Modeling Situations in Neural Chat Bots

Shoetsu Sato[1], Naoki Yoshinaga[2], Masashi Toyoda[2], and Masaru Kitsuregawa[2,3]

{shoetsu, ynaga, toyoda, kitsure}@tkl.iis.u-tokyo.ac.jp

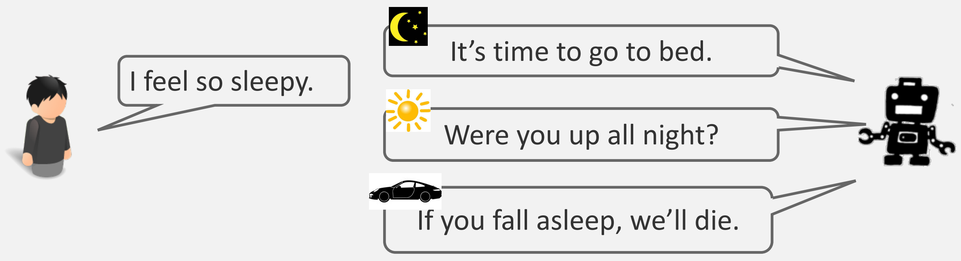
[1] The University of Tokyo, [2] IIS, the University of Tokyo, [3] NII, Japan

Motivation

In chat-dialogue modeling…

- existing data-driven models hesitate when choosing from various possible responses

→ They tend to generate typical, frequent responses [Li+, '15]



The appropriate response varies depending on **conversational situations**

Our approach

Baseline

- Seq2seq-based neural conversational model [Vinyal+, 15]

Idea:

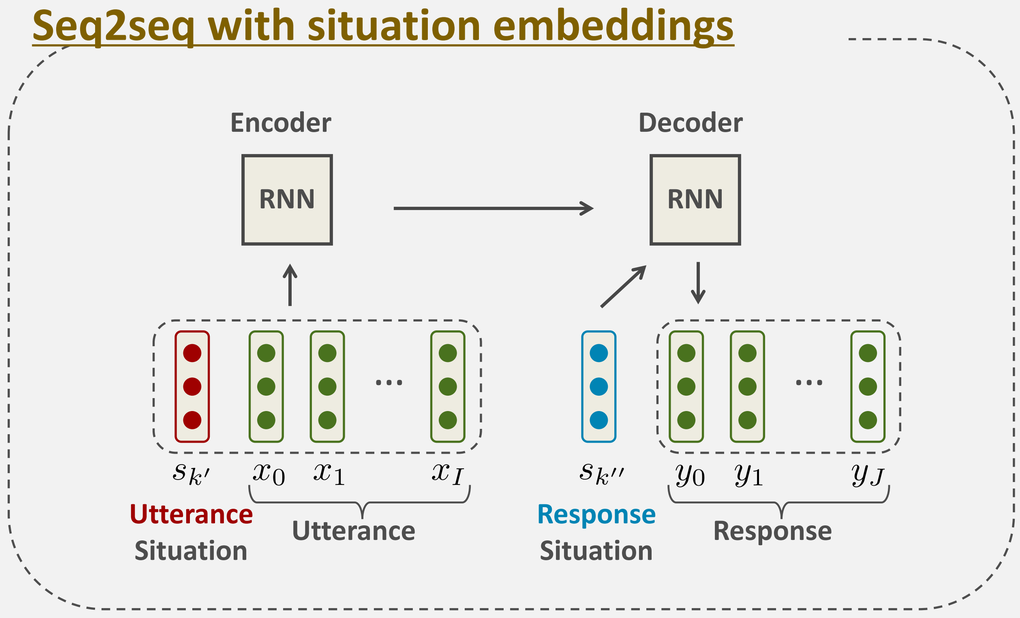
- Improve neural conversational model to handle conversational situations represented as discrete variables

Targeted situation-types and how to obtain them:

- **Time (season) :** split conversation data into 4 season types depending on timestamps
- **Utterance :** cluster utterances by their topics or speaking-styles and regard their belonging cluster's id as a situation
- **Speaker / Addressee (Profile) :** In the same way as utterance, cluster speaker's / addressee's profiles

