Neural Emoji Recommendation in Dialogue Systems

Han JunYing

Introduction

- Emoji are much more informative and flexible that could express profound meanings beyond words.
- Aim to automatically recommend appropriate emojis attached to the current reply in multi-turn dialogue system according to the contextual information.
- Formalize this task as emoji classification and use H-LSTM to construct dialogue representations followed by a softmax classifier for emoji classification.

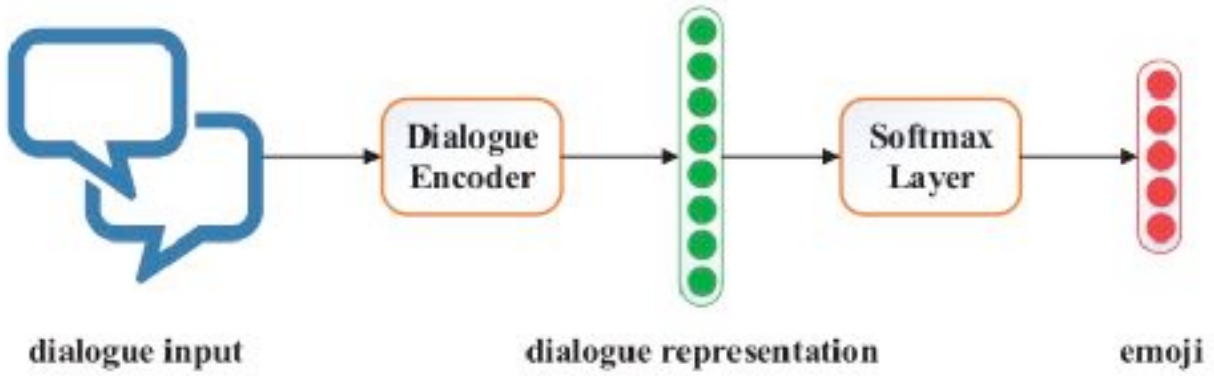
- This classification task seems to be similar with sentiment analysis, while the differences between this two tasks are still significant:
- (1) sentiment analysis typically focuses on predicting the sentiment polarities of sentences or documents, while emoji classification attempts to recommend from larger amounts of candidates, which are much more detailed and complicated to analyze.
- (2) In sentiment analysis, the sentiment polarities are relatively objective and stable.The usages of emojis in real-world dialogue systems are rather subjective and flexible, significantly influenced by user preferences and specific scenarios.

