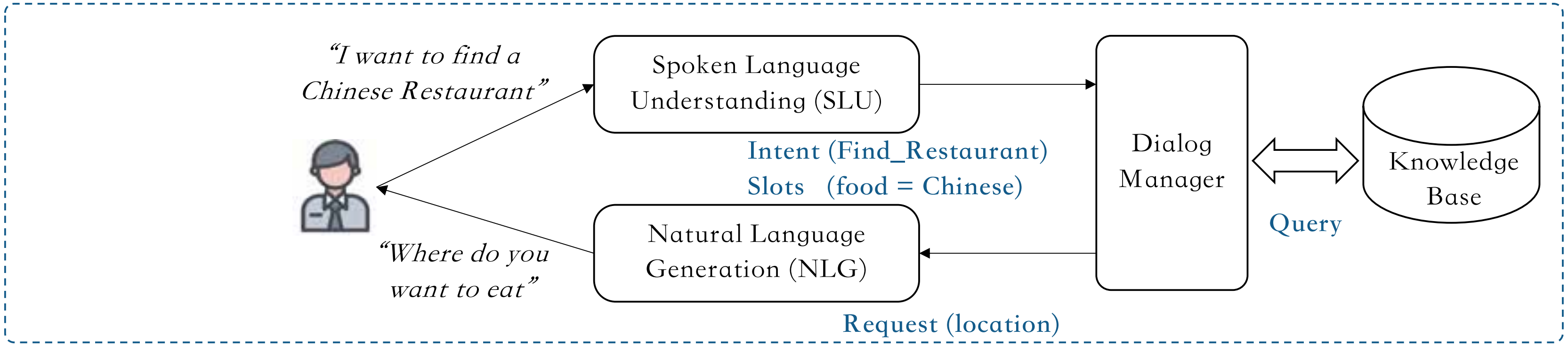
[1]Research Center for Social Computing and Information Retrieval (SCIR), Harbin Institute of Technology

[2]Microsoft Research Asia, Beijing, China

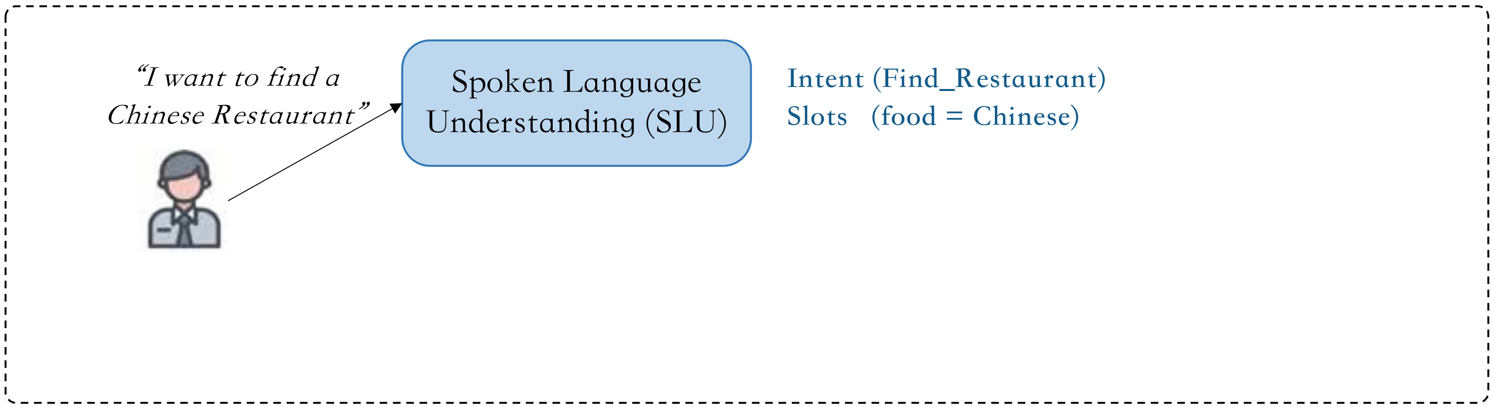
[3]Department of Computer Science, National University of Singapore

Background

Architecture of Task-oriented Dialogue

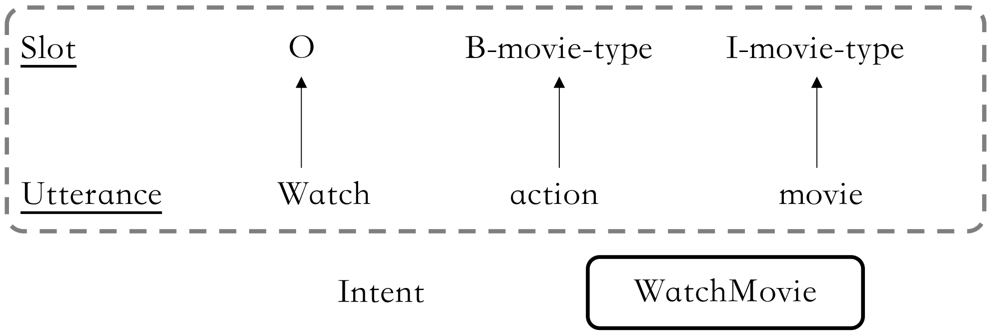


Architecture of Task-oriented Dialogue



Spoken Language Understanding SLU

- SLU
  - Slot filling -> sequence labeling
  - Intent detection -> classification task



Motivation

Progress

- Progress
  - With the success of Deep Learning technique, remarkable progress has been made in spoken language understanding.

