A Joint Multi-Task Learning Framework For Spoken Language Understanding

Changliang Li[1], Cunliang Kong[1,2], Yan Zhao[1,2]

[1]Institute of Automation, Chinese Academy of Sciences
[2]Beijing Language and Culture University

Contents

1. Introduction

A brief introduction of background and our model.

1.1 SLU

- **S**poken **L**anguage **U**nderstanding
- A crucial part of spoken dialogue system
- Two basic tasks:
  - Slot Filling
  - Intent Determination

1.1.1 Slot Filling

Sentence	*show*	*flights*	*from*	*Boston*	*To*	*New*	*York*	*today*
Slots/Concepts	O	O	O	B-dept	O	B-arr	I-arr	B-date

- Sequence labeling problem
- Traditional machine learning approaches:
  - Hidden Markov models (HMMs)
  - Conditional random field (CRF)
  - … …

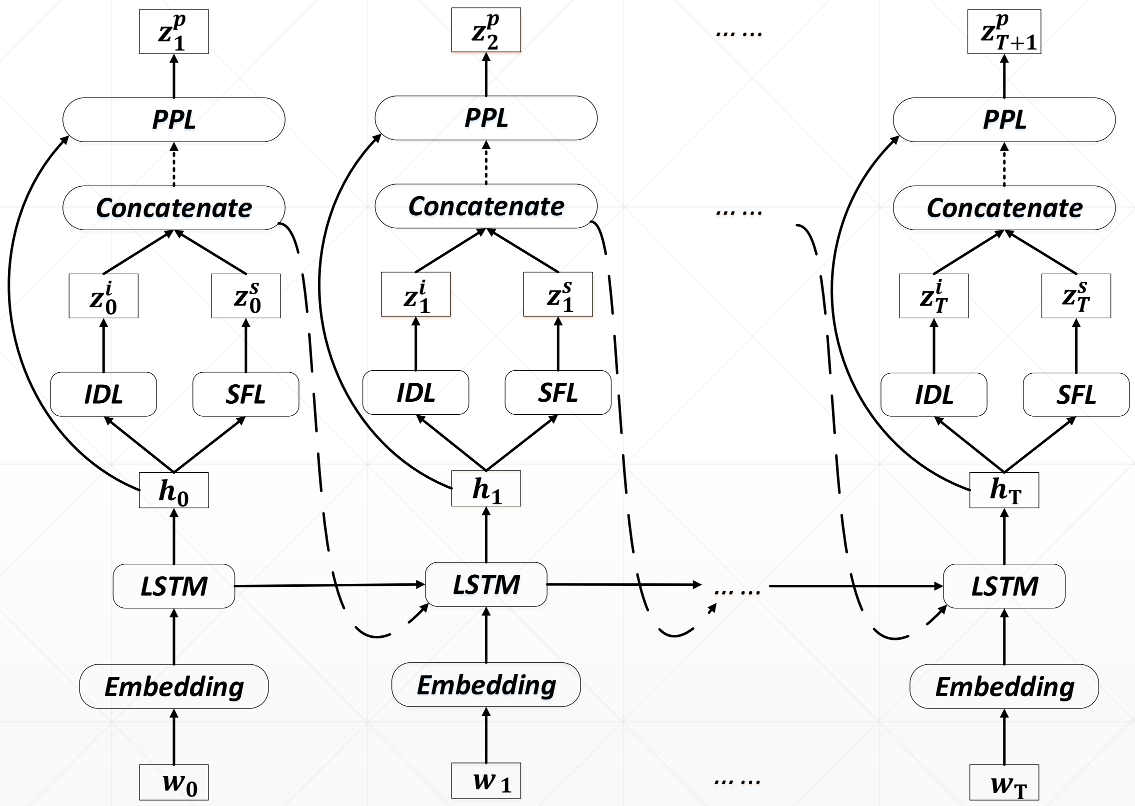
1.1.2 Intent Determination

- Sentence: Show flights from Boston to New York today.
- Intent: Find_Flight

- Classification Problem
- Traditional machine learning approaches:
  - Maximum entropy
  - Support vector machine with linear kernel (LinearSVM)
  - … …

1.4 Our Method

- Joint multi-task learning framework
  - Slot filling
  - Intent determination
  - Part of speech (POS) prediction
- Effectively use of correlation among three tasks
- Additional linguistic information