Overall Architecture

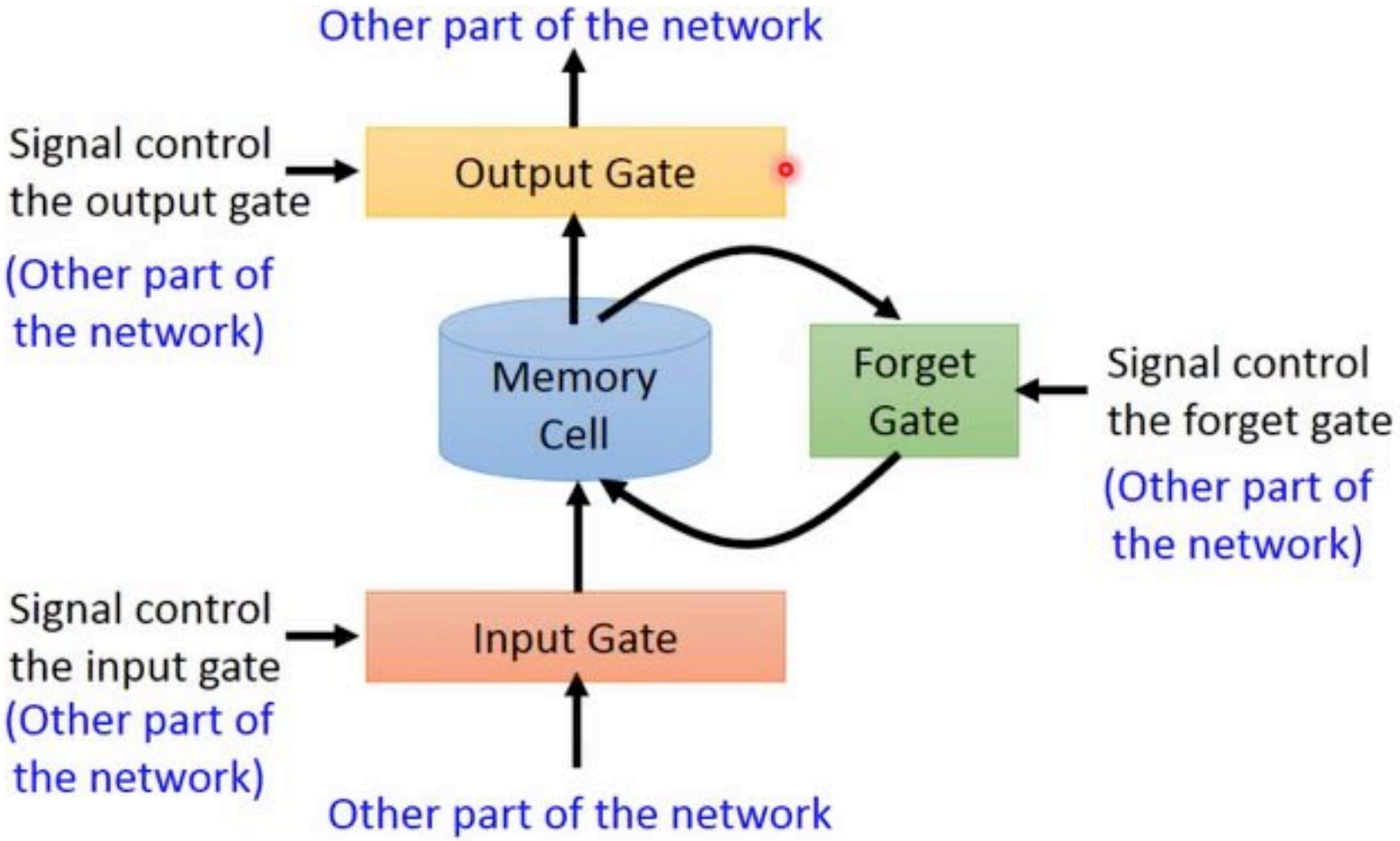
1.All dialogue instances are considered as inputs after data preprocessing.

2.Design two neural dialogue encoders to construct the dialogue representations.(F-LSTM,H-LSTM)

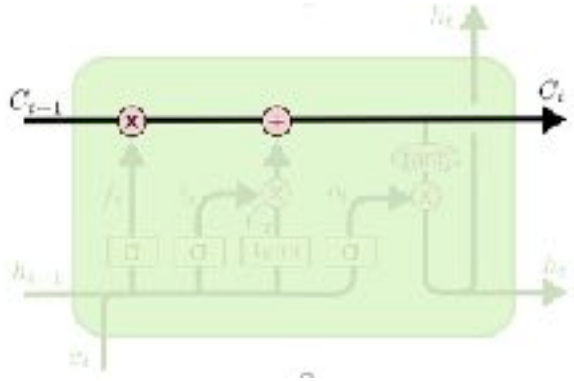
3.utilize a softmax classifier to calculate the probabilities of all emoji candidates



