

# NEURAL ATTENTION MODELS FOR SEQUENCE CLASSIFICATION : ANALYSIS AND APPLICATION TO KEY TERM EXTRACTION AND DIALOGUE ACT DETECTION

Sheng-syun Shen, Hung-yi Lee    Presenter: Yen-chen Wu  
National Taiwan University



## 1. INTRODUCTION

### Task: Sequence labeling

- Input: a sequence  $X = \{x_1, x_2, x_3, \dots, x_T\}$
- Output: label of the sequence

$$Y = \{y_1, y_2, \dots, y_M\} \quad (\text{key term extraction})$$

$$Y = y \quad (\text{dialogue act detection})$$

### Key Issue:

- Selecting the right label considering the sequence
- Label can be an element of the sequence or not within the sequence
- Elements in the sequence are noisy

### Application Example:

- Key term extraction
- Dialogue act detection

## 3. EXPERIMENTS

### Data Set

- Key term extraction
- Posts from Stack Overflow
- Training data: 250,000 posts
- Testing data: 40,000 posts
- Label: 2~6 key terms for each post
- Average length of the post: 120

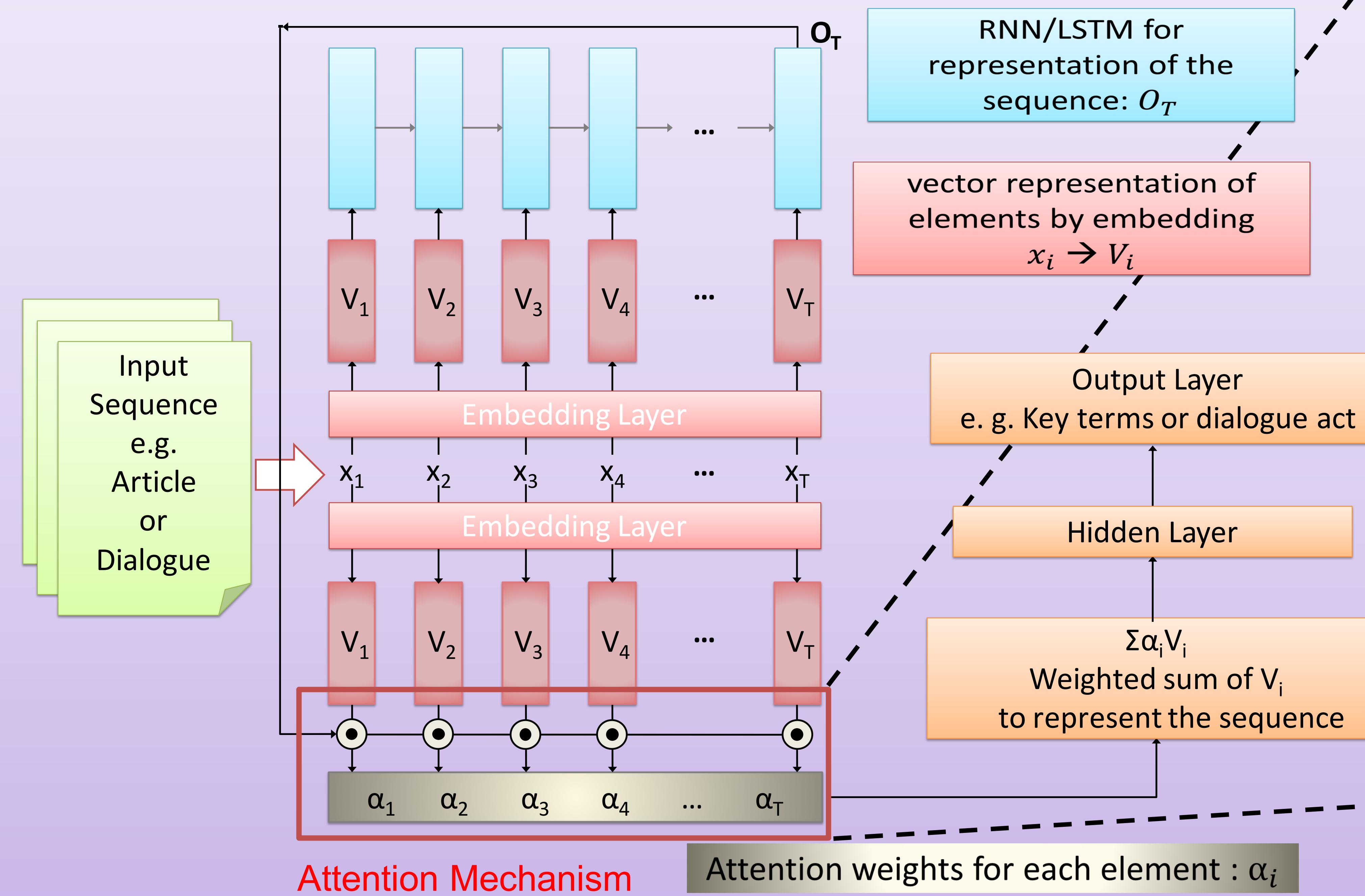
Total 24,000 kinds of labeled key term  
Select the **1,000 most frequent key terms** as candidates