- Problems
  - Rely on a considerable amount of labeled data, which is only available on the English data set[1]
  - Hard to scale to other low-resource languages

[1] Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP (IJCAI2020)

Progress

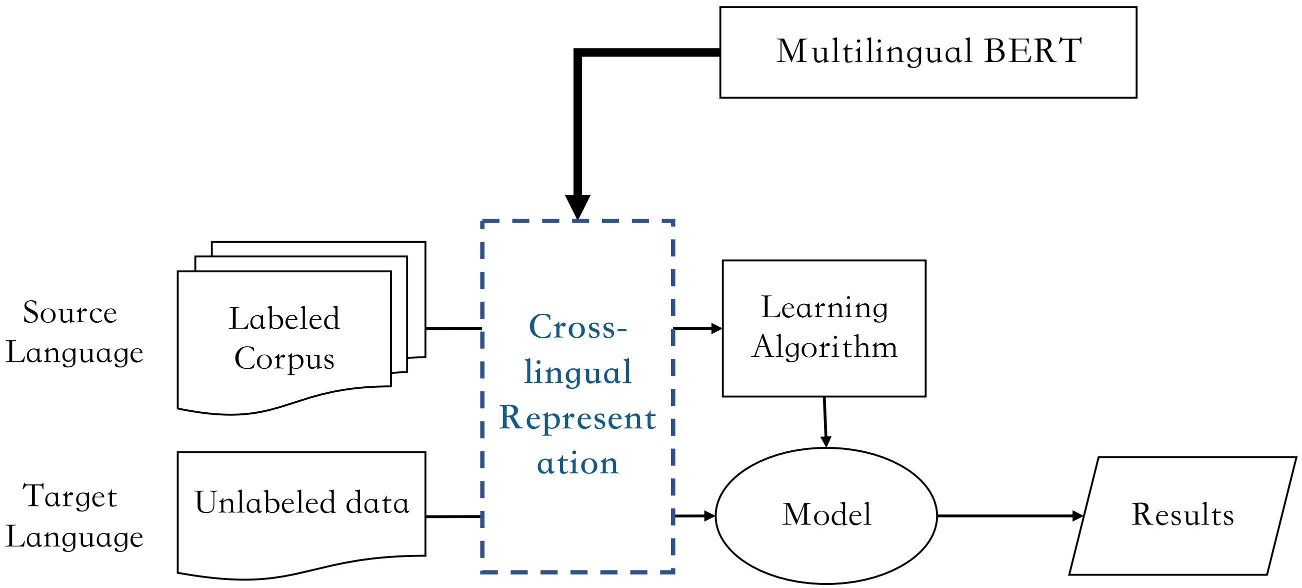
- Progress
  - With the success of DeepLearning technique, remarkable progress has been made in spoken language understanding.

- Problems
  - Rely on a considerable amount of labeled data, which is only available on the English data set[1]
  - Hard to scale to other low-resource languages

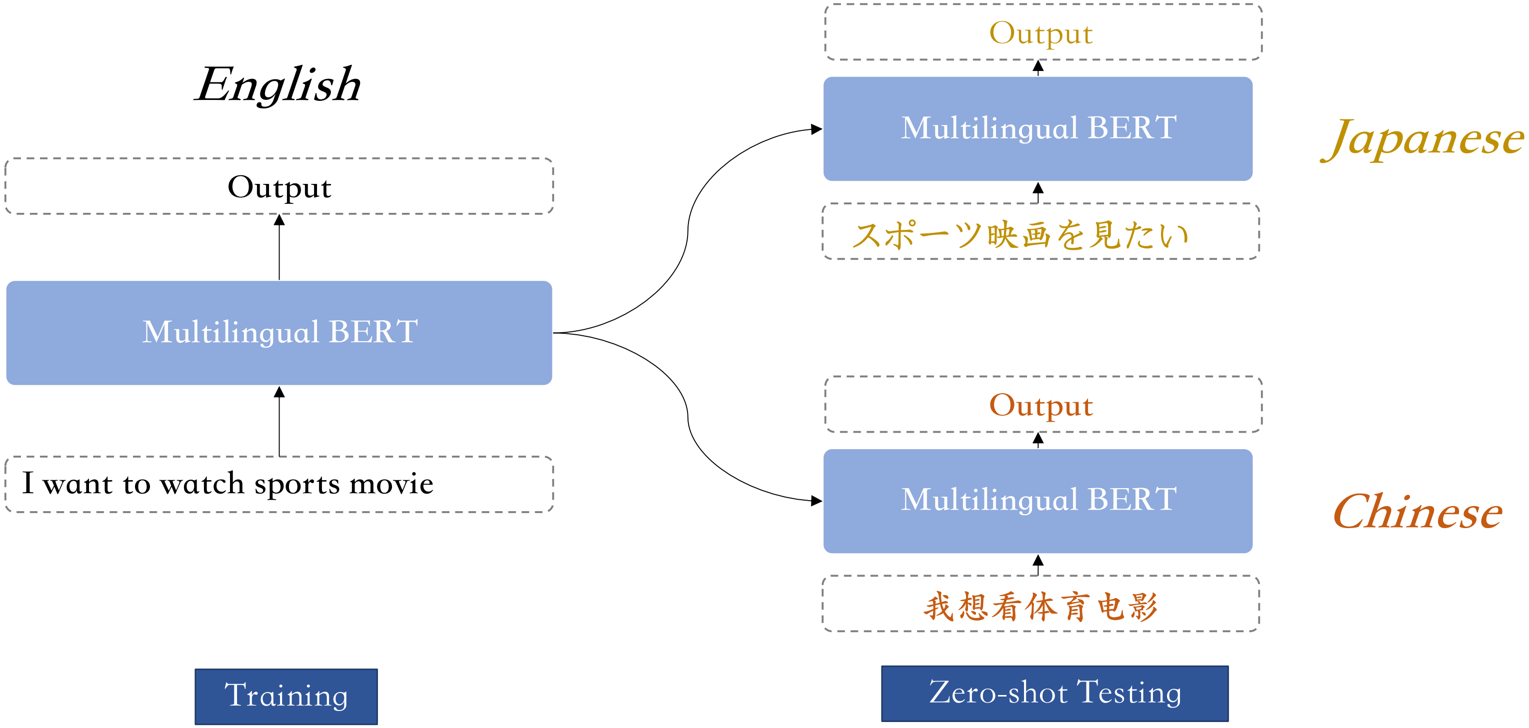
Zero-shot SLU have gained increasing attention