Dialogue encoder-Long Short-Term Memory

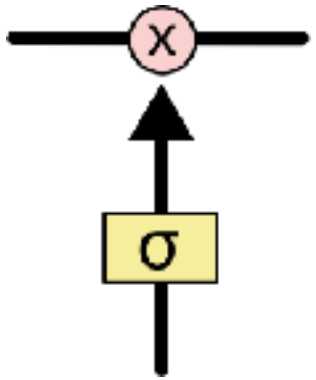


Memory cell

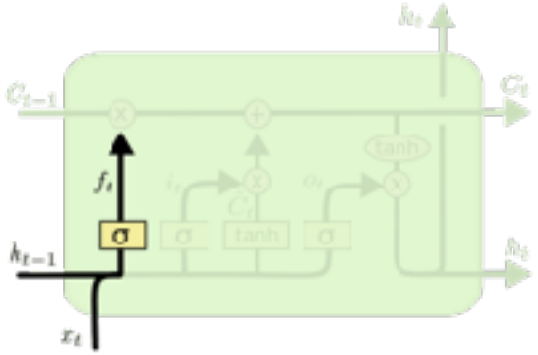


The output of hidden layer are stored in the memory

Gate

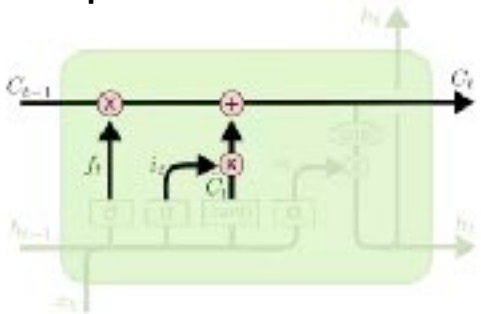


Forget information

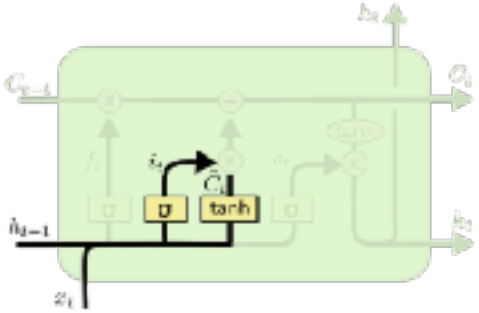


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Update cell state

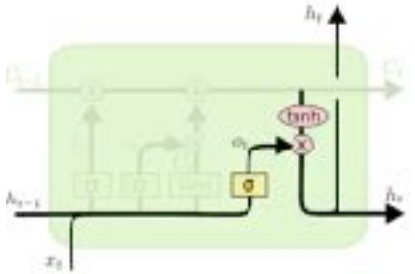


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$



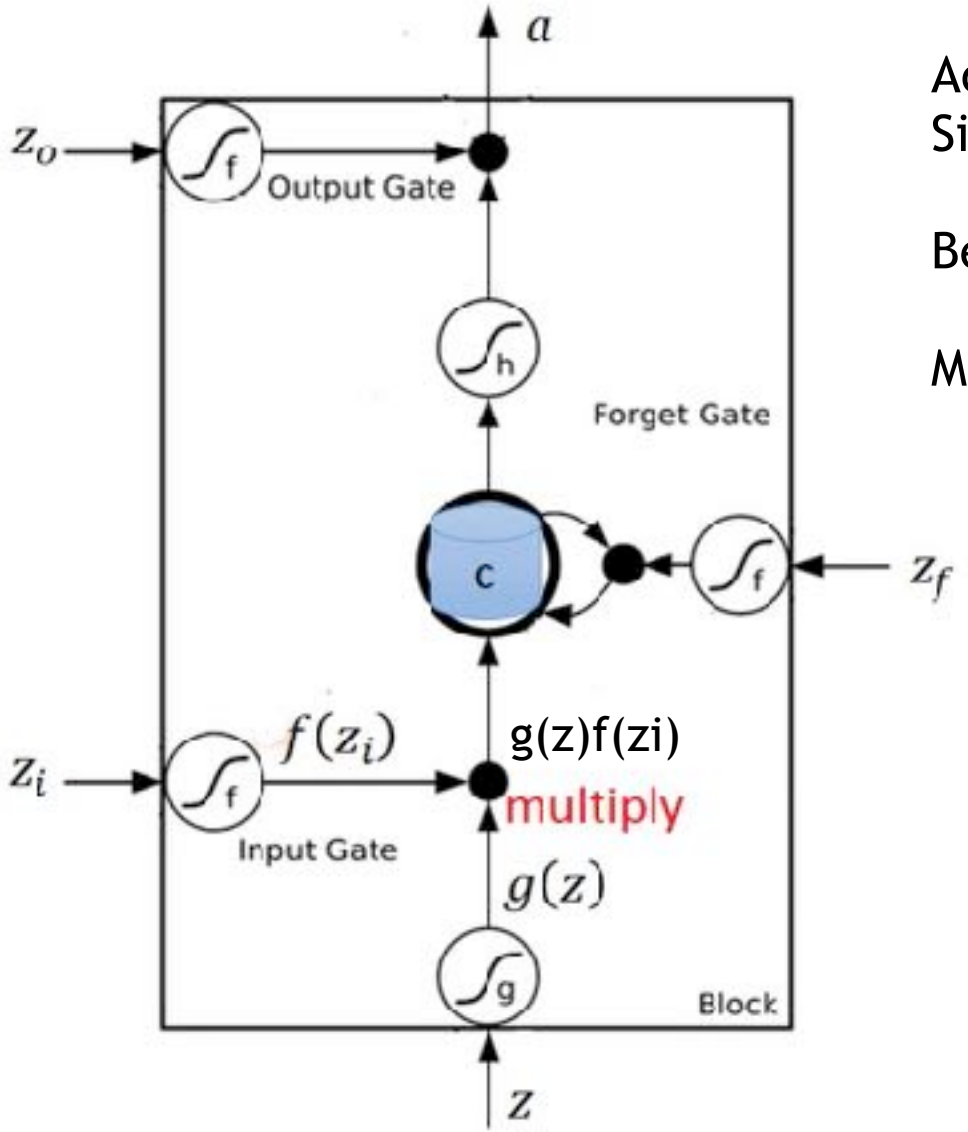
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \sigma\left(W_C \cdot [h_{t-1}, x_t] \; + \; b_C\right)$$