	Model	MAP (%)	P@R (%)
Baseline	(a) Oracle	77.2	
	(b) Tf-idf Sorting	9.9	8.9
	(c) Multi-layer Perceptron	33.1	29.7
	(d) Long Short Term Memory	43.1	40.2
Proposed Method	Neural Attention Model (e) Sharpening	39.3	36.2
	Neural Attention Model (f) Smoothing	<b>50.5</b>	<b>46.4</b>

Neural Attention Model + Smoothing is the best

## 2. PROPOSED APPROACH

### Neural Attention Model Structure



### Attention Mechanism

- Cosine similarity score:  
 $e = (e_1, e_2, \dots, e_T)$  and  $e_i = O_T \odot V_i$   
where  $\odot$  denotes cosine similarity between two vectors

- Attention weight  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_T)$  come from the **normalized score**  $e$

### Sharpening Normalization

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^T \exp(e_i)}$$

### Smoothing Normalization

$$\alpha_i = \frac{\sigma(e_i)}{\sum_{i=1}^T \sigma(e_i)}$$

### Visualization of Attention Mechanism

### Dialogue Act Detection

Switchboard Dialog Act (SwDA)  
Training data: 1,115 Conversation  
Testing data: 19 Conversation  
43 dialogue act label

	Model	Accuracy (%)
Baseline	(a) Support vector Machine	65.8
	(b) Multi-layer Perceptron	67.3
	(c) Long Short Term Memory	69.7
	(d) LSTM with context information	71.7
Proposed Method	Neural Attention Model (e) Sharpening	69.9
	Neural Attention Model (f) Smoothing	70.4
	Neural Attention Model with context information (g) Sharpening	69.8
	Neural Attention Model with context information (h) Smoothing	<b>72.6</b>

Neural Attention Model + Smoothing + Context Information is the best

Ground truth: ios, facebook

5-best predict: ios, facebook-graph-api, facebook, objective-c, iphone

I have an **iOS** application that already using some methods of **Facebook Graph API**, but I need to implement **sending private message to friend** by **Facebook** from my application. As I know, there is no way to **sending private messages by Graph API**, but it **maybe possible** by help **Facebook Chat API**. I already read **documentation** but it do not help me. If anybody has some kind of example or **tutorial**, how to implement **Facebook Chat API in iOS** application, how **sending requests** or something, it will be very helpfull. Thanks.

The darker the color, the higher the weights.

Context information: appended  $n$  previous utterances to the utterance being classified

## 4. CONCLUSIONS

The proposed model is able to highlight important part of long input sequence to produce the good labels