[1] Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP (IJCAI2020)

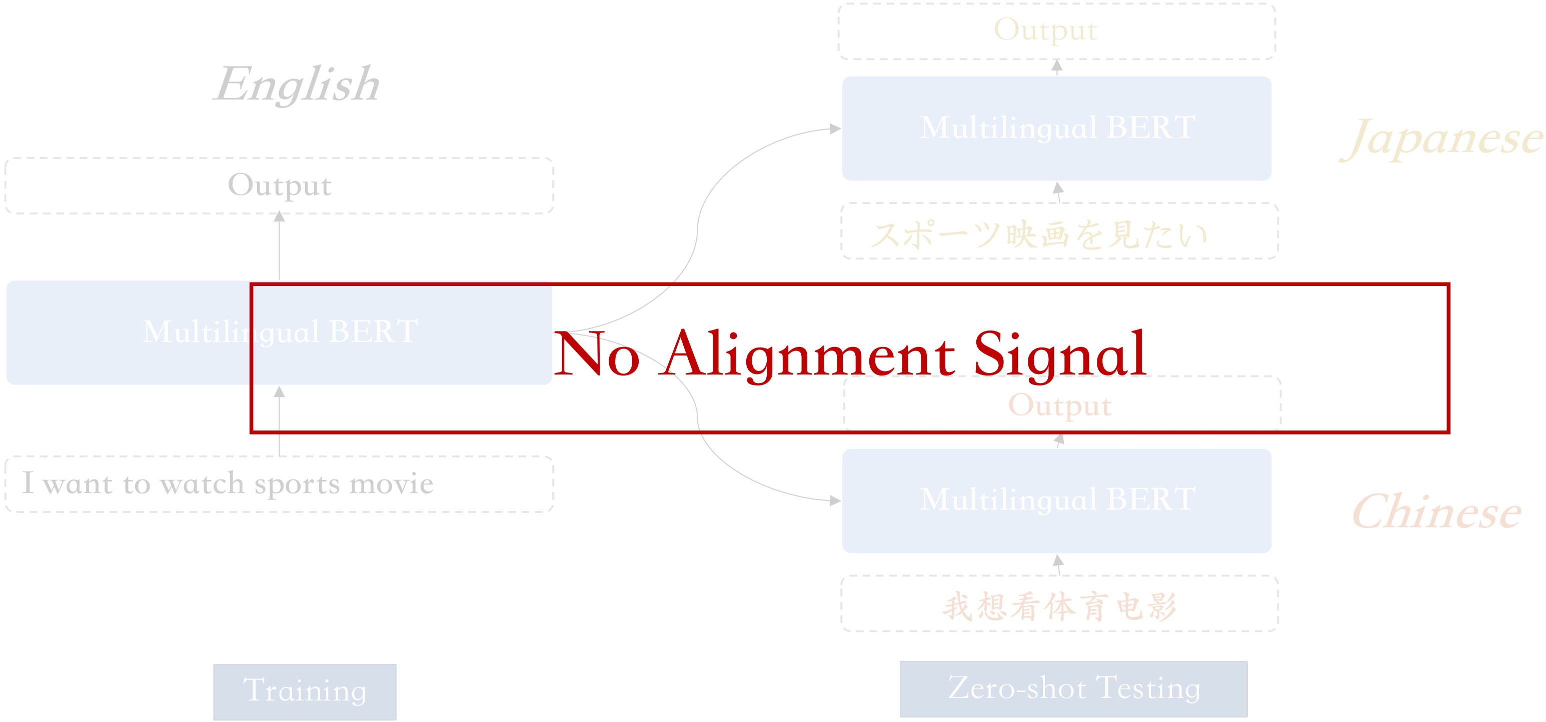
Zero-shot Cross-lingual Framework



Multilingual BERT

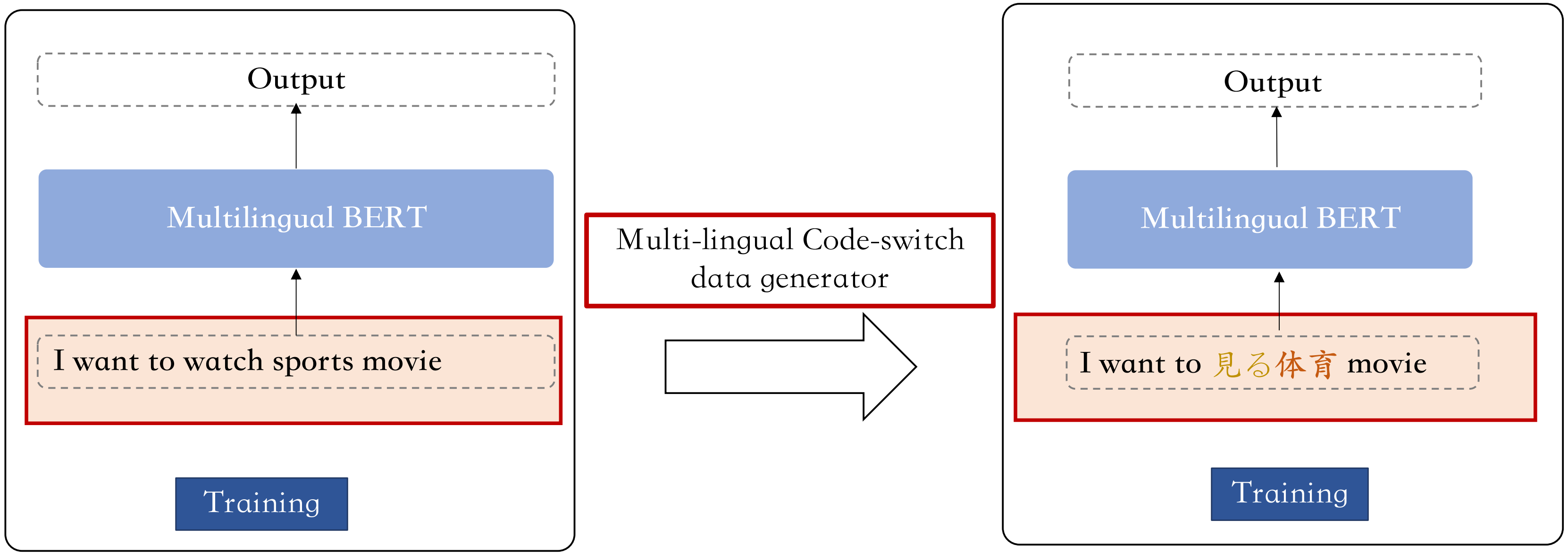


Multilingual BERT



③ Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. (IJCAI2020)

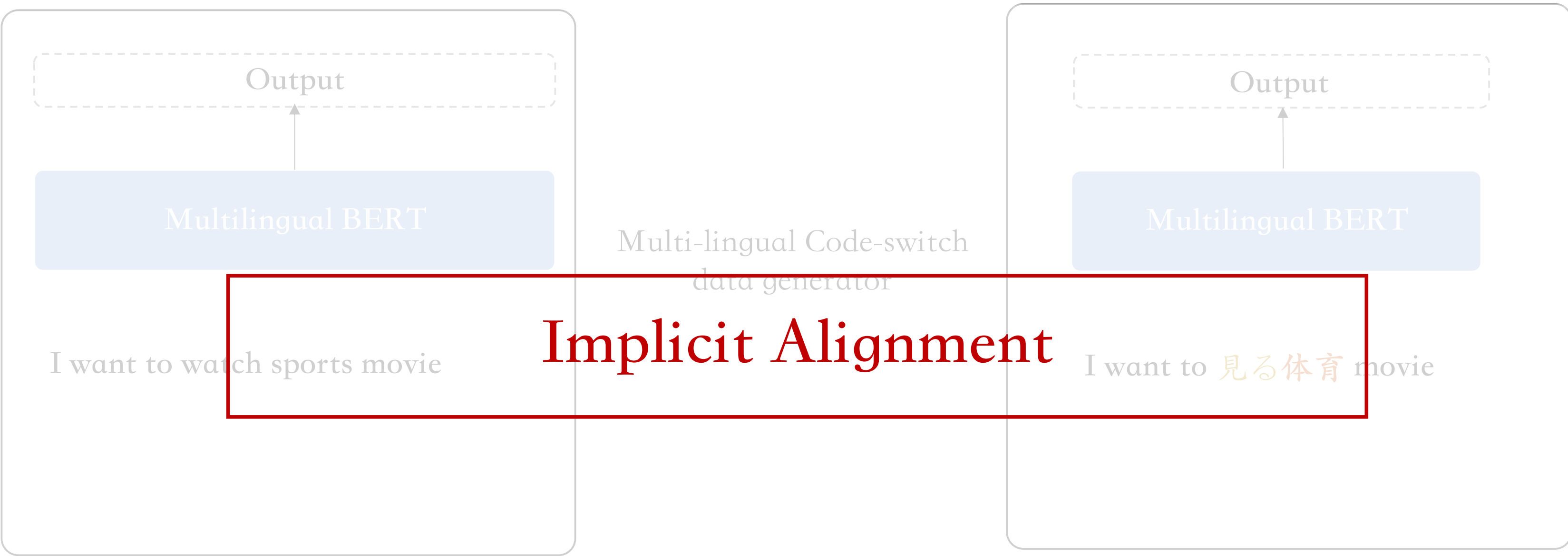
CoSDA-ML



Automatically and implicitly align the replaced word vectors in the source and all target languages by mixing their context information.

Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. (IJCAI2020)

CoSDA-ML



Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. (IJCAI2020)

Challenges of CoSDA-ML

- First Challenge
  - Such implicit alignment process seems to be a **black box**, which not only seriously affects the alignment representation but also makes it hard to analyze the alignment mechanism.