Add new information



$$o_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] \; + \; b_o\right)$$
$$h_t = o_t * \tanh(C_t)$$

Output information

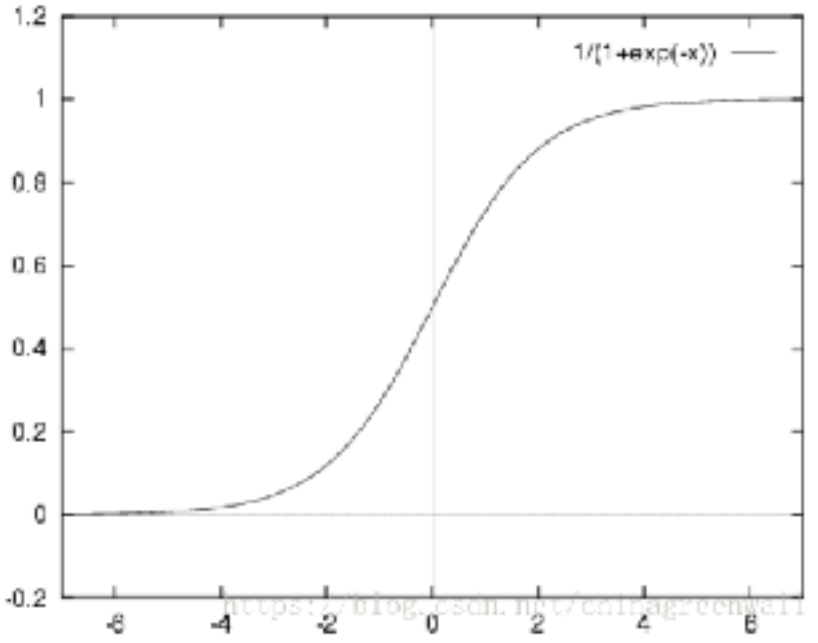


Activation function is usually a Sigmoid function

Between 0 and 1

Mimic open and close gate

Activation function



Dialogue encoder-Three LSTM

Single LSTM

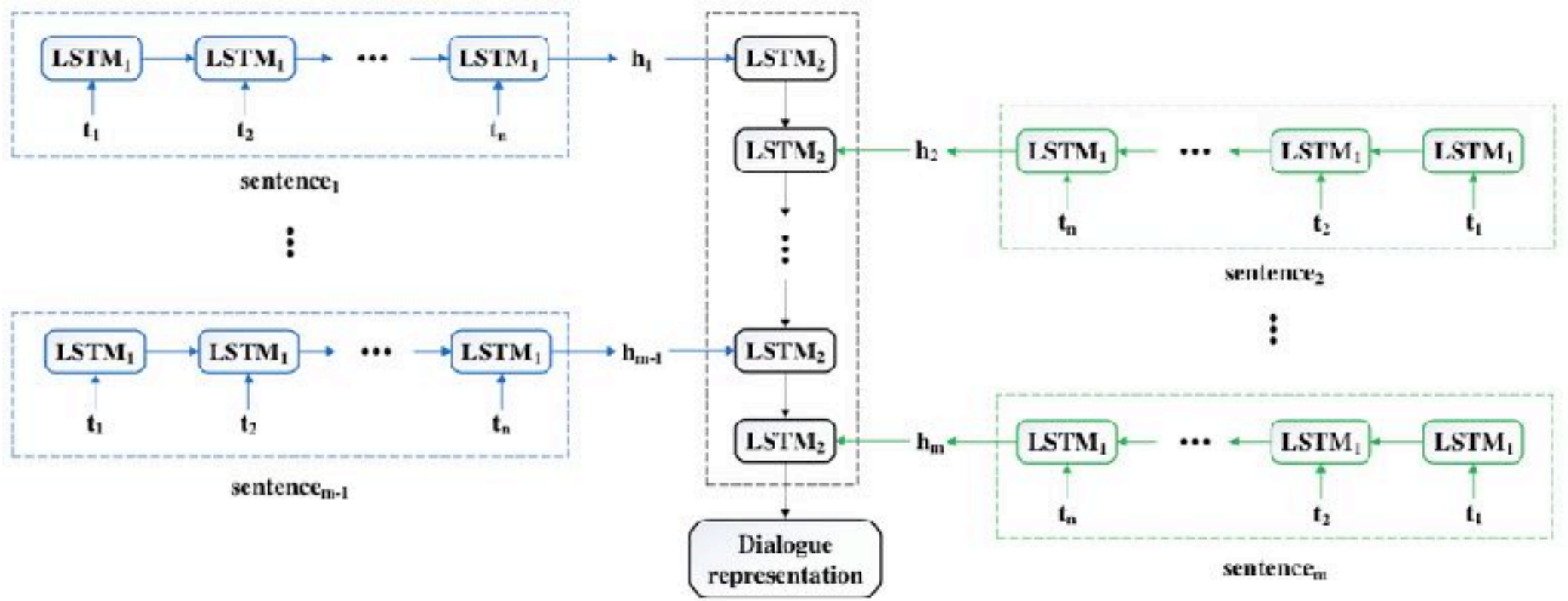
1.Considers the reply sentences as inputs, regardless of the rich information in the contexts, which may harm the performance of recommendation.

Flattened Long Short-Term Memory

concatenates all sentences in each dialogue sequentially to form a long sequence. Specifically, we define si = {xi1, xi2, · · · , xin} as the i-th sentence in dialogue, and the input sequence of F-LSTM after flattening will be {x11,· · · ,x1n,· · · ,xm1,· · · ,xmn}.

Hierarchical Long Short-Term Memory

买



Hierarchical Long Short-Term Memory

First attempt to learn each sentence's meaning, and then further understand the whole dialogue through all sentences, which is exactly what human do in real–world conversations.

Word layer: $$\mathbf{h}_t^{(1)} = LSTM_1(\mathbf{x}_t, \mathbf{h}_{t-1}^{(1)}).$$

Sentence layer $$\mathbf{h}_t^{(2)} = LSTM_2(\mathbf{h}_{n_t}^{(1)}, \mathbf{h}_{t-1}^{(2)}).$$

Objective Formalization

- Softmax

$$S_i = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

$$p(e_i|d) = \frac{\exp(\mathbf{W}_{s_i}\mathbf{d} + \mathbf{b}_{s_i})}{\sum_{j=1}^{n_e} \exp(\mathbf{W}_{s_j}\mathbf{d} + \mathbf{b}_{s_j})},$$

p(ei|d) stands for the probability of i–th emoji given the dialogue d

Ws is a projection matrix

bs is the bias.

