IJCAI ECAI VIENNA 22

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

SCIR
Research Center for Social Computing and Information Retrieval
理解语言 认知社会

COLUMBIA UNIVERSITY

Spoken Language Understanding: Recent Advances and New Frontiers

Libo Qin
PhD Student
Harbin Institute of Technology

Dr. Wanxiang Che
Professor
Harbin Institute of Technology

Dr. Zhou Yu
Assistant Professor
Columbia University

Outline

- Introduction
- New taxonomy
- New Frontiers
- Conclusion and Highlight

Introduction

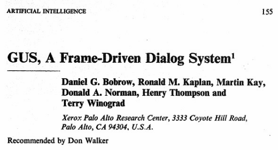
Dr. Zhou Yu
Assistant Professor
Columbia University

History of Dialogue System

1966

MIT Eliza

1977


ARTIFICIAL INTELLIGENCE 155

GUS, A Frame-Driven Dialog System[1]

Daniel G. Bobrow, Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson and Terry Winograd

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, U.S.A.

Recommended by Don Walker

GUS
Frame-Driven
Dialog System

1995

Alice
Rule-based
Dialog System

2010

IBM Watson

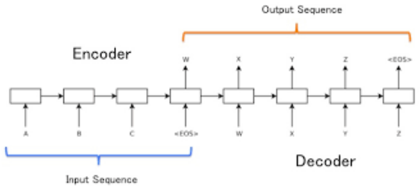
2011

Apple Siri

2014

Microsoft Xiaoice

2015



Seq2Seq Dialog System

2016

Google Assistant
Amazon Alexa

2016-2018

- Alibaba: ALIME
- Baidu: Duer
- Xiami: Xiaoai
- Tmall Bot
- JD.COM: DingDong

2020

- Google: Meena
- FAIR: Blender
- OpenAI: GPT-3
- MSR& Columbia: GODEL

http://coai.cs.tsinghua.edu.cn/hml/media/files/dialog.pdf

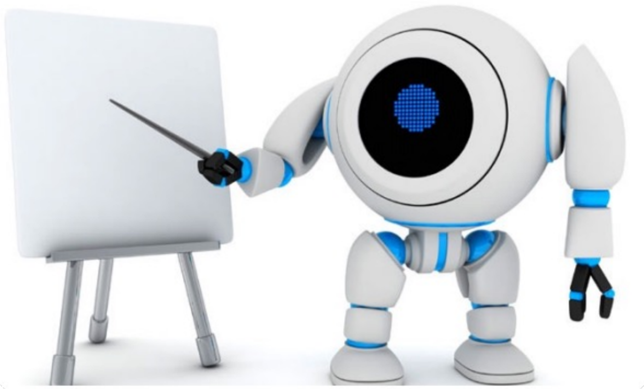
Application of Task-oriented Dialogue

Smart Home

In-car Assistant

Medical Diagnosis

Emotional Robot



Intelligent Teacher

Travel Service Robot

Pictures are selected from the Internet

Architecture of Task-oriented Dialogue

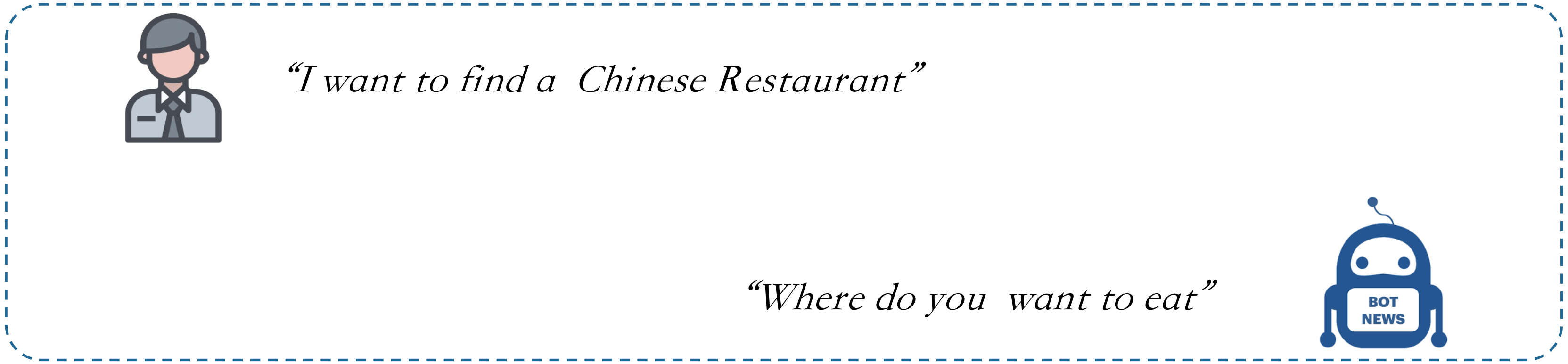


Figure: Example of Dialogue

How does a task-oriented dialogue conduct the communication?

Architecture of Task-oriented Dialogue

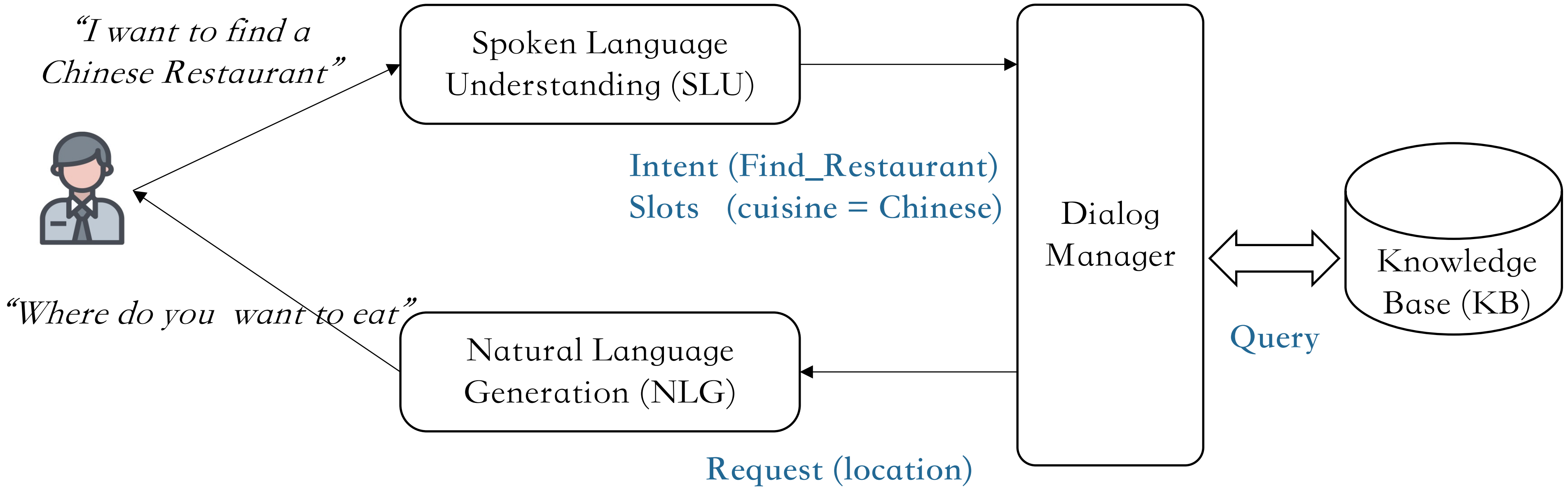
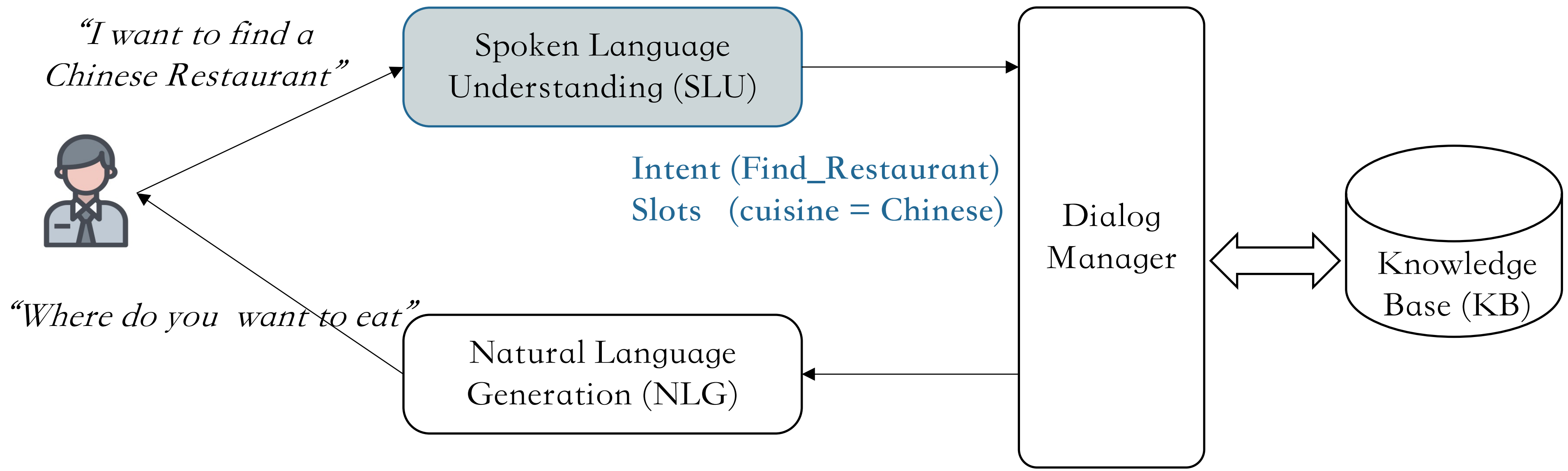


Figure: Pipeline Task-oriented Dialogue System

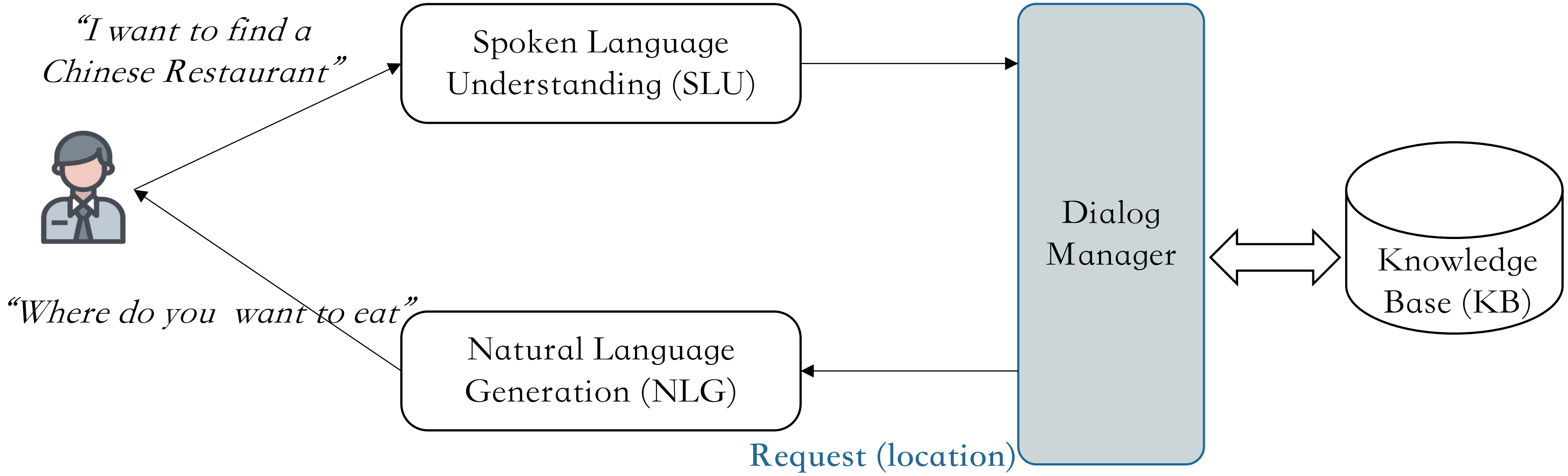
Architecture of Task-oriented Dialogue

- Spoken Language Understanding (SLU)
    - Input: user utterance
    - Output: semantic frame (intents&slots)



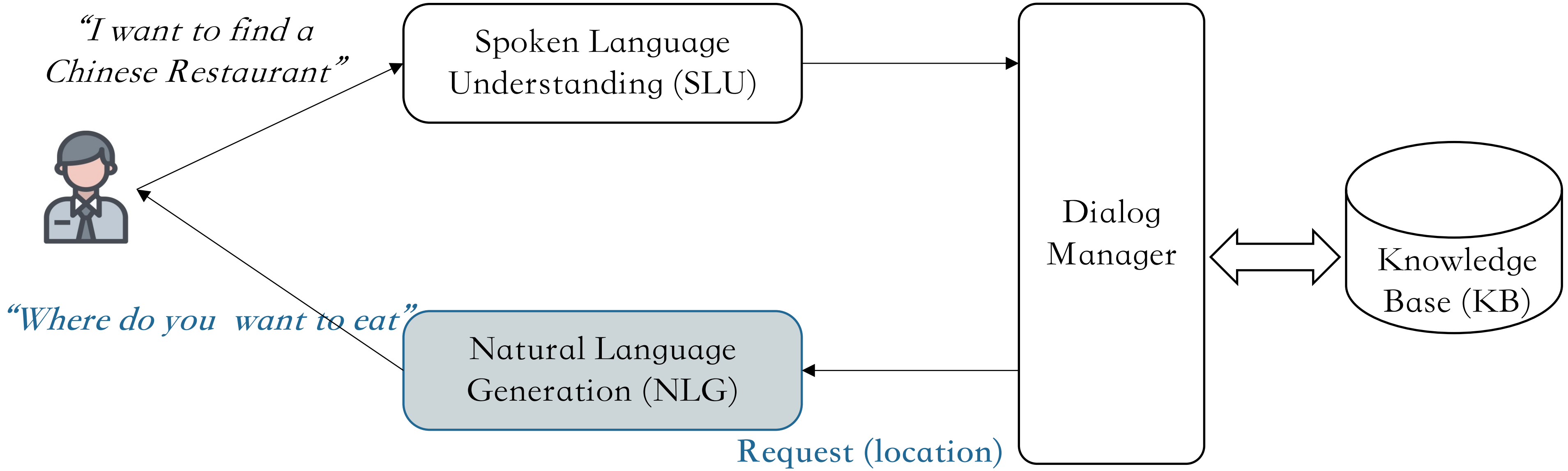
Architecture of Task-oriented Dialogue

- Dialog Manager
  - Dialogue state tracking
    - Input: a dialogue / a turn with its previous state
    - Output: dialogue state at the current turn (e.g. slot-value pairs)
  - Dialogue Policy
    - Input: dialogue state + KB results
    - Output: system action (speech-act + slot-value pairs)



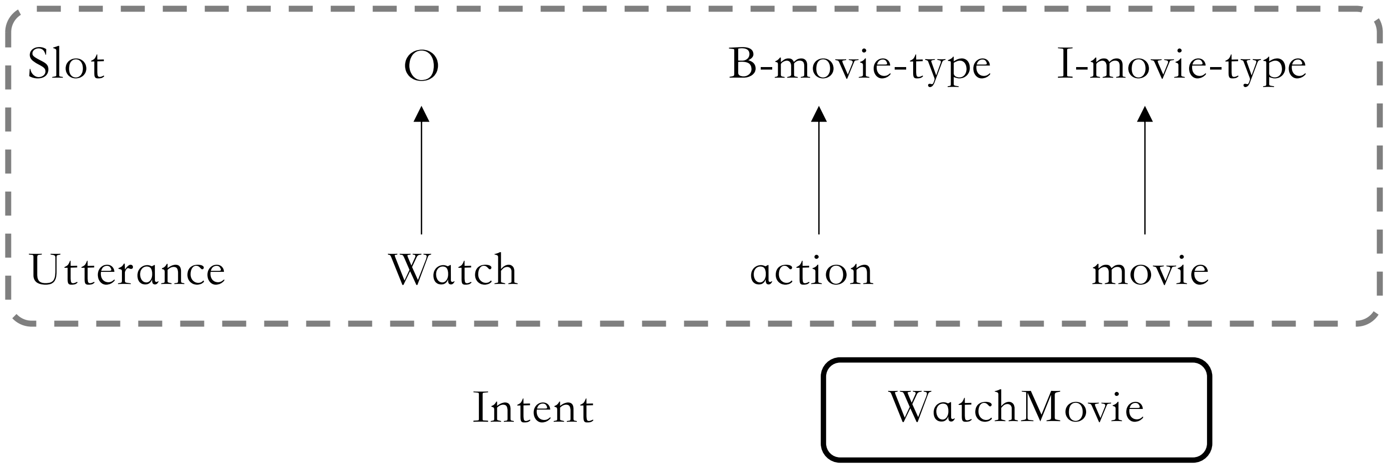
Architecture of Task-oriented Dialogue

- Natural Language Generation (NLG)
  - Input: system action (speech-act + slot-value pairs)
  - Output: natural language response



Introduction

- SLU
  - Slot filling -> sequence labeling
    - Input: $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$
    - Output: $\mathbf{S} = \{s_1, s_2, \ldots, s_n\}$
  - Intent detection -> classification task
    - Input: $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$
    - Output: I

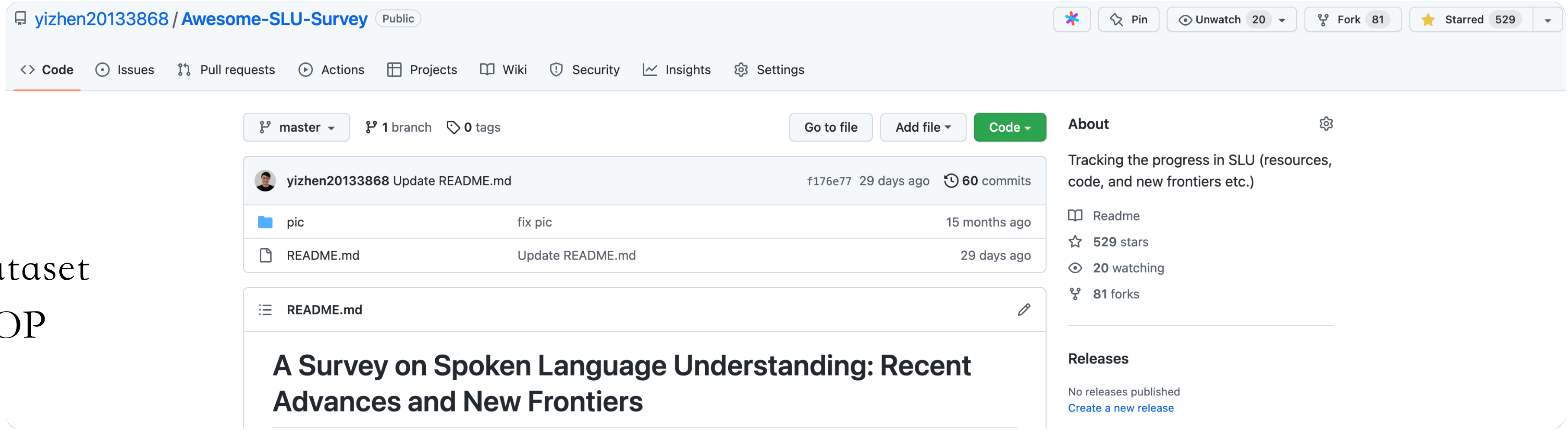


Evaluation Metrics

- F1 scores for slot filling
  - F1 scores are adopted to evaluate the performance of slot filling, which is the harmonic mean score between precision and recall
- Intent accuracy for intent detection
  - Intent Accuracy is used for evaluating the performance of intent detection, calculating the ratio of sentences for which intent is predicted correctly
- Overall accuracy for SLU
  - Overall accuracy is adopted for calculating the ratio of sentences for which both intent and slot are predicted correctly in a sentence

Resources

- Dataset
  - ATIS
  - SNIPS
  - TOP semantic parsing
  - Simulated Dialogues dataset
  - MTOP: Multilingual TOP
  - …



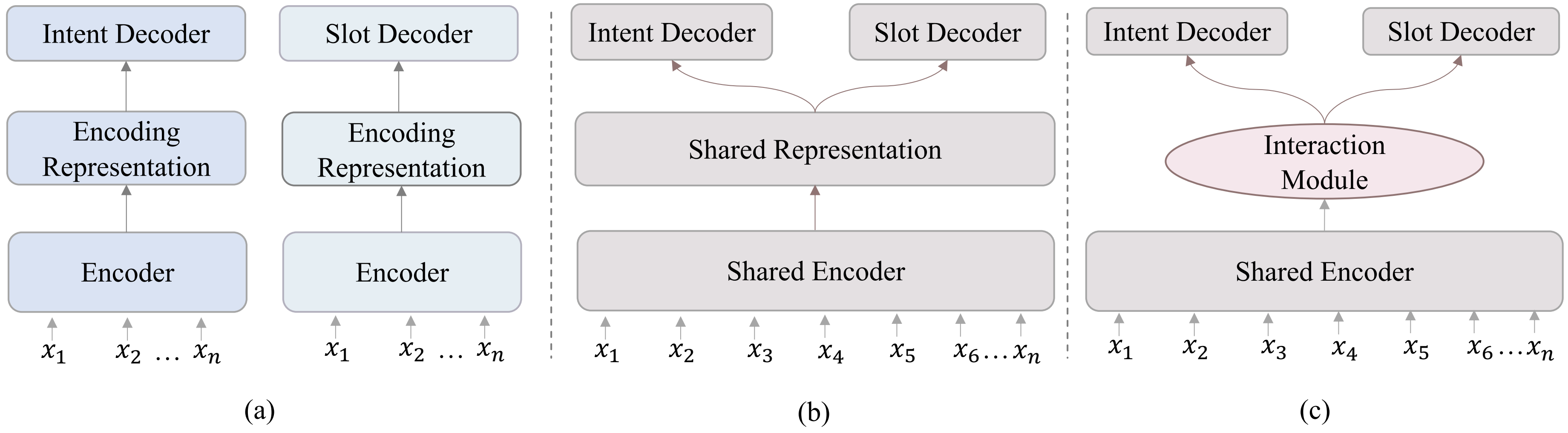
- Surveys
  - Louvan et al. Recent Neural Methods on Slot Filling and Intent Classification for Task-Oriented Dialogue Systems: A Survey. COLING2020.
  - Qin et al. A Survey on Spoken Language Understanding: Recent Advances and New Frontiers. IJCAI2021 Survey.

SLU Models

Dr. Wanxiang Che
Professor
Harbin Institute of Technology

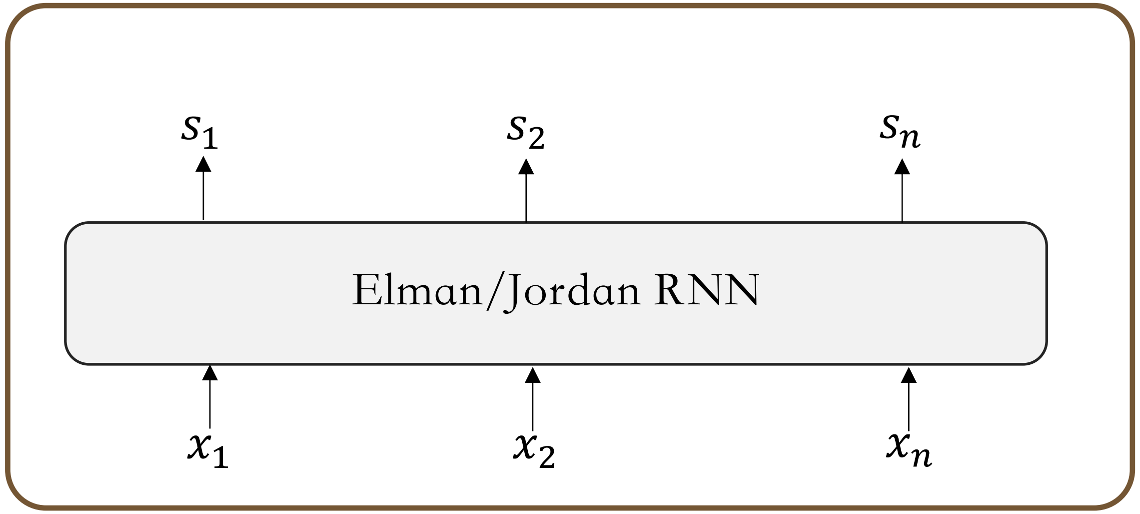
SLU Models

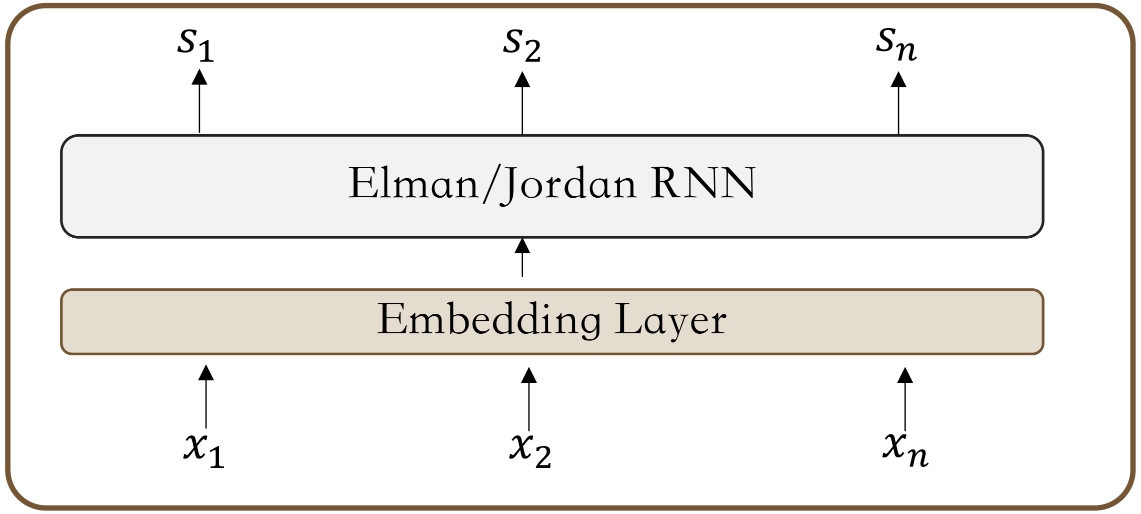
- Single Model
  - (a) Slot Filling and Intent Detection model
- Joint Model
  - (b) Implicit Joint Model and (c) Explicit Joint Model



Single Model - Slot Filling

- Compare several RNN architectures, including the Elman-type and Jordan-type networks with their variants, for slot filling
- Explore the effectiveness of word embeddings for slot filling (word embedding boost the slot filling performance)



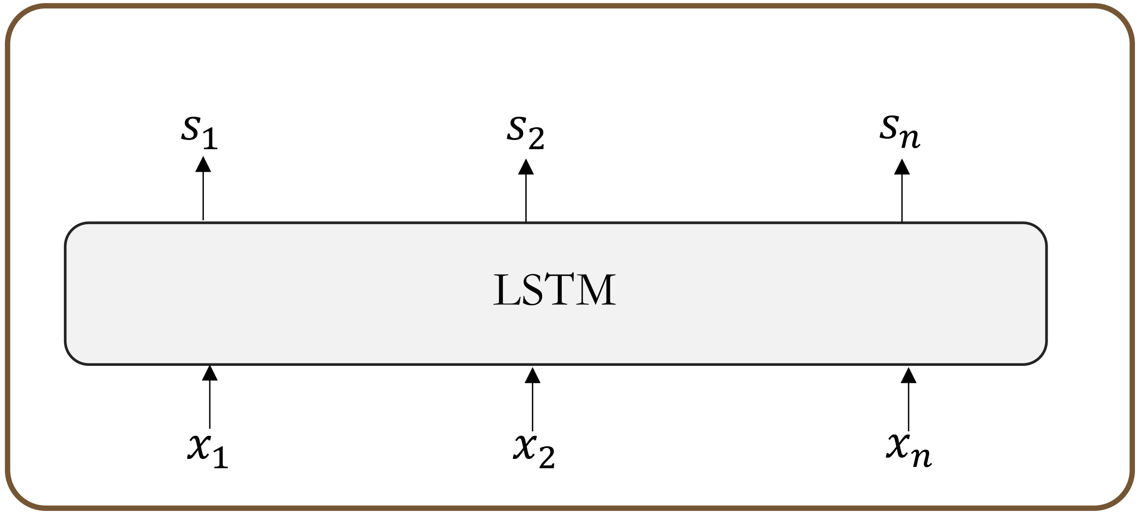


Grégoire et al. Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding. Interspeech 2013.

Single Model - Slot Filling

- An application of LSTMs to spoken language understanding.

