Neural Conversation Recommendation with Online Interaction Modeling

Xingshan Zeng[1,2], Jing Li[3], Lu Wang[4], Kam-Fai Wong[1,2]

[1]The Chinese University of Hong Kong, Hong Kong, China

[2]MoE Key Laboratory of High Confidence Software Technologies, China

[3]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

[4]Khoury College of Computer Sciences, Northeastern University, Boston, United States

1. Introduction: Conversation Recommendation

❑ Motivation:

- Social media has revolutionized people's social interactions – voice opinions and exchange ideas in online platforms!
- It exceeds any individual's capability of digesting the huge volume of online discussions formed every day!
- Urgent need for an automatic filtering tools!

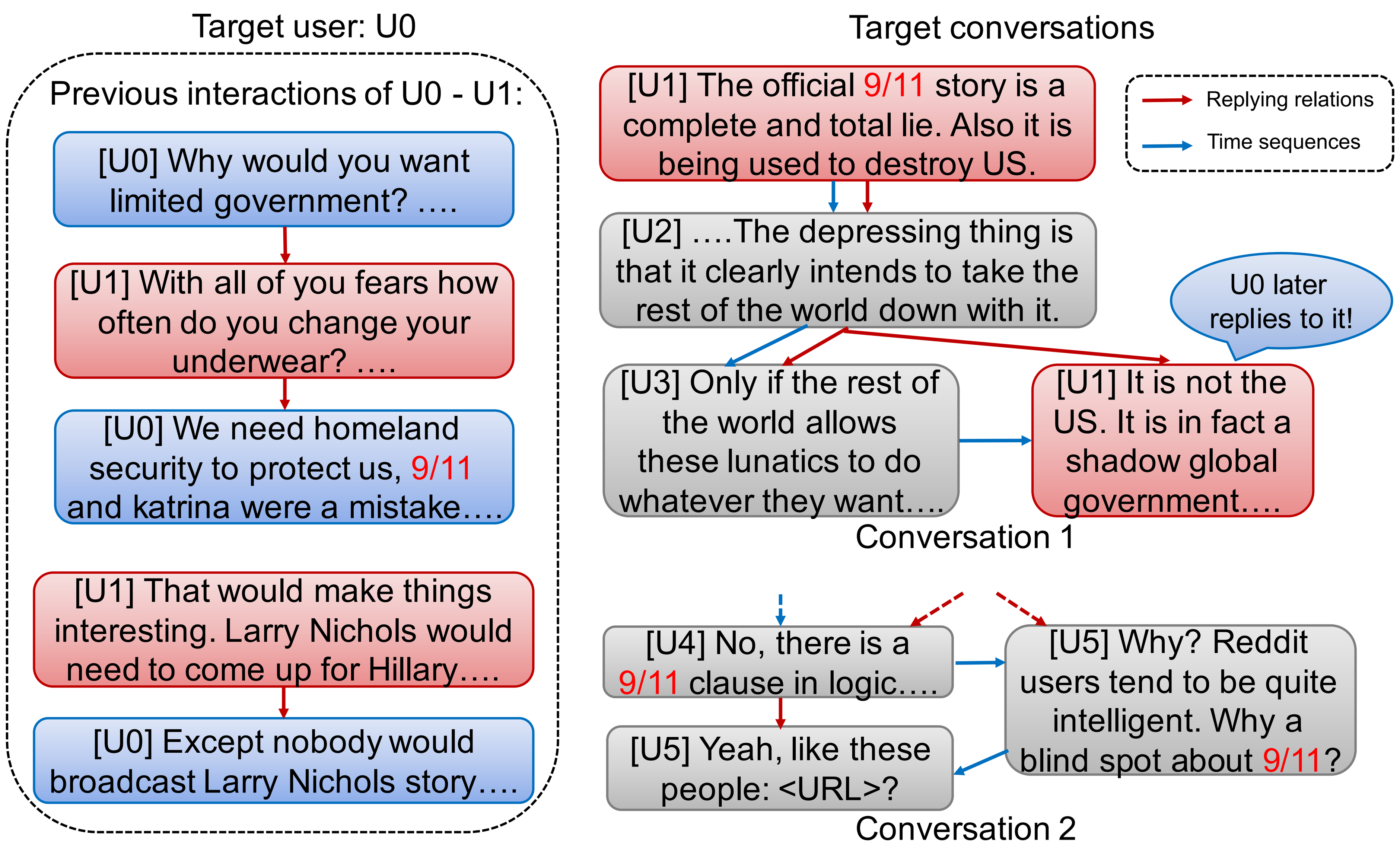
❑ Definition:

- Given: a list of conversations that are available for one user.
- Aim: to forecast which conversations the user is more likely to participate in.