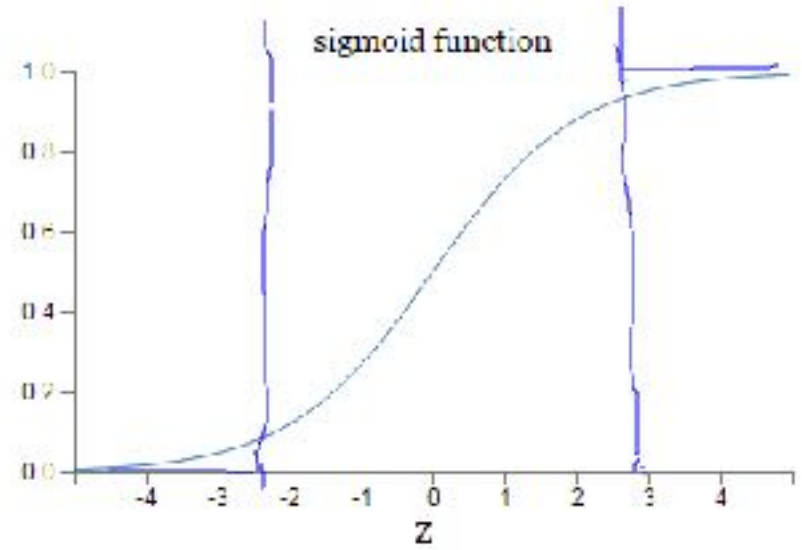
Loss function

$$C = \frac{1}{2}(a - y)^2$$

- y is Expected output
- a is Neural network actual output a=σ(Wx+b)
-

$$\frac{\partial C}{\partial W} = (a - y)\sigma'(a)x^T$$

$$\frac{\partial C}{\partial b} = (a - y)\sigma'(a)$$



Cross-entropy cost function

$$H' = \frac{1}{n}\sum(a_n - y_n) = \frac{1}{n}\sum(\sigma(z_n) - y_n)$$

$$\frac{\partial C}{\partial w_j} = \frac{1}{n}\sum_x x_j(\sigma(z) - y).$$

$$\frac{\partial C}{\partial b} = \frac{1}{n}\sum_x (\sigma(z) - y).$$

$$J(\theta) = -\frac{1}{n_d}[\sum_{i=1}^{n_d}\sum_{j=1}^{n_e} 1\{y^{(i)} = e_j\} \log p(e_i|d)].$$

n_d and n_e are the number of dialogue and emoji. $1\{y^{(i)} = e_j\}$ equals 1 only if the reply sentence of the i-th dialogue have the j-th emoji, and otherwise equals 0.

Data set

- Utilize the Weibo 2015 dialogues in Chinese as the original dataset.
- Select 10 emojis with relative high frequencies as our labels in classification.
- Implement some data cleaning pro- cedures on these extracted raw dialogues.
- Wipe out all Weibo user names, quotes and transmission information, remove all emojis for fair predictions.
- Consider to be out-of-vocabulary whose frequencies are less than 30.
- Discard the dialogues that contain the sentences whose lengths are more than 50 words or OOVs percentages are more than 25%.

Table 1: Statistics of the dataset

Dataset	#Emoji	#Train	#Valid	#Test
Weibo2015	10	1,164,694	64,732	64,271

Experiment setting

AdaDelta

- decay constant ⇢ = 0.95.
- the dropout ratio is set to be 0.5.
- select the dimension of word embeddings nx
- the dimension of hidden embeddings nh among {64, 128, 256, 384, 512}
- the mini-batch size B among {16, 32, 64, 128}.
- the optimal configurations of the models are: nx = nh = 384, B = 128.
- the max length of dialogues is 4 while the max length of sentences is 50.
-

Experiment baselines

- A bag-of-words(BOW)
- Eg:

1： Bob likes to play basketball, Jim likes too.