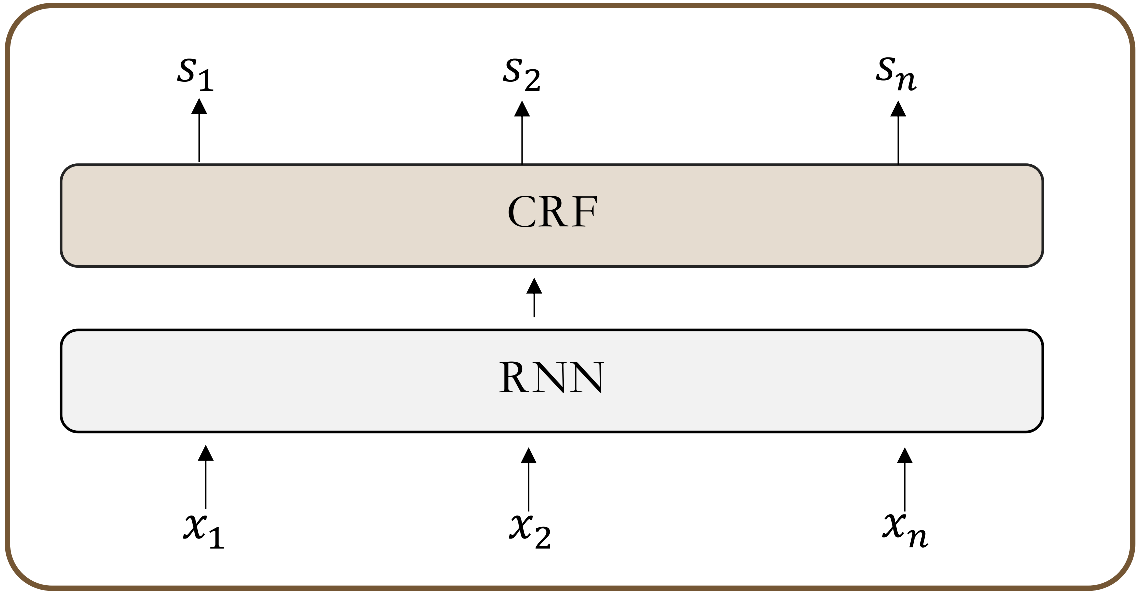
- The LSTMs achieves state-of-the-art results on the ATIS data



Yao et al. SPOKEN LANGUAGE UNDERSTANDING USING LONG SHORT-TERM MEMORY NEURAL NETWORKS. SLT 2014.

Single Model - Slot Filling

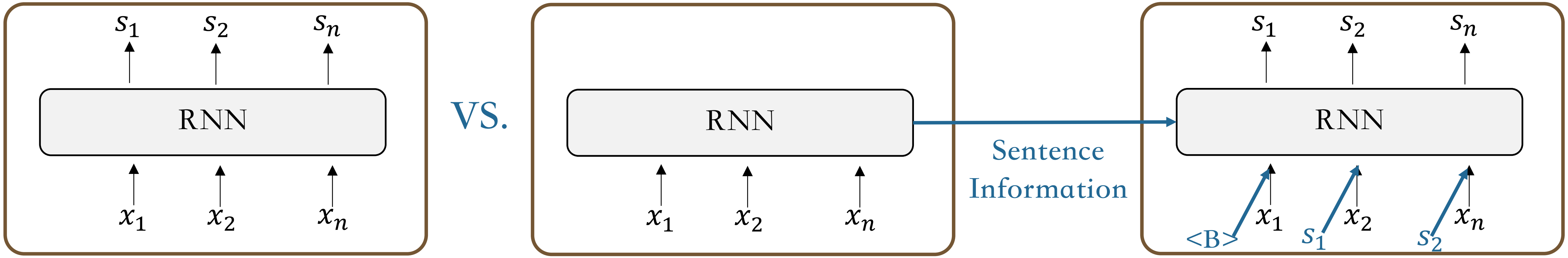
- Apply CRF for RNN-based model
- Demonstrate that RNNs can be successfully merged with CRFs to do language understanding



Yao et al. RECURRENT CONDITIONAL RANDOM FIELD FOR LANGUAGE UNDERSTANDING. ICASSP 2014.

Single Model - Slot Filling

- An encoder-labeler LSTM that can conduct slot filling conditioned on the encoded sentence-level information
- Explicitly model label dependencies

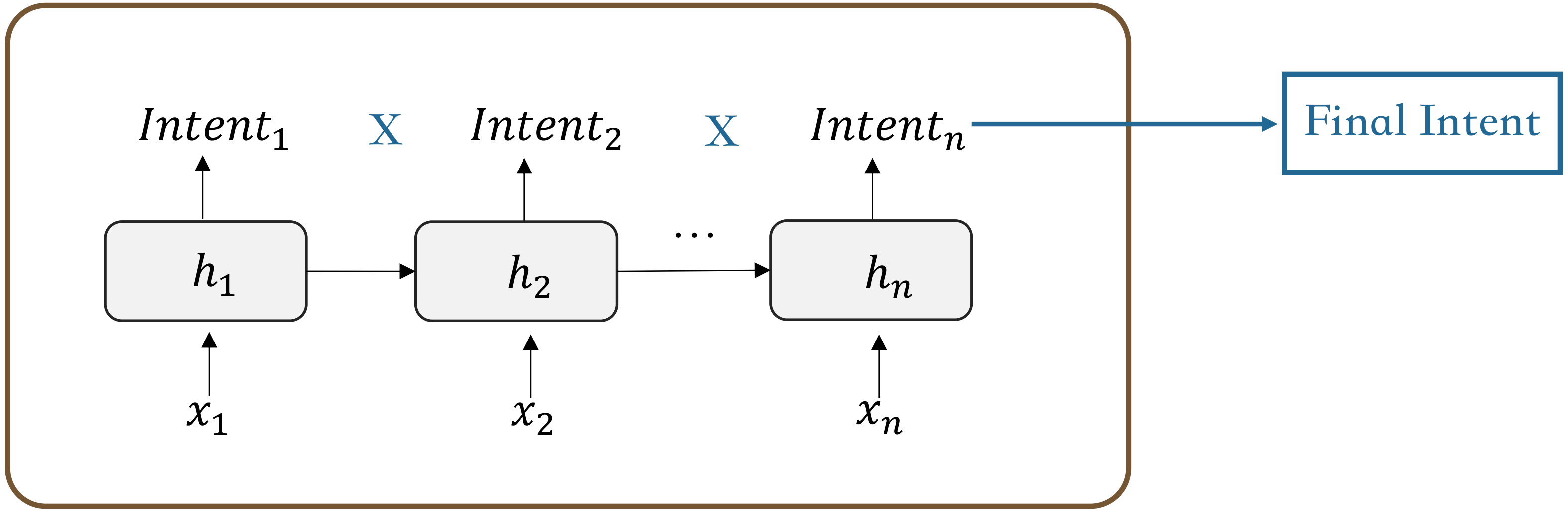


Gakuto et al. Leveraging Sentence-level Information with Encoder LSTM for Semantic Slot Filling. EMNLP 2016.

Single Model - Intent Detection

- RNN for intent detection

- The utterance intent is the product of the probability of the intent of each token

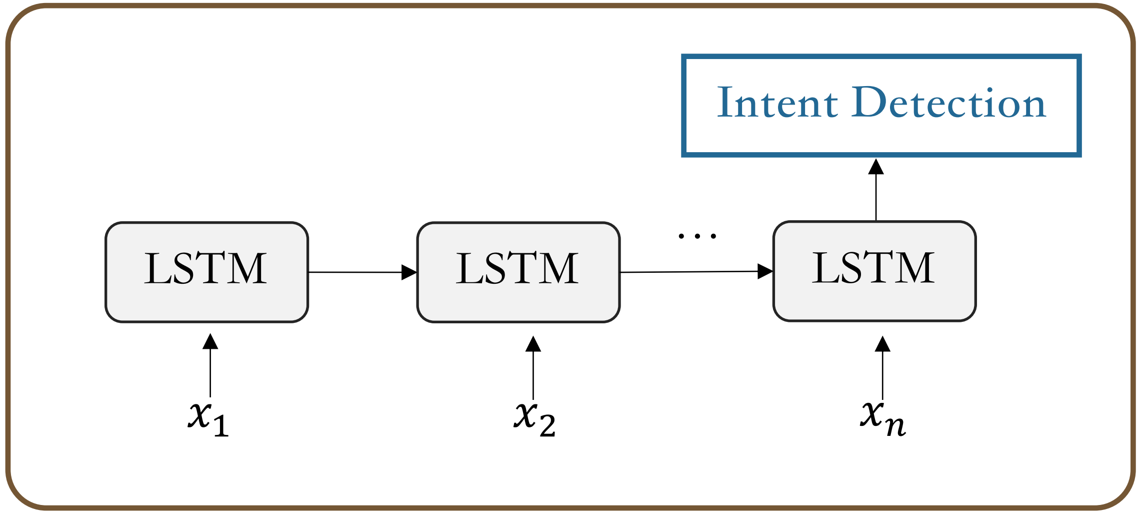


Suman et al. Recurrent neural network and LSTM models for lexical utterance classification. SLT 2014.

Single Model - Intent Detection

- LSTM for intent detection (leverage the final state for intent detection)

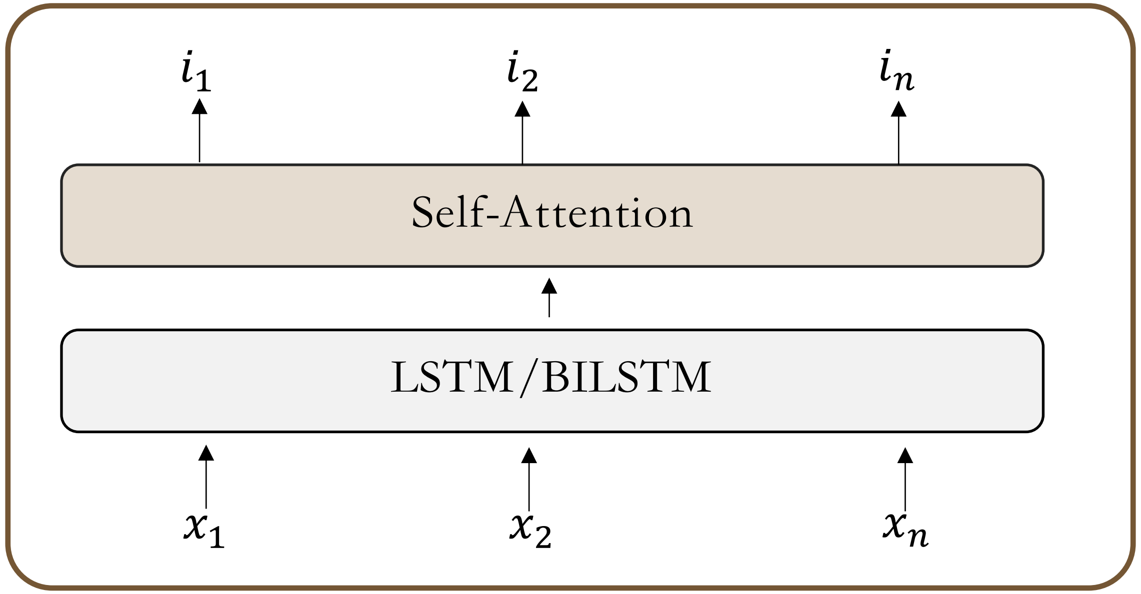
- Obtain better performance than RNN



Suman et al. Recurrent neural network and LSTM models for lexical utterance classification. SLT 2014.

Single Model - Intent Detection

- Self-attention network for intent detection

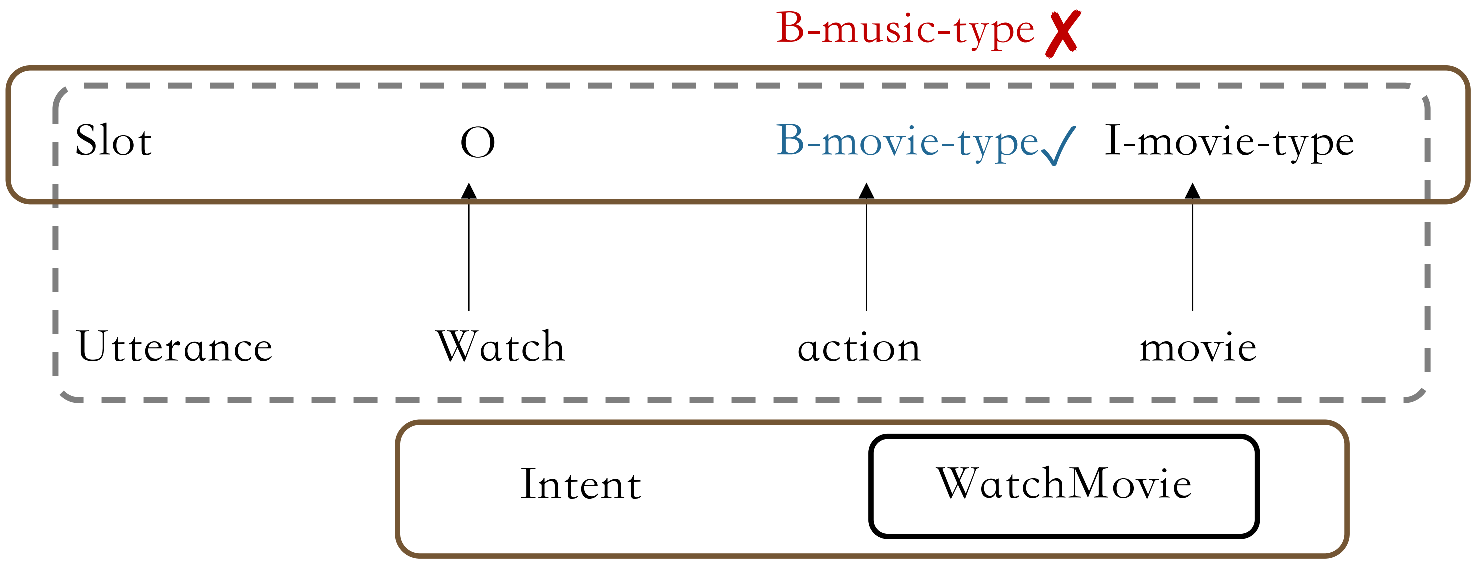


$$\mathbf{C} = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}$$

Yolchuyeva et al. Self-Attention Networks for Intent Detection. RANLP 2019.

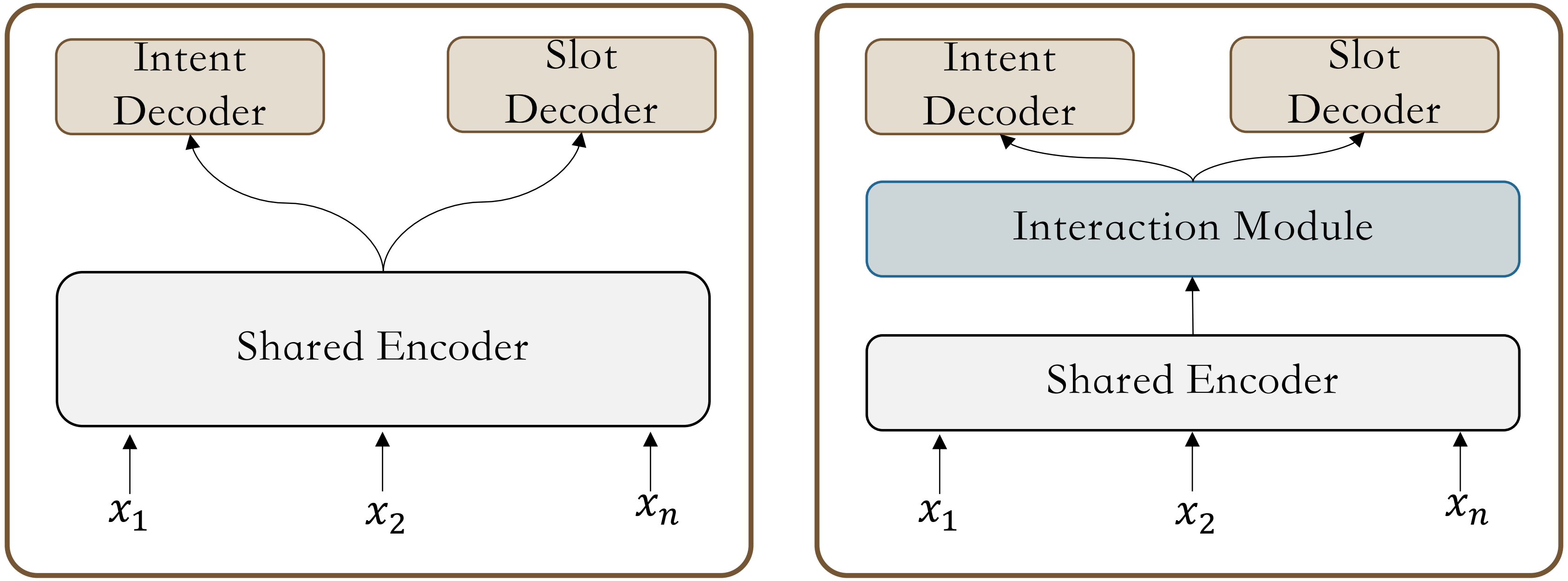
Joint Model

- Intent detection and slot filling is highly correlated.



Joint Model

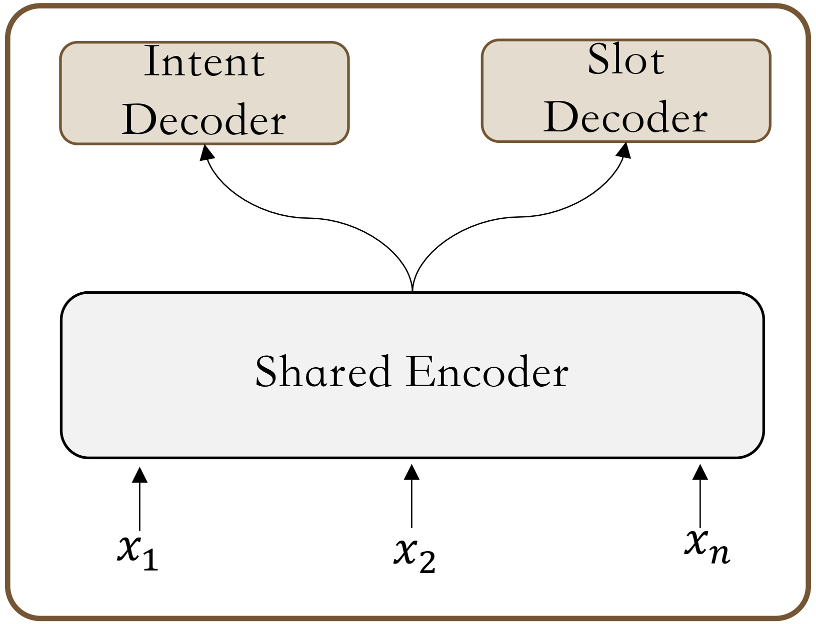
- Implicit Joint Model
- Explicit Joint Model



(a) Implicit Joint Model (b) Explicit Joint Model

Joint Model-Implicit Joint Modeling

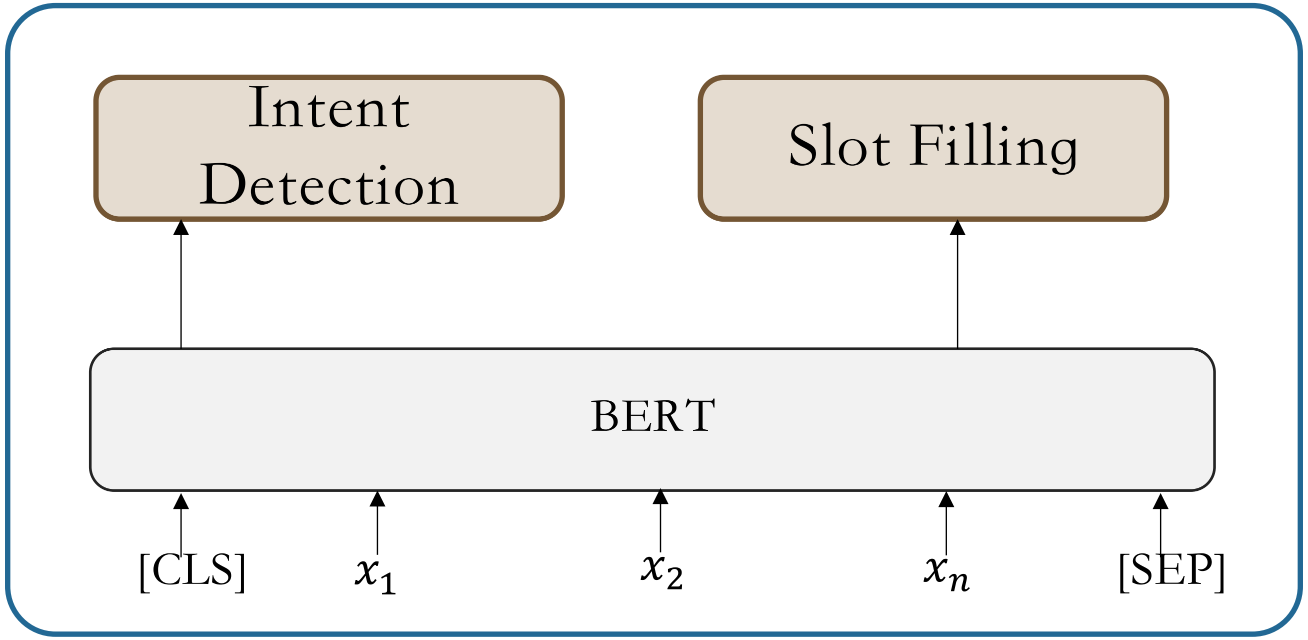
- A joint model where a shared encoder and two separate decoders for slot filling and intent detection



Zhang et al. A Joint Model of Intent Determination and Slot Filling for Spoken Language Understanding. IJCAI 2016.

Joint Training with Pre-trained Model

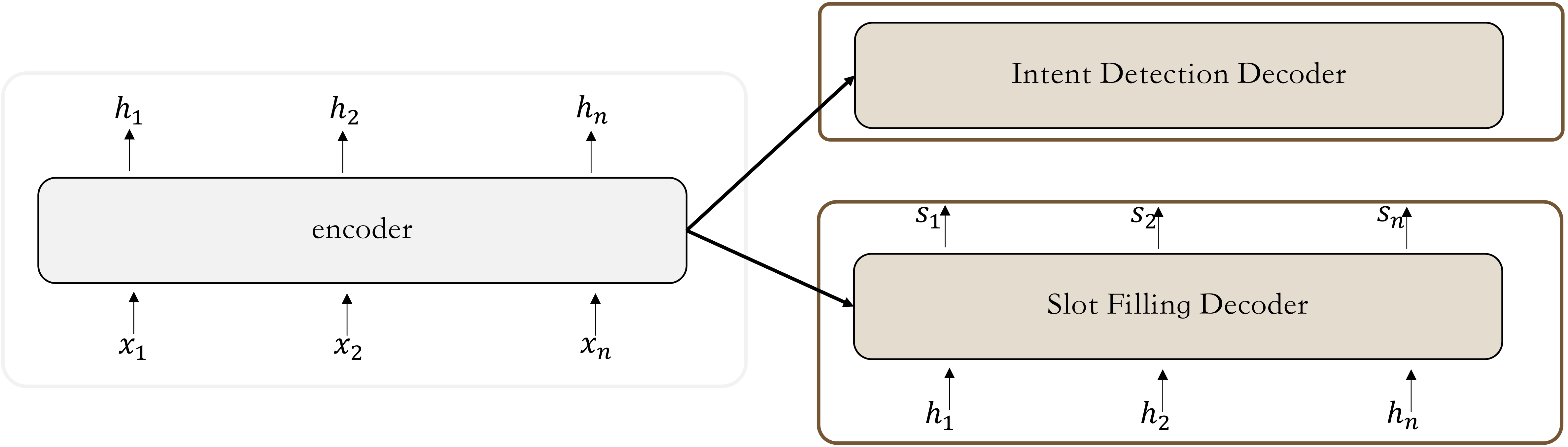
- The first joint work based on BERT (obtains 6%~17% improvement with non pre-trained model)



Chen et al. BERT for Joint Intent Classification and Slot Filling. Arxiv 2019.

Joint Model-Implicit Joint Modeling

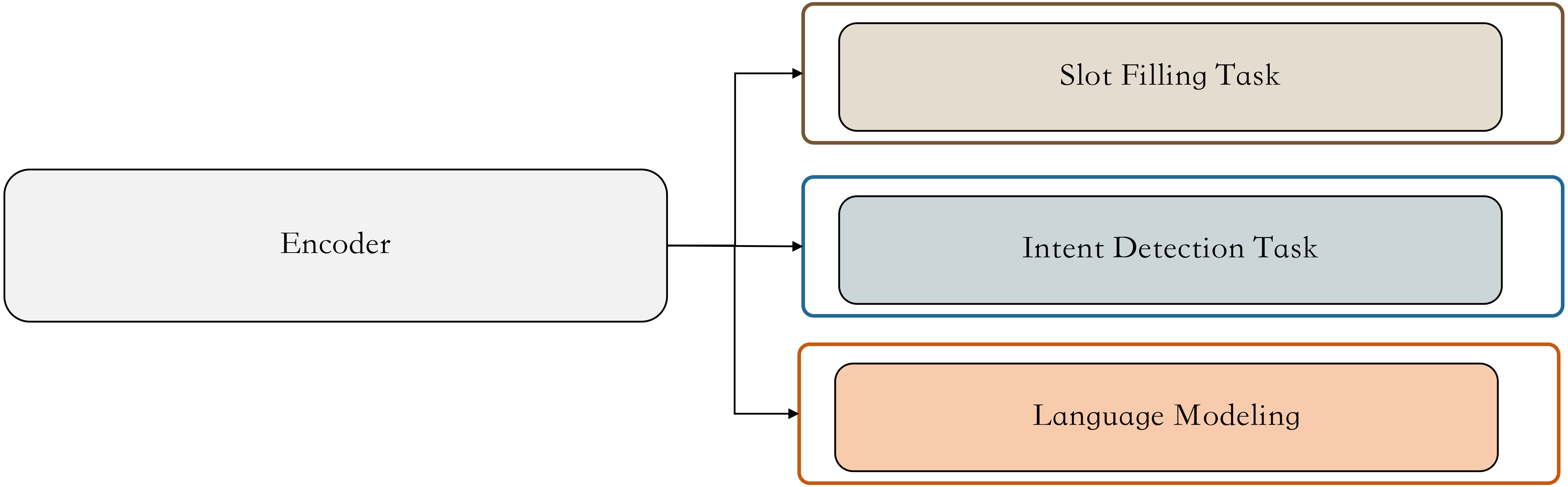
- An encoder-decoder neural network for slot filling and intent detection



Liu et al. Attention-based recurrent neural network models for joint intent detection and slot filling. Interspeech 2016.

Joint Model-Implicit Joint Modeling

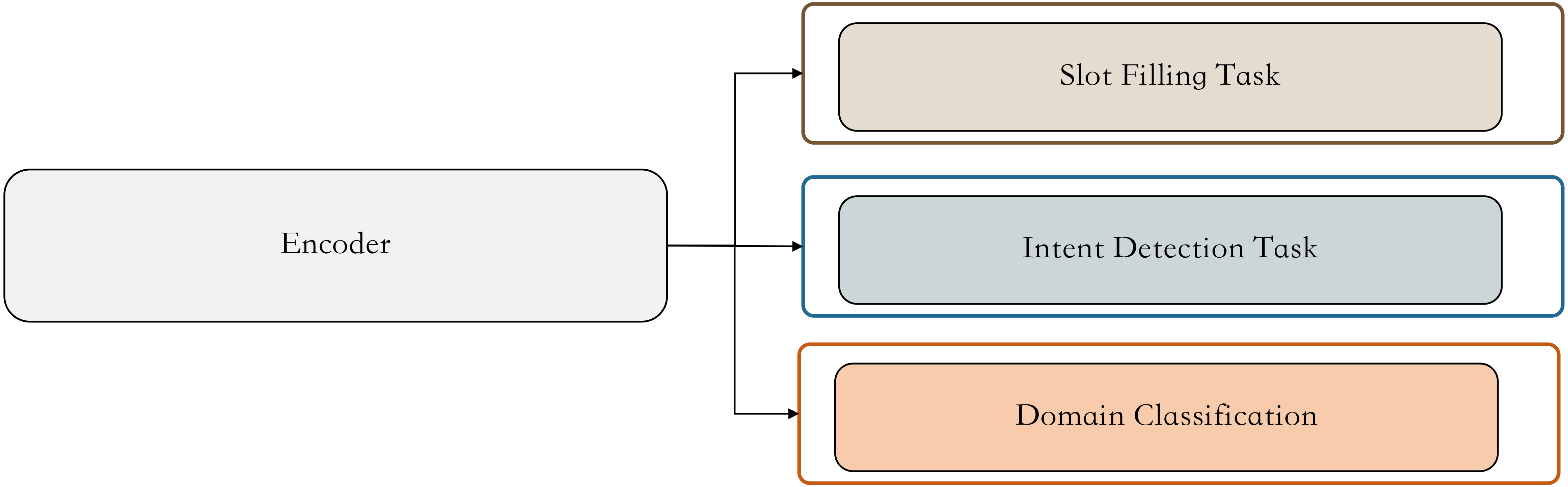
- A conditional RNN model that can be used to jointly perform online spoken language understanding and language modeling
- SLU can benefit Language Modeling task



Liu et al. Joint Online Spoken Language Understanding and Language Modeling with Recurrent Neural Networks. SIGDIAL 2016.