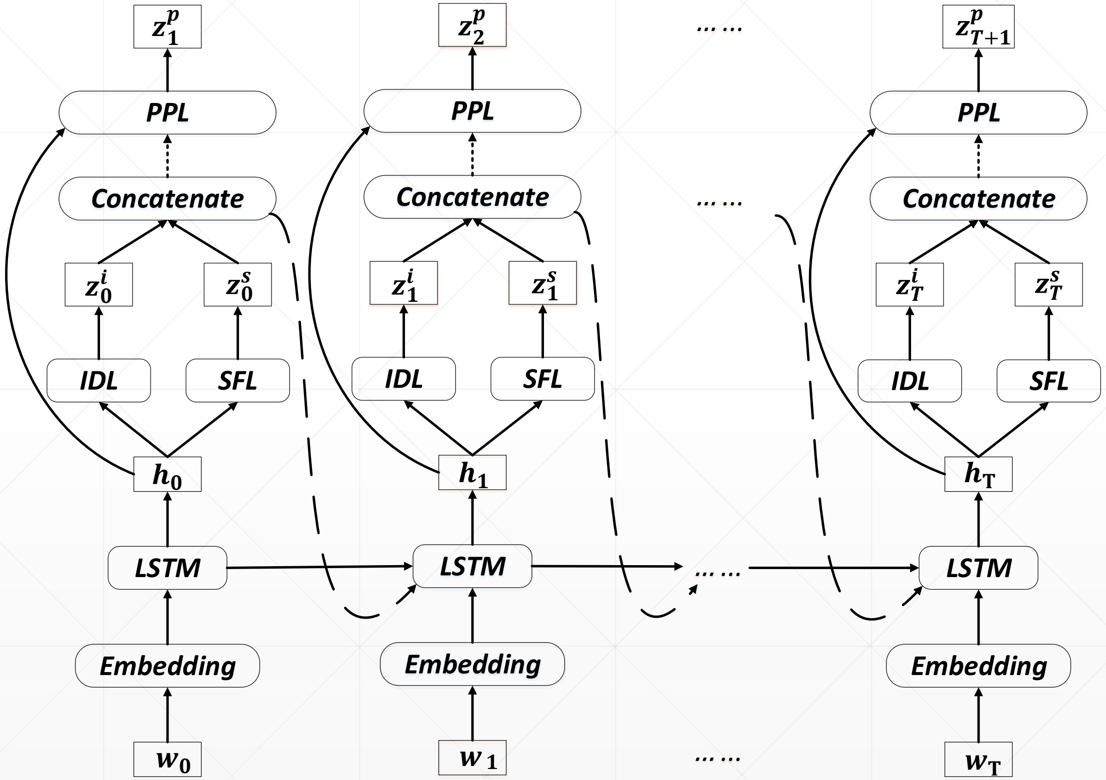


2. Method

Our proposed model and joint learning method.

2.1 Our Proposed Model

- The model consists of five layers
  - Embedding Layer
  - LSTM Layer
  - NLU Module
    - Slot Filling Layer (SFL)
    - Intent Determination Layer (IDL)
  - POS Prediction Layer

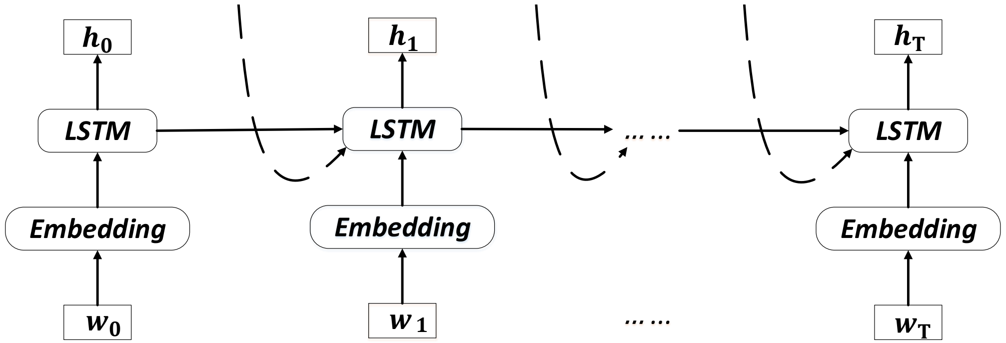


2.1.1 Embedding Layer

- Map input words as word embeddings
- Input
  - A sequence of input words
  - $\mathbf{w} = (w_0, \dots \dots, w_{T+1})$
- Output
  - A sequence of vectors
  - $\mathbf{v} = (v_0, \dots \dots, v_{T+1})$