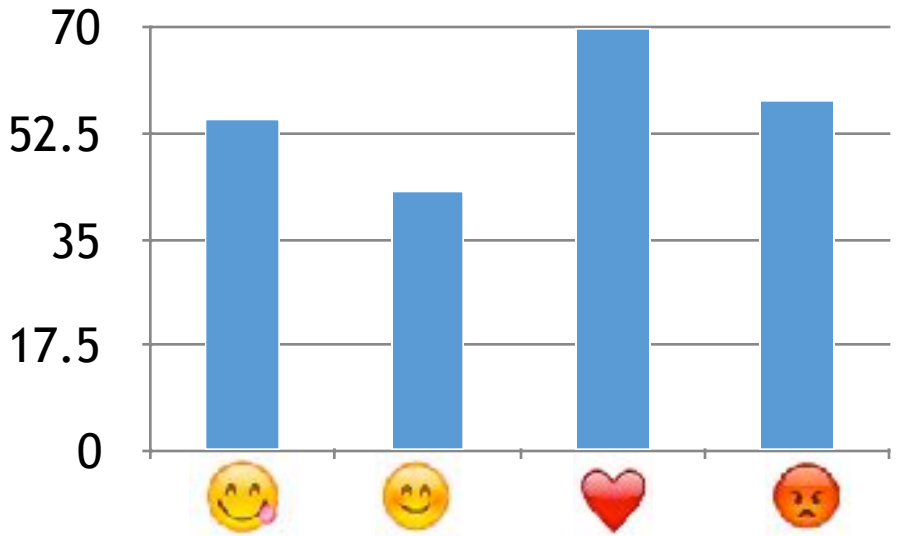
2： Bob also likes to play football games.

Dictionary = {1:"Bob", 2. "like", 3. "to", 4. "play", 5. "basketball", 6. "also", 7. "football", 8. "games", 9. "Jim", 10. "too"}。

1： [1, 2, 1, 1, 1, 0, 0, 0, 1, 1]

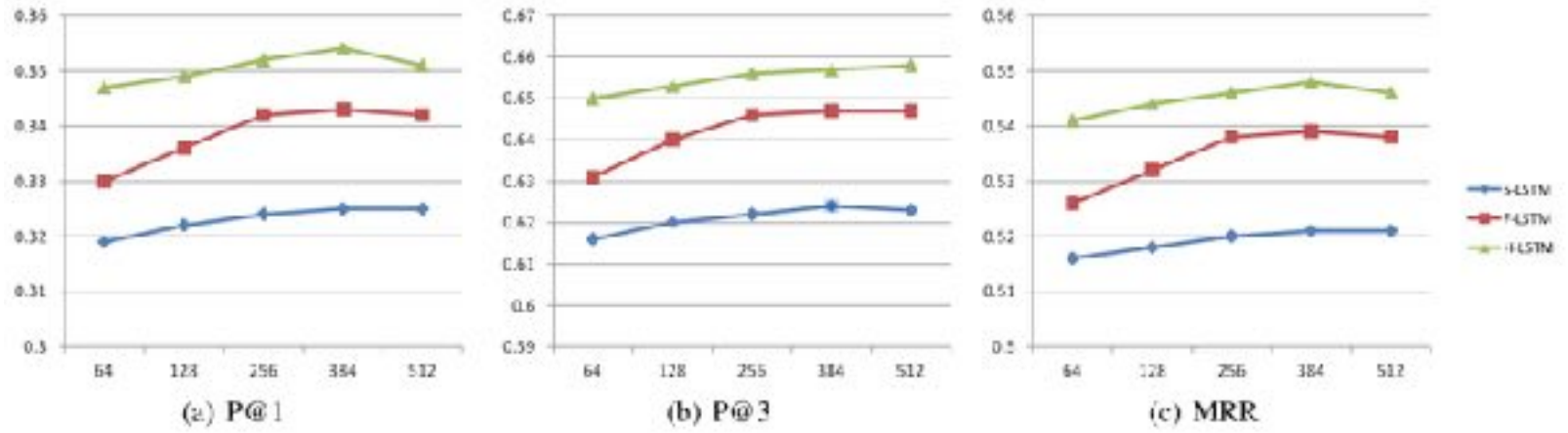
2： [1, 1, 1, 1 ,0, 1, 1, 1, 0, 0]

