Addressee and Response Selection in Multi-Party Conversation with Speaker Interaction RNNs

Rui Zhang[1], Honglak Lee[2], Lazaros Polymenakos[3], Dragomir Radev[1]

[1]Yale University, [2]University of Michigan, Ann Arbor, [3]IBM T. J. Watson Research Center

LILY Lab

Introduction

In this paper, we study the problem of addressee and response selection in multi-party conversations. Understanding multi-party conversations is challenging because of complex speaker interactions: multiple speakers exchange messages with each other, playing different roles (sender, addressee, observer), and these roles vary across turns. To tackle this challenge, we propose the Speaker Interaction Recurrent Neural Network (SI-RNN). Whereas the previous state-of-the-art system updated speaker embeddings only for the sender, SI-RNN uses a novel dialog encoder to update speaker embeddings in a role-sensitive way. Additionally, unlike the previous work that selected the addressee and response separately, SI-RNN selects them jointly by viewing the task as a sequence prediction problem. Experimental results show that SI-RNN significantly improves the accuracy of addressee and response selection, particularly in complex conversations with many speakers and responses to distant messages many turns in the past.

An example of Addressee and Response Selection

	Sender	Addressee	Utterance
1	codepython	wafflejock	thanks
1	wafflejock	codepython	yup np
2	wafflejock	theoletom	you can use sudo apt-get install packagename – reinstall, to have apt-get install reinstall some package/metapackage and redo the configuration for the program as well
3	codepython	-	i installed ubuntu on a separate external drive. now when i boot into mac, the external drive does not show up as bootable. the blue light is on. any ideas?
4	Guest54977	-	hello there. wondering to anyone who knows, where an ubuntu backup can be retrieved from.
2	theoletom	wafflejock	it's not a program. it's a desktop environment.
4	Guest54977	-	did some searching on my system and googling, but couldn't find an answer
2	theoletom	-	be a trace of it left yet there still is.
2	theoletom	-	i think i might just need a fresh install of ubuntu. if there isn't a way to revert to default settings
5	releaf	-	what's your opinion on a $500 laptop that will be a dedicated ubuntu machine?
5	releaf	-	are any of the pre-loaded ones good deals?
5	releaf	-	if not, are there any laptops that are known for being oem-heavy or otherwise ubuntu friendly?
3	codepython	-	my usb stick shows up as bootable (efi) when i boot my mac. but not my external hard drive on which i just installed ubuntu. how do i make it bootable from mac hardware?
3	Jordan_U	codepython	did you install ubuntu to this external drive from a different machine?
5	Umeaboy	releaf	what country you from?
5	wafflejock		
Model Prediction		**Addressee**	**Response**
Direct-Recent+TF-IDF		theoletom	ubuntu install fresh
Dynamic-RNN		codepython	no prime is the replacement
SI-RNN		⋆ releaf	⋆ there are a few ubuntu dedicated laptop providers like umeaboy is asking depends on where you are

Figure 1. An example of addressee and response selection. SI-RNN chooses to engage in a new sub-conversation by suggesting a solution to "releaf" about Ubuntu dedicated laptops.

Notation and Problem Formulation

Given a responding speaker a_{res} and a dialog context C, the task is to select a response and an addressee. C is a list ordered by time step:

$$\mathcal{C} = [(a^{(t)}_{sender}, a^{(t)}_{addressee}, u^{(t)})]_{t=1}^{T}$$

	Data	Notation
Input	Responding Speaker	a_{res}
	Context	$\mathcal{C}$
	Candidate Responses	$\mathcal{R}$
Output	Addressee	$a \in \mathcal{A}(\mathcal{C})$
	Response	$r \in \mathcal{R}$
Sender ID at time t		$a^{(t)}_{sender}$
Addressee ID at time t		$a^{(t)}_{addressee}$
Utterance at time t		$u^{(t)}$
Utterance embedding at time t		$\mathbf{u}^{(t)}$
Speaker embedding of a_i at time t		$\mathbf{a}^{(t)}_i$

Table 1: Notations for the task and model.

Speaker Interaction RNNs

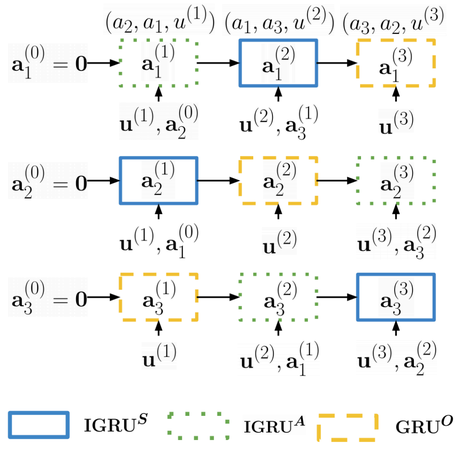

Figure 2. Dialog encoders SI-RNN for an example context at the top.