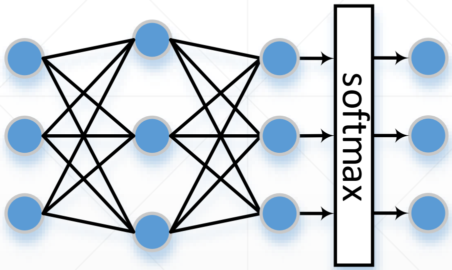
2.1.2 LSTM Layer

- Concatenate these vectors as x_t
  - Current word vector
  - Previous intent label
  - Previous slot label
- Encode these information as h_t
  - $h_t = \mathrm{LSTM}(h_{t-1}, x_t)$

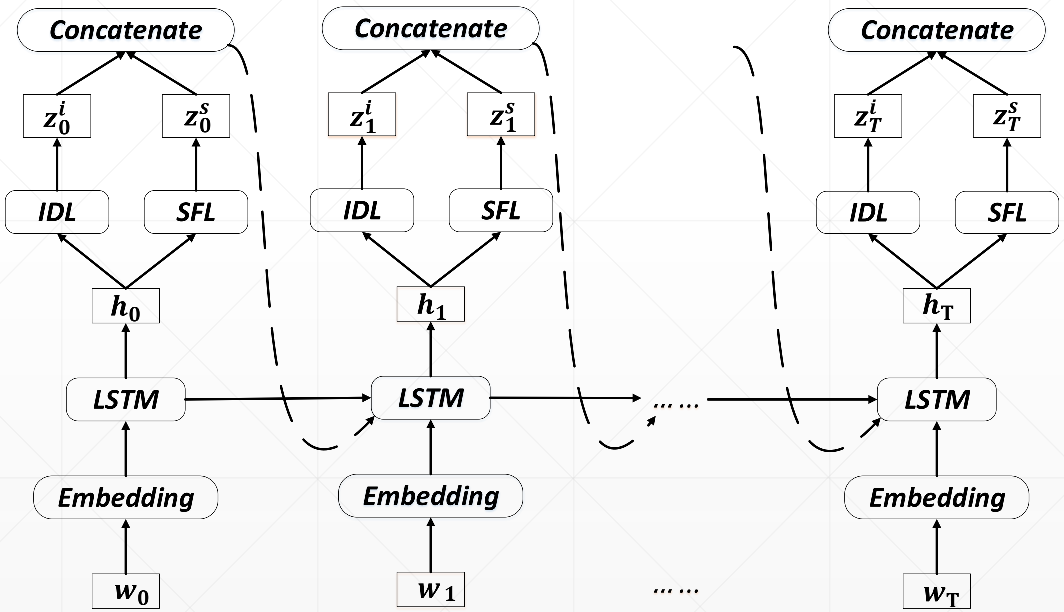


2.1.3 NLU Module

- IDL and SFL have similar architecture
  - A multilayer feedforward neural network
  - A softmax function