Experiment baselines



Evaluation Protocol

two metrics to evaluate our models:
(1) the precision in top k emoji candidates (P@k)
(2) the mean reciprocal rank (MRR)



(a) P@1 (b) P@3 (c) MRR

Figure 4: Evaluation results with different dimensions.

Experimental Results

1.All models with multi–turn dialogue inputs significantly outperform all baselines on both evaluation metrics including P@k and MRR.

2.The H–LSTM model achieves the best performance among all models, which confirms the improvements introduced by the hierarchical structure in conversations when constructing dialogue representations.

3.The performances on emoji classification still seem to be far from perfectness.

4.indeed existing emojis that are more confusing and harder to be predicted than other emojis.

5.predicting emojis such as tears of joy and thinking are more challenging, for these emojis are more ambiguous and complicated.

Table 2: Evaluation results on emoji classification

Method	P@1 (%)	P@3 (%)	MRR (%)
S-BOW	29.6	57.9	49.1
F-BOW	24.6	51.3	44.3
S-LSTM	32.5	62.4	52.1
F-LSTM	34.3	64.7	53.9
H-LSTM	**35.4**	**65.7**	**54.8**

Table 3: Evaluation results of P@1 on different emojis

Emoji	Definition	S-LSTM	H-LSTM
😂	*tears of joy*	16.5	21.6
🤔	*thinking*	21.6	22.7
😄	*laugh*	17.5	24.1
😰	*nervous*	23.2	27.1
😏	*sly*	23.5	28.5
😋	*delicious*	33.1	32.7
😭	*cry*	35.6	38.9
😲	*astonished*	46.6	47.4
😡	*angry*	49.3	51.0
❤️	*heart*	60.3	62.2

Table 4: Examples of different models on emoji classification

No.	Dialogue	S-LSTM	H-LSTM	Answer
1	A: 别哭了出去吃！ (Stop crying, and let's hang out for eating!) B: 去哪吃 (To where?) A: 我之前收藏了一大堆宁波吃的的链接，随便找一家！ (I've collected lots of recommendations on eating in Ning Bo, we can choose from them!) B: 好呀好呀你啥时候有空 (Great! When will you be free?)	*shy*	*delicious*	*delicious*
2	A: 太过分了啊啊啊啊 (It's so unacceptable!) B: 气气啊啊啊啊 (I'm really angry!) A: 你生谁气 (Who are you mad at?) B: 那个提香蕉的 (The person who mentioned bananas!)	*delicious*	*nervous*	*nervous*
3	A: 芭比娃娃一样?? (Just like a barbie doll?) B: 太好看 (It's so beautiful!) A: 哈哈哈谢谢 (LOL, thank you!) B: 等等你的短发呢 (Wait! Where is your short hair?)	*thinking*	*shy*	*thinking*
4	A: 越画越好 (Your paintings are getting better since you draw more.) B: 谢谢姐姐鼓励，画画真的让人开心 (Thanks for your encourage, my sister. Drawing really makes me happy.) A: 是，这是个很好的爱好 (I agree, that's a good hobby.) B: 跳舞也是 (And so is dancing.)	*laugh*	*heart*	*shy*

Dango

This is a floating assistant that can be run on a mobile phone (a mobile app that can be run in other App fronts). It can predict the emoji that will be used based on what you and your friends have written in any application. , textures, and GIFs. Make rich conversations everywhere: Messenger, Snapchat, and more.Currently, Dango is only available to Android platform users, but Whirlscape says the iOS version will be available soon.