Findings on Conversation Disentanglement

Rongxin Zhu, Jey Han Lau, Jianzhong Qi

School of Computing and Information System, The University of Melbourne

source code

Introduction

1.1 Problem Definition

Conversation disentanglement aims at identifying threads in multi-party conversations. The figure below shows two threads in different colors, with reply-to links between utterances.

[12:05] <ydnar> for what reason would a dvd not play if i have libdvdcss2 installed?

[12:05] <gourdin> we will we be able to access an edgy repo ?

[12:05] <Ng> ydnar: what are you using to play it?

[12:06] <Anfangs> Edgy Eft is the next codename for Ubuntu dapper+1. See https://ubuntu.com/0064.html.

[12:06] <holycow> because it couldn't crack the encoding for the particual portion of the dvde

[12:06] <ydnar> tried vlc. holycow, do you have any

[12:06] <gourdin> I don't think the link works

1.2 Limitation of previous methods

- transformer-based models are not systematically compared with respect to performance, memory consumption and speed
- previous methods don't leverage **dialogue history** effectively
- greedy decoding algorithm recovers threads by finding the parent utterance for each utterance of interest (UOI) **independently**

Methodology

2.1 Pairwise Model

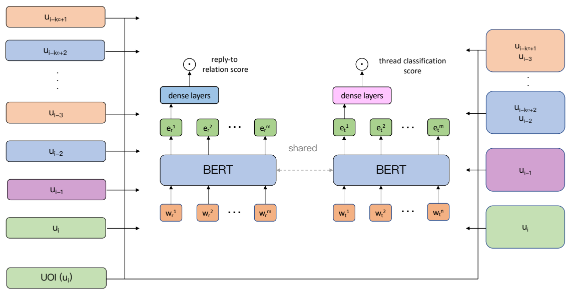
Pairwise models measure the similarity between UOI and each candidate separately. BERT+MF (manual features) is still a strong baseline.

Model	Link Prediction Precision	Recall	F1	Ranking R@1	R@5	R@10	Clustering 1-1	VI	F
Last Mention	37.1	35.7	36.4	-	-	-	21.4	60.5	4.0
GLOVE+MF	71.5	68.9	70.1	70.2	95.8	98.6	76.1	91.5	34.0
MF	71.1	68.5	69.8	70.2	94.0	97.3	75.0	91.3	31.5
POLY-BATCH	39.3	37.9	38.6	40.8	69.8	80.8	52.3	80.8	9.8
POLY-INLINE	42.2	40.7	41.4	42.8	70.8	81.3	62.0	84.4	13.6
ALBERT	46.1	44.4	45.3	46.8	77.3	88.4	68.6	87.9	22.4
BERT	48.2	46.4	47.3	48.8	75.4	84.7	74.3	89.3	26.3
BERT+TD	67.9	65.4	66.6	66.9	90.6	95.3	76.0	91.1	34.9
BERT+MF	**73.9**	**71.3**	**72.6**	**73.9**	**95.8**	**98.6**	**77.0**	**92.0**	**40.9**

Model	GPU Mem (GB)	Speed (ins/s)
BERT	18.7	9.4
ALBERT	14.6	9.4
POLY-INLINE	9.9	16.8
POLY-BATCH	5.1	36.4

Poly-encoder is the fastest and most memory efficient model, with a sacrifice of performance.

2.2 Context Expansion using Multi-task Learning



Model	Link Prediction Precision	Recall	F1	Ranking R@1	R@5	R@10	Clustering 1-1	VI	F
BERT	48.2	46.4	47.3	48.8	75.4	84.7	74.3	89.3	26.3
BERT+MF	73.9	71.3	72.6	73.9	95.8	98.6	77.0	92.0	40.9
MULTI ($\alpha = 1$)	65.6	63.2	64.4	66.7	91.8	95.6	64.6	87.7	24.3
MULTI ($\alpha = 5$)	66.9	64.5	65.7	65.4	91.8	95.6	68.7	88.8	27.4
MULTI ($\alpha = 10$)	65.2	62.9	64.0	64.4	91.4	95.6	70.3	89.5	28.1
MULTI ($\alpha = 20$)	64.7	62.4	63.5	63.9	91.0	95.0	68.3	88.8	26.7
MULTI+MF ($\alpha = 1$)	72.8	70.2	71.5	71.9	94.0	96.4	76.3	91.8	36.1
MULTI+MF ($\alpha = 5$)	73.3	70.7	72.0	72.4	94.0	96.5	72.8	90.8	33.1
MULTI+MF ($\alpha = 10$)	72.2	69.6	70.8	70.4	93.4	96.4	71.8	90.2	29.9
MULTI+MF ($\alpha = 20$)	70.8	68.2	69.5	69.4	93.4	97.3	73.2	90.6	28.6

We conduct utterance-to-utterance and utterance-to-thread classification at the same time, which outperforms pairwise models when manual features are unavailable.

2.3 Bipartite Graph Matching for Conversation Disentanglement

U51 [20:24] <groob> corba: I tried virtualization but the performance just isn't that great on my hardware sadly.

⋮

U54 [20:28] <groob> corba: Thanks for the help! I'll try using a dedicated boot partition instead. The biggest downside that I can think of is that the partition order in the menu will change depending on which OS ran update-grub last. If that could be solved, then it would be perfect.

⋮

U58 [20:29] <groob> corba: Err, I mean the order of the operating systems. OS1 appears first currently, but if OS2 ran update-grub, it would cause OS2 to be listed first.

U59 [20:29] <corba> groob, you could change that with grub-customizer easily

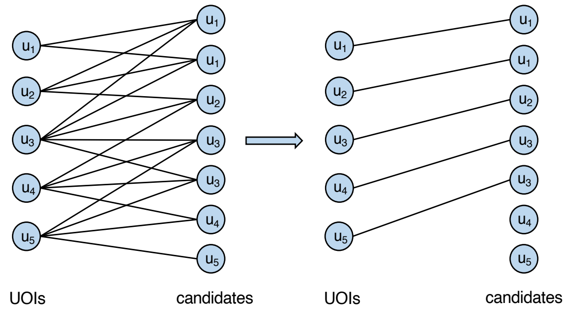
U60 [20:30] <groob> corba: Oh, cool. Good to know! Thanks!

⋮

U67 [20:32] <Bashing-om> groob: There can be only one boot control authority per hard drive . What I do is disable 30_os-prober in the seconday system,' sudo update-grub' amd in the primary also 'sudo apt-update-grub' to propogate the changes to the system(s) .

U_{67} chooses U_{54} as parent in global decoding algorithm but chooses U_{58} in greedy algorithm.

Bipartite matching-based algorithm recovers threads by identifying the parent utterance of a set of UOIs jointly.



	Precision	Recall	F1
Oracle	88.4	85.2	86.8
Rule-Based	73.7	70.9	72.3
FFN	73.8	71.0	72.3
BERT+FFN	72.9	70.3	71.5

We frame conversation disentanglement as a **maximum-weight bipartite matching** problem. It has the potential to outperform greedy approaches.

Conclusion

- BERT combined with manual features is still a strong baseline for conversation disentanglement
- The multi-task learning framework that conducts utterance-to-utterance and utterance-to-thread classification at the same time outperforms pairwise models when manual features are not available
- Bipartite graph matching-based conversation disentanglement shows potential to outperform greedy approaches.

The 19th Annual Workshop of the Australasian Language Technology Association. 8-10 December 2021

* Please contact rongxinz1@student.unimelb.edu.au for further information.