Joint Model-Implicit Joint Modeling

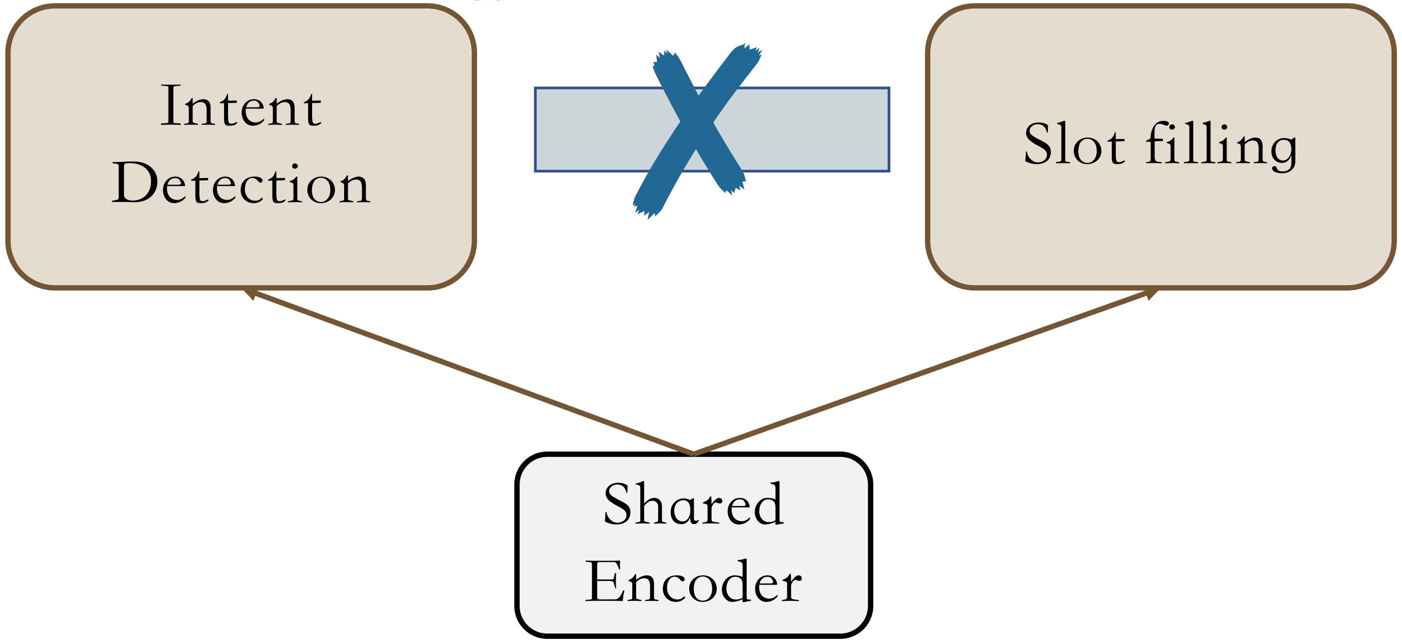
- A multi-domain, multi-task framework to jointly model the domain classification, intent detection and slot filling simultaneously



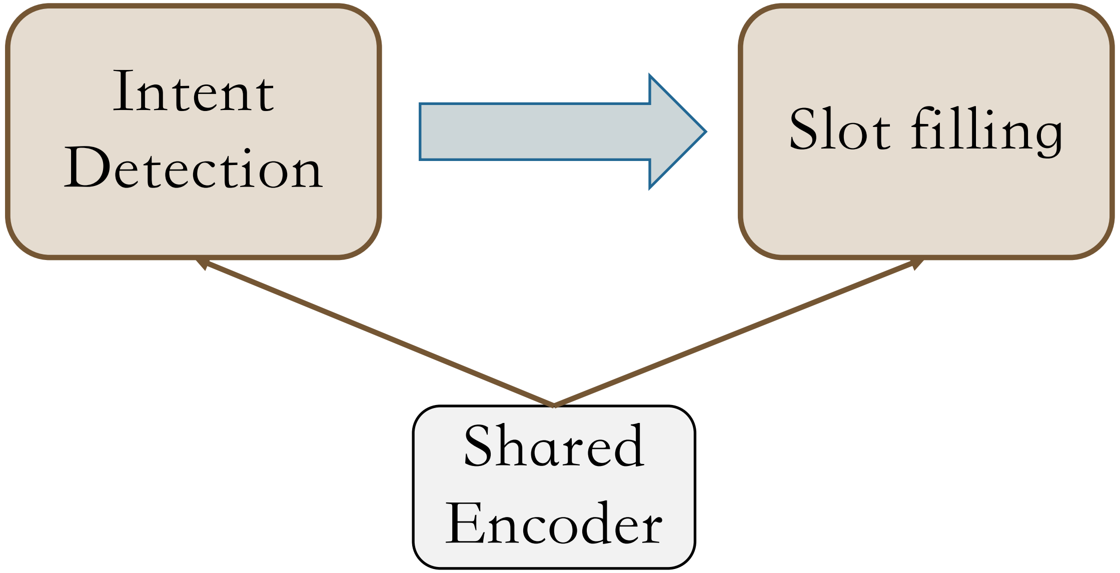
Hakkani-Tur et al. Multi-Domain Joint Semantic Frame Parsing using Bi-directional RNN-LSTM. Interspeech 2016.

Drawback-Implicit Joint Modeling

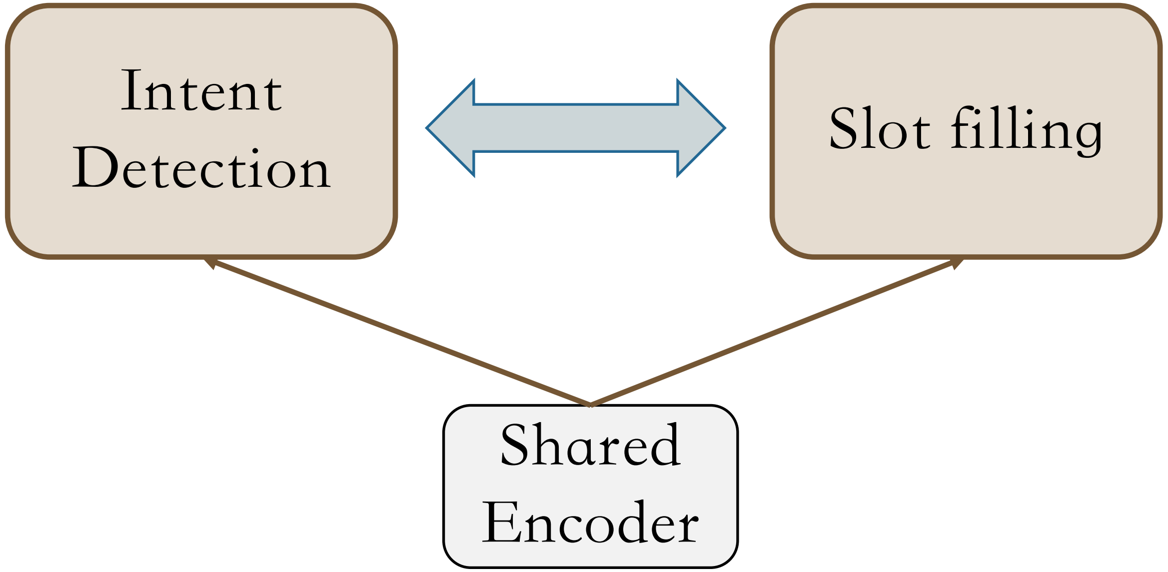
Rely on a set of shared parameters



Explicit Joint Modeling

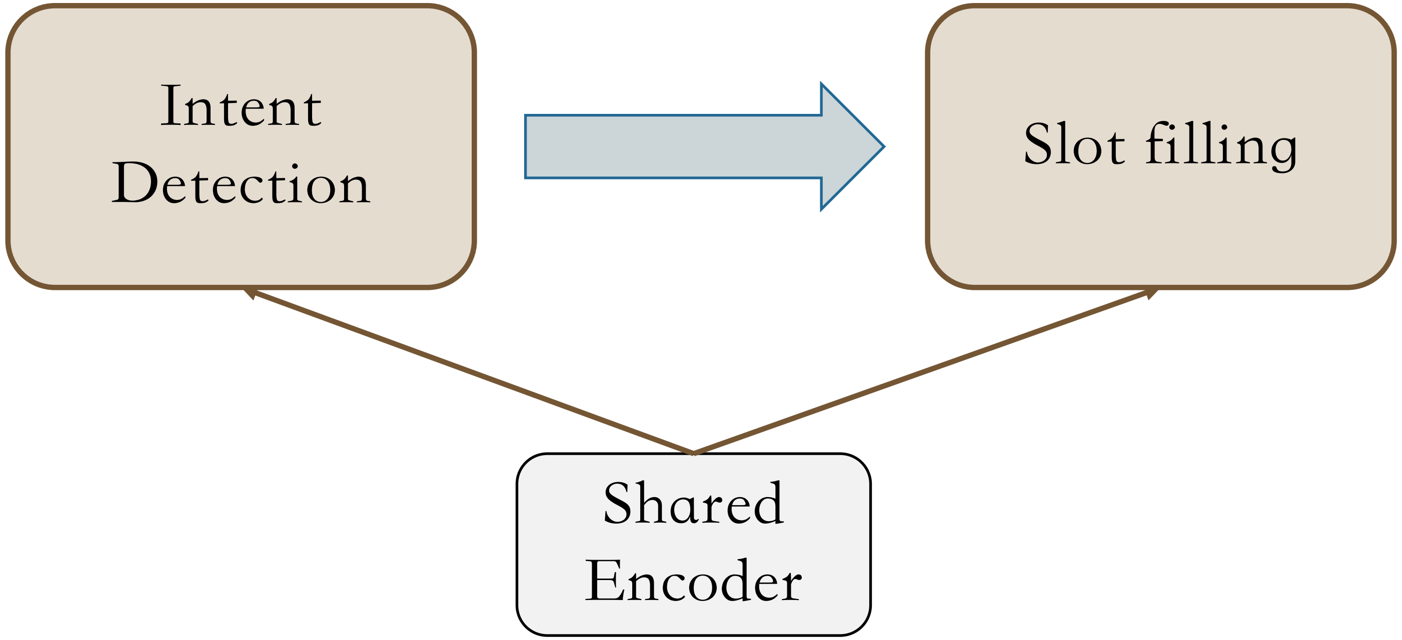


(a) Single Flow Interaction



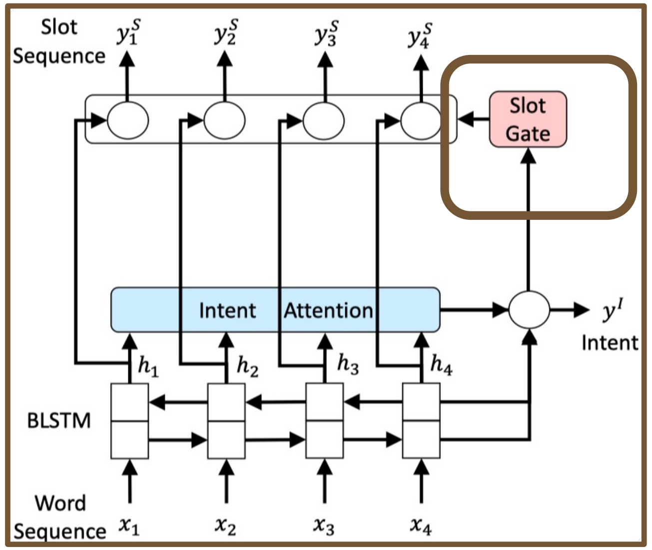
(b) Bidirectional Flow Interaction

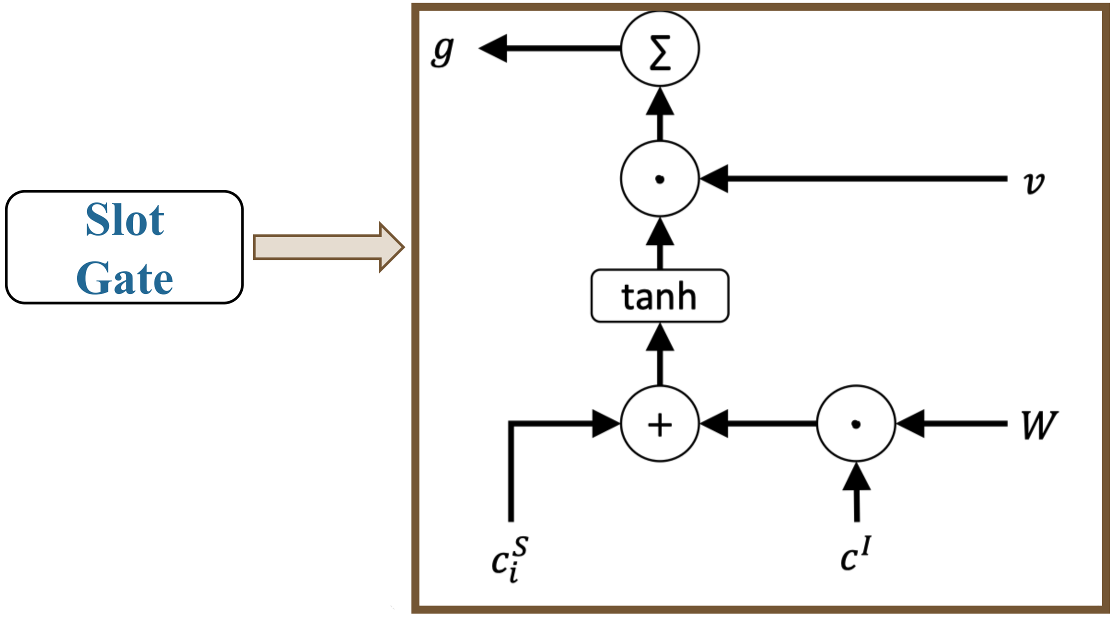
Joint Model - Single Flow Interaction



Joint Model - Single Flow Interaction

- Slot-gated mechanism to allow the slot filling can be conditioned on the intent information
- Information flow from intent to slot



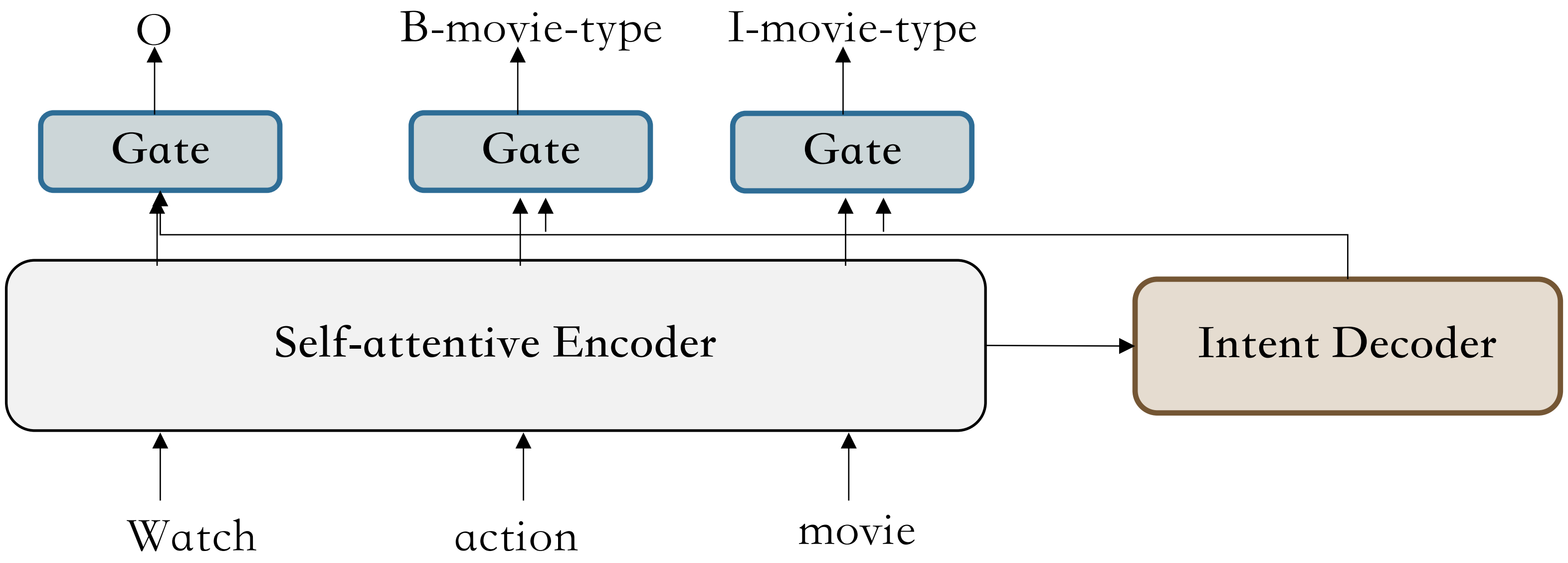


G will be larger if slot and intent are better related

Goo et al. Slot-Gated Modeling for Joint Slot Filling and Intent Prediction. NAACL 2018.

Joint Model - Single Flow Interaction

- A novel self-attentive model with gate mechanism to inject intent information for slot filling
- Information flow from intent to slot

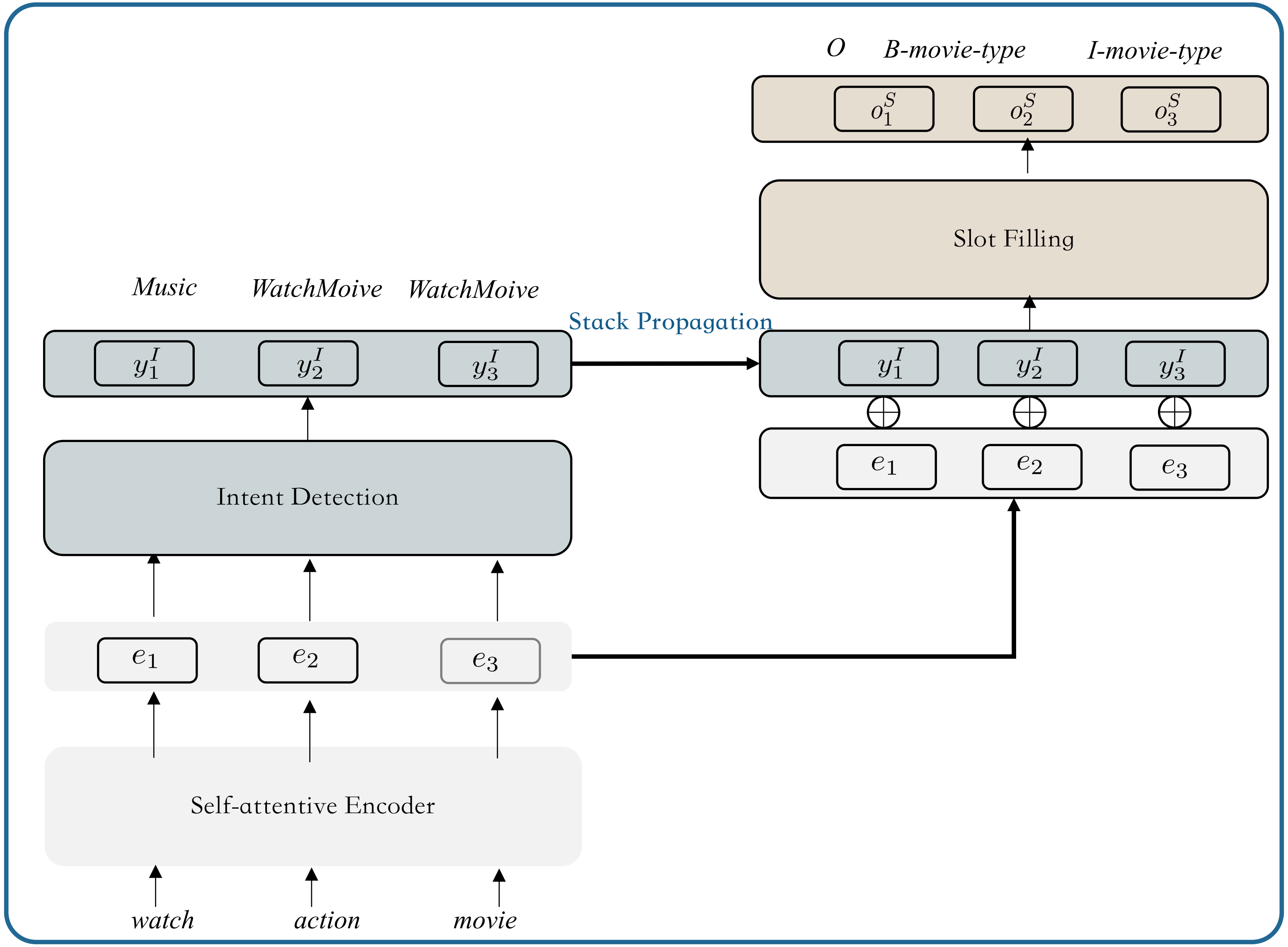


Li et al. A Self-Attentive Model with Gate Mechanism for Spoken Language Understanding. EMNLP 2018.

Joint Model - Single Flow Interaction

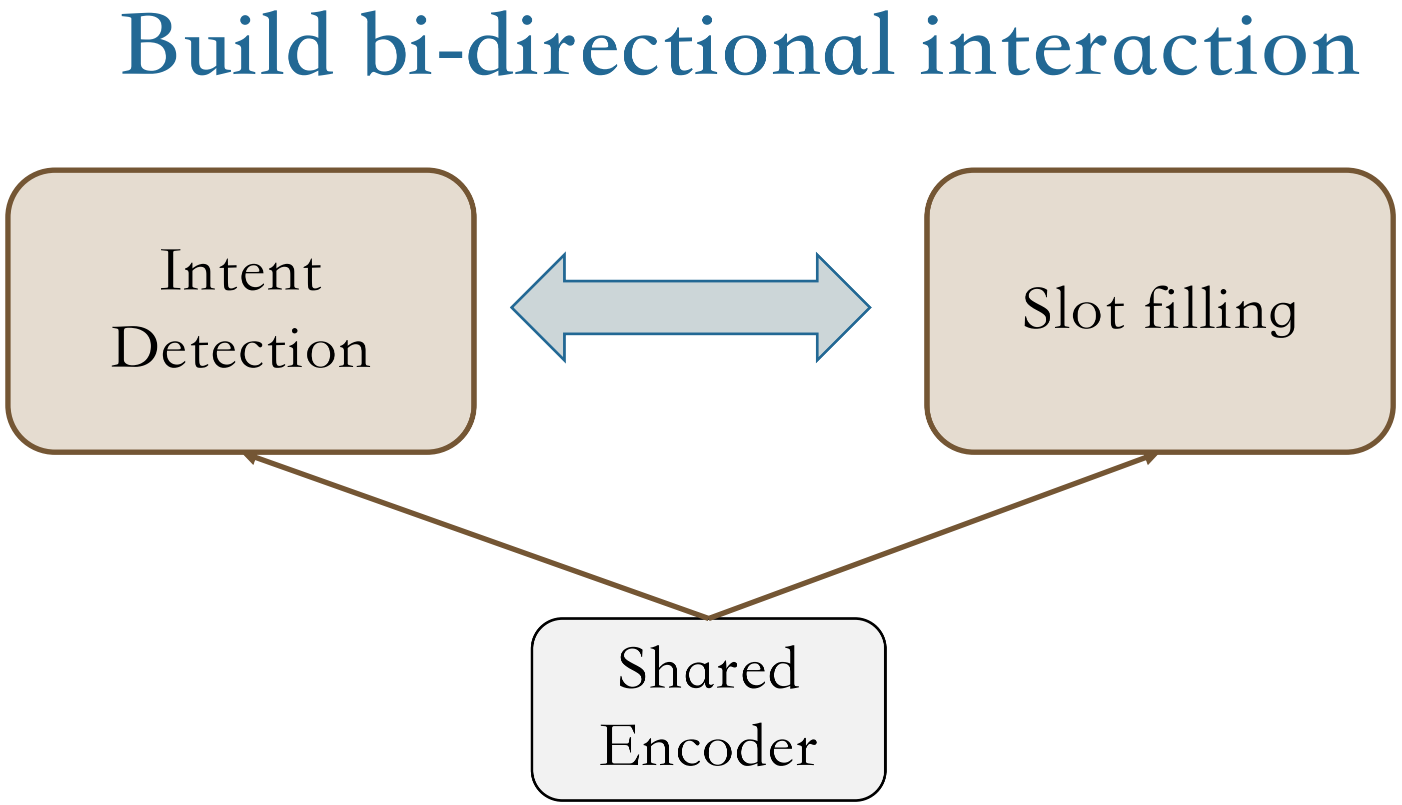
- Directly leverage the **intent output result** for slot filling

- Information flow from **intent to slot**



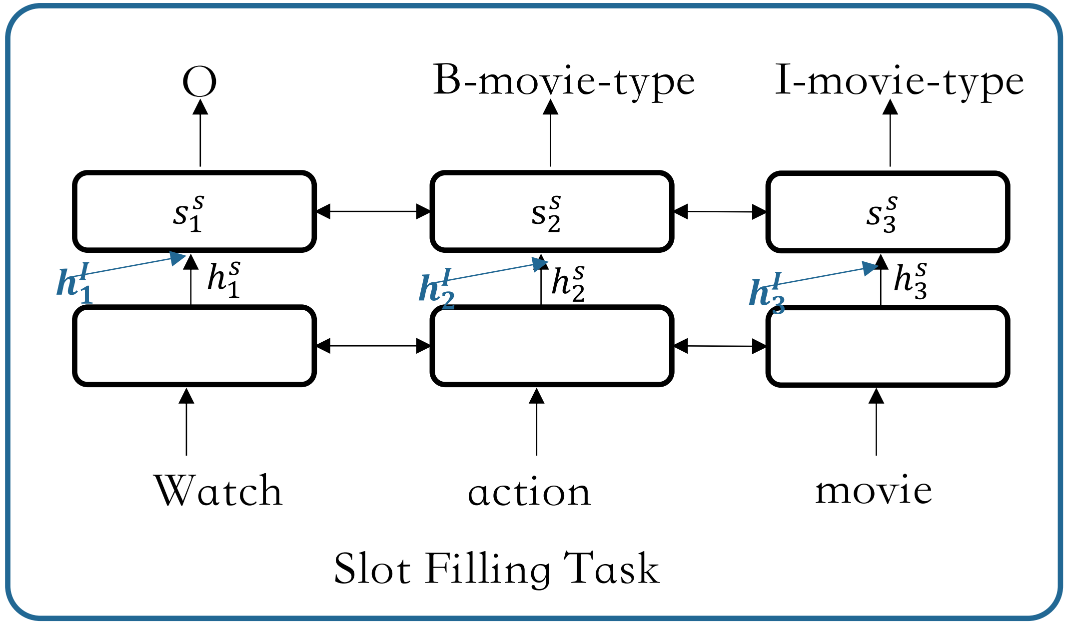
Qin et al. A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding. EMNLP 2019.

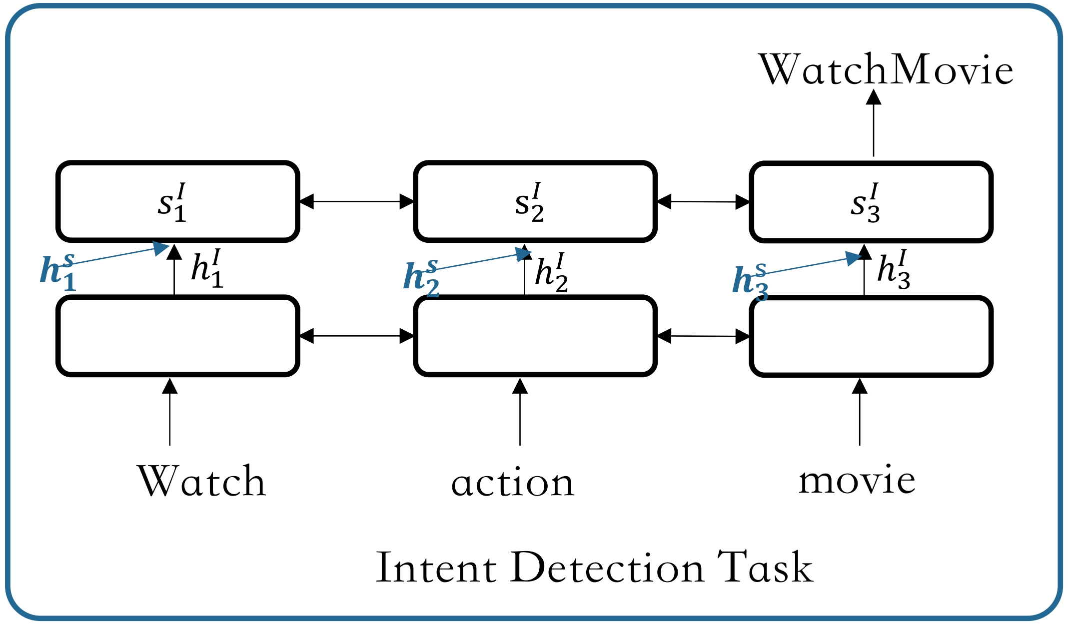
Joint Model - Bi-directional Flow Interaction



Joint Model - Bi-directional Flow Interaction

- A Bi-model based RNN structures to model cross-impact between two tasks
- Information flow from both intent to slot and slot to intent

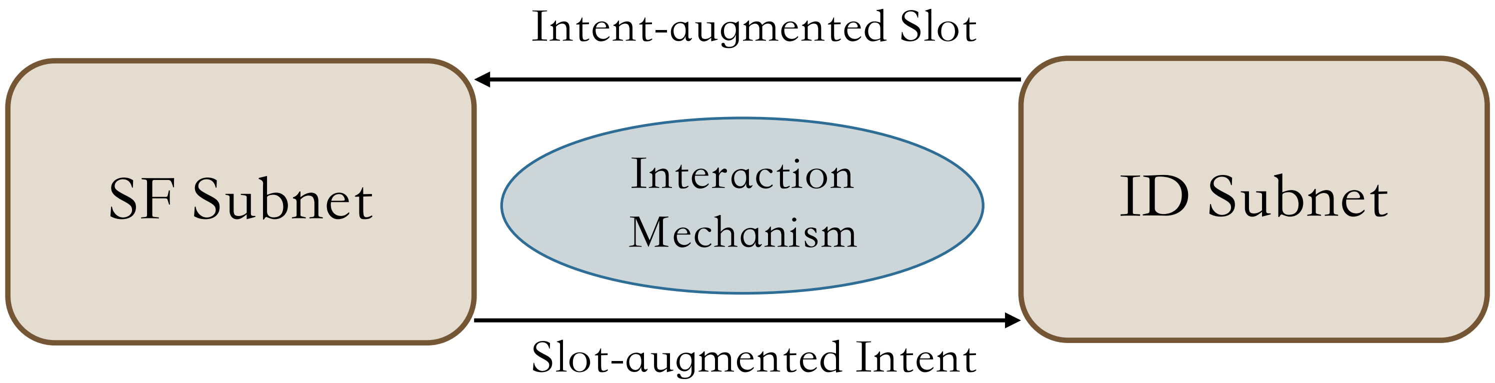




Wang et al. A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling. NAACL 2018.