- Second Challenge
  - Simply relying on shared parameters does not distinguish between the **varying granularities** of the tasks: the intent detection is **sentence-level** and the slot filling is **token-level**, which does not offer fine-grained cross-lingual transfer for token-level slot filling.

GL-CLEF: A Global – Local Contrastive Learning Framework for Cross-lingual Spoken Language Understanding

- **To solve the first challenge**
  - We employ contrastive learning (CL) to **explicitly** align representations of similar sentences across different languages.
  - The key insight in GL-CLEF is to encourage representations of similar sentences to be more similar than negative example pairs via contrastive learning.

GL-CLEF: A Global – Local Contrastive Learning Framework for Cross-lingual Spoken Language Understanding

- **To solve the second challenge**
  - We first introduce a ***Local module*** in GL-CLEF to learn different granularity alignment representations (i.e., sentence-level Local intent CL and token-level local slot CL).
    - To be specific, ***sentence-level local intent CL*** and ***token-level local slot CL*** are introduced for aligning similar sentence and token representations across different languages for intent detection and slot filling, respectively.
  - We further propose a ***Global module*** named semantic-level global intent – slot CL to bring the representations of slot and intents within a sentence closer together.
    - We further argue that slot and intent are highly correlated and have similar semantic meanings in a sentence

Positive and Negatives samples of CL

- Positive Samples Generation
  - We employ CoSDA-ML to generate multi-lingual code-switched data, which is considered as the positive samples.

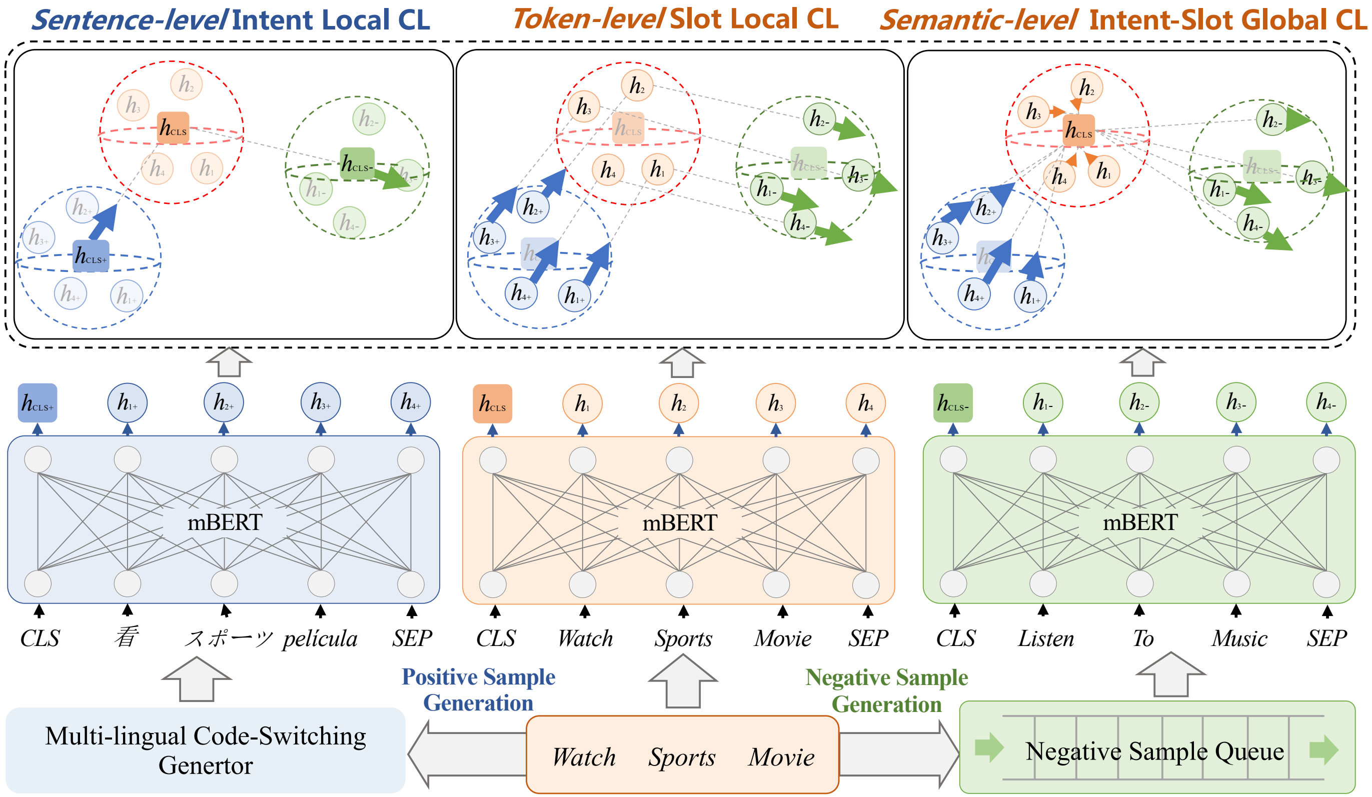
$[CLS_{ori}]$ watch sports movie ⇨ *Positive Samples Generation* ⇨ $[CLS_{pos}]$ watch 运动 movie

- Negatives Samples Generation
  - Other different queries in a batch can be considered as negative samples.

