# PIN: A Novel Parallel Interactive Network for Spoken Language Understanding

Peilin Zhou, Zhiqi Huang, Fenglin Liu, Yuexian Zou*

ADSPLAB, School of ECE, Peking University

# Introduction

## Spoken Language Understanding

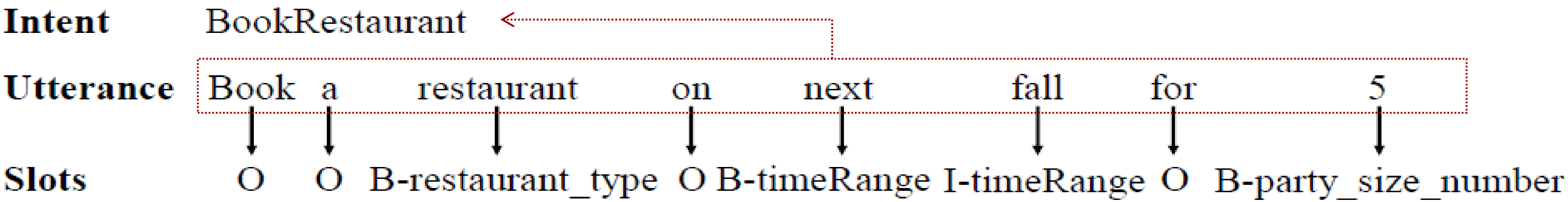
- It typically involves intent detection (ID) and slot filling (SF) tasks.

## Intent Detection

- ID is a semantic **text classification** problem
- Learning: $f: X \rightarrow Y$ that maps an input sequence $x$ to its label category $y$

## Slot Filling

- SF is often modeled as a **sequence labeling task** with explicit alignment
- Learning: $f: X \rightarrow Y$ that maps an input sequence $x$ to its label sequence $y$

