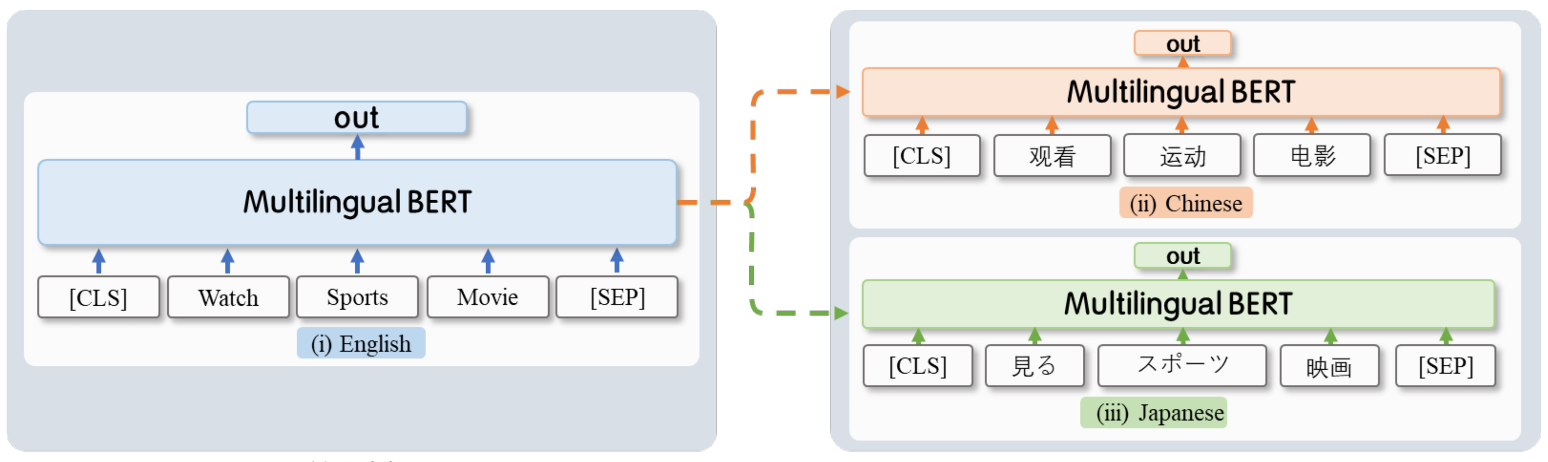
GL-CLEF: A Global-Local Contrastive Learning Framework for Cross-lingual Spoken Language Understanding

ACL 2022 Main Conference

Libo Qin[1], Qiguang Chen[1], Tianbao Xie[1], Qixin Li[1], Jian-Guang Lou[2], Wanxiang Che[1], Min-Yen Kan[3]
[1]Harbin Institute of Technology, China; [2]Microsoft Research Asia, Beijing, China;
[3]Department of Computer Science, National University of Singapore

❖ Introduction

- **Zero-shot Cross-lingual SLU**
  - Model is trained in a source language and directly applied to other target languages.