SI-RNN solves the task in two phases:

1. the dialog encoder maintains a set of speaker embeddings to track each speaker status, which dynamically changes with time step t. The dialog encoder in SIRNN updates embeddings for all the speakers besides the sender at each time step. Speaker embeddings are updated depending on their roles: the update of the sender is different from the addressee, which is different from the observers. Furthermore, the update of a speaker embedding is not only from the utterance, but also from other speakers. These are achieved by designing variations of GRUs for different roles.
2. then SI-RNN produces the context embedding from the speaker embeddings and joint selects the addressee and response based on embedding similarity among context, speaker, and utterance.

$$\begin{aligned}\hat{a}, \hat{r} &= \underset{a_p, r_q \in \mathcal{A}(\mathcal{C}) \times \mathcal{R}}{\arg\max}\ \mathbb{P}(r_q, a_p|\mathcal{C}) \\ &= \underset{a_p, r_q \in \mathcal{A}(\mathcal{C}) \times \mathcal{R}}{\arg\max}\ \mathbb{P}(r_q|\mathcal{C}) \cdot \mathbb{P}(a_p|\mathcal{C}, r_q) \\ &\quad + \mathbb{P}(a_p|\mathcal{C}) \cdot \mathbb{P}(r_q|\mathcal{C}, a_p)\end{aligned}$$

Data Set and Evaluation Metric

We use the Ubuntu Multiparty Conversation Corpus built from the Ubuntu IRC chat room where a number of users discuss Ubuntu-related technical issues. The log is organized as one file per day corresponding to a document. Each document consists of (Time, SenderID, Utterance) lines. If users explicitly mention addressees at the beginning of the utterance, the addresseeID is extracted. Then a sample, namely a unit of input (the dialog context and the current sender) and output (the addressee and response prediction) for the task, is created to predict the ground-truth addressee and response of this line.
Metrics include accuracy of addressee selection (ADR), response selection (RES), and pair selection (ADR-RES).

Result and Analysis

		RES-CAND = 2				RES-CAND = 10			
		DEV	TEST			DEV	TEST		
	T	ADR-RES	ADR-RES	ADR	RES	ADR-RES	ADR-RES	ADR	RES
Chance	-	0.62	0.62	1.24	50.00	0.12	0.12	1.24	10.00
Recent+TF-IDF	15	37.11	37.13	55.62	67.89	14.91	15.44	55.62	29.19
Direct-Recent+TF-IDF	15	45.83	45.76	67.72	67.89	18.94	19.50	67.72	29.40
Static-RNN (Ouchi and Tsuboi 2016)	5	47.08	46.99	60.39	75.07	21.96	21.98	60.26	33.27
	10	48.52	48.67	60.97	77.75	22.78	23.31	60.66	35.91
	15	49.03	49.27	61.95	78.14	23.73	23.49	60.98	36.58
Static-Hier-RNN (Zhou et al. 2016) (Serban et al. 2016)	5	49.19	49.38	62.20	76.70	23.68	23.75	62.24	34.51
	10	51.37	51.76	64.61	78.28	25.46	25.83	64.86	36.94
	15	52.78	53.04	65.84	79.08	26.31	26.62	65.89	37.85
Dynamic-RNN (Ouchi and Tsuboi 2016)	5	49.38	49.80	63.19	76.07	23.44	23.72	63.28	33.62
	10	52.76	53.85	66.94	78.16	25.44	25.95	66.70	36.14
	15	54.45	54.88	68.54	78.64	26.73	27.19	68.41	36.93
SI-RNN (Ours)	5	60.57	60.69	74.08	78.14	30.65	30.71	72.59	36.45
	10	65.34	65.63	78.76	80.34	34.18	34.09	77.13	39.20
	15	**67.01**	**67.30**	**80.47**	**80.91**	**35.50**	**35.76**	**78.53**	**40.83**
SI-RNN w/ shared IGRUs	15	59.50	59.47	74.20	78.08	28.31	28.45	73.35	36.00
SI-RNN w/o joint selection	15	63.13	63.40	77.56	80.38	32.24	32.53	77.61	39.73

Table 2. Addressee and response selection results on the Ubuntu Multiparty Conversation Corpus. RES-CAND: the number of candidate responses. T: the context length.

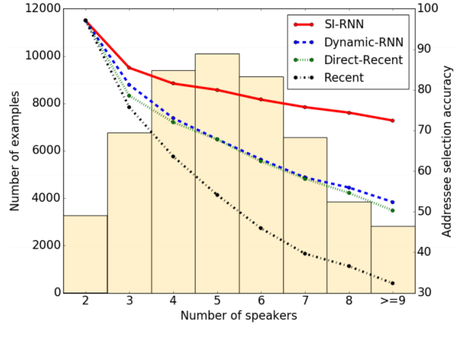

Figure 3. Effect of the number of speakers in the context.

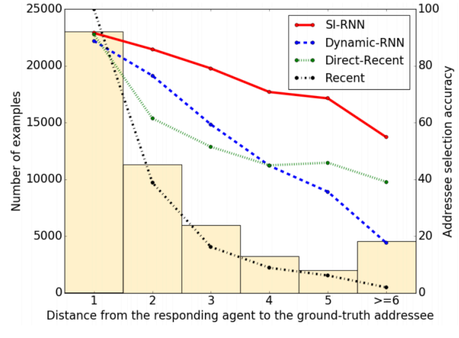

Figure 4. Effect of the addressee distance