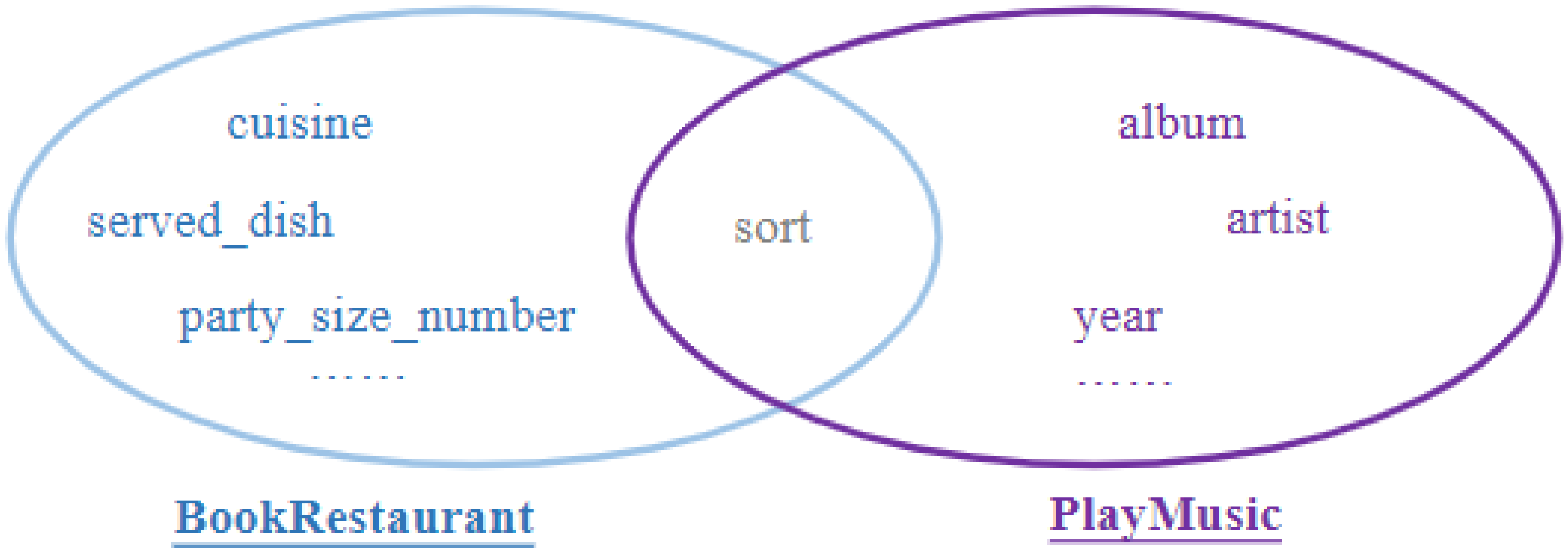



PEKING UNIVERSITY

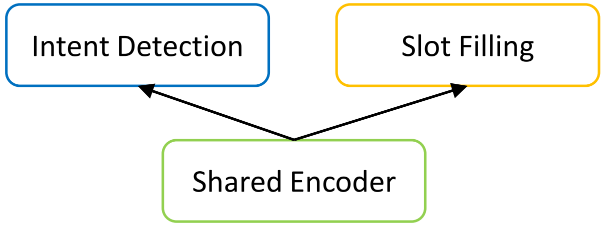
# Introduction

## Spoken Language Understanding

- Intent detection and slot filling are associated with each other



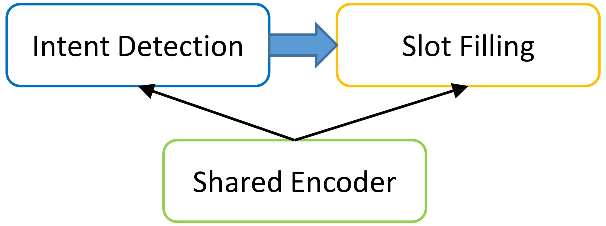
# Introduction




Liu and Lane (2016)

## Limitation & Challenge:

- Local context information is not fully exploited in their models, ignoring the intuition that local context is a useful architectural inductive prior for SF.
- Many methods fail to take full advantage of the supervised signals due to their implicit or unidirectional modeling style of the intent-slot relations.




Goo et al.(2018)
Qin et al.(2019)

## Motivation of our work:

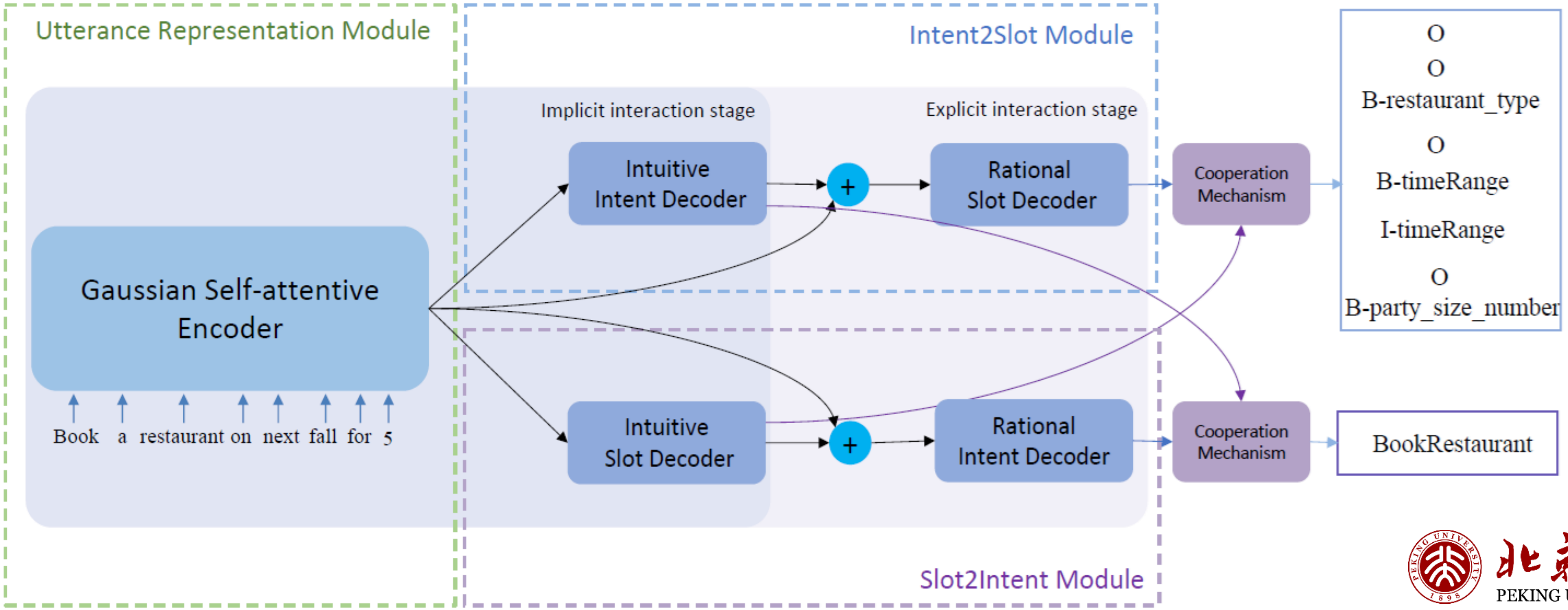
We propose a novel Parallel Interactive Network (PIN) to address above issues:

- A Gaussian self-attentive encoder is introduced to better capture local structure and contextual information at each token
- We design a Intent2Slot module and a Slot2Intent module to model the bidirectional information flow between SF and ID.