$[CLS_{ori}]$ watch sports movie ⇨ *Negative Samples Generation* ⇨ $[CLS_{neg}]$ listen to music

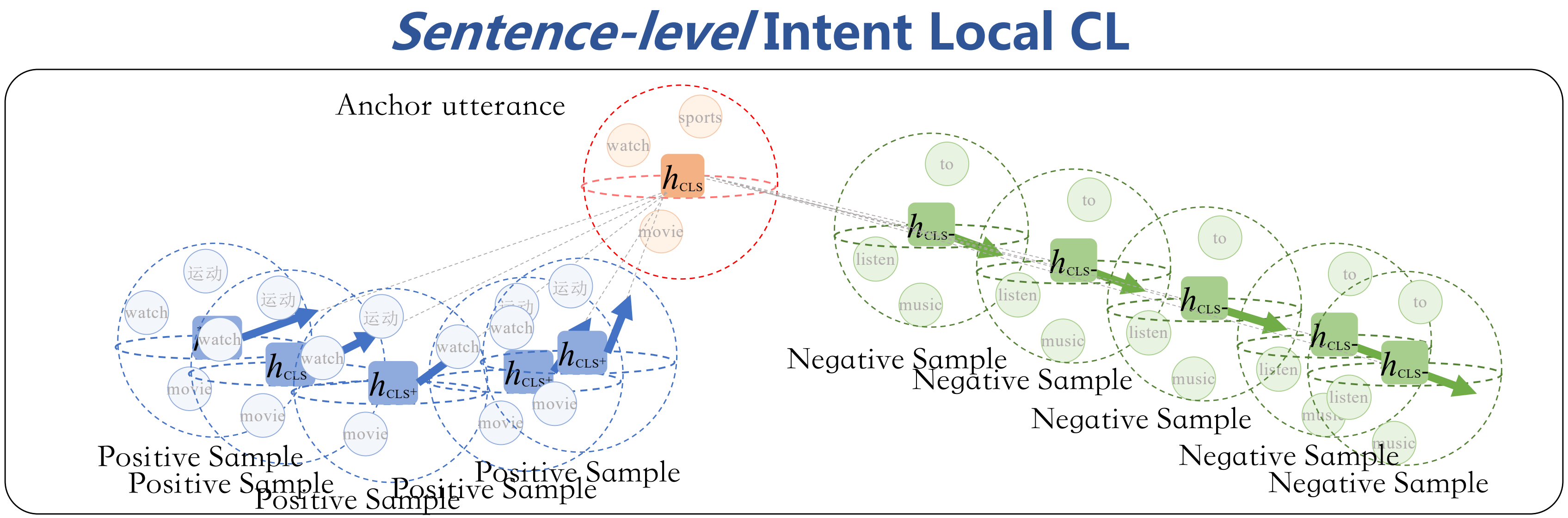
Models

Overall Framework



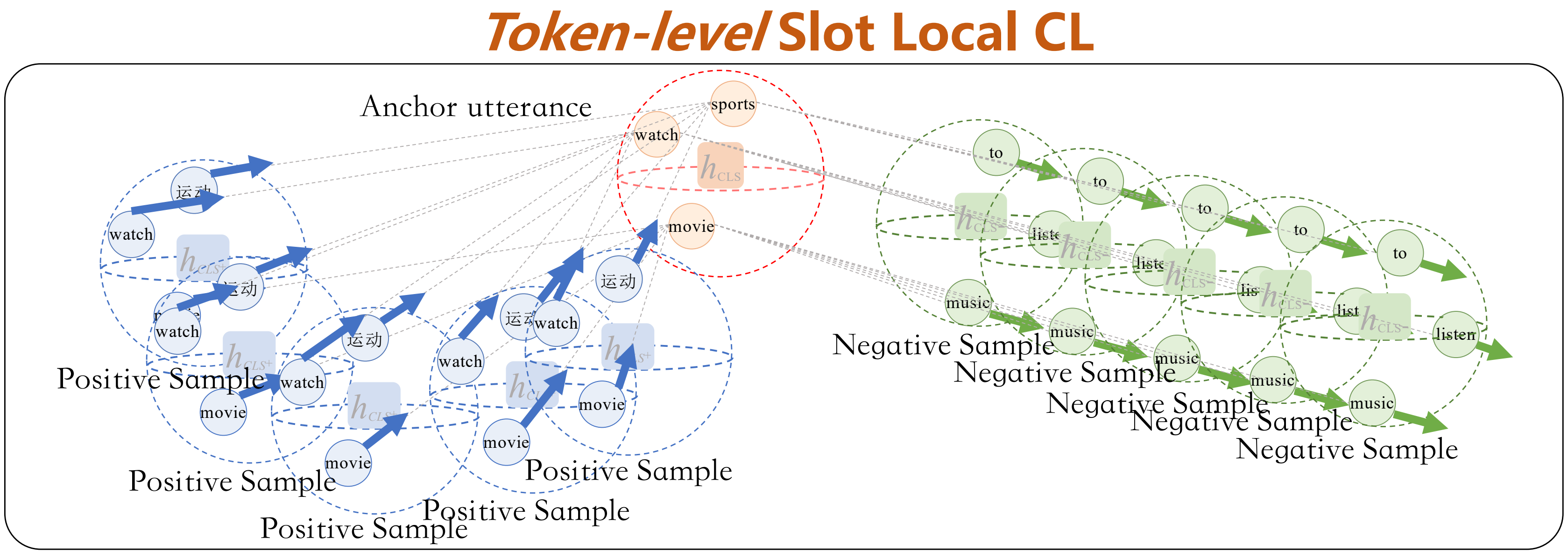
Sentence-level Local Intent CL

- A sentence-level local intent CL loss is introduced to explicitly encourage the model to align similar sentence representations into the same local space across languages for intent detection.