Proposal: Situation-aware Neural Conversational Models

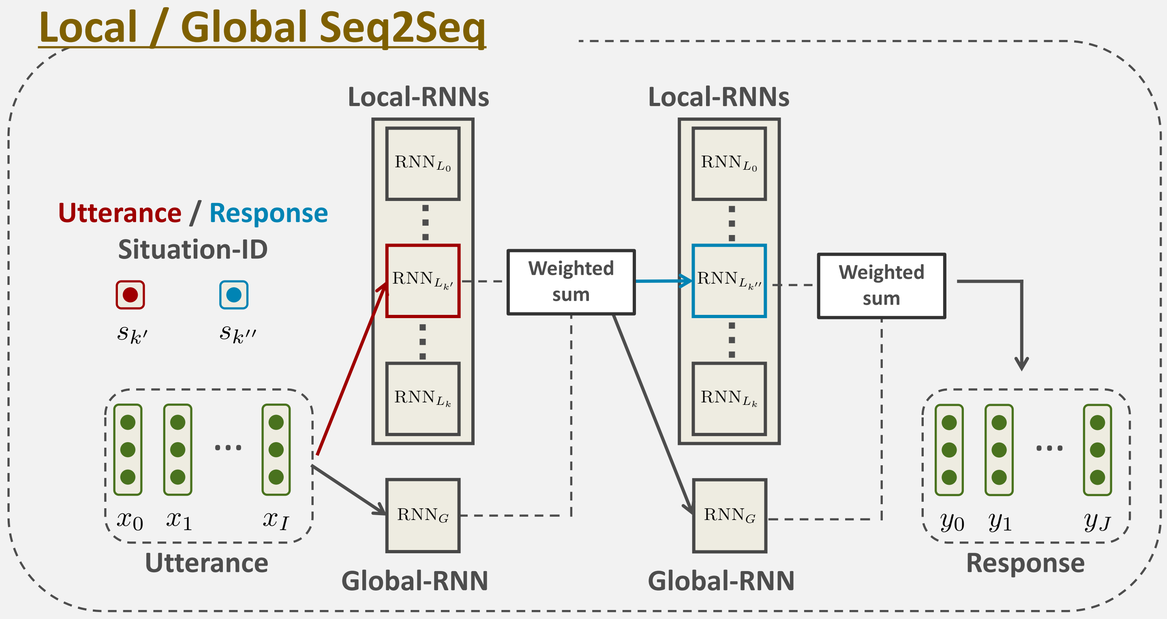
Seq2seq with situation embeddings



Situations as additional features:

- Prepend an Utterance / Response situation embedding to the encoder's / decoder's input sequence

Local / Global Seq2Seq



Situations as independent tasks:

- Train multiple local-RNNs by the given situations and one global-RNN to avoid data sparseness

Experiments

Dataset: Japanese Twitter archive

- A tweet and a mention to it are considered as an utterance-response pair
- About 23M pairs for training and 6K pairs for test

Evaluation: Response selection task

- Scoring utterance-candidate pairs by cross entropy
- **Metric: 1 in t P@k**
  the percentage of correct utterances in top **k** responses chosen from **t** candidates
  (t-1 random dummy responses, 1 actually replied)

Hyperparameters

- **RNN:** 3-layer LSTM [Zaremba+, '14]
- **optimizer**: Adam [Kingma+, '15] (learning rate: 1e-4)
- **hidden/embedding layer, and RNN:** 100 dims
- **vocabulary size** : 100,000 words

Model	1 in 2P@1	1 in 5P@1	1 in 5P@2
Baseline	64.5%	33.9%	56.6%
Situation: time (season)			
Seq2Seq emb	67.3%	37.6%	60.7%
L/G Seq2Seq	65.9%	35.8%	58.1%
Situation: utterance			
Seq2Seq emb	65.6%	35.4%	58.2%
L/G Seq2Seq	**68.5%**	**38.2%**	**62.1%**
Situation: speaker / addressee (profiles)			
Seq2Seq emb	67.8%	37.5%	61.1%
L/G Seq2Seq	66.4%	36.4%	59.2%

The situation-aware conversational models…

- **are better at selecting ground-truth responses for situation-specific conversations**
- **avoid typical responses such as "Yes !" or "You've gotta be tired."**

Examples:

Situation: season (summer)	
Input	7月になって、流石にパーカーは暑くなってきた (July is too warm to wear a hoodie.)
Baseline	そうなんです！ (Yes!)
Seq2Seq emb	まだ着てたの!? (Do you still wear one?)

Situation: utterance (opinions, questions)	
Input	ちょっと最近BOTのフォロー多いんですけど (I've recently been followed by many bot accounts.)
Baseline	お疲れ様やで (You've gotta be tired.)
L/G Seq2Seq	ブロックしちゃいましょう (Let's block them.)