Joint Model - Bi-directional Flow Interaction

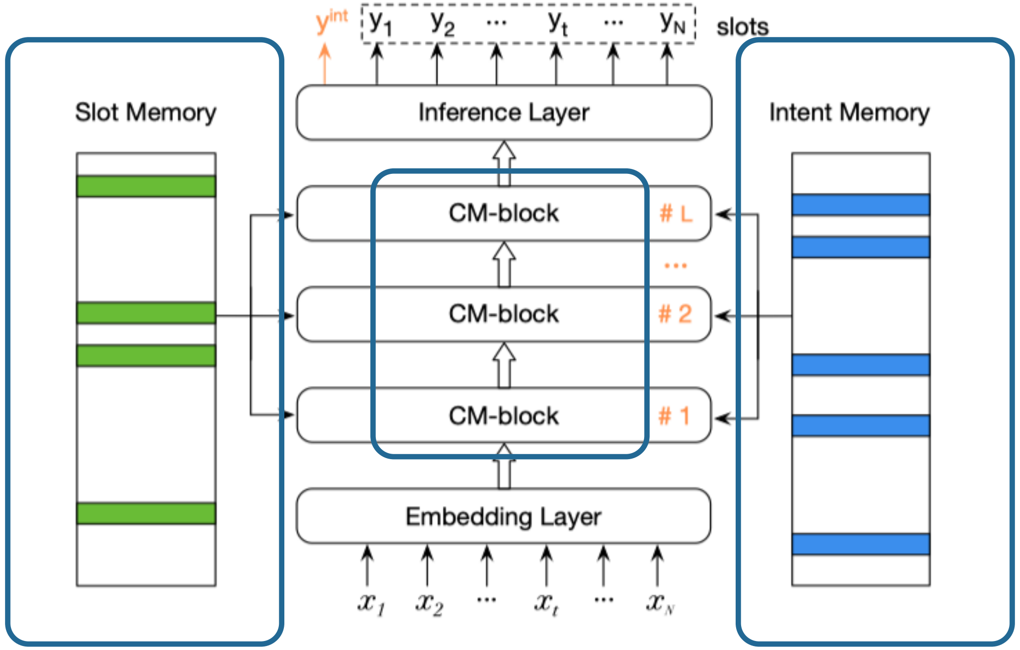
- A bi-directional extended version of gated mechanism [1] (slot-gate), including **SF Subnet and ID Subnet**
- Information flow from both intent to slot and slot to intent



[1] Goo et al. Slot-Gated Modeling for Joint Slot Filling and Intent Prediction. NAACL 2018.
[2] E et al. A Novel Bi-directional Interrelated Model for Joint Intent Detection and Slot Filling. ACL 2019.

Joint Model - Bi-directional Flow Interaction

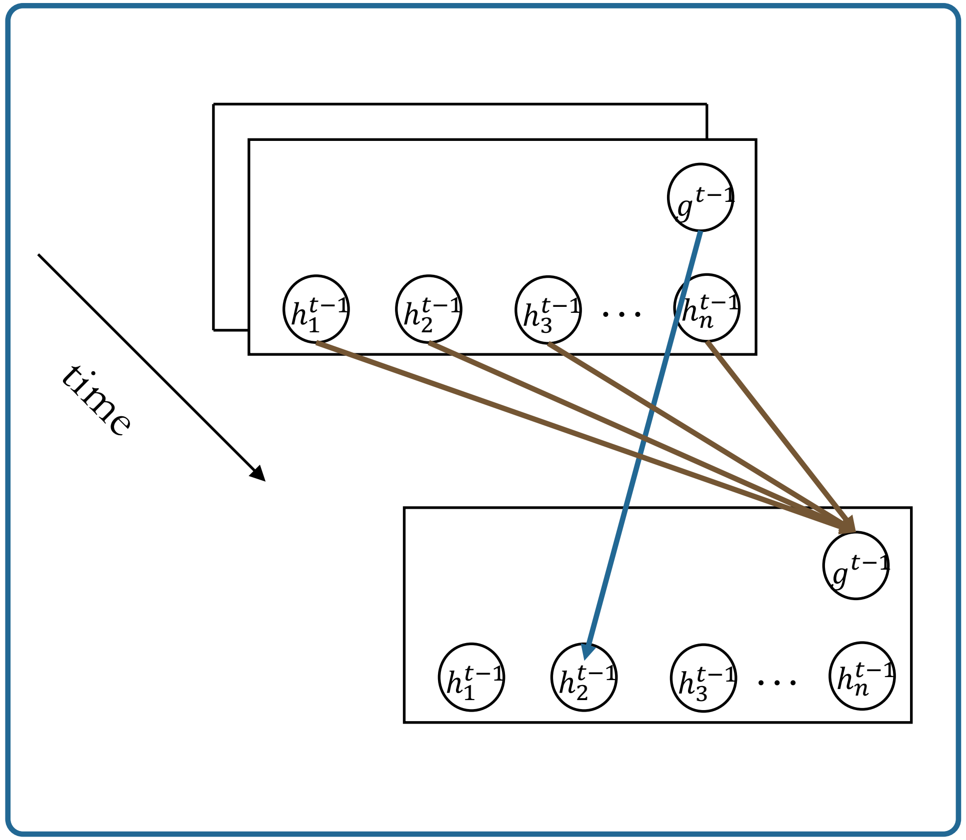
- Slot memory as well as intent memory and utilize CM-block to achieve the bi-directional of information flow
- Information flow from both intent to slot and slot to intent



Joint Model - Bi-directional Flow Interaction

- The first work to explore graph LSTM to model the semantic correlation between intent and slot

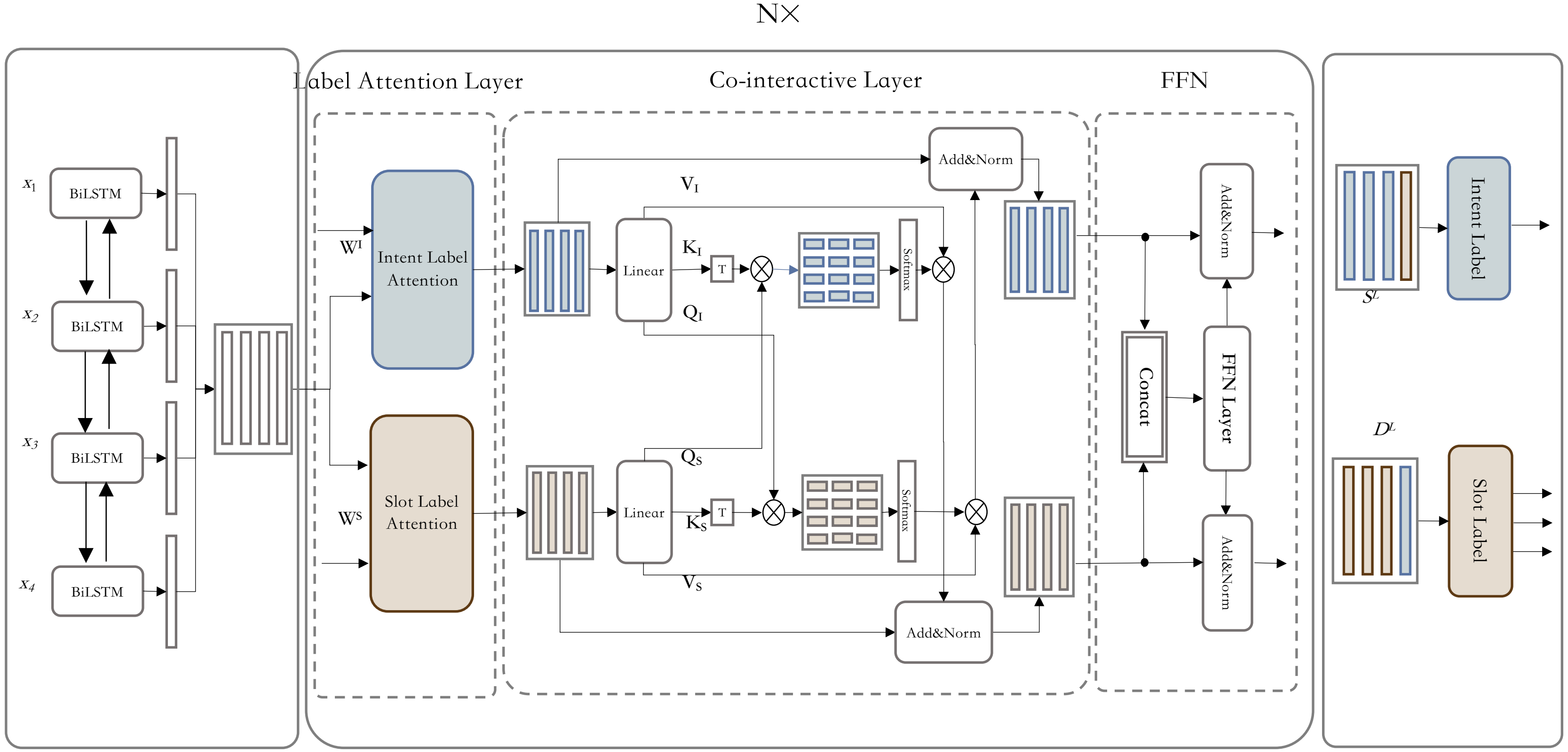
- Information flow from both intent to slot and slot to intent



Zhang et al. Graph LSTM with Context-Gated Mechanism for Spoken Language Understanding. AAAI 2020.

Joint Model - Bi-directional Flow Interaction

- Propose a co-interactive transformer for jointly modeling slot filling and intent detection
- Information flow from both intent to slot and slot to intent



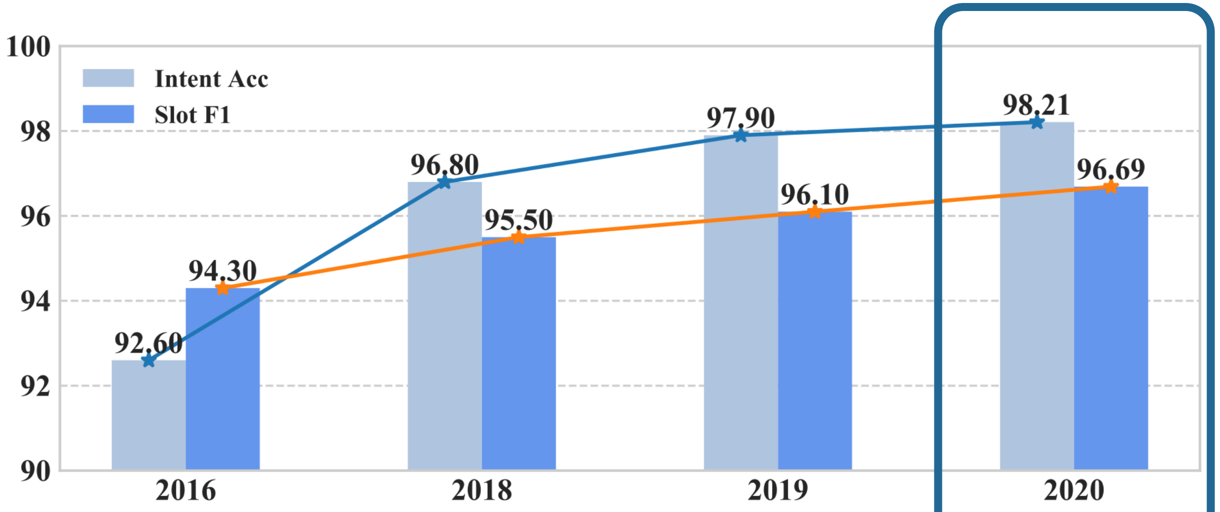
$$C_S = \text{softmax}\left(\frac{\mathbf{Q}_S \mathbf{K}_I^\top}{\sqrt{d_k}}\right)\mathbf{V}_I$$

$$C_I = \text{softmax}\left(\frac{\mathbf{Q}_I \mathbf{K}_S^\top}{\sqrt{d_k}}\right)\mathbf{V}_S$$

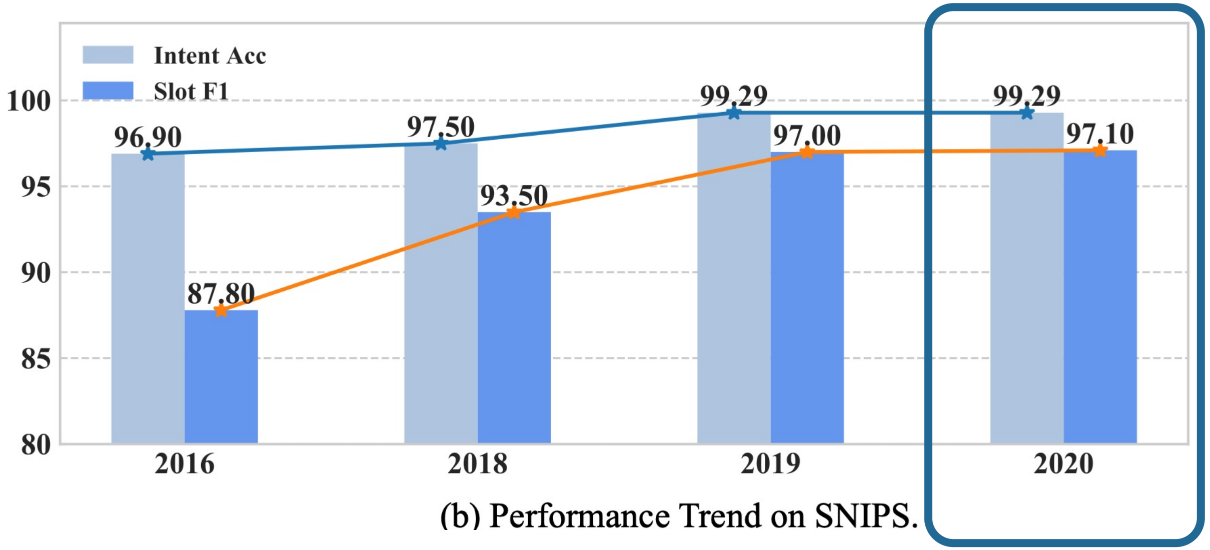
Qin et al. A CO-INTERACTIVE TRANSFORMER FOR JOINT SLOT FILLING AND INTENT DETECTION. ICASSP 2021.

Progress

- SLU direction has made significant progress in recent years
- However, the mainstream work mainly focus on the simple setting: single domain and single turn



(a) Performance Trend on ATIS.



(b) Performance Trend on SNIPS.

Qin et al. A Survey on Spoken Language Understanding: Recent Advances and New Frontiers. IJCAI2021 Survey.

IJCAI2022-Tutorial

New frontiers

Libo Qin

PhD Student
Harbin Institute of
Technology

New frontiers

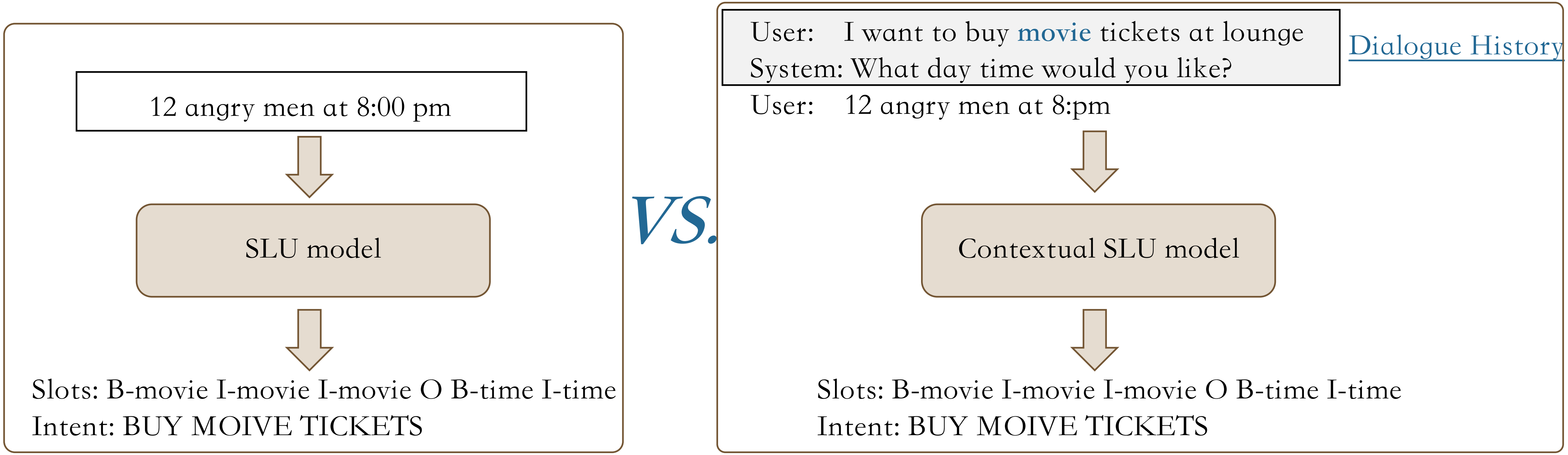
- ① Complex SLU
  - (i) Contextual SLU
  - (ii) Multi-Intent SLU

- ② Low-resource SLU
  - (i) Cross-domain SLU
  - (ii) Cross-lingual SLU
  - (iii) Few-shot SLU
  - (iv) Unsupervised SLU

- ③ Real-world SLU
  - (i) Profile-based SLU
  - (ii) Lifelong Learning
  - (iii) Robust SLU
  - (iv) Chinese SLU

- ④ Unified and Pre-training Paradigm
  - (i) Unified Model
  - (ii) Pre-training Paradigm

① Complex SLU (Contextual SLU)

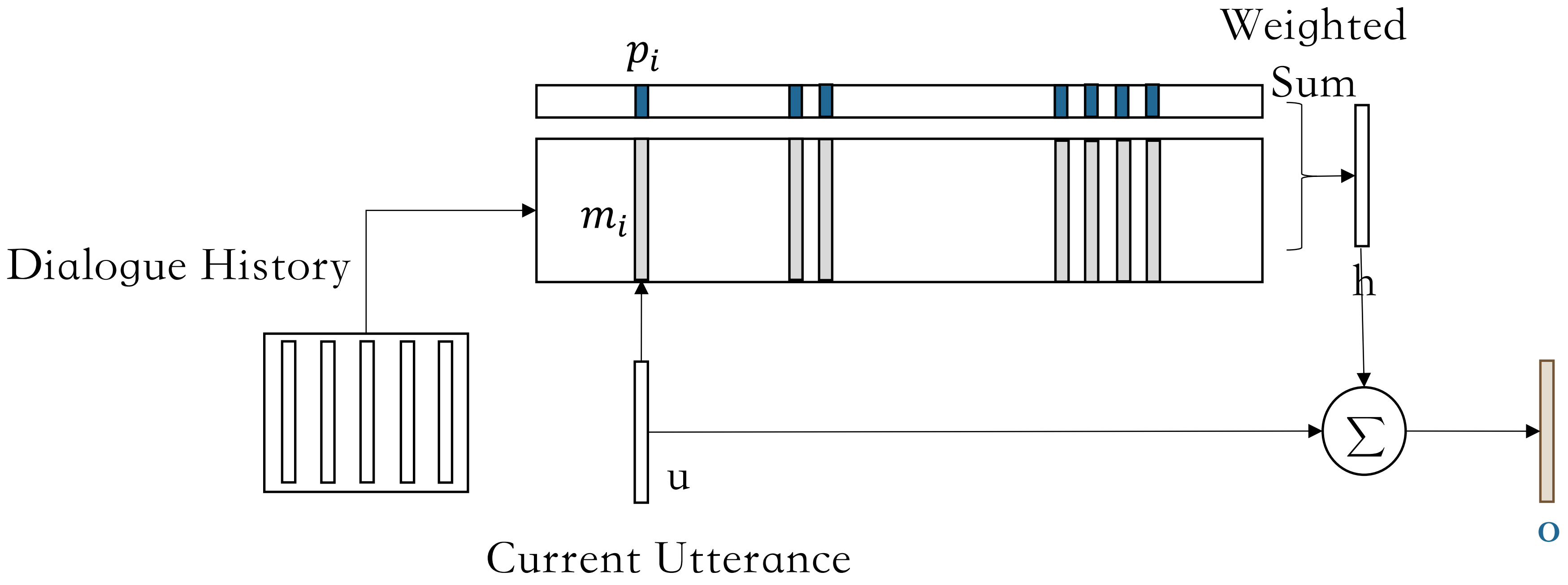


(a) Traditional SLU (b) Contextual SLU

How to leverage relevant dialogue history to make the correct prediction?

① Complex SLU (Contextual SLU)

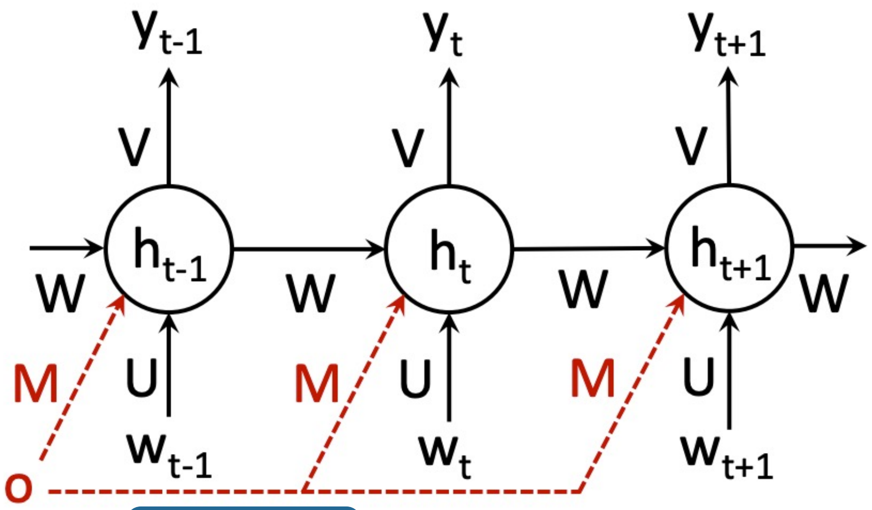
- A memory network is introduced to encode dialogue history



Chen et al. End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding. Interspeech 2016.

① Complex SLU (Contextual SLU)

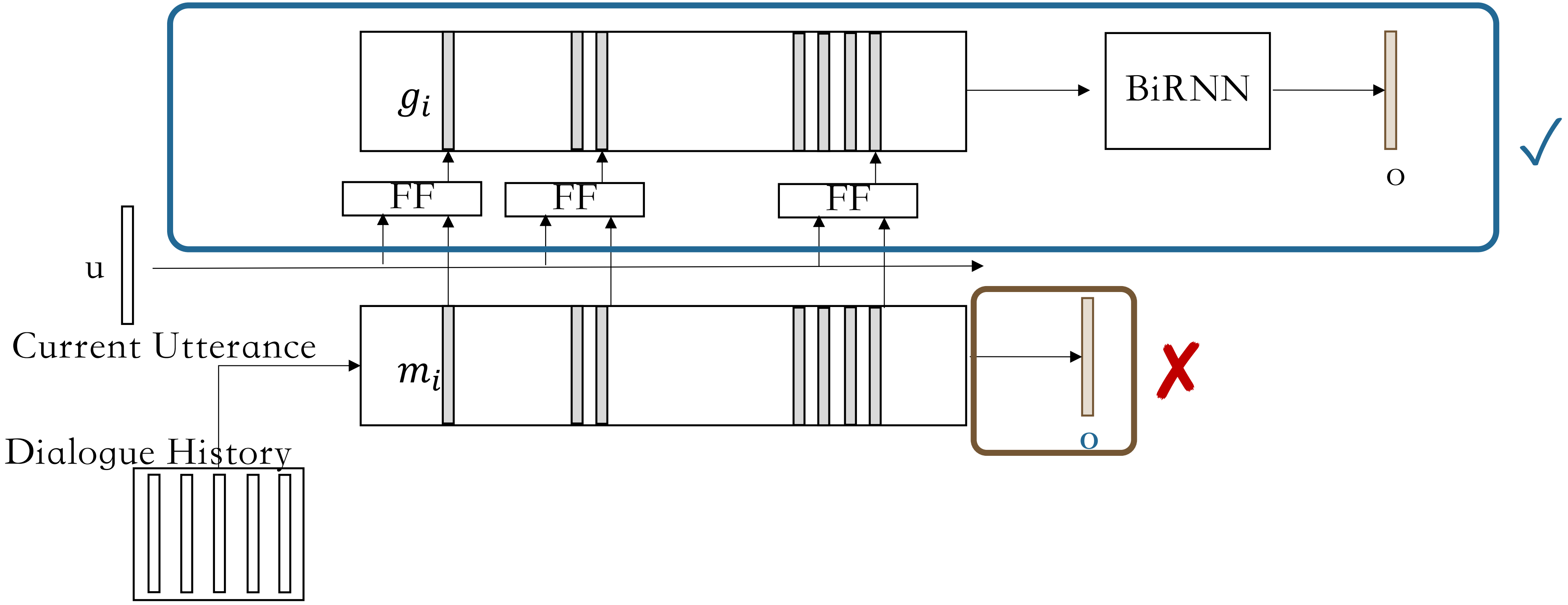
- Dialogue history representation is regarded as the input for the RNN slot tagger



$$h_t = \phi(\boxed{Mo} + Ww_t + Uh_{t-1})$$

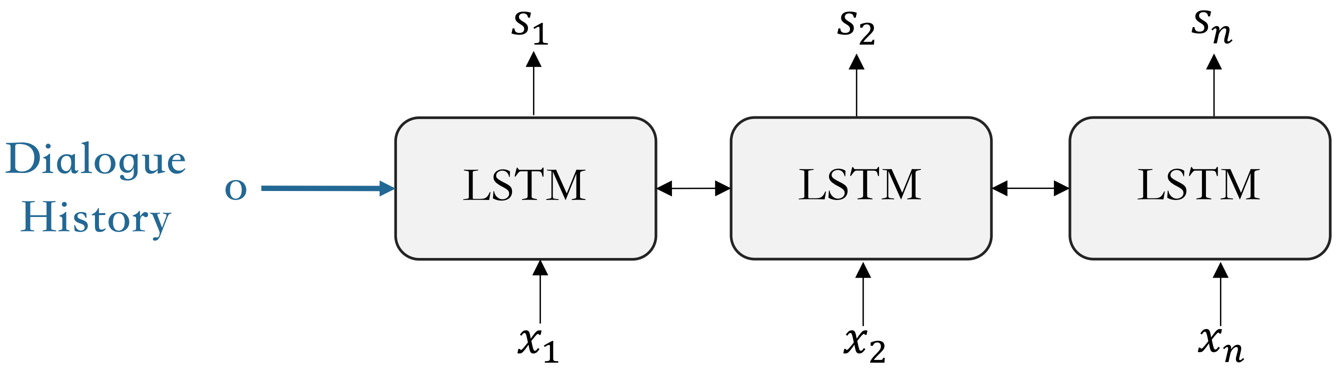
① Complex SLU (Contextual SLU)

- Enhance the memory networks by considering temporal order of utterances in the memory.



① Complex SLU (Contextual SLU)

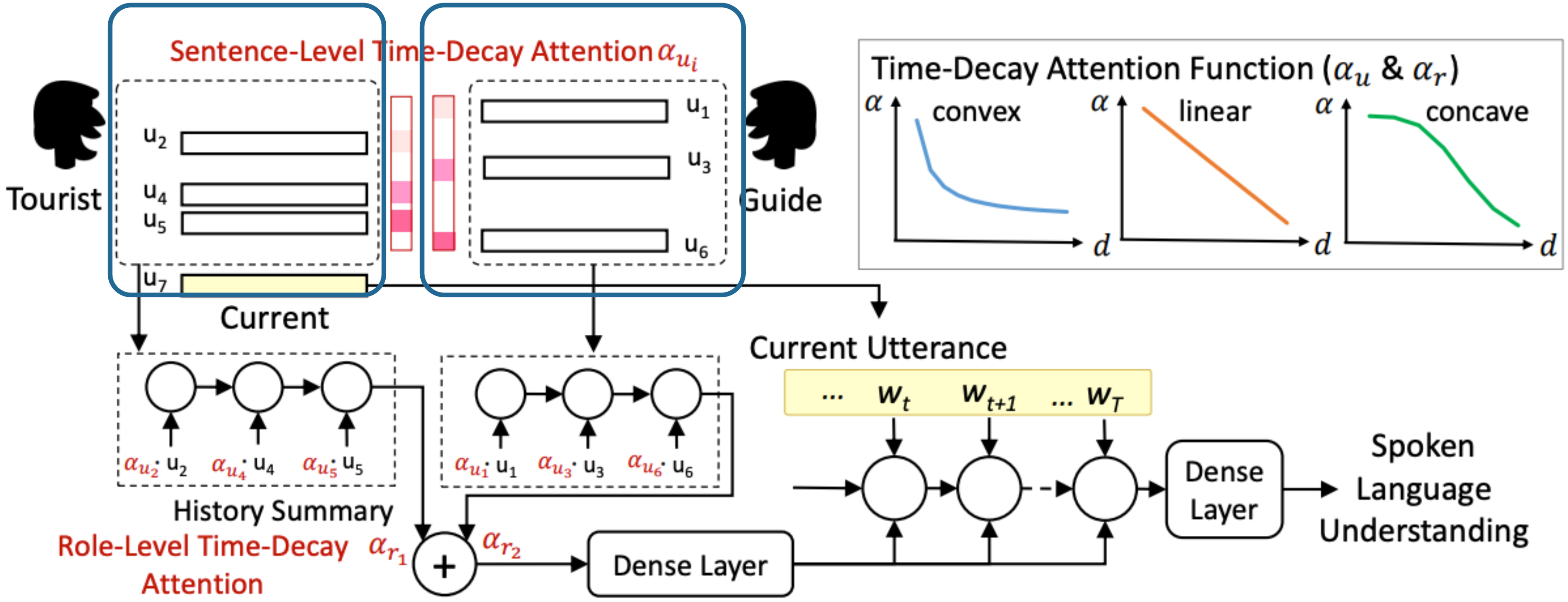
- Dialogue history encoding o is used as the initial states of both forward and backward LSTMs



Bapna et al. Sequential Dialogue Context Modeling for Spoken Language Understanding. SIGDIAL 2017.