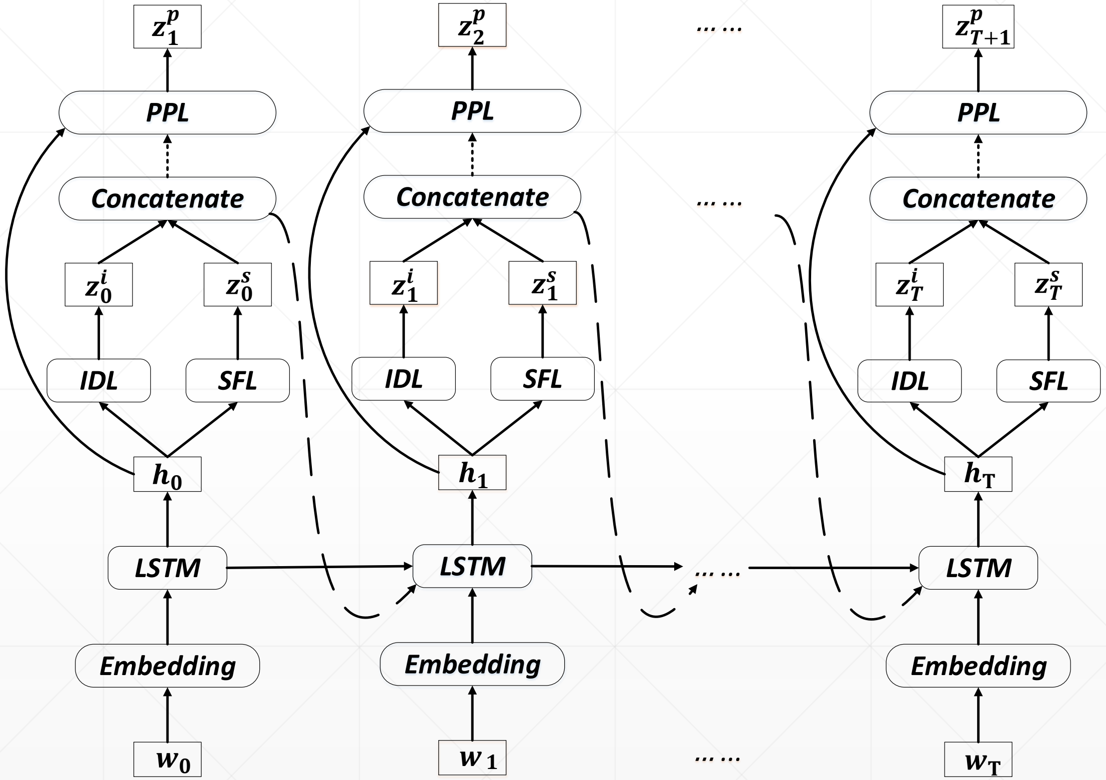


- After the obtain of labels
  - Concatenation
  - Fed into LSTM layer



2.1.4 POS Prediction Layer

- Variation of RNN language model
- Predict the POS tag of next word
- $P(\boldsymbol{y}^{\boldsymbol{p}}|\mathbf{w}) = \prod_{t=0}^{T} P(y_{t+1}^{p}|w_{\leq t}, y_{\leq t}^{i}, y_{\leq t}^{s})$
- Similar architecture with slot filling layer



2.2 Joint Learning

- Cross-entropy loss function
- For three tasks:
- $$\begin{cases} \mathcal{L}^s = -P(\boldsymbol{y}^s)\log P(\boldsymbol{d}^s) \\ \mathcal{L}^i = -P(\boldsymbol{y}^i)\log P(\boldsymbol{d}^i) \\ \mathcal{L}^p = -P(\boldsymbol{y}^p)\log P(\boldsymbol{d}^p) \end{cases}$$
- Joint training:
- $\mathcal{L} = \sum_{\mathcal{D}}\left[\mathcal{L}^s + \mathcal{L}^i + \mathcal{L}^p\right] - \lambda R(\theta)$

3. Experiment

Experiments conducted on ATIS benchmark.

3.1 Dataset and Metrics

- ATIS (Airline Travel Information Systems) dataset
  - 4978 utterances
  - 127 distinct slot labels
  - 18 different intent types
- Metrics
  - Intent error rate
  - Slot filling F1 score
  - Language model perplexity

3.2 Results

	Model	Intent Error	F1 Score	LM PPL
1	RecNN	4.60	93.22	-
2	RecNN+Viterbi	4.60	93.96	-
3	Independent training RNN intent model	2.13	-	-
4	Independent training RNN slot filling model	-	**94.91**	-
5	Independent training RNN language model	-	-	11.55
6	Basic joint training model	2.02	94.15	11.33
7	Joint model with *local* intent context	1.90	94.22	11.27
8	Joint model with *recurrent* intent context	1.90	94.16	10.21
9	Joint model with *local & recurrent* intent context	1.79	94.18	10.22
10	Joint model with *local* slot label context	1.79	94.14	11.14
11	Joint model with *recurrent* slot label context	1.79	94.64	11.19
12	Joint model with *local & recurrent* slot label context	1.68	94.52	11.17
13	Joint model with *local* intent + slot label context	1.90	94.13	11.22
14	Joint model with *recurrent* intent + slot label context	1.57	94.47	10.19
15	Joint model with *local & recurrent* intent + slot label context	1.68	94.45	10.28
16	SLU-LM-POS model with *recurrent* intent + slot label context	1.68	94.57	**2.89**
17	SLU-LM-POS model with *local & recurrent* intent + slot label context	**1.46**	94.81	2.92

4. Conclusion

The advantages of our proposed model.

4. Conclusion

- Joint Learning
  - Make full use of the correlation among multi tasks
  - Obtain additional linguistic information
- State-of-the-art among multi tasks on ATIS dataset.

Q&A

Thanks for Your Attention!