# Proposed method

## Parallel Interactive Network (PIN)

- The PIN consists of the **Utterance Representation Module**, the **Intent2Slot Module**, the **Slot2Intent Module** and a **Cooperation Mechanism**.



# Parallel Interactive Network (PIN)

## Utterance Representation Module

- We use BiLSTM with Gaussian self-attention mechanism to leverage both advantages of local structure and contextual information for a given utterance, which are useful for ID and SF tasks.

$\mathbf{E} = \mathbf{H} \oplus \mathbf{C}$

$\mathbf{H} = (h_1, h_2, \ldots, h_T)$ $\quad \mathbf{C} = (c_1, c_2, \ldots, c_T)$

$\overrightarrow{h}_i = \overrightarrow{\mathrm{LSTM}}\left(\phi^{\mathrm{emb}}(x_i), \overrightarrow{h}_{i-1}\right)$ $\quad \overleftarrow{h}_i = \overleftarrow{\mathrm{LSTM}}(\phi^{\mathrm{emb}}(x_i), \overleftarrow{h}_{i+1})$

$h_i = \overrightarrow{h}_i \oplus \overleftarrow{h}_i$

$c_i = \sum_i \mathrm{Softmax}(-|wd_{i,j}^2 + b| + (x_i \cdot x_j))x_j$

## Slot2Intent Module

- Intuitive Slot Decoder

$\mathbf{h}_t^{IS} = LSTM(\mathbf{h}_{t-1}^{IS}, \mathbf{y}_{t-1}^{IS} \oplus \mathbf{e}_t)$

$\mathbf{y}_t^{IS} = \mathrm{softmax}(\mathbf{W}_h^{IS} \mathbf{h}_t^{IS})$

- Rational Intent Decoder

$\mathbf{h}_t^{RI} = LSTM(\mathbf{h}_{t-1}^{RI}, \mathbf{y}_{t-1}^{RI} \oplus \mathbf{y}_t^{IS} \oplus \mathbf{e}_t)$

$\mathbf{y}_t^{RI} = \mathrm{softmax}(\mathbf{W}_h^{RI} \mathbf{h}_t^{RI})$

## Intent2Slot Module

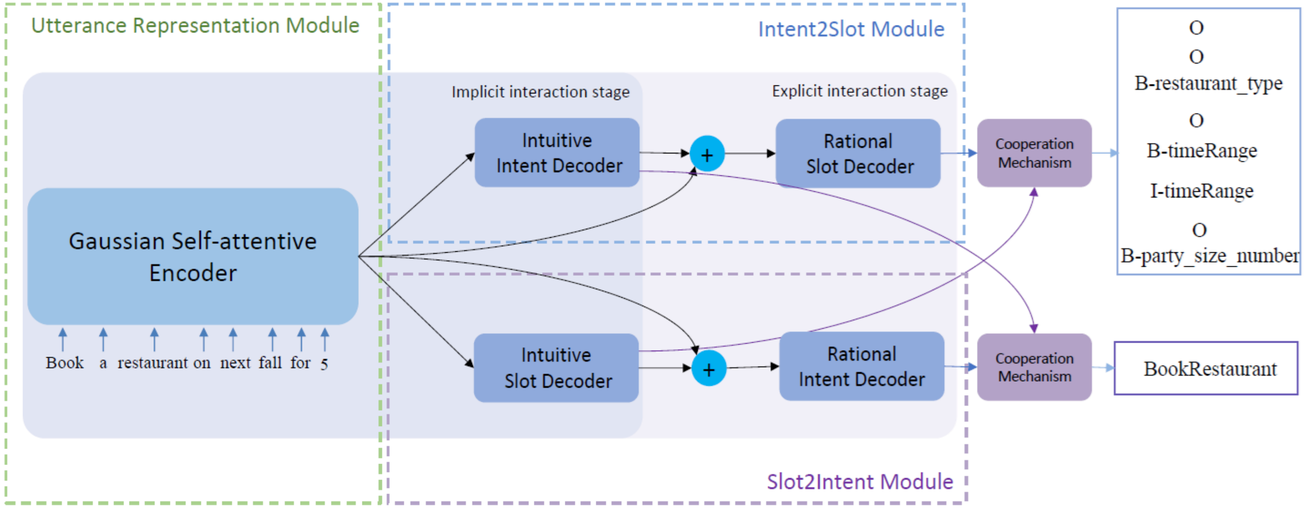
The Intent2Slot Module has the similar structure as the Slot2Intent Module but switches the tasks for the two decoders.