① Complex SLU (Contextual SLU)

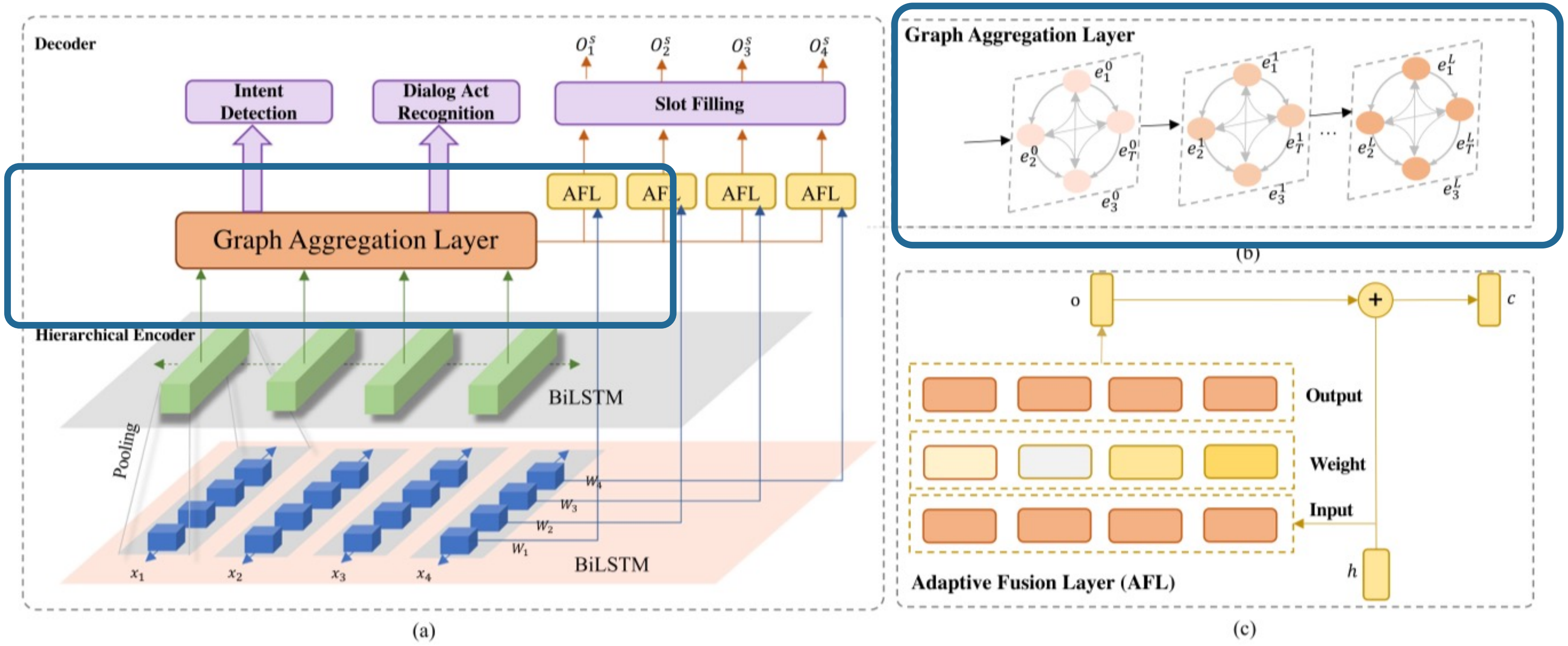
- Different time-decay attention mechanisms for leveraging contextual information.
- Convex; Linear; concave



Su et al. How Time Matters: Learning Time-Decay Attention for Contextual Spoken Language Understanding in Dialogues. NAACL 2018.

① Complex SLU (Contextual SLU)

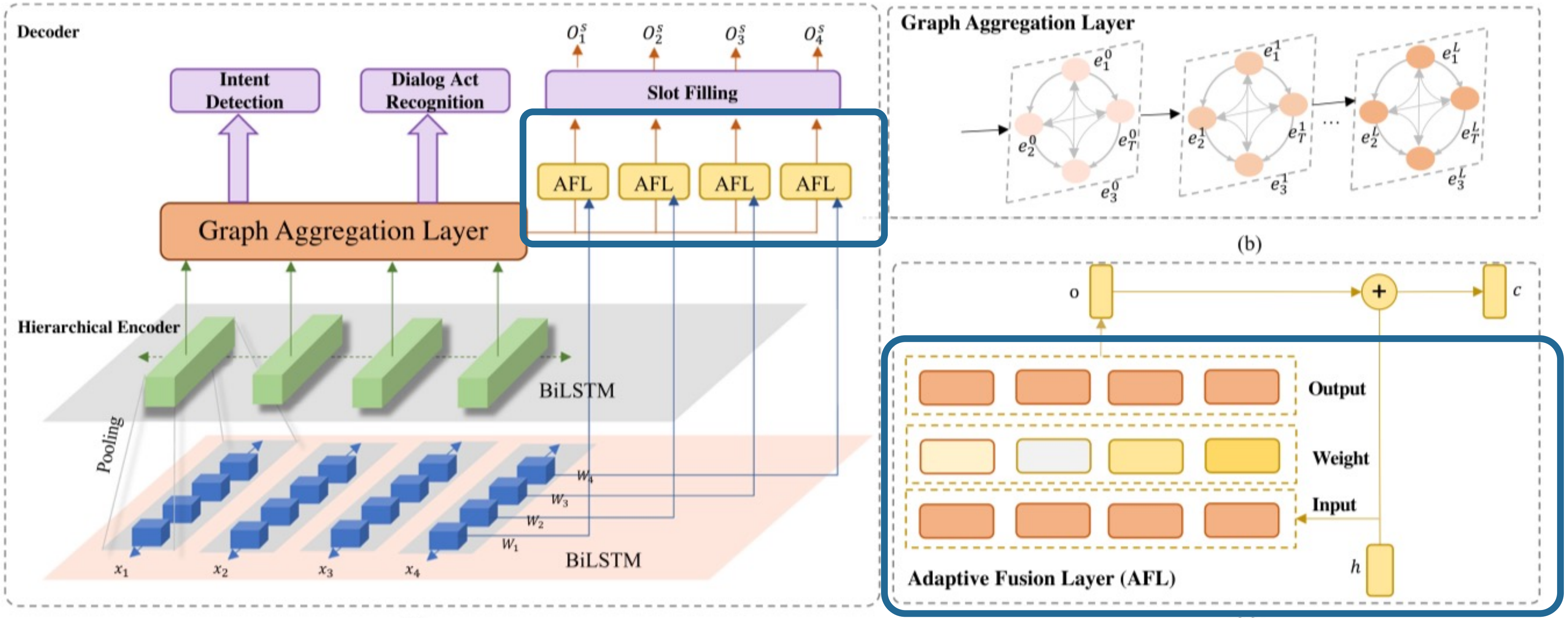
- A context-aware graph convolutional network to automatically incorporate contextual information,



Qin et al. Knowing Where to Leverage: Context-Aware Graph Convolutional Network With an Adaptive Fusion Layer for Contextual Spoken Language Understanding. TASLP 2021.

① Complex SLU (Contextual SLU)

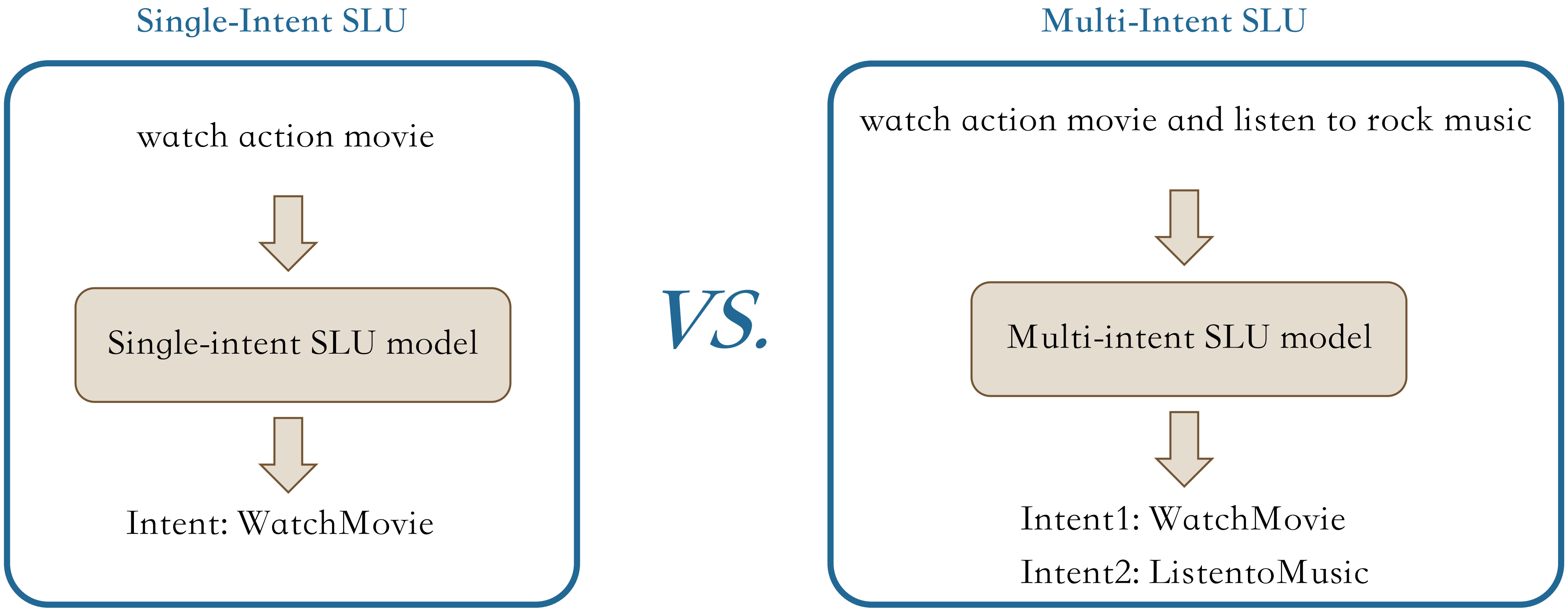
- An adaptive fusion layer to dynamically leverage the contextual information for each token, achieving a fine-grained contextual information integration for the token-level slot prediction.



(a)
(b)
(c)

Qin et al. Knowing Where to Leverage: Context-Aware Graph Convolutional Network With an Adaptive Fusion Layer for Contextual Spoken Language Understanding. TASLP 2021.

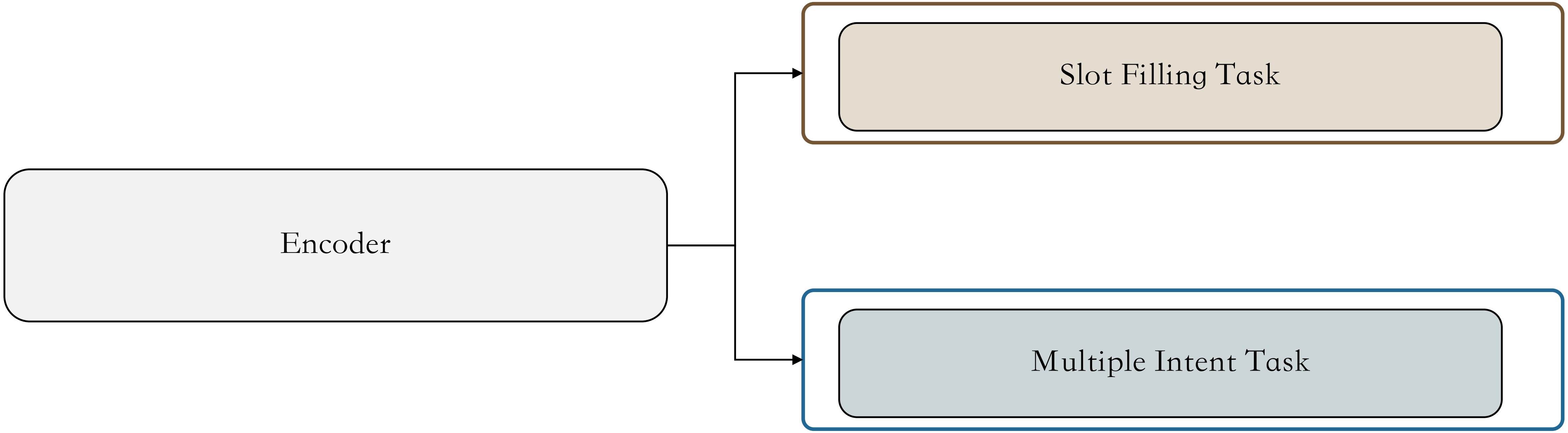
① Complex SLU (Multi-intent SLU)



How to model the interaction between slots and multiple intents?

① Complex SLU (Multi-intent SLU)

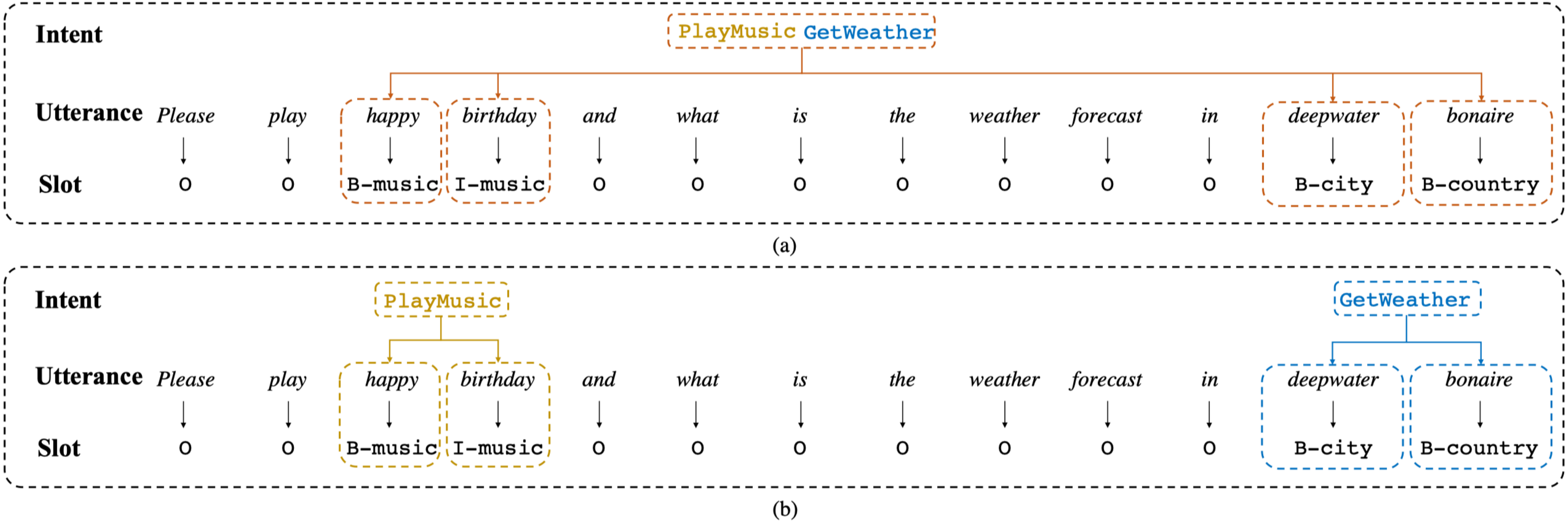
- This is the first joint work to jointly model slot filling and multiple intent detection, which achieves to **implicitly model the relationship between slots and multiple intents.**



Gangadharaiah et al. Joint multiple intent detection and slot labeling for goal oriented dialog. NAACL 2019.

① Complex SLU (Multi-intent SLU)

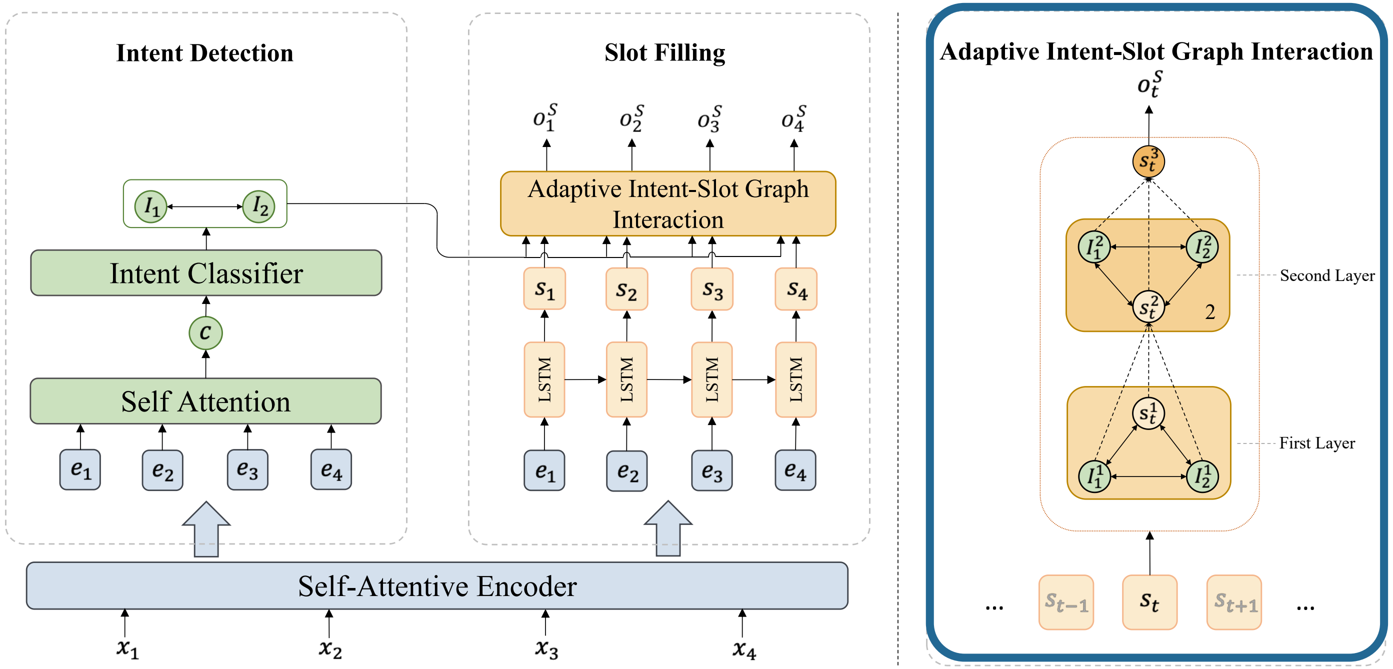
- (a) Prior model simply treat multiple intents as an overall intent information
- (b) Fine-grained multiple intents integration method



Gangadharaiah et al. Joint multiple intent detection and slot labeling for goal oriented dialog. NAACL 2019.

① Complex SLU (Multi-intent SLU)

- Propose an Adaptive Graph Interactive Framework (AGIF) to achieve to **explicitly integrate multiple** intent information for token-level slot filling

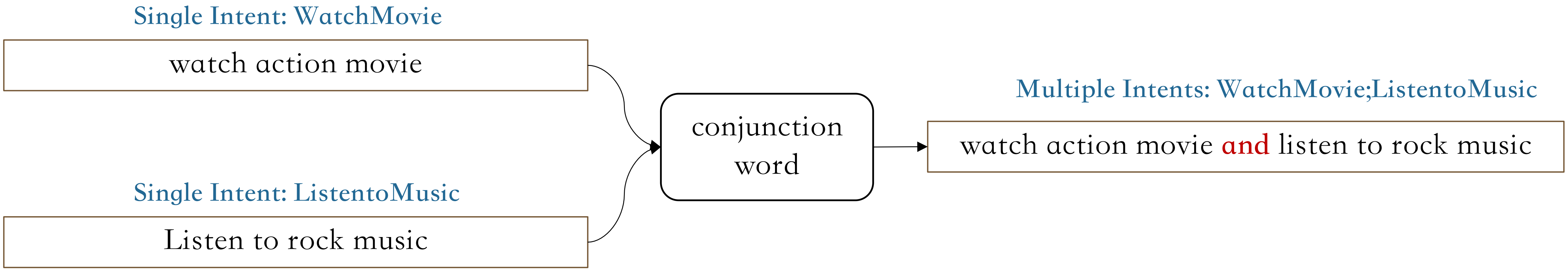


(a) (b)

Qin et al. AGIF: An Adaptive Graph-Interactive Framework for Joint Multiple Intent Detection and Slot Filling. EMNLP 2020.

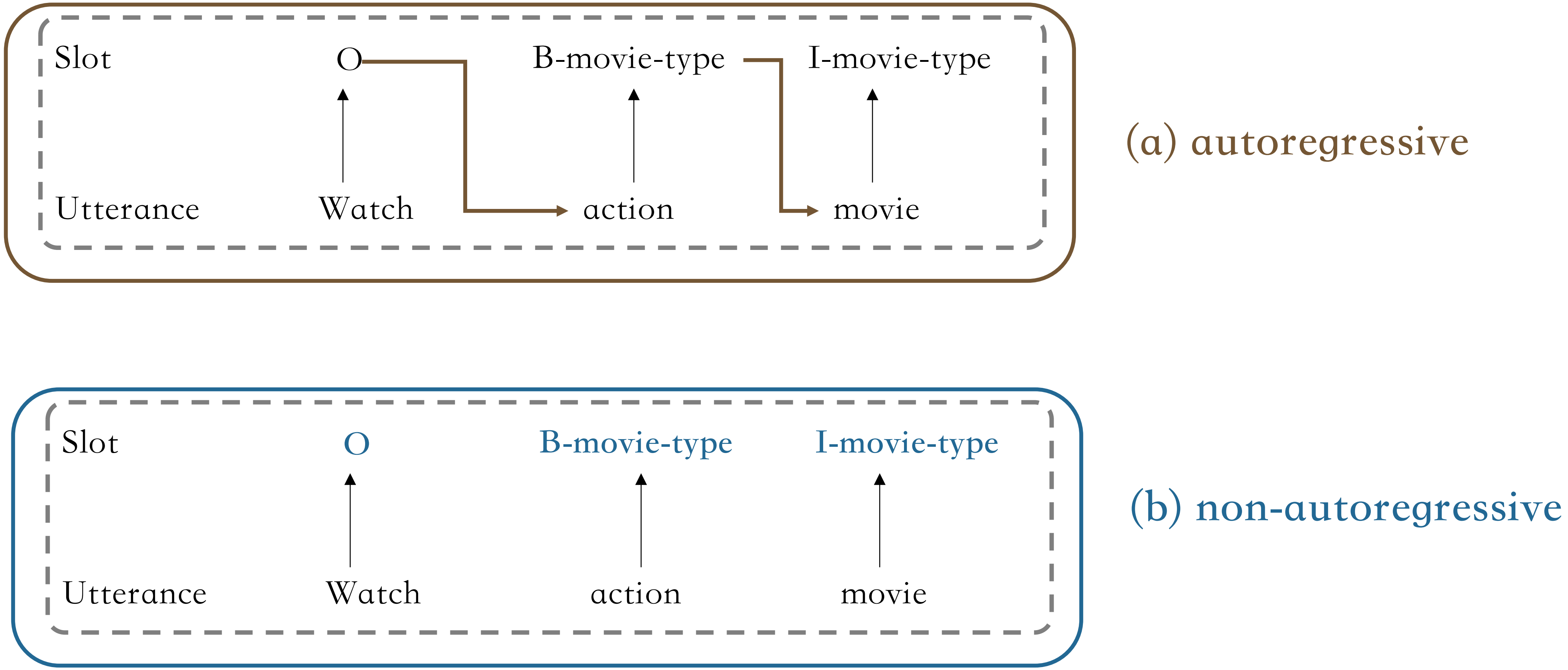
① Complex SLU (Multi-intent SLU)

- In addition, this work construct two multiple intent SLU datasets by using conjunctions, e.g., "and", to connect single-intent sentences, which facilitate the research



Qin et al. AGIF: An Adaptive Graph-Interactive Framework for Joint Multiple Intent Detection and Slot Filling. EMNLP 2020.

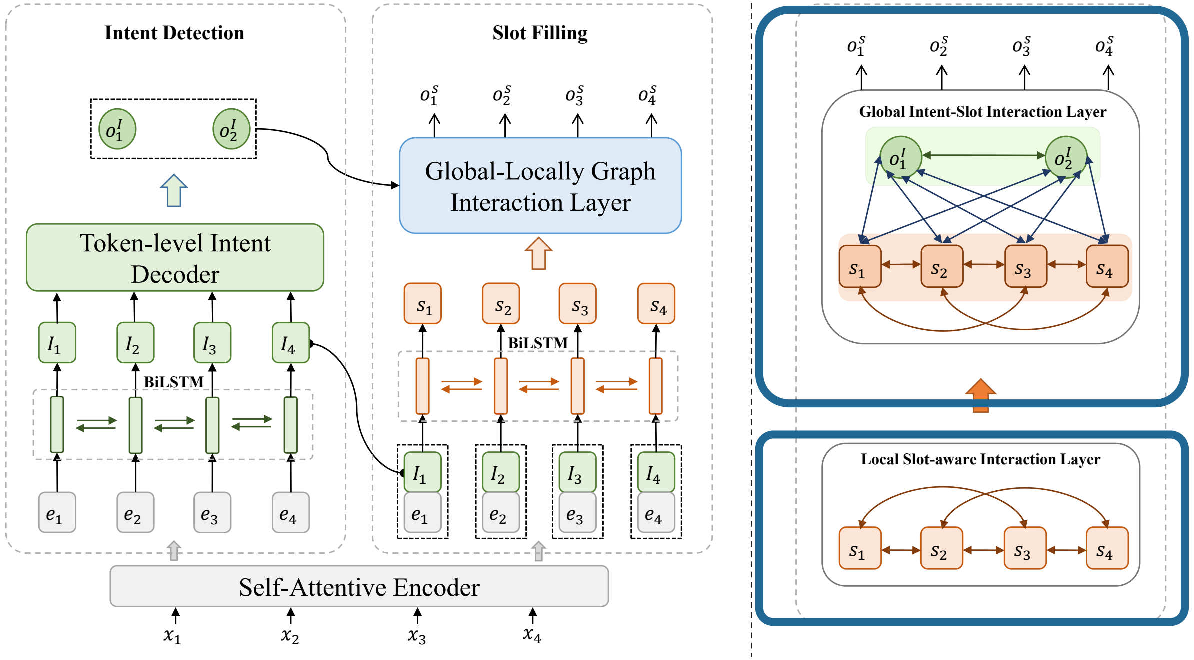
① Complex SLU (Multi-intent SLU)



Qin et al. GL-GIN: Fast and Accurate Non-Autoregressive Model for Joint Multiple Intent Detection and Slot Filling. ACL 2021.

① Complex SLU (Multi-intent SLU)

- A global-locally graph-interaction network
  - A local graph is used to handle uncoordinated slots problem
  - A global graph is introduced to model sequence-level intent-slot interaction to perform non-autoregressive slot sequence generation



(a) Model Framework (b) Global-Locally Graph Interaction Layer

Qin et al. GL-GIN: Fast and Accurate Non-Autoregressive Model for Joint Multiple Intent Detection and Slot Filling. ACL 2021.

② Low-resource SLU (Cross-domain SLU)

Source Domain → Targe Domain

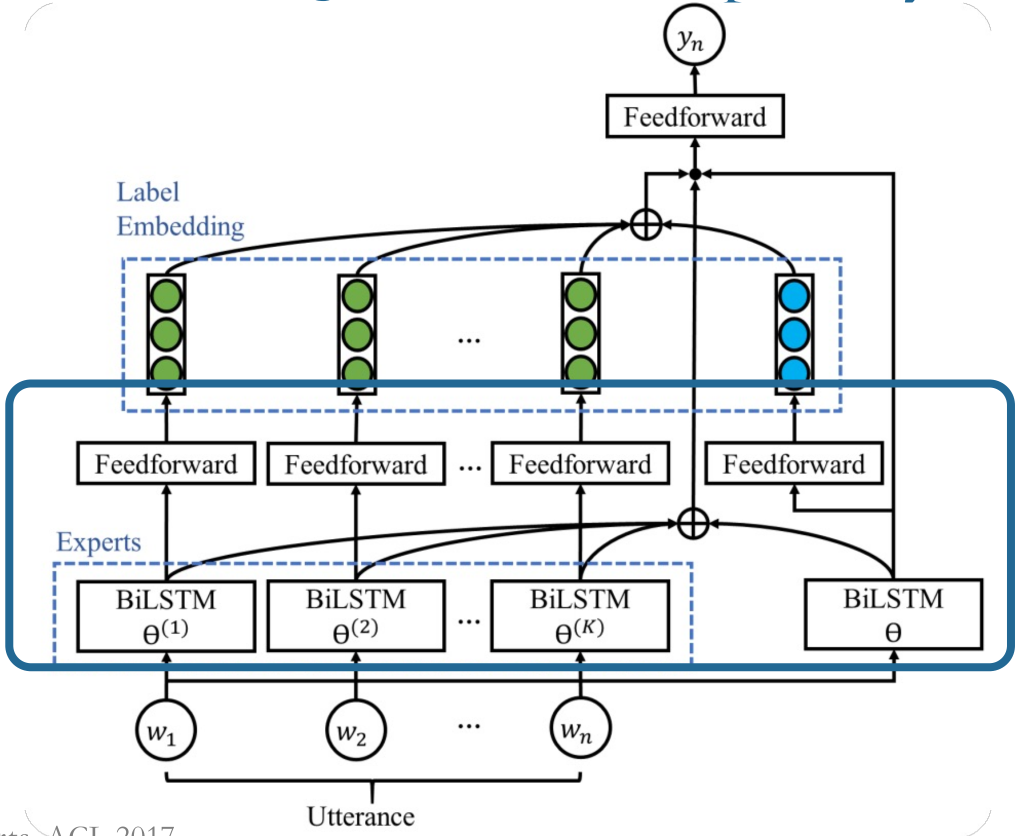
with sufficient data

with limited/no data

How to transfer knowledge from source domain to target domain?