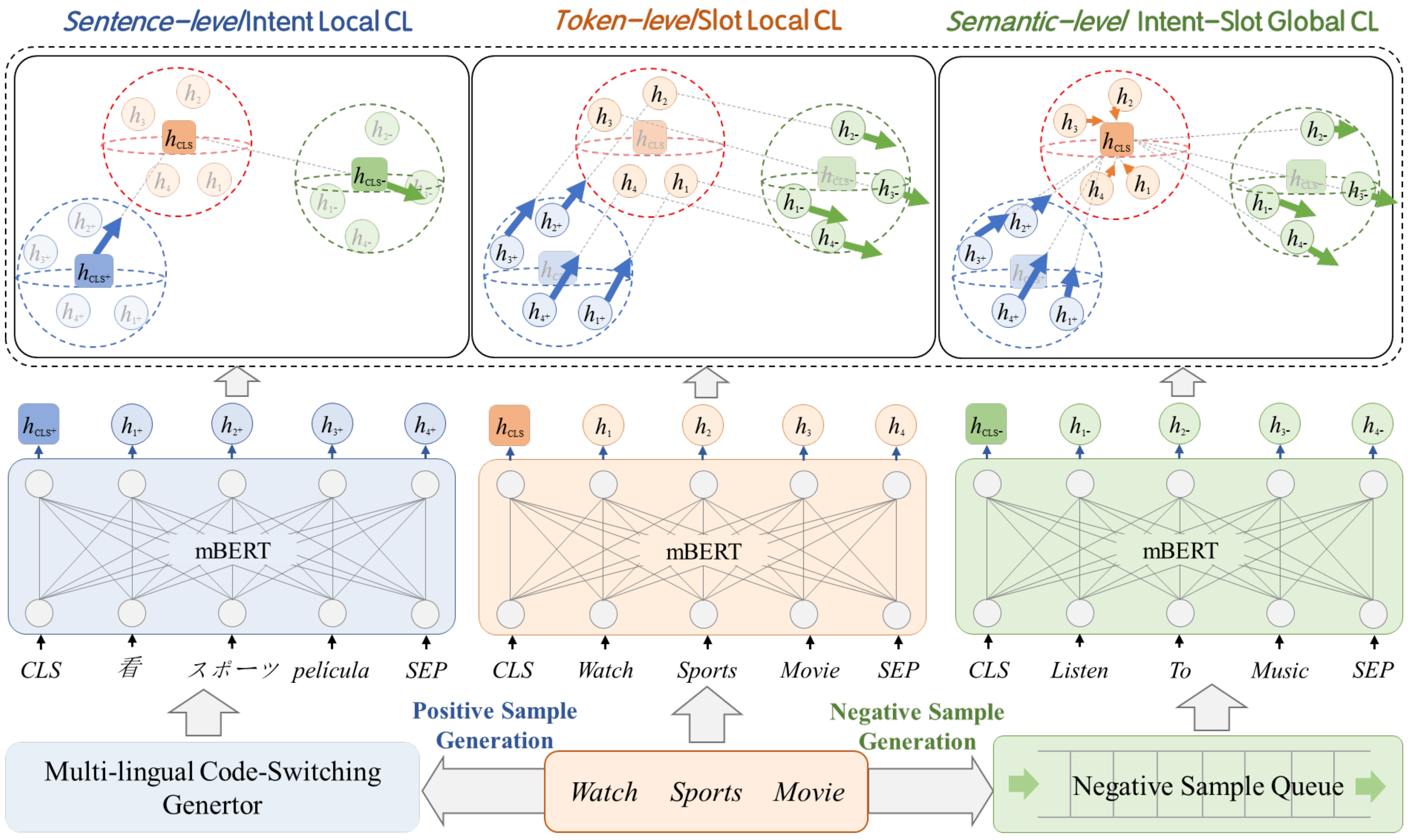


(a) Training　　(b) Zero-shot Testing

- **Challenges of current methods**
  - **First Challenge:**
    - Current methods perform an implicit alignment process that seems to be a ***black box***, which affects the alignment representation and makes it hard to analyze the alignment mechanism.
  - **Second Challenge:**
    - Current methods do not offer cross-lingual transfer between the ***varying granularities*** of the tasks: the intent detection is ***sentence-level*** and the slot filling is ***token-level***.

❖ Model

- **To solve the first challenge:**
  - Exploring contrastive learning (CL) to ***explicitly*** align representations of similar sentences across different languages.
- **To solve the second challenge:**
  - We first introduce a ***Local module*** to learn different granularity alignment representations.
    - ***Sentence-level* Local CL Module** to align sentence representation across languages for intent detection.
    - ***Token-level* Local CL Module** to align token representations across languages for slot filling.
  - We further propose a ***Global module***, as slots and intent are often highly related semantically when they belong to the same query.
    - ***Semantic-level* Global CL Module** to align representations between a slot and an intent.



The main architecture of GL-CLEF

❖ Experiments

- **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)	90.26	92.50	96.64	95.18	77.88	77.04	95.30	75.04	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**	**84.31**	**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	-
IT-S2S-PTR* (Zhu et al., 2020)	87.23	39.46	50.06	46.78	11.42	12.60	39.30	–	28.72	-
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**

1. Explicit alignment method GL-CLEF outperforms the implicit alignment method (CoSDA)
2. Our framework achieves the state-of-the-art performance

- **Experimental results on BiLSTM and XLM-R**

Intent Acc.	en	de	es	fr	hi	ja	pt	tr	zh	AVG
BiLSTM (Hochreiter and Schmidhuber, 1997)	72.56	70.96	70.35	60.05	64.50	64.33	71.75	56.22	60.13	65.65
BiLSTM+GL-CLEF	**84.77**	**74.44**	**71.09**	**69.53**	**65.29**	**66.14**	**77.02**	**63.36**	**67.08**	**70.97**
XLM-R (Conneau et al., 2020)	98.32	97.19	98.03	94.94	88.91	88.50	96.41	72.45	91.15	93.02
XLM-R+GL-CLEF	**98.66**	**98.43**	**98.04**	**97.85**	**93.84**	**88.83**	**97.76**	**81.68**	**91.38**	**94.05**
Slot F1	en	de	es	fr	hi	ja	pt	tr	zh	**AVG**
BiLSTM (Hochreiter and Schmidhuber, 1997)	75.43	15.81	34.97	33.38	5.83	4.98	43.89	9.51	27.51	27.92
BiLSTM+GL-CLEF	**87.45**	**38.40**	**46.06**	**46.16**	**20.28**	**29.53**	**59.67**	**37.25**	**42.48**	**45.25**
XLM-R (Conneau et al., 2020)	94.58	72.35	76.72	71.81	60.51	9.31	70.08	45.21	13.44	57.38
XLM-R+GL-CLEF	**95.88**	**84.91**	**82.47**	**80.99**	**61.11**	**55.57**	**77.27**	**54.55**	**80.50**	**74.81**
Overall Acc.	en	de	es	fr	hi	ja	pt	tr	zh	**AVG**
BiLSTM (Hochreiter and Schmidhuber, 1997)	37.06	0.78	3.08	0.63	0.22	0.00	10.20	0.00	0.03	5.80
BiLSTM+GL-CLEF	**61.37**	**4.60**	**9.10**	**4.30**	**0.34**	**2.03**	**16.82**	**2.80**	**2.46**	**11.53**
XLM-R (Conneau et al., 2020)	87.45	43.05	42.93	43.74	19.42	5.76	40.80	9.65	6.60	33.31
XLM-R+GL-CLEF	**88.24**	**64.91**	**53.51**	**58.28**	**19.49**	**13.77**	**52.35**	**14.55**	**52.07**	**46.35**

1. Contributions from contrastive learning and pre-trained model use are complementary.
2. GL-CLEF still obtains gains over BiLSTM.

- **Paper & Code**

Paper　　Code

Email: lbqin@ir.hit.edu.cn
Homepage: http://ir.hit.edu.cn/~lbqin/