Token-level Local Slot CL

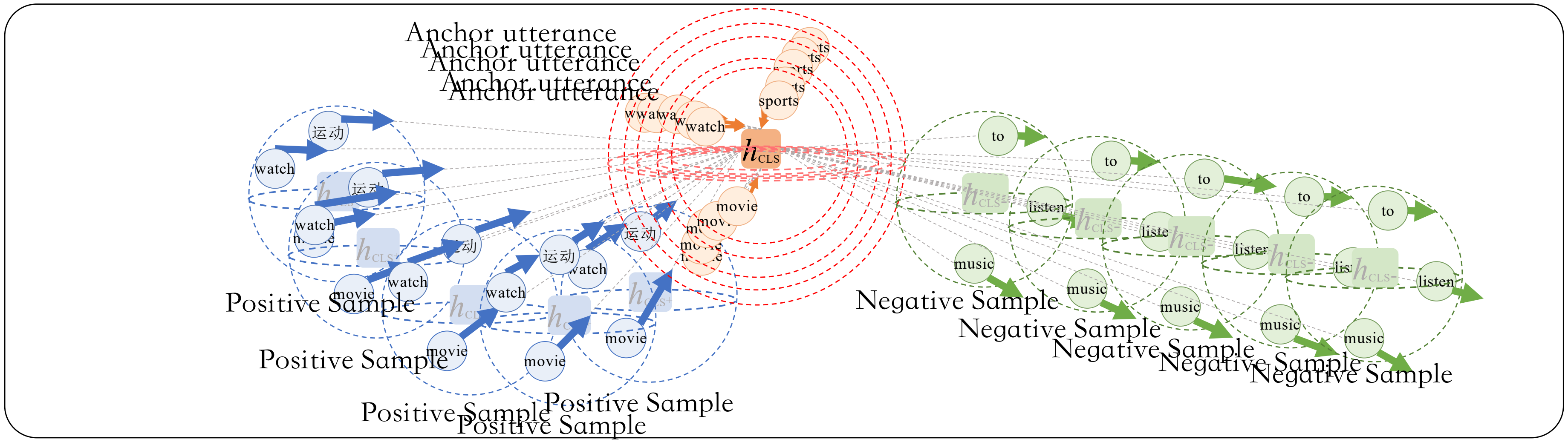
- We propose a token-level local slot CL loss to help the model to consider token alignment for slot filling, achieving fine-grained cross-lingual transfer. In this situation, token-level CL is applied to all tokens in the query.



Semantic-level Global Intent-slot CL

- Further, we introduce a semantic-level global intent-slot CL loss to model the semantic interaction between slots and intent, which may further enhance cross-lingual transfer between them.

Semantic-level **Intent-Slot Global CL**



Training

- The overall objective in GL-CLEF is a tuned linear combination of the individual losses:

$$\mathcal{L} = \lambda_I \mathcal{L}_I + \lambda_S \mathcal{L}_S + \lambda_{\mathrm{LI}} \mathcal{L}_{\mathrm{LI}} + \lambda_{\mathrm{LS}} \mathcal{L}_{\mathrm{LS}} + \lambda_{\mathrm{GIS}} \mathcal{L}_{\mathrm{GIS}},$$

where L_I denotes intent detection loss; L_S is slot filling loss; L_{LI} denotes sentence-level local intent CL loss; L_{LS} denotes token-level local slot CL loss; L_{GIS} denotes semantic-level global intent-slot CL loss and λ_* are tuning parameters for each loss component.

Experiments

Datasets

- Multi-ATIS++
  - 9 languages including English (en), Spanish (es), Portuguese (pt), German (de), French (fr), Chinese (zh), Japanese (ja), Hindi (hi), and Turkish (tr).

Language	Utterances			Intents	Slots
	Train	Dev	Test		
English	4,488	490	893	18	84
Spanish	4,488	490	893	18	84
Portuguese	4,488	490	893	18	84
German	4,488	490	893	18	84
French	4,488	490	893	18	84
Chinese	4,488	490	893	18	84
Japanese	4,488	490	893	18	84
Hindi	1440	160	893	17	75
Turkish	578	60	715	17	71

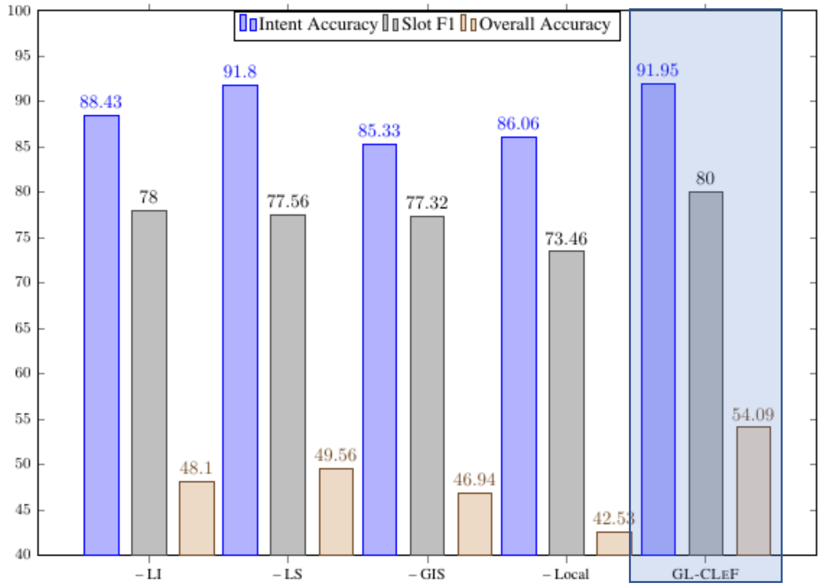
Table 1: MultiATIS++ datasets and statistics.