② Low-resource SLU (Cross-domain SLU)

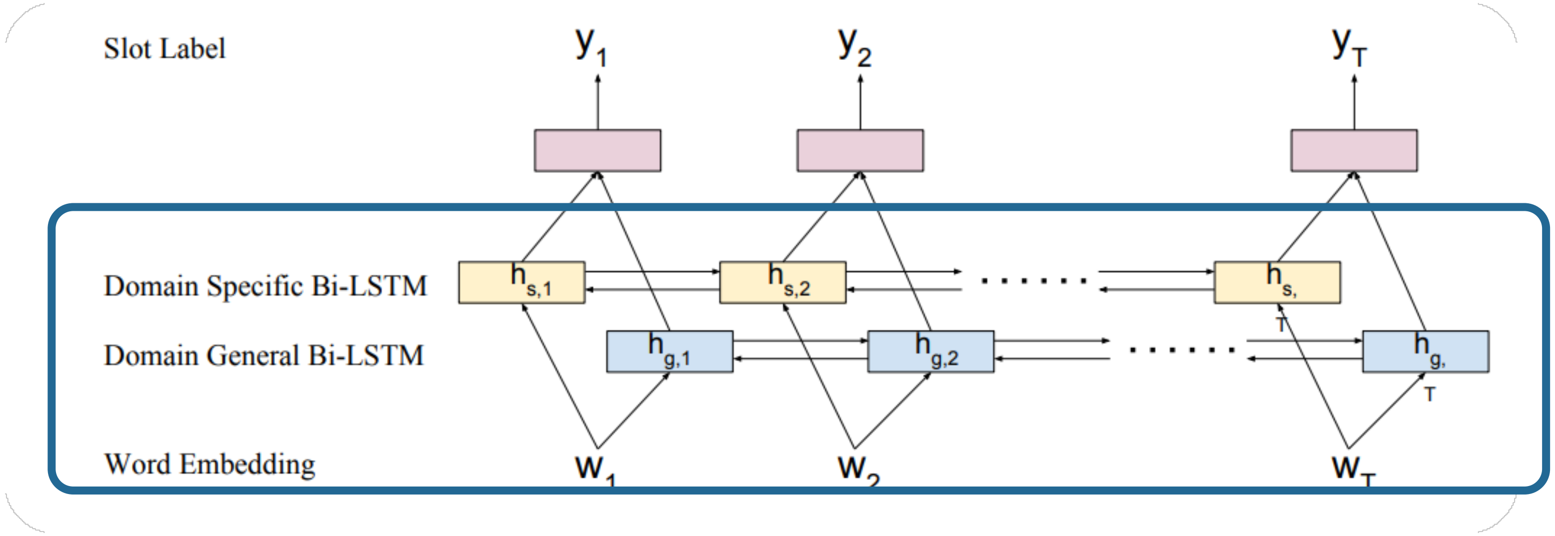
- An attention mechanism to make full use of previous trained domain expert knowledge for new domain, which has the advantage of transferring domain knowledge without explicitly re-training on all domains together



Kim et al. Domain attention with an ensemble of experts. ACL 2017.

② Low-resource SLU (Cross-domain SLU)

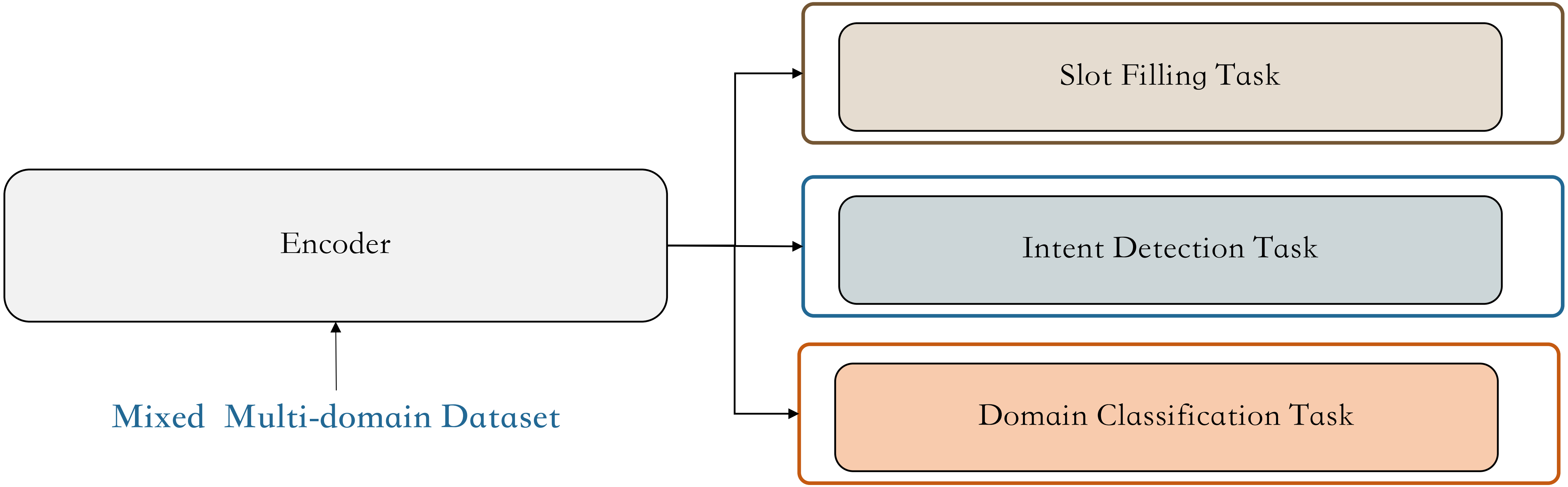
- Domain general Bi-LSTM to capture domain-shared knowledge
- Domain specific Bi-LSTM to incorporate domain-specific knowledge
- Combine domain-shared and domain-specific features for multi-domain slot filling



Liu et al. Multi-domain adversarial learning for slot filling in spoken language understanding. NIPS Workshop, 2017.

② Low-resource SLU (Cross-domain SLU)

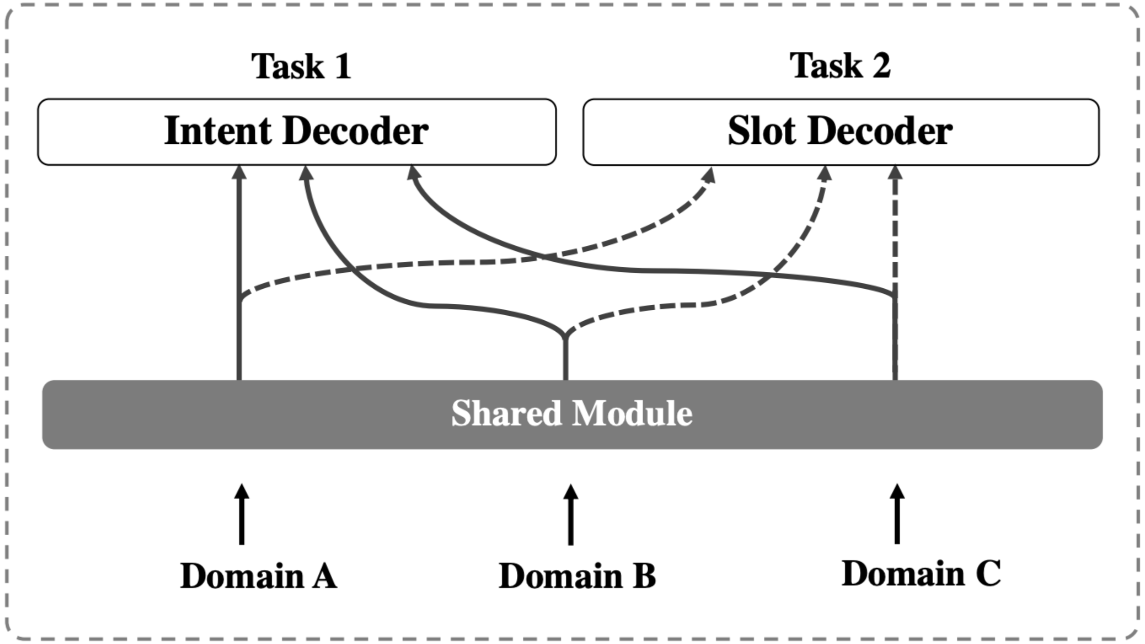
- One network to jointly model slot filling, intent detection and domain classification, implicitly learning the domain-shared information



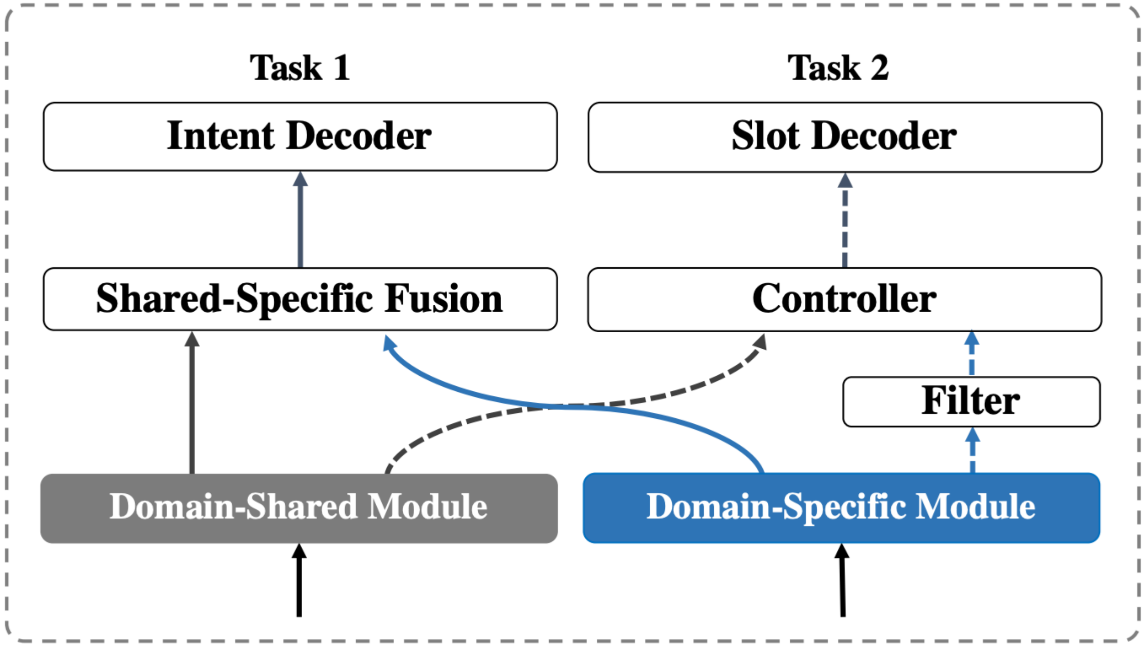
Kim et al. ONENET: JOINT DOMAIN, INTENT, SLOT PREDICTION FOR SPOKEN LANGUAGE UNDERSTANDING. ASRU 2017.

② Low-resource SLU (Cross-domain SLU)

- (a) Prior work trains a single model on a mixed dataset
- (b) The domain-aware and task-aware model



(a)



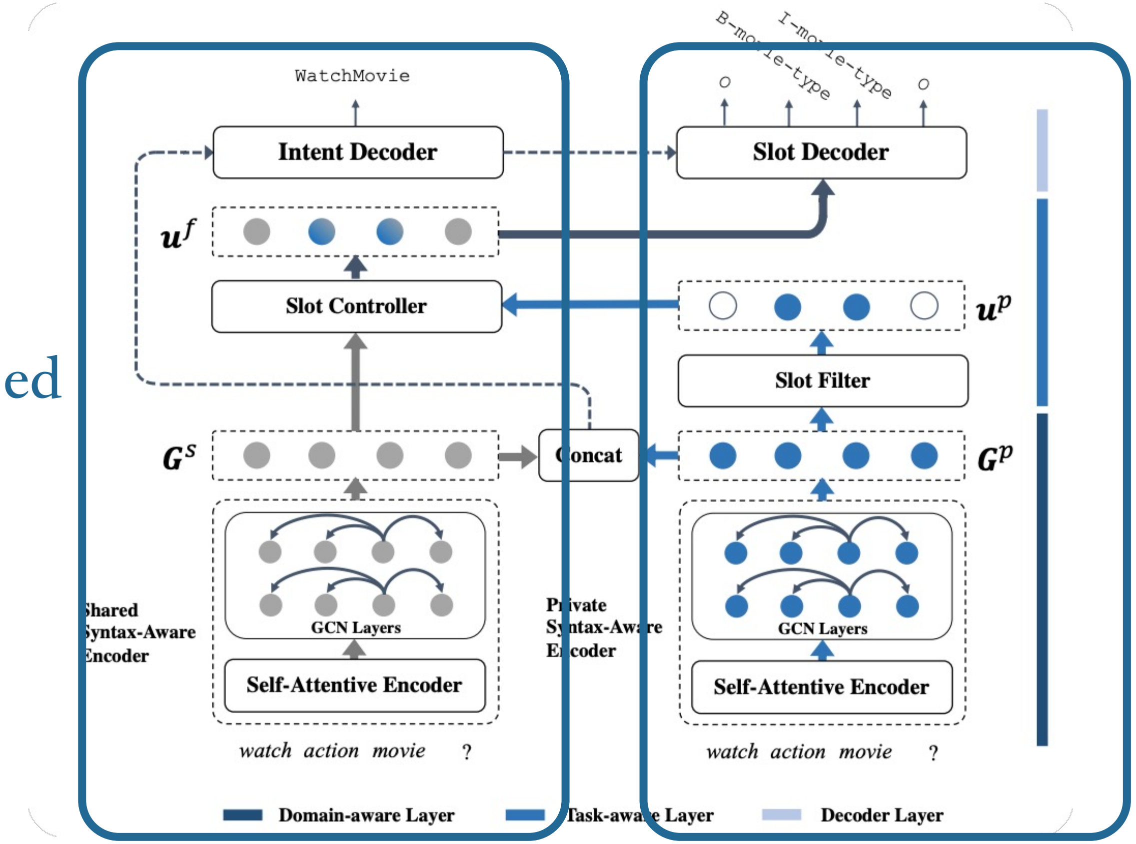
(b)

Qin et al. Multi-Domain Spoken Language Understanding Using Domain- and Task-Aware Parameterization. TALLIP 2021.

② Low-resource SLU (Cross-domain SLU)

- A model with separate domain- and task-specific parameters, which enables model to capture the task-aware and domain-aware features for multi-domain SLU.

Domain-Shared Module

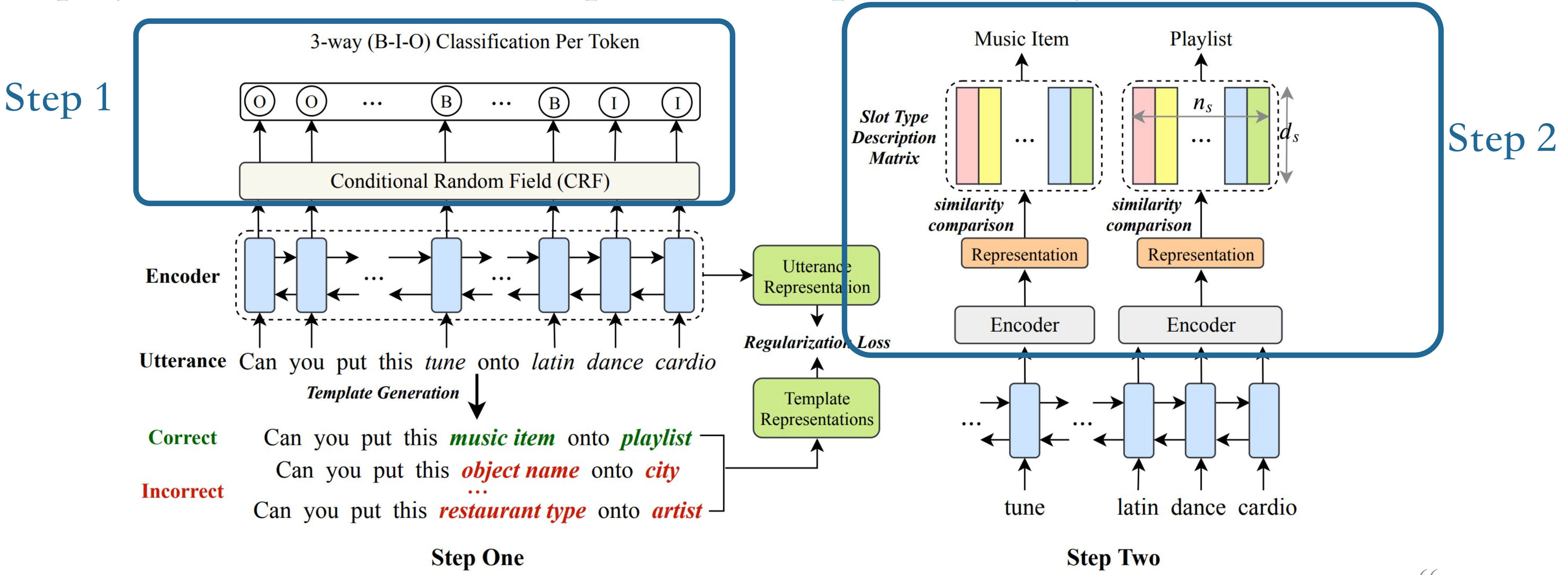


Domain-Private Module

Qin et al. Multi-Domain Spoken Language Understanding Using Domain- and Task-Aware Parameterization. TALLIP 2021.

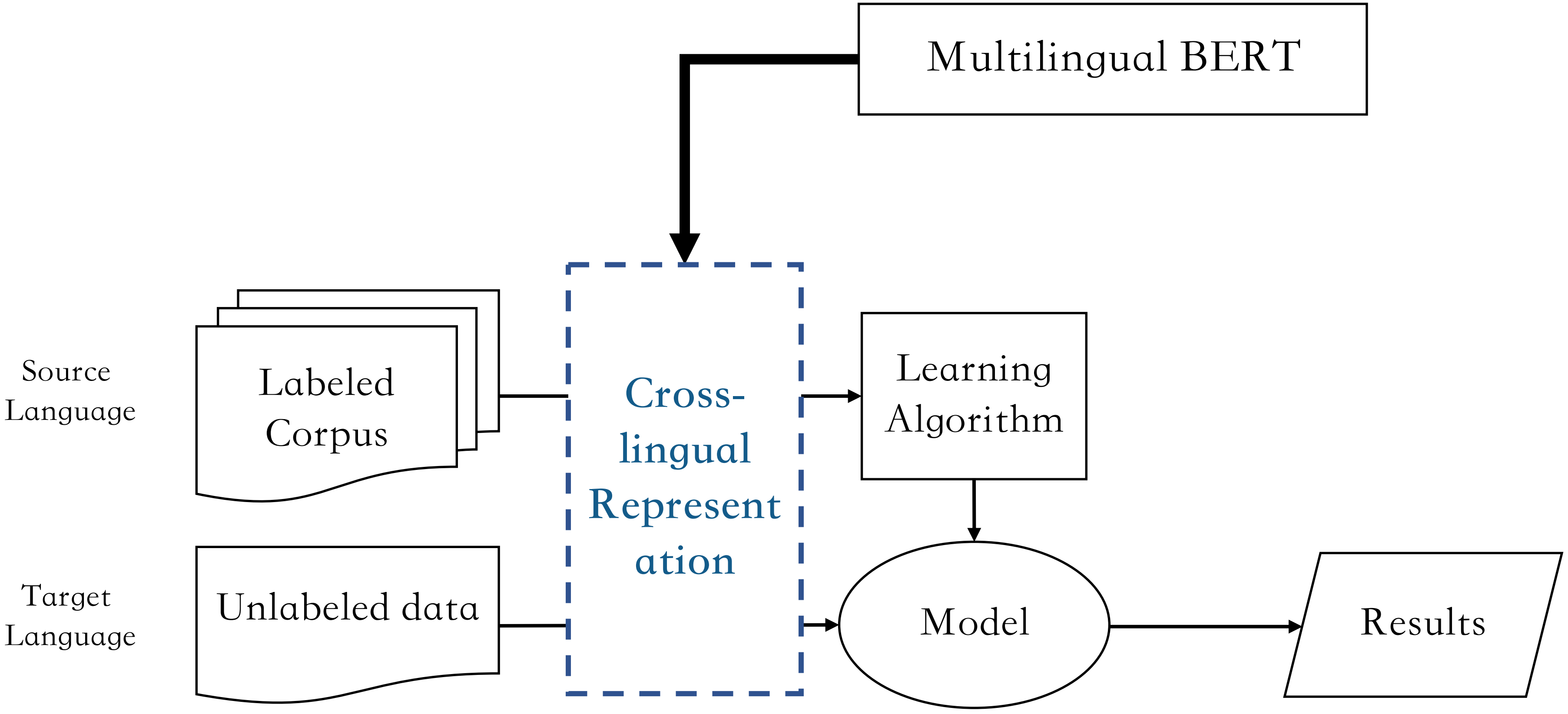
② Low-resource SLU (Cross-domain SLU)

- Fist predict the entity type, and then predict slot type based on the similarity with the representation of each slot type description
- Employ regularization loss to improve the adaptation ability



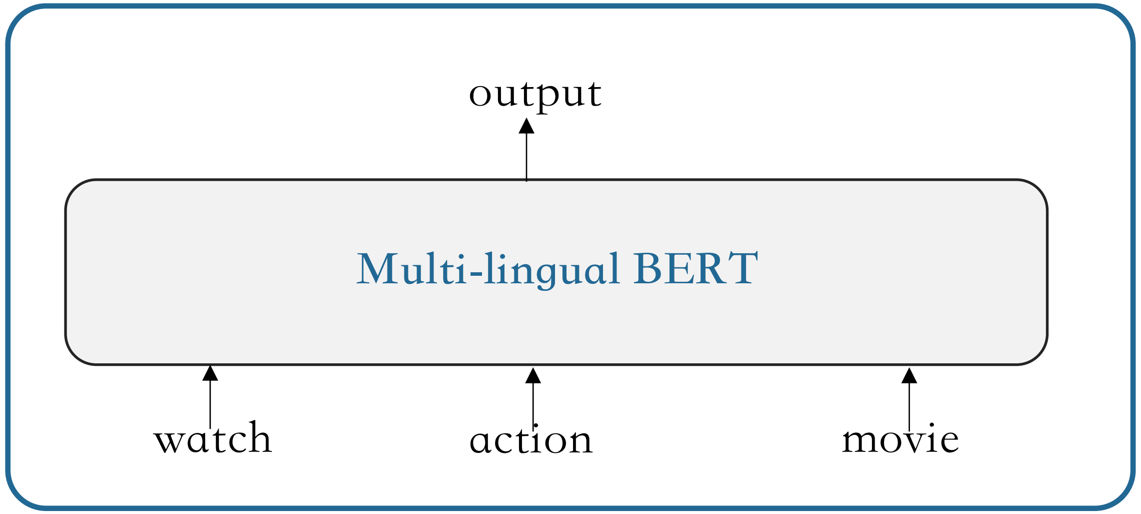
Liu et al. Coach: A Coarse-to-Fine Approach for Cross-domain Slot Filling. ACL 2020.

② Low-resource SLU (Cross-lingual SLU)



② Low-resource SLU (Cross-lingual SLU)

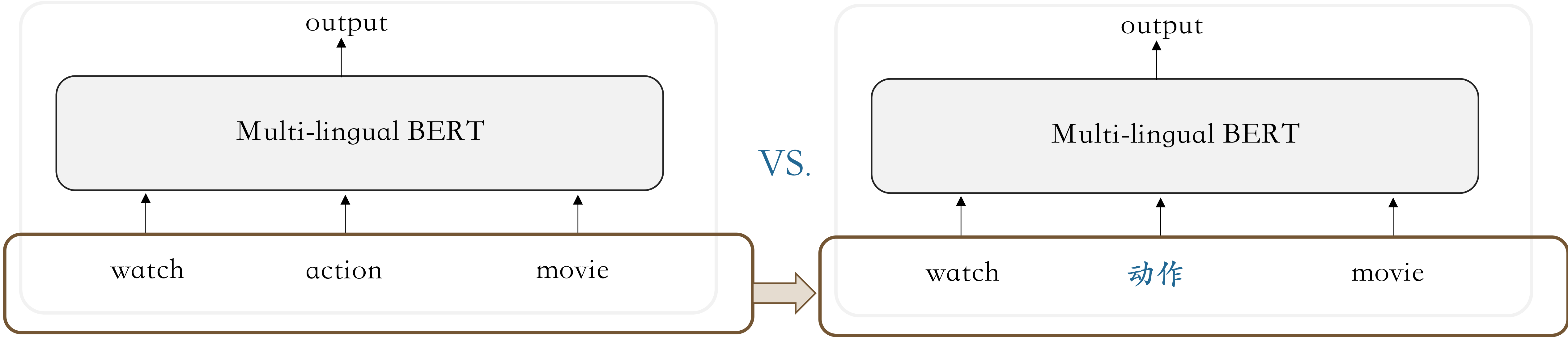
- Employ multi-lingual BERT for cross-lingual SLU



Jacob et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. AAAI 2020.

② Low-resource SLU (Cross-lingual SLU)

- Attention-informed mixed-language training (MLT) for cross-lingual SLU by using code-switching sentences for fine-tuning, which can implicitly align representations between source language and target language

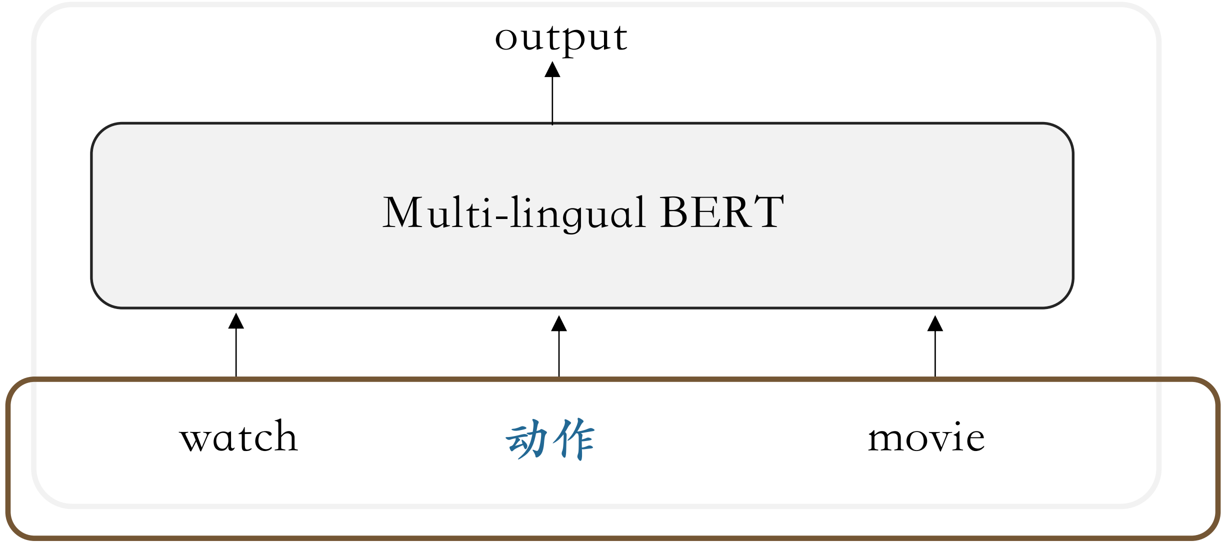


Liu et al. Attention-Informed Mixed-Language Training for Zero-shot Cross-lingual Task-oriented Dialogue Systems. AAAI 2020.

② Low-resource SLU (Cross-lingual SLU)

- How to choose the replaced word?
  - Attention method

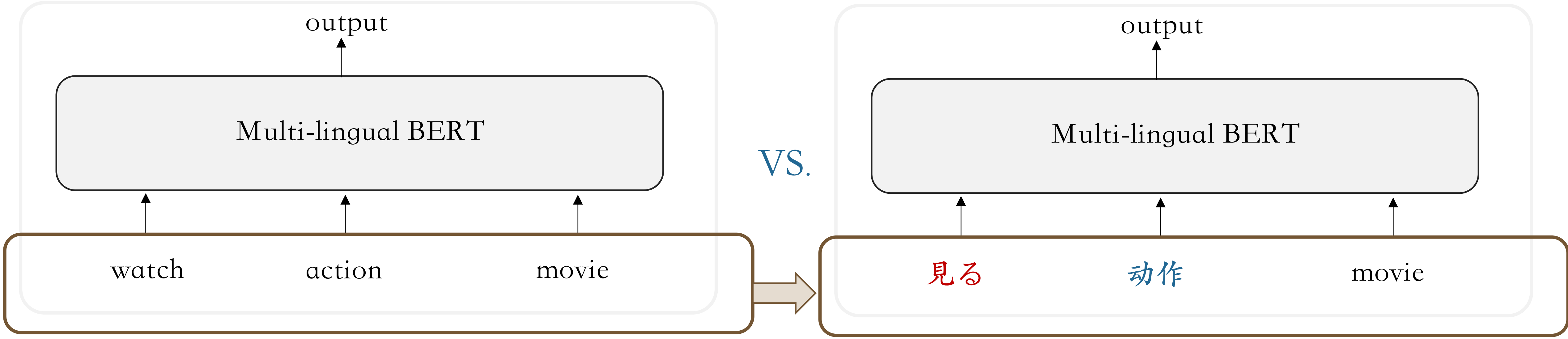
- Only contain one target language



Liu et al. Attention-Informed Mixed-Language Training for Zero-shot Cross-lingual Task-oriented Dialogue Systems. AAAI 2020.