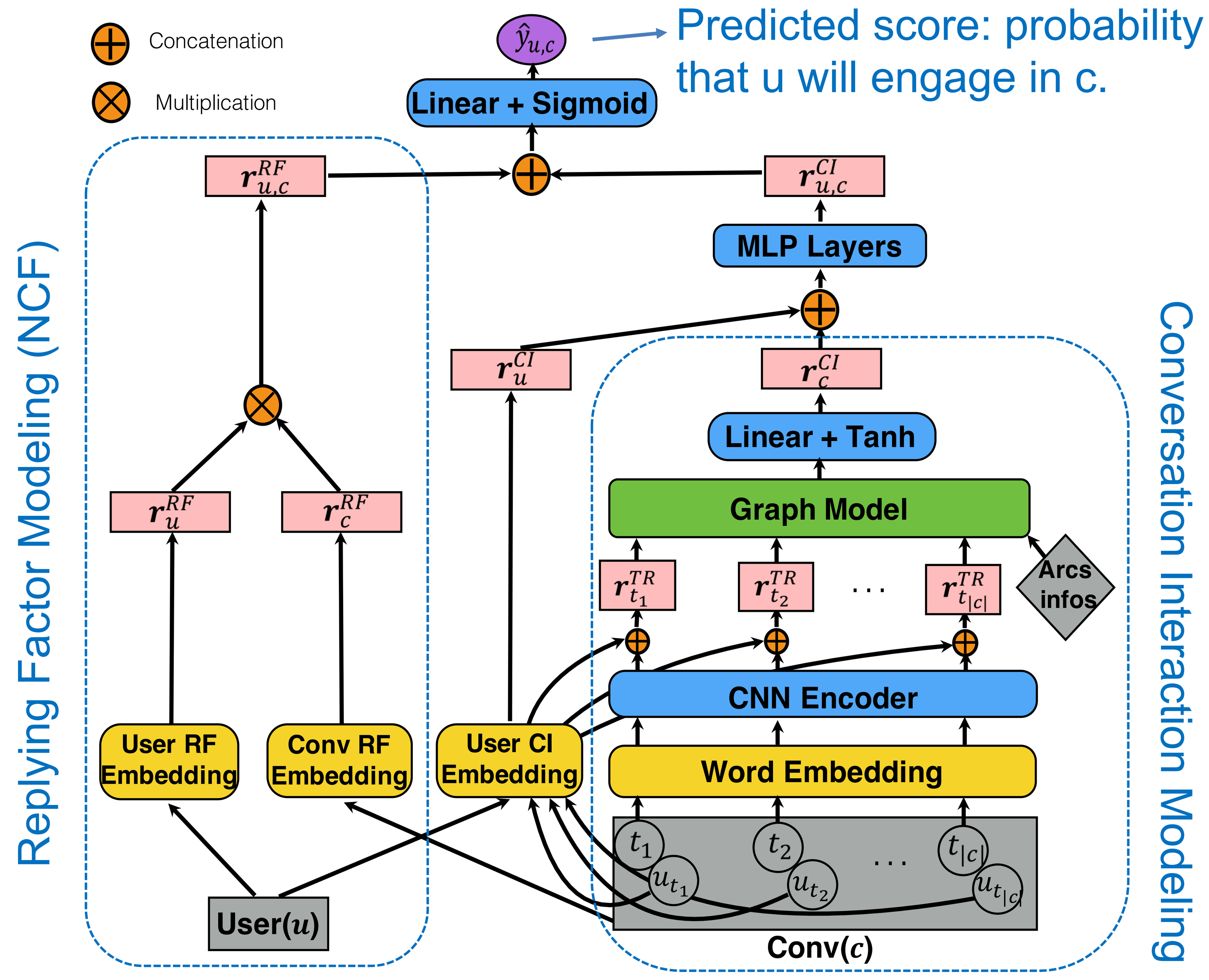
❑ Previous work:

- Post-level: not considering context information;
- Conversation-level: only capture topic information or word co-occurrence patterns, failure to leverage replying information.