Main Results

Intent Accuracy.	en	de	es	fr	hi	ja	pt	tr	zh	AVG
mBERT* (Xu et al., 2020)	-	95.27	96.35	95.92	80.96	79.42	94.96	69.59	86.27	-
mBERT† (Devlin et al., 2019)	98.54	95.40	96.30	94.31	82.41	76.18	94.95	75.10	82.53	88.42
Ensemble-Net* (Razumovskaia et al., 2021)	92.50	90.26	96.64	95.18	77.88	77.04	95.30	75.40	84.99	87.20
CoSDA† (Qin et al., 2020)	95.74	94.06	92.29	77.04	82.75	73.25	93.05	80.42	78.95	87.32
GL-CLEF	**98.77**	**97.53**	**97.05**	**97.72**	**86.00**	**82.84**	**96.08**	**83.92**	**87.68**	**91.95**
Slot F1	en	de	es	fr	hi	ja	pt	tr	zh	**AVG**
Ensemble-Net* (Razumovskaia et al., 2021)	85.05	82.75	77.56	76.19	14.14	9.44	74.00	45.63	37.29	55.78
mBERT* (Xu et al., 2020)	-	82.61	74.98	75.71	31.21	35.75	74.05	23.75	62.27	-
mBERT† (Devlin et al., 2019)	95.11	80.11	78.22	82.25	26.71	25.40	72.37	41.49	53.22	61.66
CoSDA† (Qin et al., 2020)	92.29	81.37	76.94	79.36	64.06	66.62	75.05	48.77	77.32	73.47
GL-CLEF	**95.39**	**86.30**	**85.22**	**84.32**	**70.34**	**73.12**	**81.83**	**65.85**	**77.61**	**80.00**
Overall Accuracy.	en	de	es	fr	hi	ja	pt	tr	zh	**AVG**
AR-S2S-PTR* (Zhu et al., 2020)	86.83	34.00	40.72	17.22	7.45	10.04	33.38	–	23.74	23.77
IT-S2S-PTR* (Zhu et al., 2020)	87.23	39.46	50.06	46.78	11.42	12.60	39.30	–	28.72	32.69
mBERT† (Devlin et al., 2019)	87.12	52.69	52.02	37.29	4.92	7.11	43.49	4.33	18.58	36.29
CoSDA† (Qin et al., 2020)	77.04	57.06	46.62	50.06	26.20	28.89	48.77	15.24	46.36	44.03
GL-CLEF	**88.02**	**66.03**	**59.53**	**57.02**	**34.83**	**41.42**	**60.43**	**28.95**	**50.62**	**54.09**

Our framework achieves the state-of-the art performance by beating CoSDA-ML with 10.06% performance

Ablation Analysis



All components contribute a lot for the final performance

Visualization Analysis

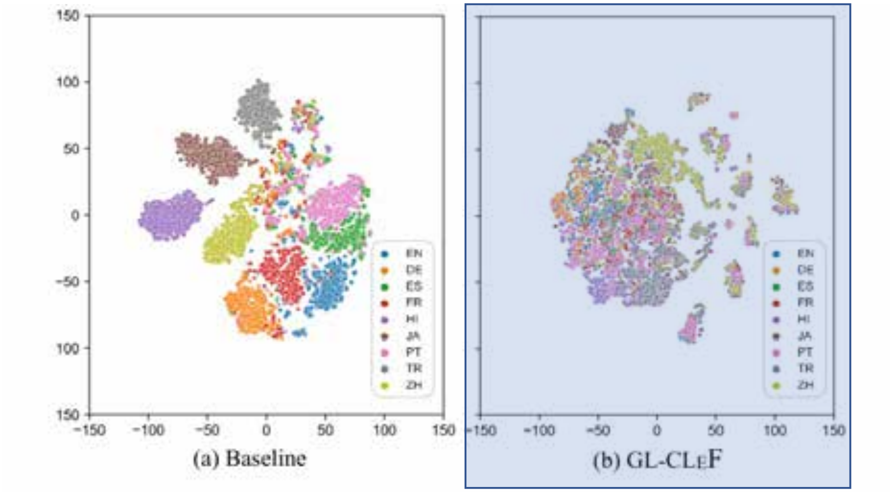


Figure 5: t-SNE visualization of sentence vectors from (a) mBERT and (b) GL-CLEF. Different colors represents different languages.

GL-CLEF successfully pulls representations closer across different languages.

Conclusion

- We introduced a global-local contrastive learning (CL) framework (GL-CLEF) to **explicitly** align representations across languages for zero-shot cross-lingual SLU.

- Besides, the proposed Local CL module and Global CL module achieves to **learn different granularity alignment** (i.e., sentence-level local intent alignment, token-level local slot alignment, semantic-level global intent-slot alignment).

- Experiments on MultiATIS++ show that GL-CLEF obtains **best performance** and extensive analysis indicate GL-CLEF successfully pulls closer the representations of similar sentence across languages.

Thanks&QA

Paper

Code