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 $y$

<table>
<thead>
<tr>
<th>Intent</th>
<th>BookRestaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utterance</td>
<td>Book a restaurant on next fall for 5</td>
</tr>
<tr>
<td>Slots</td>
<td>O O B-restaurant_type O B-timeRange I-timeRange O B-party_size_number</td>
</tr>
</tbody>
</table>
Introduction

Spoken Language Understanding

- Intent detection and slot filling are associated with each other
**Introduction**

**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.

**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.

\[
E = H \oplus C \\
H = (h_1, h_2, ..., h_T) \\
C = (c_1, c_2, ..., c_T)
\]

\[
h_i = \text{LSTM}(\phi^\text{emb}(x_i), h_{i-1}) \\
h_t = \text{LSTM}(\phi^\text{emb}(x_i), h_{t+1}) \\
c_i = \sum_j \text{Softmax}(x(-w_0d f_j + b + (x_1 \cdot x_j))x_j)
\]

**Slot2Intent Module**

- Intuitive Slot Decoder
  \[ h_i^S = \text{LSTM}(h_{t-1}^S, y_{t-1}^S \oplus e_i) \]
  \[ y_t^S = \text{softmax}(W_i^h h_t^S) \]

- Rational Intent Decoder
  \[ h_t^R = \text{LSTM}(h_{t-1}^R, y_{t-1}^R \oplus y_t^S \oplus e_i) \]
  \[ y_t^R = \text{softmax}(W_i^h h_t^R) \]

**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**

\[ r_t^S = \text{softmax}(\text{MLP}(h_t^S)) \]
\[ r_t^I = \text{softmax}(\text{MLP}(h_t^I)) \]
\[ h_t^S = h_t^S \odot r_t^S + h_t^S \odot (1 - r_t^S) \]
\[ h_t^I = \sum_{t=1} h_t^I \odot r_t^I + h_t^I \odot (1 - r_t^I) \]
# Experiment Results

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

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

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

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