- Intuitive Intent Decoder
- Rational Slot Decoder

## Cooperation Mechanism

$r_t^S = \mathrm{softmax}(\mathrm{MLP}(\mathbf{h}_t^{RS}))$ $\quad r_t^I = \mathrm{softmax}(\mathrm{MLP}(\mathbf{h}_t^{RI}))$

$h_t^S = \mathbf{h}_t^{RS} \odot r_t^S + \mathbf{h}_t^{IS} \odot (1 - r_t^S)$ $\quad h^I = \sum_{t=1}^{T} \mathbf{h}_t^{RI} \odot r_t^I + \mathbf{h}_t^{II} \odot (1 - r_t^I)$



# Experiment Results

| Model | SNIPS | | | ATIS | | |
|---|---|---|---|---|---|---|
| | Intent (Err) | Slot (F1) | Overall (Acc) | Intent (Err) | Slot (F1) | Overall (Acc) |
| Recursive NN [43] | 2.7 | 88.3 | - | 4.6 | 94.0 | - |
| Dilated CNN, Label-Recurrent [44] | 1.7 | 93.1 | - | **1.9** | 95.5 | - |
| Attention Bi-RNN [5] | 3.3 | 87.8 | 74.1 | 8.9 | 94.2 | 78.9 |
| Joint Seq2Seq [7] | 3.1 | 87.3 | 73.2 | 7.4 | 94.2 | 80.7 |
| Slot-Gated Model [4] | 3.0 | 88.8 | 75.5 | 6.4 | 94.8 | 82.2 |
| Stack-Propagation [36] | 2.0 | 94.2 | 86.9 | 3.1 | 95.9 | 86.5 |
| SF-ID,SF first [38] | 2.6 | 91.4 | 80.6 | 2.2 | 95.8 | 86.8 |
| SF-ID,ID first [38] | 2.7 | 92.2 | 80.4 | 2.9 | 95.8 | 86.9 |
| Graph LSTM [45] | 2.3 | 93.8 | 85.6 | 3.6 | 95.8 | 86.2 |
| PIN (our model) | 0.9 | 94.5 | 88.0 | 2.8 | 95.9 | 87.1 |
| Joint BERT [46] | 1.4 | 97.0 | 92.8 | 2.5 | 96.1 | 88.2 |
| Graph LSTM + ELMo [45] | 1.7 | 95.3 | 89.7 | 2.8 | 95.9 | 87.6 |
| Stack-Propagation + BERT [36] | 1.0 | 97.0 | 92.9 | 2.5 | 96.1 | 88.6 |
| PIN(our model) + BERT | **0.8** | **97.1** | **93.2** | 2.2 | **96.3** | **88.8** |

TABLE 1: Experiment results of our model and the baselines on two benchmark datasets.

| Model | SNIPS | | | ATIS | | |
|---|---|---|---|---|---|---|
| | Intent (Err) | Slot (F1) | Overall (Acc) | Intent (Err) | Slot (F1) | Overall (Acc) |
| w/o Slot2Intent module | 3.1 | **95.8** | 86.5 | 3.1 | 95.7 | 86.5 |
| w/o Intent2Slot module | 2.0 | 94.3 | 87.0 | 3.0 | 95.7 | 86.7 |
| w/o Gaussian self-attention | 2.3 | 92.9 | 84.4 | 3.1 | 94.9 | 85.0 |
| w/o cooperation mechanism | 1.4 | 94.3 | 87.4 | 3.4 | 95.9 | 87.0 |
| Full PIN model | **0.9** | 94.5 | **88.0** | **2.8** | **95.9** | **87.1** |

TABLE 2 Ablation experiments on two benchmarks to investigate the impacts of various components.

# Conclusion

- We propose a novel parallel interactive network (PIN) for spoken language understanding
- PIN could support bidirectional and explicit information exchange between ID and SF while reduce the prediction bias.
- The experimental results demonstrate the effectiveness of our approach, which outperforms all comparison methods in terms of most metrics on the two publicly benchmark datasets.

# Thank you!

If you have any questions about the paper, you can send an email to zhoupl@pku.edu.cn