❑ Our idea: online interactions indicated by replying structures could benefit users' future conversation behavior prediction!



2.1 Model Overall Structure



2.2 Model Details

❑ Replying Factor Modeling:

- Neural Collaborative Filtering (NCF) (He et al. 2017)

$$\boldsymbol{r}_{u,c}^{RF} = \boldsymbol{r}_u^{RF} \odot \boldsymbol{r}_c^{RF}$$

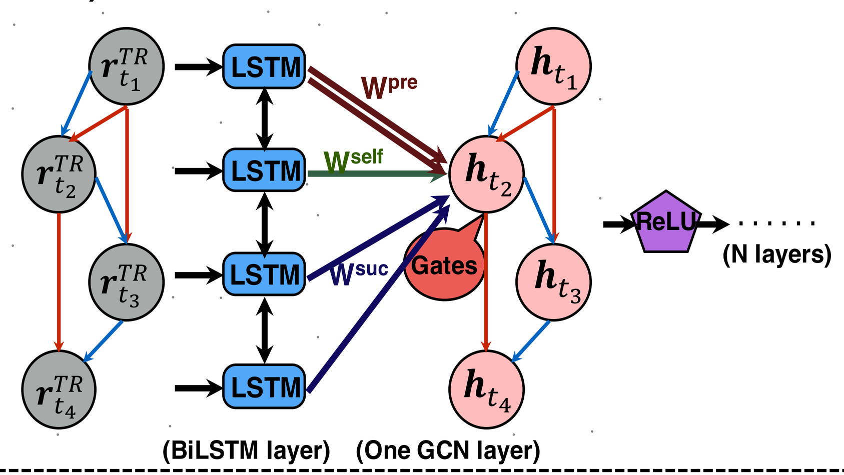
❑ Conversation Interaction Modeling:

- Turn-level Modeling
  - ✓ Pretrained Word Embedding + CNN Encoder
- Turn Interaction Modeling
  - ✓ Graph-State LSTM (GLSTM)

$$g_t^{GLSTM} = \sigma(W^p \boldsymbol{x}_t^p + W^s \boldsymbol{x}_t^s + U^p \boldsymbol{h}_t^p + U^s \boldsymbol{h}_t^s + b)$$

  - ✓ Graph Convolutional Networks (GCN)

$$h_t^{GCN} = \sum_{i\in E^p(t)} \omega_{i,t}(W^{pre} h_i^{LSTM} + b^{pre}) + \sum_{j\in E^s(t)} \omega_{j,t}(W^{suc} h_j^{LSTM} + b^{suc}) + \omega_{t,t}(W^{self} h_t^{LSTM} + b^{self})$$



3. Experiments and Results

■ Statistics of Datasets

Dataset	# of users	# of convs	# of turns	convs/ user	turns/ conv
Twitter	10,122	7,500	38,999	1.7	5.2
Reddit	13,134	29,477	109,774	5.9	3.7

■ Experiment Setup:

✓ First 75% as observation history for training, rest equally divided into test and development;

✓ Negative sampling to alleviate imbalance.