② Low-resource SLU (Cross-lingual SLU)

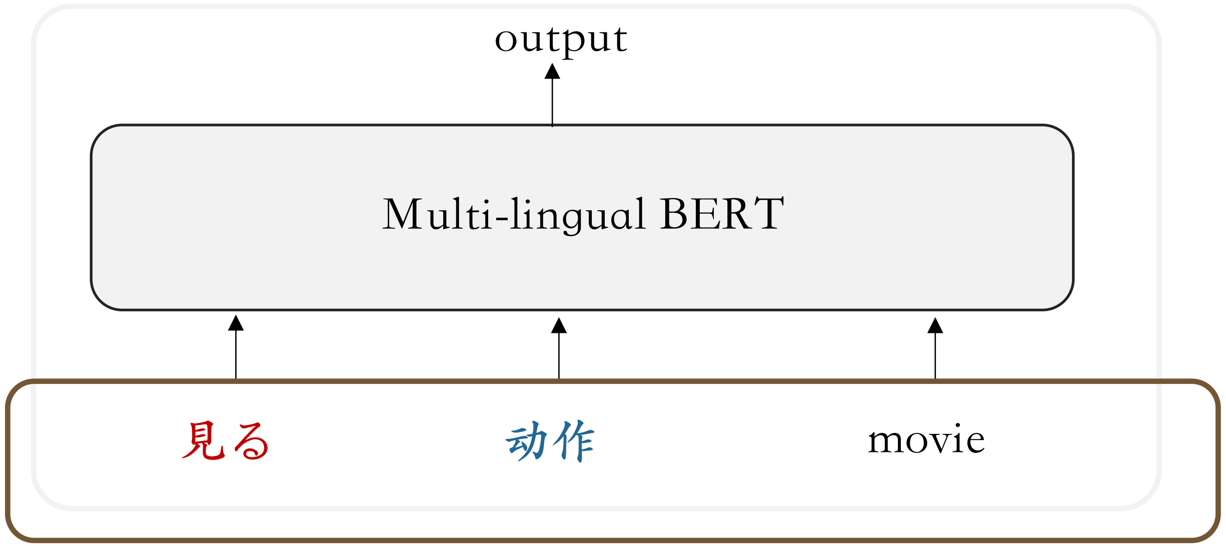
- An augmentation framework (CoSDA-ML) to generate multilingual code-switching data to fine-tune mBERT for aligning representations from source and multiple target languages.



Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

② Low-resource SLU (Cross-lingual SLU)

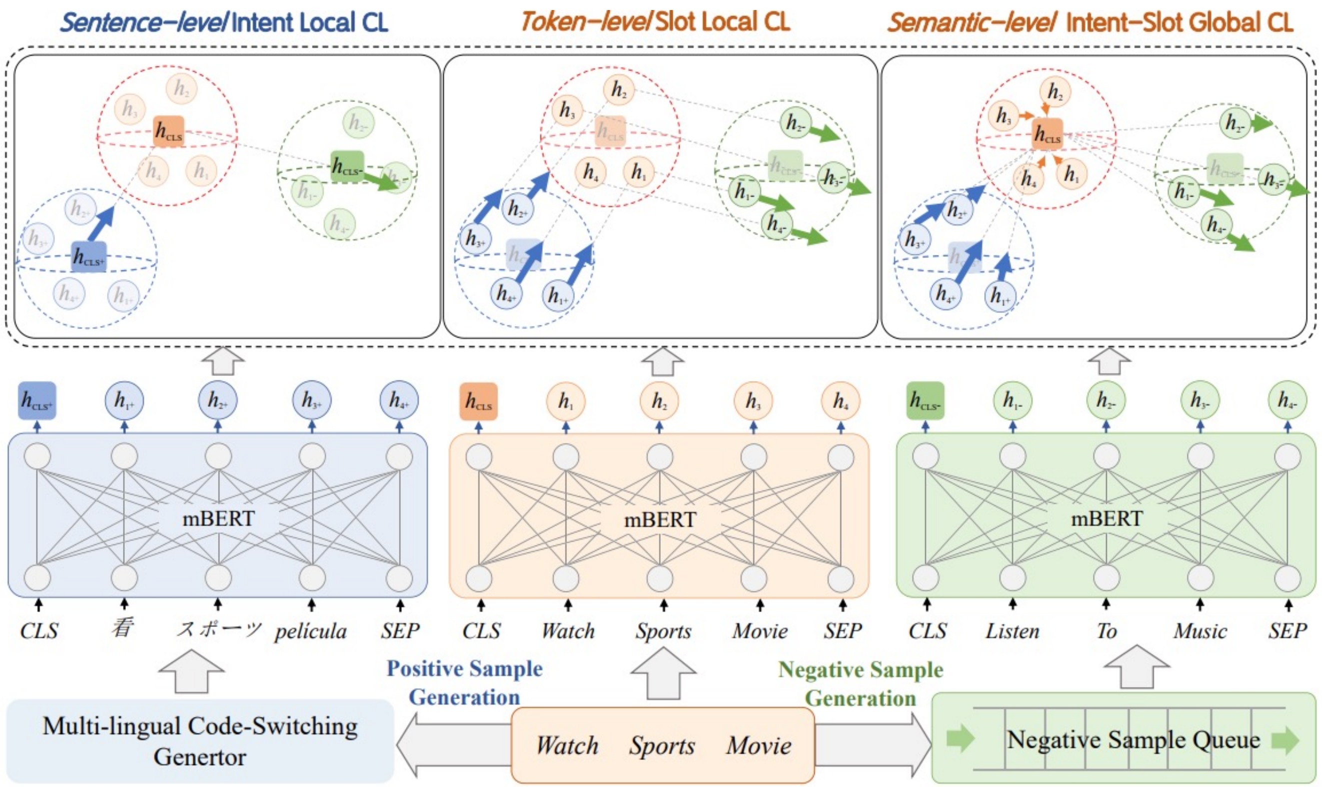
- How to choose the replaced word?
  - Randomly select each word in a sentence
- Align representation across source language and multiple target languages at once



Qin et al. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

② Low-resource SLU (Cross-lingual SLU)

- A global – local contrastive learning (CL) framework (GL-CLeF) to explicitly align representations across languages for zero-shot cross-lingual SLU



② Low-resource SLU (Cross-lingual SLU)

- Positive Samples Generation
  - Employ code-switching technique to generate multi-lingual code-switched data, which is considered as the positive samples.

$[CLS_{ori}]$ watch action movie ⇒ *Positive Samples Generation* ⇒ $[CLS_{pos}]$ watch 动作 movie

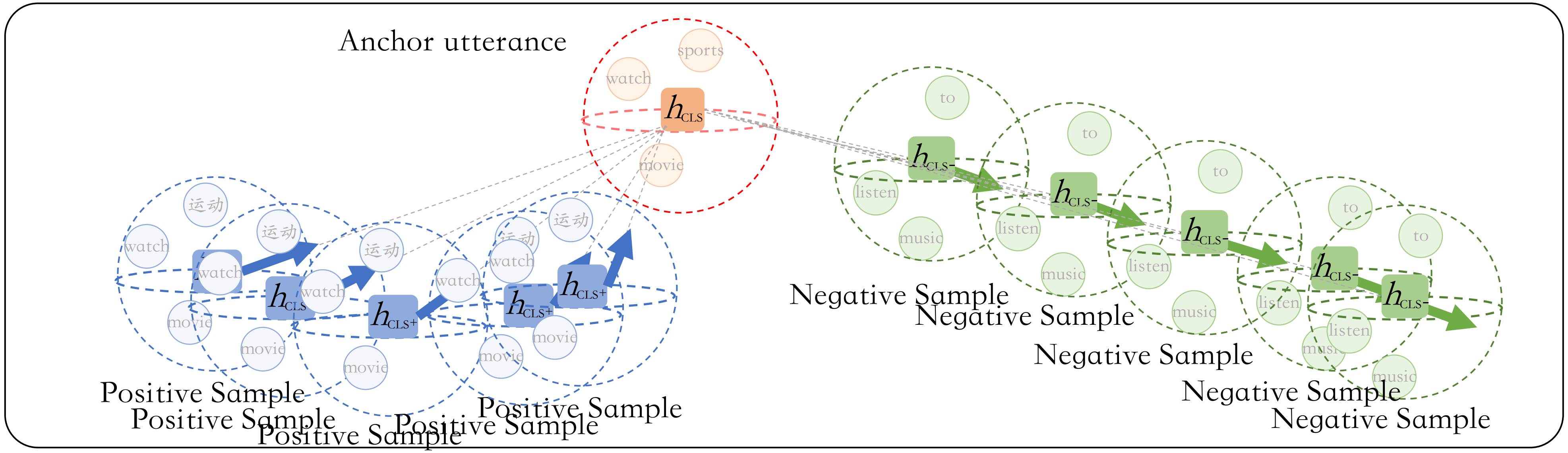
- Negatives Samples Generation
  - Other different queries in a batch can be considered as negative samples.

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

② Low-resource SLU (Cross-lingual SLU)

- 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.

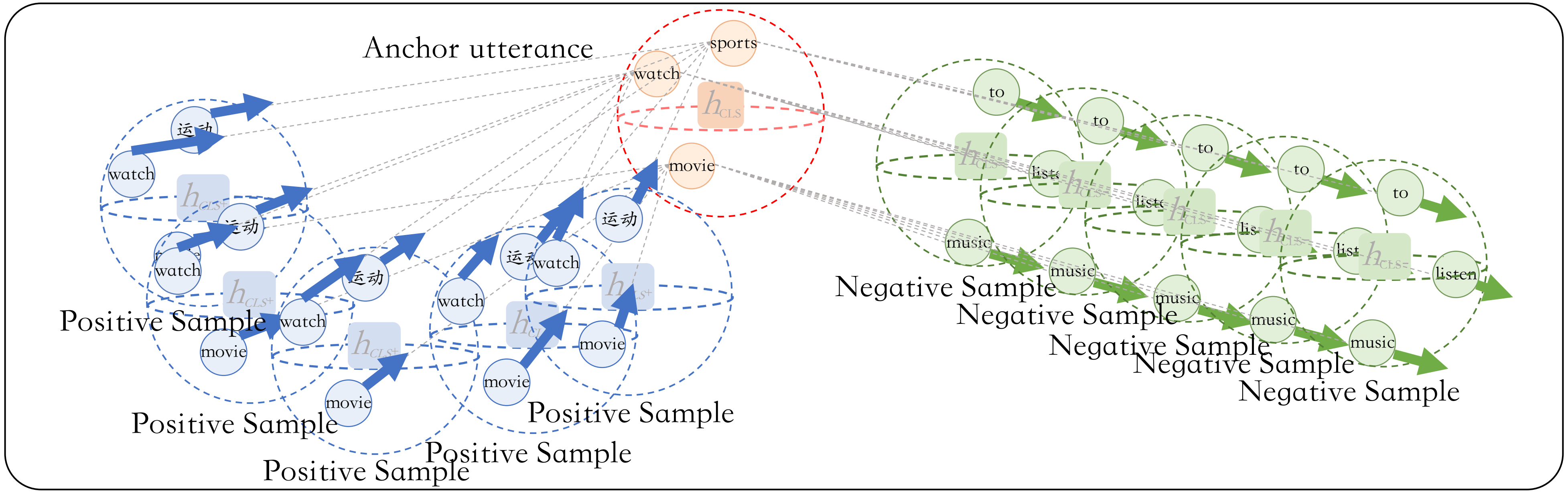
Sentence-level Intent Local CL



② Low-resource SLU (Cross-lingual SLU)

- 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.

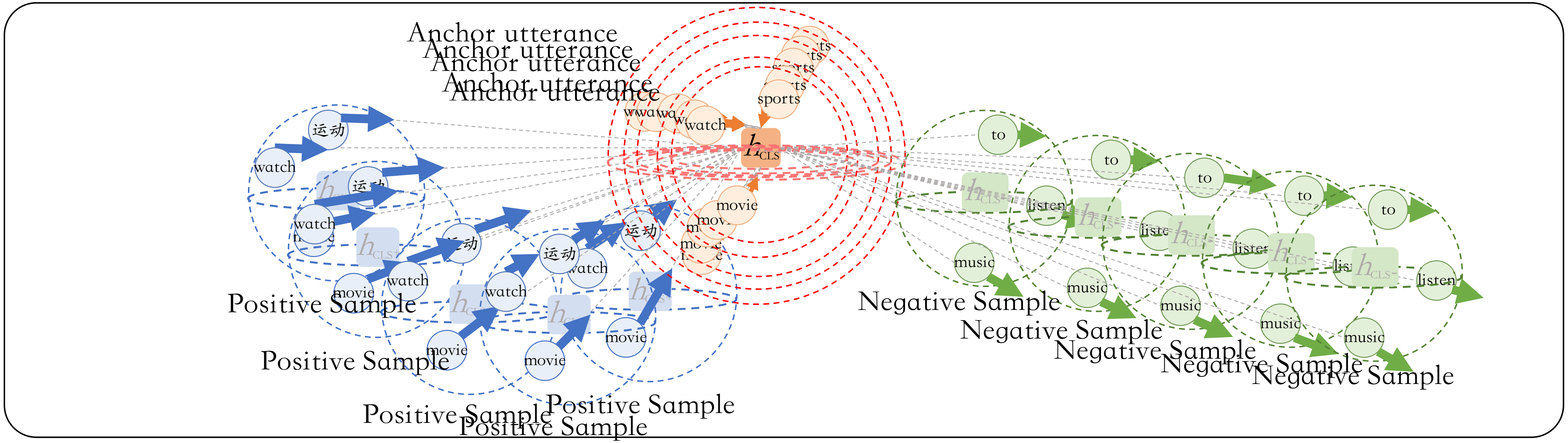
Token-level Slot Local CL



② Low-resource SLU (Cross-lingual SLU)

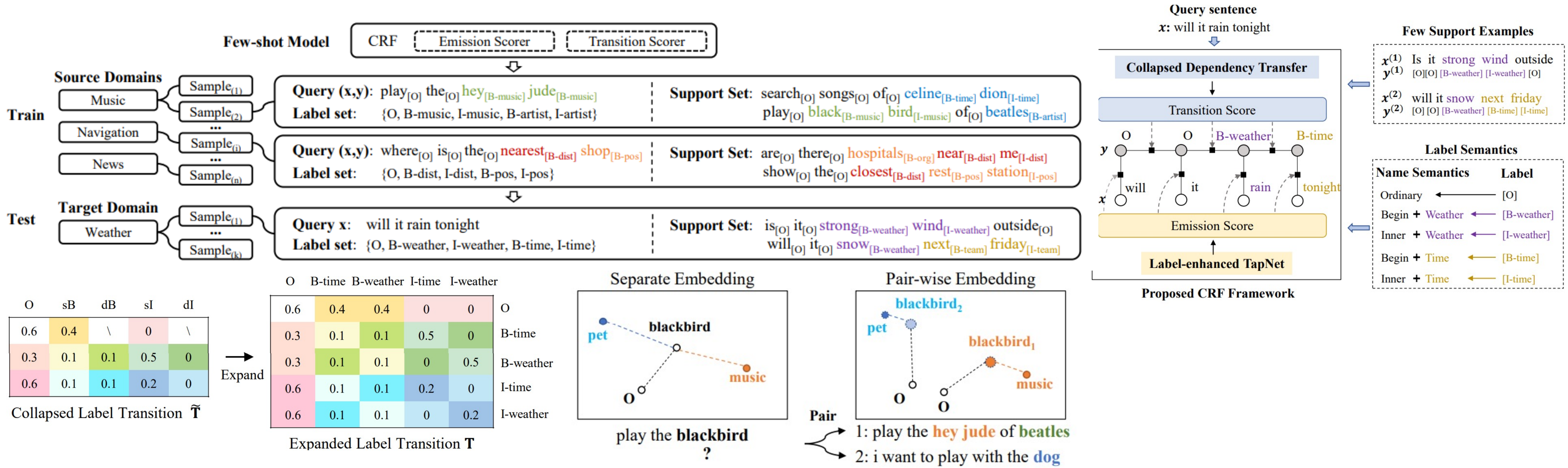
- A semantic-level global intent-slot CL loss to model the semantic interaction between slots and intent, which may further enhance cross-lingual transfer between them.

Semantic-level Intent-Slot Global CL



② Low-resource SLU (Few-shot SLU)

- A CRF model for few-shot slot filling
  - A collapsed dependency transfer mechanism to transfer abstract label dependency patterns across domains
  - L-TapNet to leverage label name semantics in representing labels
  - A pair-wise embedding mechanism to obtain better word representation



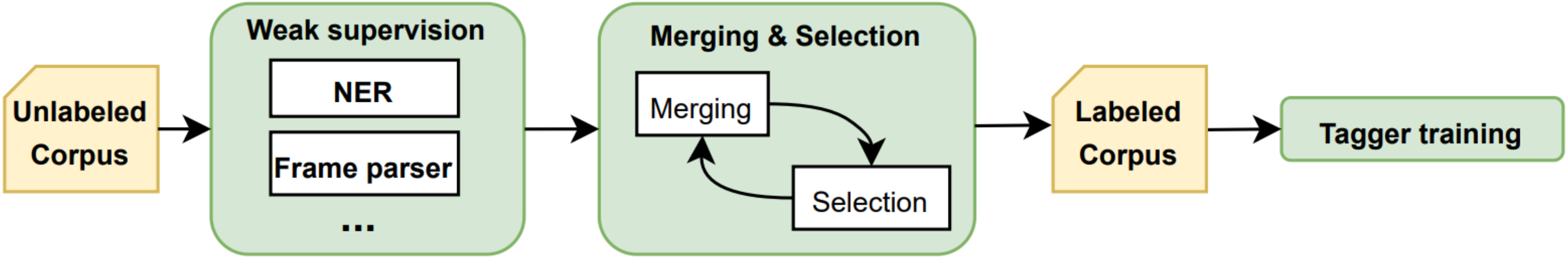
	O	sB	dB	sI	dI
O	0.6	0.4	\	0	\
B	0.3	0.1	0.1	0.5	0
I	0.6	0.1	0.1	0.2	0

O	B-time	B-weather	I-time	I-weather	
0.6	0.4	0.4	0	0	O
0.3	0.1	0.1	0.5	0	B-time
0.3	0.1	0.1	0	0.5	B-weather
0.6	0.1	0.1	0.2	0	I-time
0.6	0.1	0.1	0	0.2	I-weather

Hou et al. Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network. ACL 2020.

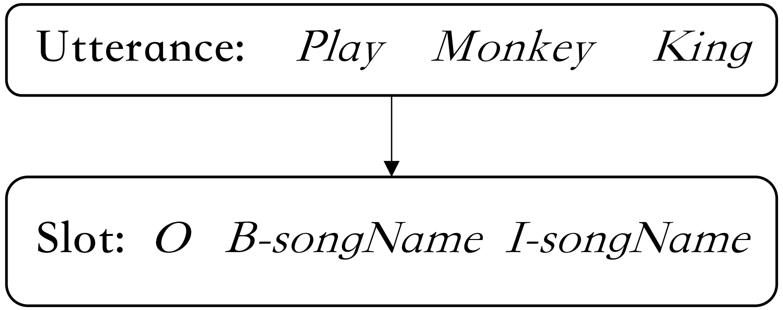
② Low-resource SLU (Unsupervised SLU)

- First, use the existing linguistic annotation models (e.g., NER) to identify potential slot candidates
- Then, automatically identify domain-relevant slots by using clustering algorithms.
- Finally, use the resulting slot annotation to train a slot filling model that achieves to perform slot tagging with no human intervention



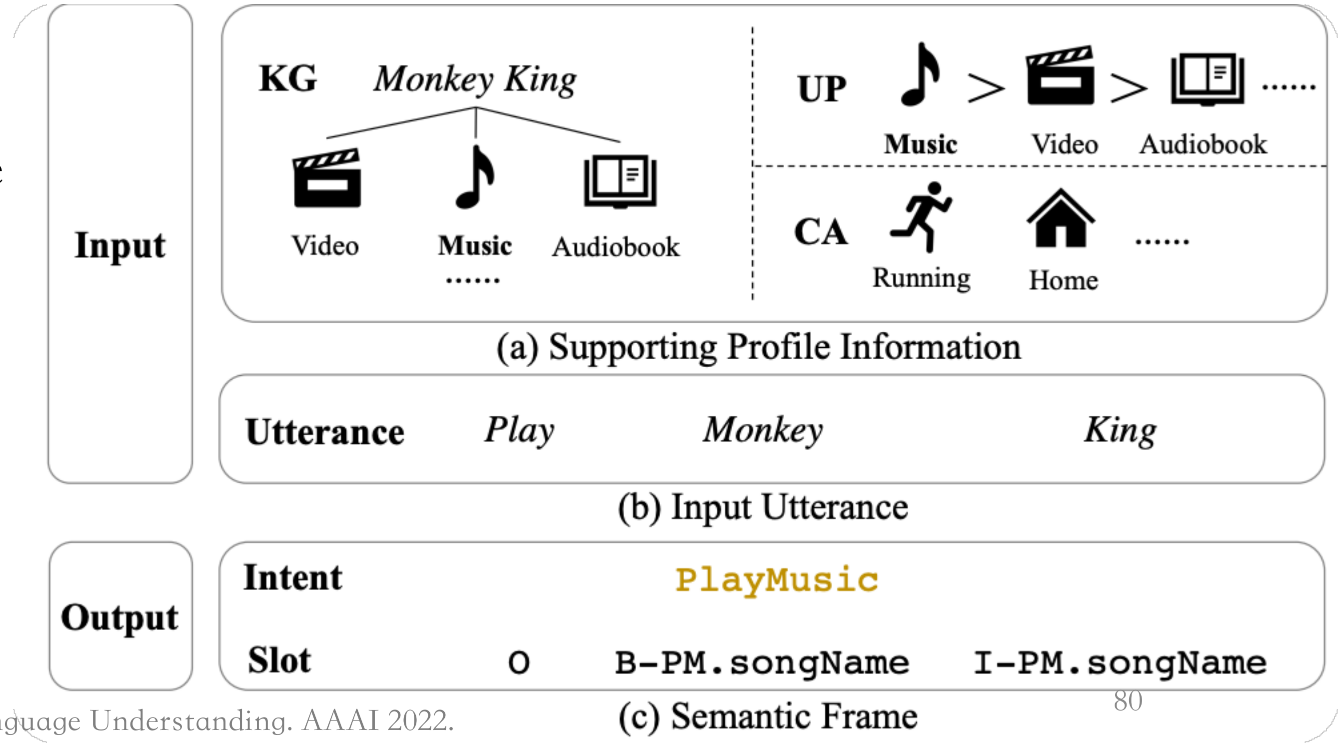
③ Real-world SLU (Profile-based SLU)

- Traditional SLU
  - Input: Plain Text
  - Output: Semantic Parsing Result



VS.

- Profile-based SLU
  - Input: Plain Text + Supporting Profile Information
  - Output: Semantic Parsing Result



Xu et al. Text is no more Enough! A Benchmark for Profile-based Spoken Language Understanding. AAAI 2022.

③ Real-world SLU (Profile-based SLU)

- Profile-based SLU requires a model to rely not only on the surface utterance but also on the supporting information
- Supporting Information
  - Knowledge Graph (KG): large amounts of interlinked entities and their corresponding rich attributes
  - User Profile (UP): a collection of settings and information (items) associated with the user
  - Context Awareness (CA): denotes the user state and environmental information, including the user's movement state, posture, geographic location, etc

Input	
Utterance	*Play Monkey King*
KG	`Mention` "*Monkey King*": {music, video and audiobook}, ...
UP	`Preference` for [music, video, audiobook]: [0.5, 0.3, 0.2], ...
CA	`Movement State`: Running, `Geographic Location`: Home, ...
Output	
Intent	`PlayMusic`
Slot	`O B-PlayMusic.songName I-PlayMusic.songName`

Xu et al. Text is no more Enough! A Benchmark for Profile-based Spoken Language Understanding. AAAI 2022.

③ Real-world SLU (Profile-based SLU)

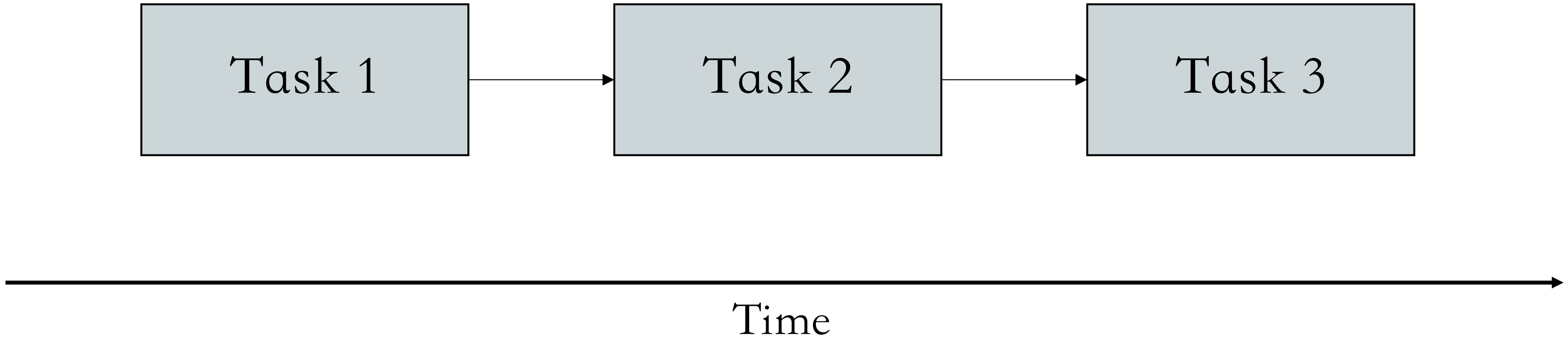
- All baseline including non pre-trained and pre-trained models significantly drop a lot, which show the existing models can not support the profile-based SLU
- It can be seen that the performance of all models improve significantly by a large margin by incorporating supporting profile information

Model	w/o Profile			w/ Profile		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Non Pre-trained SLU Models						
Slot-Gated (Goo et al. 2018)	36.53	41.24	32.02	74.18	83.24	69.11
Bi-Model (Wang, Shen, and Jin 2018)	37.37	44.63	32.58	77.76	82.30	73.45
SF-ID (E et al. 2019)	39.63	42.37	30.89	73.70	83.24	68.36
Stack-Propagation (Qin et al. 2019)	39.29	39.74	36.35	81.08	83.99	78.91
General SLU Model	42.24	43.13	37.85	83.27	85.31	79.10
Pre-trained-based SLU Models						
BERT (Devlin et al. 2019)	44.80	45.76	42.18	82.51	84.56	80.98
XLNet (Yang et al. 2019)	**46.92**	**48.59**	**43.88**	83.39	85.88	81.73
RoBERTa (Liu et al. 2019)	45.92	47.83	43.13	82.90	85.31	81.17
ELECTRA (Clark et al. 2020)	46.48	47.46	42.56	**84.38**	**86.63**	**82.30**

Table 3: Slot Filling and Intent Detection results on the PROSLU dataset.

Xu et al. Text is no more Enough! A Benchmark for Profile-based Spoken Language Understanding. AAAI 2022.

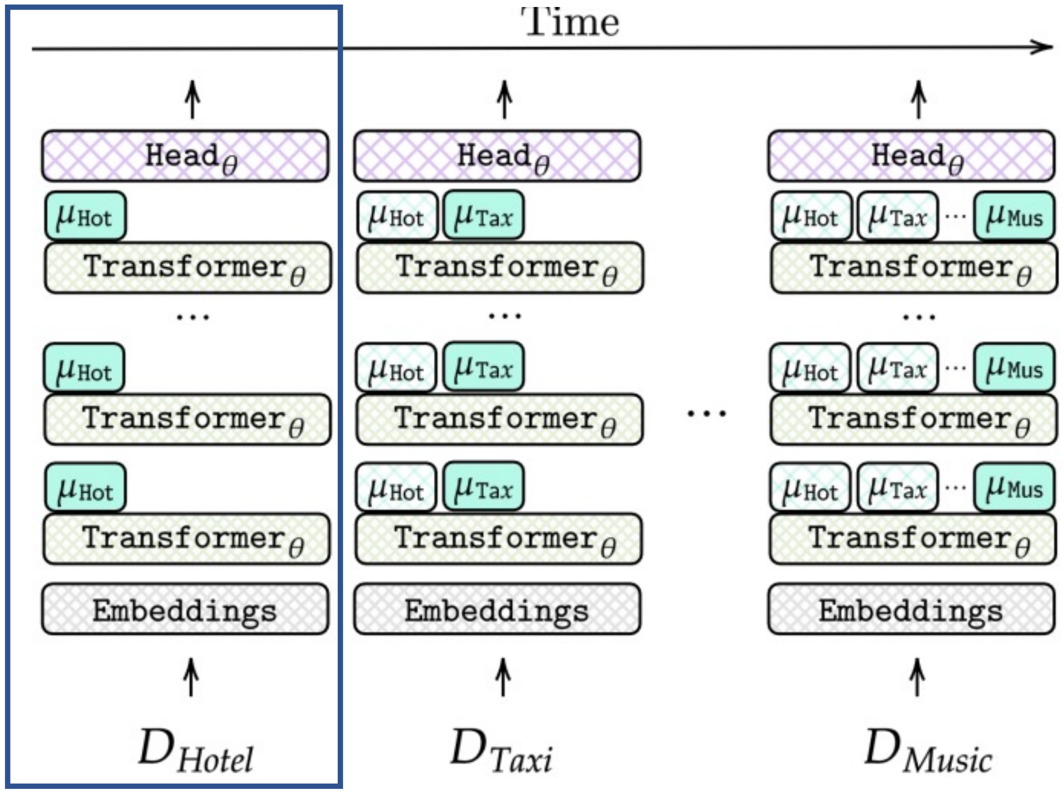
③ Real-world SLU (Lifelong Learning)



How to learn from new task without forgetting knowledge learned from previous task?