■ Loss Function

✓ weighted binary cross-entropy loss:

$$\mathcal{L} = -\sum_{(u,c)\in\mathcal{T}} \left[\lambda\cdot y_{u,c}\log(\hat{y}_{u,c}) + (1-y_{u,c})\log(1-\hat{y}_{u,c})\right]$$

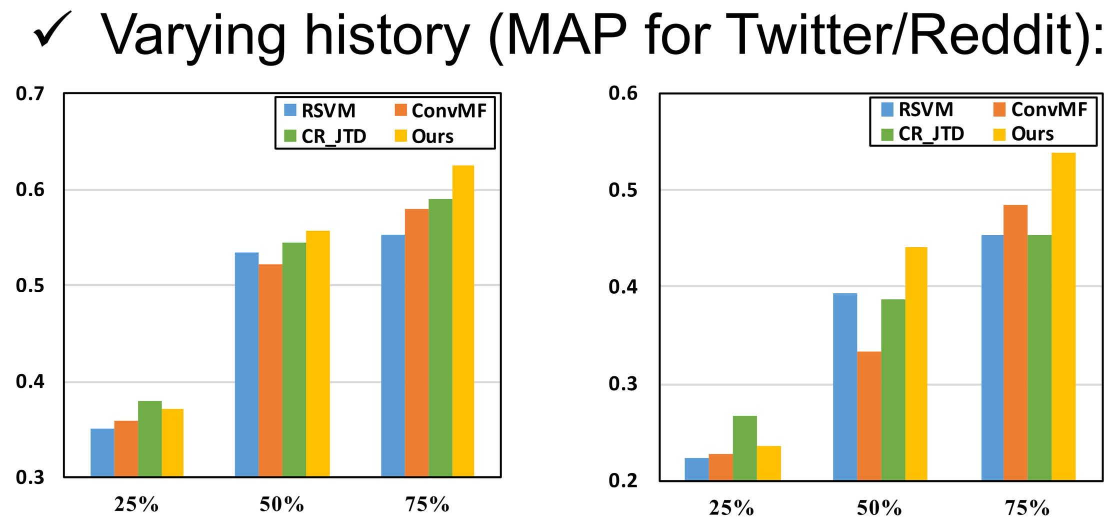
■ Different Interaction Modeling Comparison

Models	Train time	MAP in dev. set	
		Twitter	Reddit
BiLSTM	0.94	0.617	0.498
GLSTM	1.25	0.617	0.528
GCN (W/O BiLSTM)	1.03	0.619	0.530
GCN (With BiLSTM)	1.00	**0.620**	**0.533**

■ Comparison with Previous Work

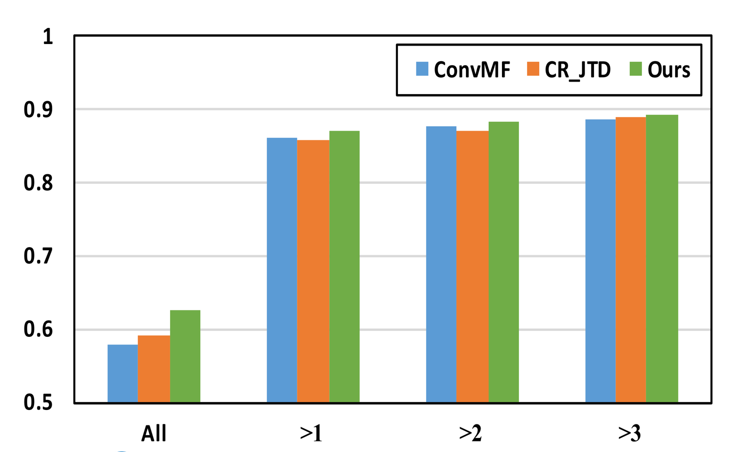
Models	Twitter			Reddit		
	MAP	P@1	nDCG@5	MAP	P@1	nDCG@5
Popularity	0.023	0.005	0.010	0.082	0.033	0.063
RSVM	0.554	0.575	0.559	0.453	0.457	0.466
NCF	0.573	0.593	0.576	0.412	0.544	0.461
ConvMF	0.579	0.596	0.583	0.485	0.532	0.520
CR_JTD	0.591	0.591	0.600	0.453	0.559	0.485
Ours	**0.625**	**0.632**	**0.626**	**0.538**	**0.670**	**0.590**

✓ Varying history (MAP for Twitter/Reddit):



Better with longer history!

✓ Varying user sparsity:



Sparsity severely harms!

✓ First Time Replies Prediction (MAP):

Models	RSVM	NCF	ConvMF	CR_JTD	Ours
Twitter	0.002	0.033	0.049	0.090	**0.160**
Reddit	0.049	0.038	0.210	0.075	**0.212**