③ Real-world SLU (Lifelong Learning)

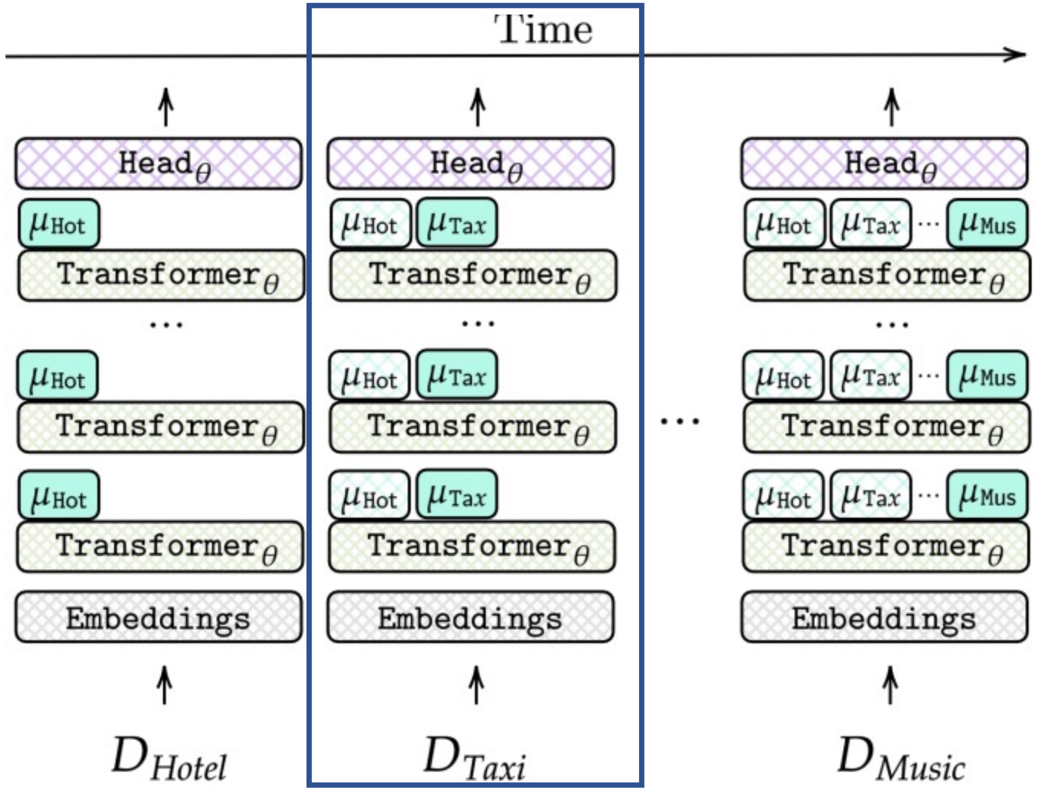
- Only train task-specific parameters and the original weights are left frozen



Madotto et al. Continual Learning in Task-Oriented Dialogue Systems. EMNLP 2020.

③ Real-world SLU (Lifelong Learning)

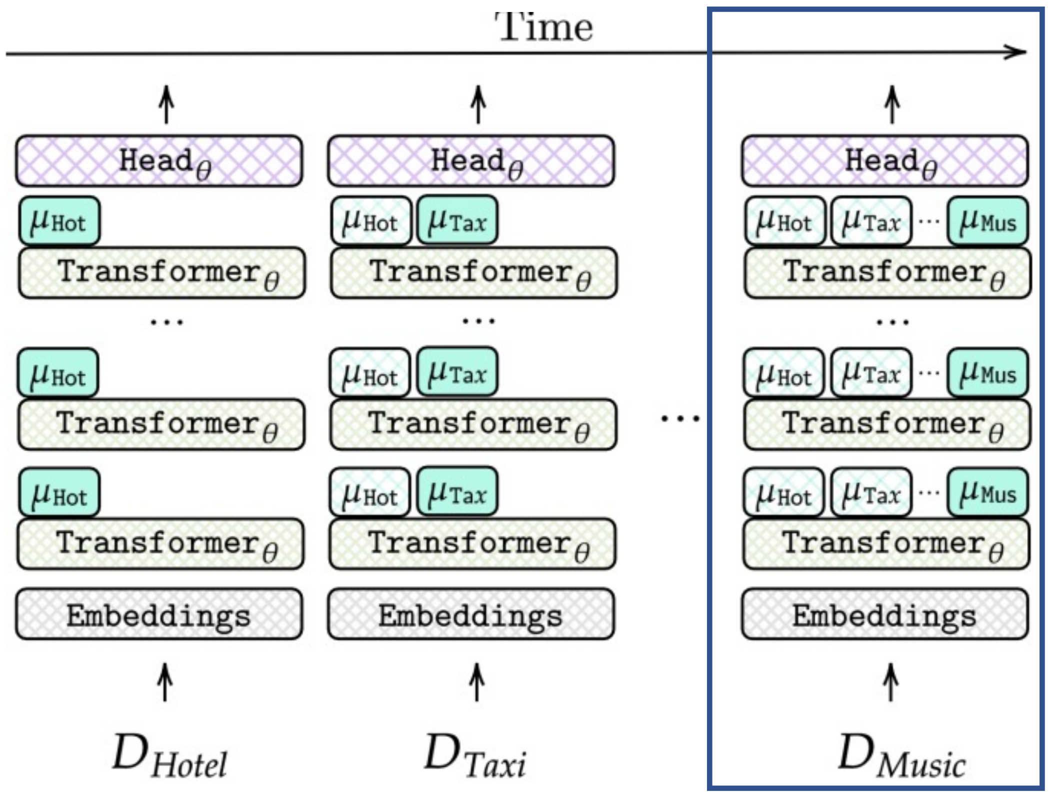
- Only train task-specific parameters and the original weights are left frozen



Madotto et al. Continual Learning in Task-Oriented Dialogue Systems. EMNLP 2020.

③ Real-world SLU (Lifelong Learning)

- Only train task-specific parameters and the original weights are left frozen

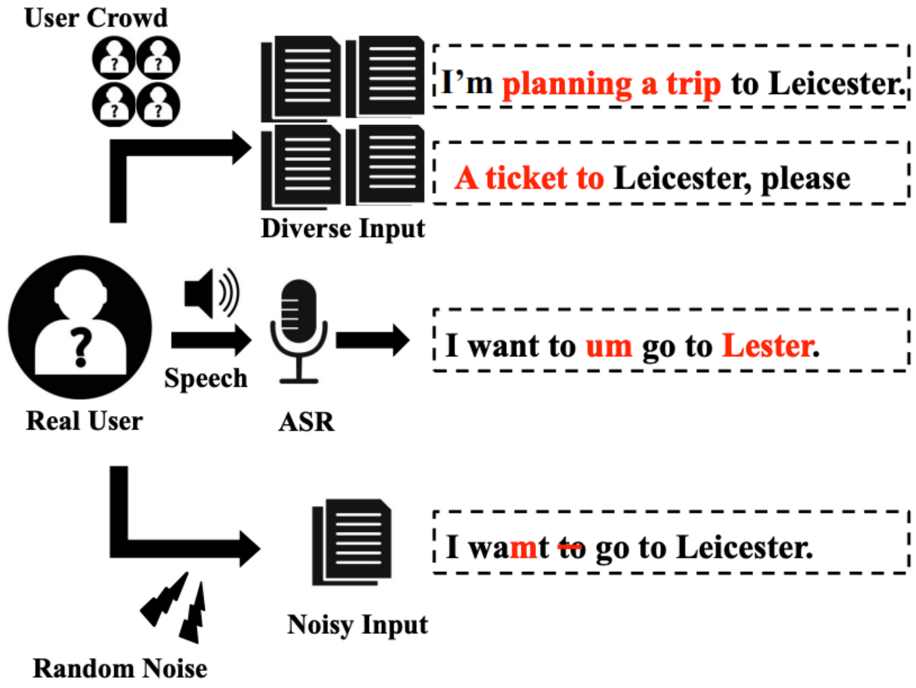


Madotto et al. Continual Learning in Task-Oriented Dialogue Systems. EMNLP 2020.

③ Real-world SLU (Robust SLU)

- This work make a comprehensive study on robustness of task-oriented dialogue system including SLU task

Training Step



Test Step

Liu et al. Robustness Testing of Language Understanding in Task-Oriented Dialog. ACL 2021.

③ Real-world SLU (Robust SLU)

- The current models lack robustness

Model	Train	Ori.	WP	TP	SR	SD	Avg.	Drop	Recov.
MILU	Original	74.15	71.05	69.58	61.53	65.27	66.86	-7.29	/
	Augmented	75.78	72.49	71.96	64.76	70.92	70.03	-5.75	+3.17
BERT	Original	78.82	75.92	74.57	70.31	70.31	72.78	-6.04	/
	Augmented	78.21	76.70	75.63	72.04	77.34	75.43	-2.78	+2.65
ToD-BERT	Original	**80.61**	77.30	76.19	70.88	71.94	74.08	-6.53	/
	Augmented	80.37	**77.32**	**77.26**	**72.54**	**79.04**	**76.54**	-3.83	+2.46
CopyNet	Original	67.84	63.90	61.41	56.11	59.26	60.17	-7.67	/
	Augmented	69.35	67.10	65.90	60.98	67.71	65.42	-3.93	**+5.25**
GPT-2	Original	78.78	74.96	72.85	69.00	69.19	71.50	-7.28	/
	Augmented	79.15	75.25	73.86	71.37	74.19	73.67	-5.48	+2.17

(a) Frames

How to improve the robustness of SLU model is a very important research topic

Liu et al. Robustness Testing of Language Understanding in Task-Oriented Dialog. ACL 2021.

③ Real-world SLU (Chinese SLU)

English SLU

watch action movie

English SLU model

Semantic Parsing Results

VS.

Chinese SLU

周冬雨 (Dongyu Zhou) / 有 (has) / 哪些 (which) / 电影 (movies)”

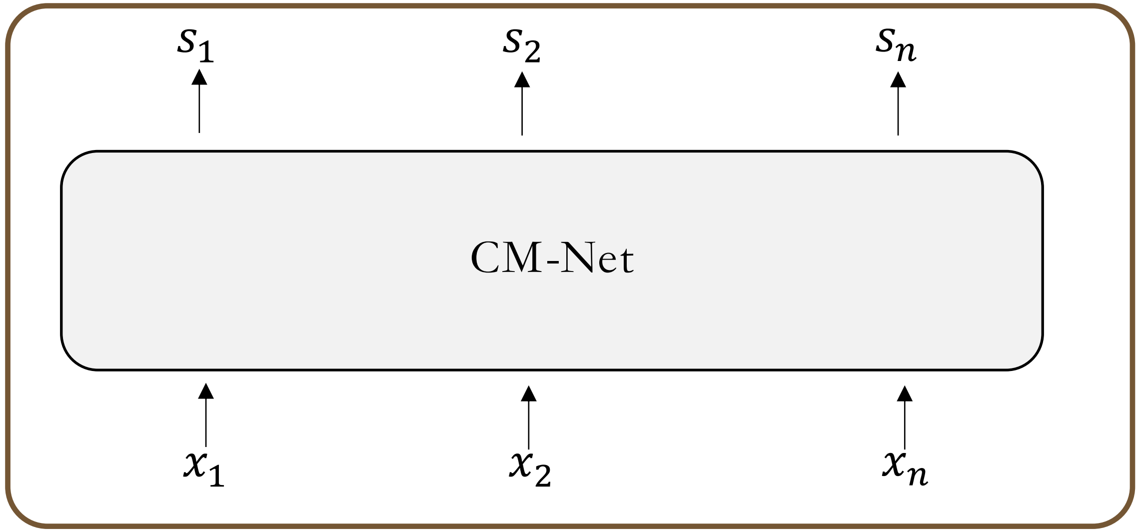
Chinese SLU model

Semantic Parsing Results

How to leverage both character and word feature for Chinese SLU

③ Real-world SLU (Chinese SLU)

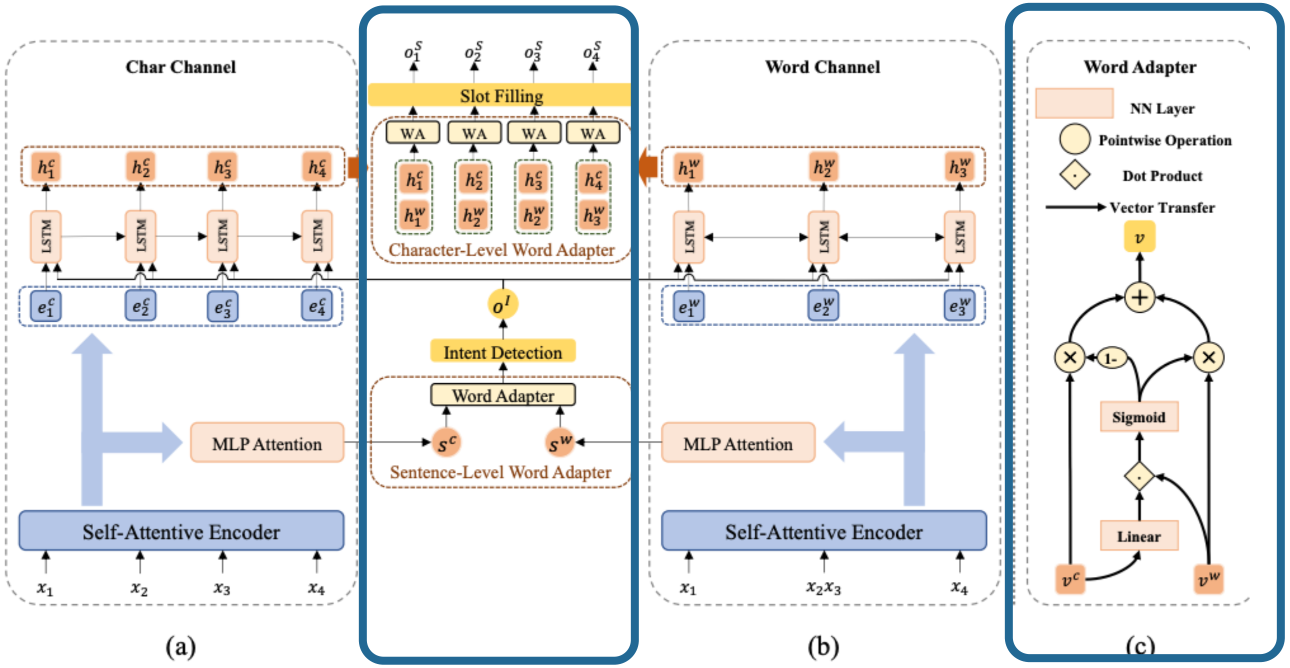
- Character-level model for Chinese SLU, which is able to utilize the character information
- Ignore word information for Chinese SLU



Liu et al. CM-Net: A Novel Collaborative Memory Network for Spoken Language Understanding. EMNLP 2019.

③ Real-world SLU (Chinese SLU)

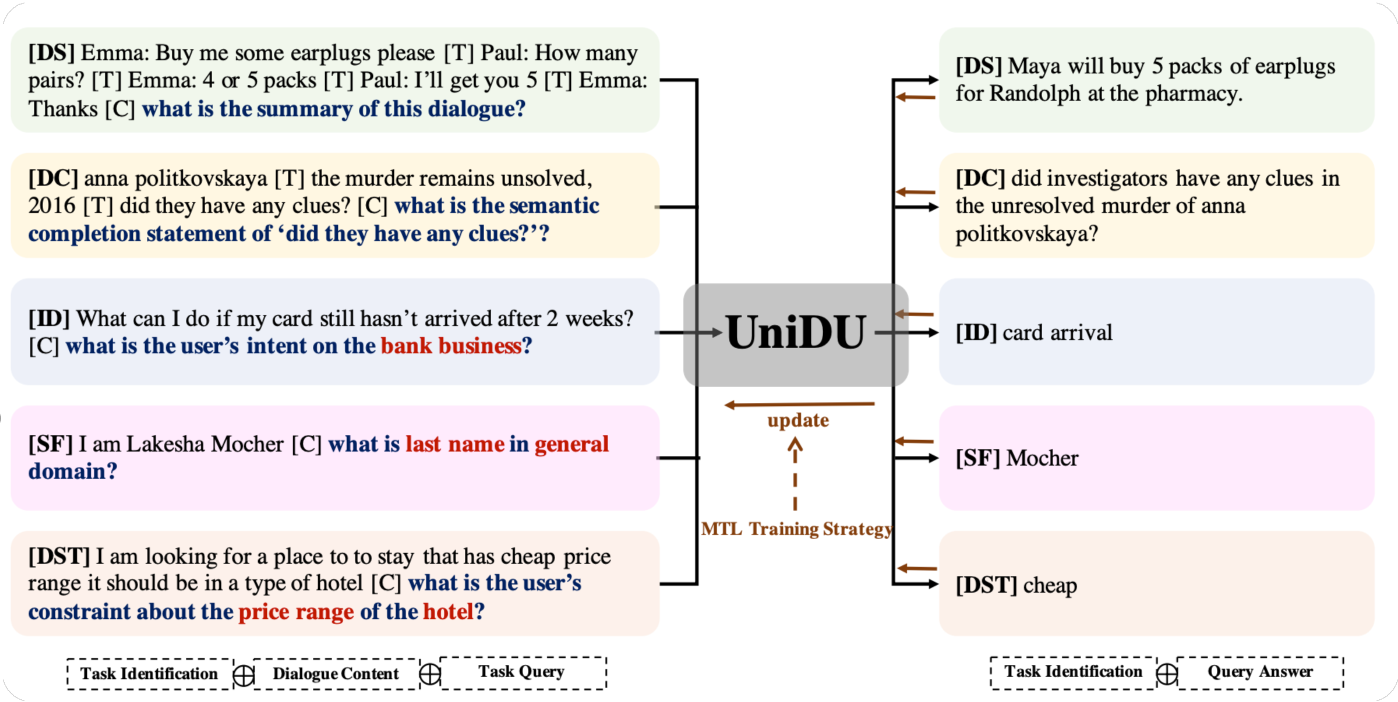
- Multi-level adapter for injecting both character and word information
  - Character-level word adapter for slot filling
  - Sentence-level word adapter for intent detection



Teng et al. INJECTING WORD INFORMATION WITH MULTI-LEVEL WORD ADAPTER FOR CHINESE SPOKEN LANGUAGE UNDERSTANDING. ICASSP 2021.

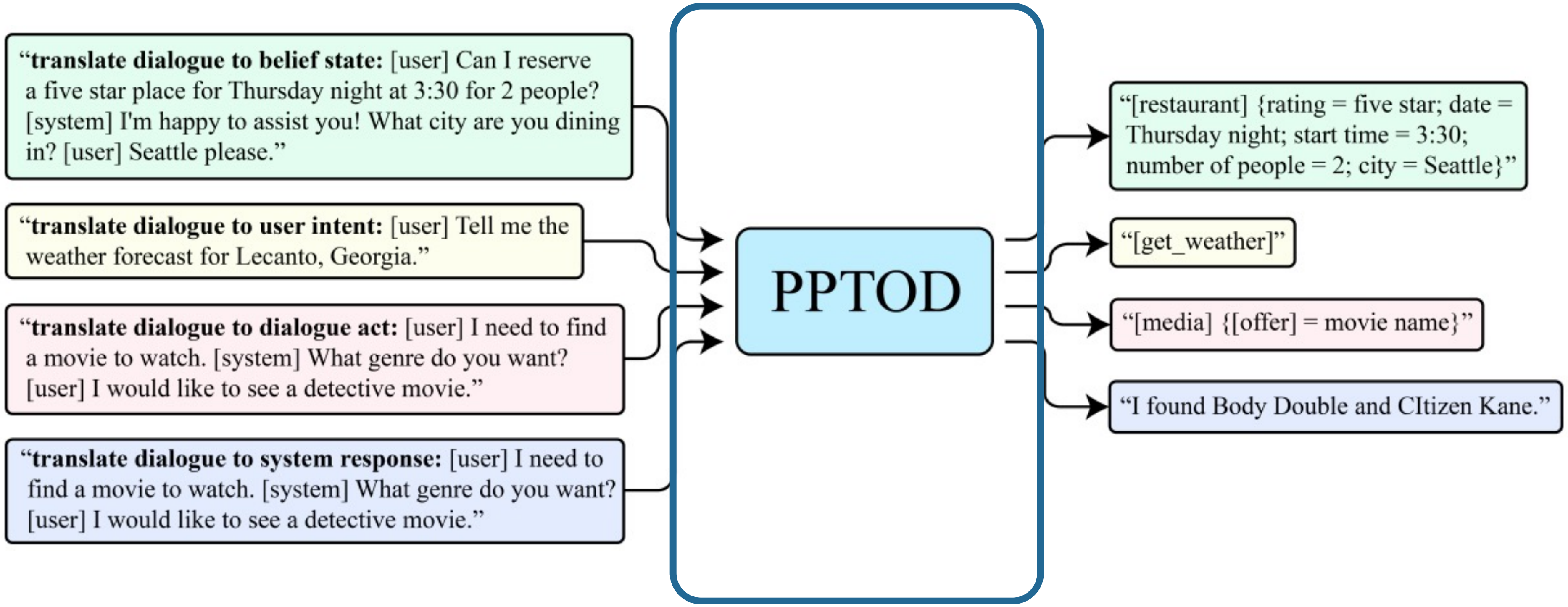
④ Unified and Pre-training Paradigm (Unified Model)

- Different dialogue understanding (DU) tasks as the unified generation including five DU tasks
  - Dialogue Summary (DS)
  - Dialogue Completion (DC)
  - Slot Filling (SF)
  - Intent Detection (ID)
  - Dialogue State Tracking (DST)



④ Unified and Pre-training Paradigm (Pre-training Model)

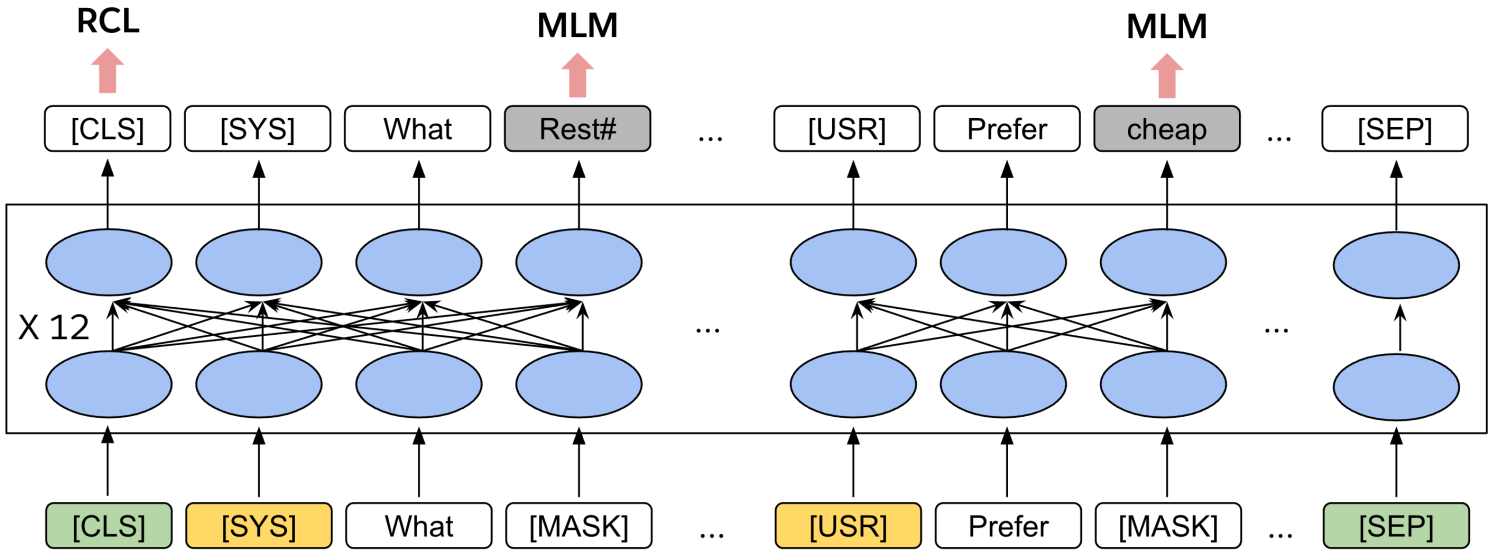
- A unified model to model all components in pipeline task-oriented dialogue including dialogue state tracking, intent detection, dialogue act recognition and response generation



Su et al. Multi-Task Pre-Training for Plug-and-Play Task-Oriented Dialogue System. ACL 2022.

④ Unified and Pre-training Paradigm (Pre-training Model)

- A pre-trained dialogue BERT model for task-oriented dialogue using two pre-training task
  - Masked language modeling
  - Response contrastive loss



Wu et al. TOD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogue. EMNLP 2020.

④ Unified and Pre-training Paradigm (Pre-training Model)

- Masked Language Modeling
  - help model to capture the dialogue domain feature

- Response Contrastive Loss
  - encourage the model to capture underlying dialogue sequential order, structure information, and response similarity.

	Domain (acc)	**Intent** (acc)	**Dialogue Act** (F1-micro)
GPT2	63.5%	74.7%	85.7%
DialoGPT	63.0%	65.7%	84.2%
BERT	60.5%	71.1%	85.3%
TOD-BERT-mlm	63.9%	70.7%	83.5%
TOD-BERT-jnt	**68.7%**	**77.8%**	**86.2%**

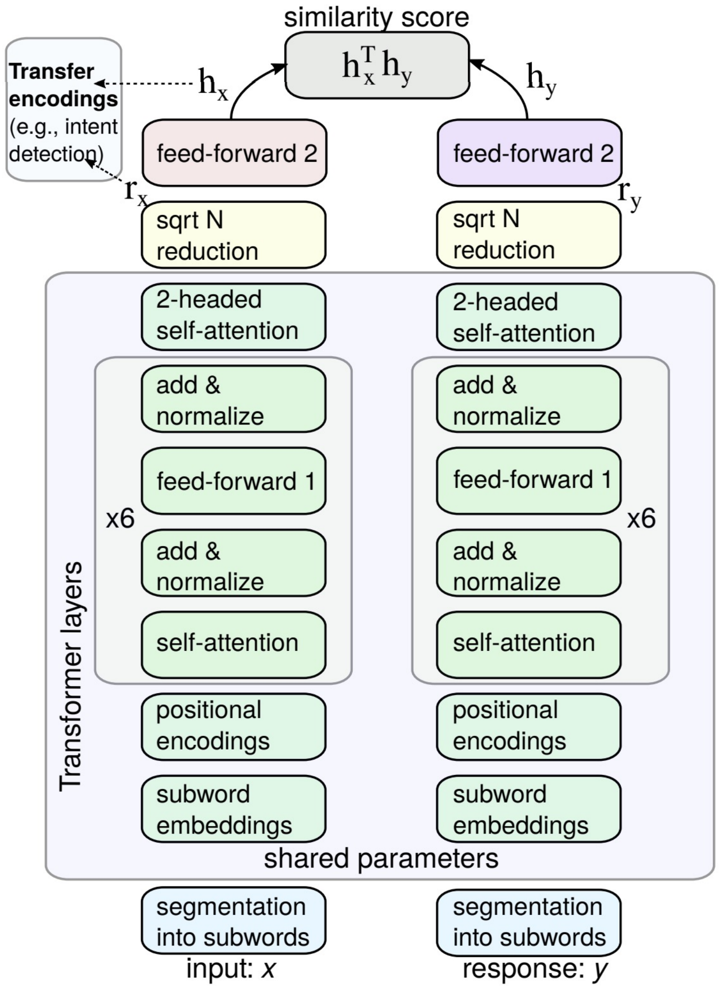
Wu et al. TOD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogue. EMNLP 2020.

④ Unified and Pre-training Paradigm (Pre-training Model)

- A pretraining framework for conversational tasks that is effective, affordable, and quick to train
  - Response selection task

	Banking	Shopping	Company FAQ
USE-LARGE	92.2	94.0	62.4
BERT-LARGE	93.2	94.3	61.2
ConveRT	92.7	94.5	64.3

Table 6: Intent classification results.



Henderson et al. ConveRT: Efficient and Accurate Conversational Representations from Transformers. EMNLP 2020 Findings.

Conclusion and Highlight

Libo Qin

PhD Student

Harbin Institute of Technology

Conclusion and Highlight

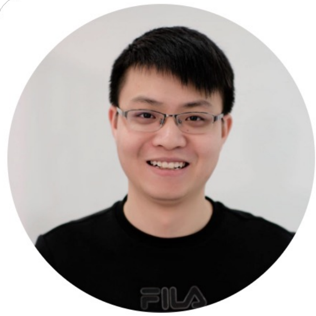
- Spoken Language Understanding has attracted more and more attention in both academic and industry.

- Remarkable success have been achieved on single-turn, single domain and single-intent SLU direction.

- Multi-intent SLU, Robust SLU, Profile-based SLU and other real-world scenario applications gradually become the future trend.

- Resource: https://github.com/yizhen20133868/Awesome-SLU-Survey

Thanks & QA



Libo Qin
PhD Student
Harbin Institute of Technology

Dr. Wanxiang Che
Professor
Harbin Institute of Technology

Dr. Zhou Yu
Assistant Professor
Columbia University