Conversational Response Re-ranking Based on Event Causality and Role Factored Tensor Event Embedding

Shohei Tanaka[1], Koichiro Yoshino[12], Katsuhito Sudoh[1], Satoshi Nakamura[1]

[1]Nara Institute of Science and Technology [2]PRESTO, Japan Science and Technology Agency

Abstract | Method to select coherent and diverse responses based on event causality

Neural Conversational Model (NCM) struggles with maintaining response **coherency** and **diversity** to **continue dialogues**.
To solve the problem, we proposed a method to re-rank response candidates by using **Event Causality** and **Event Embedding**.
Experimental results (dist, human evaluation, etc.) show that our method improves response coherency and diversity to continue dialogues.

1 Introduction | Improving dialogue continuity of NCM

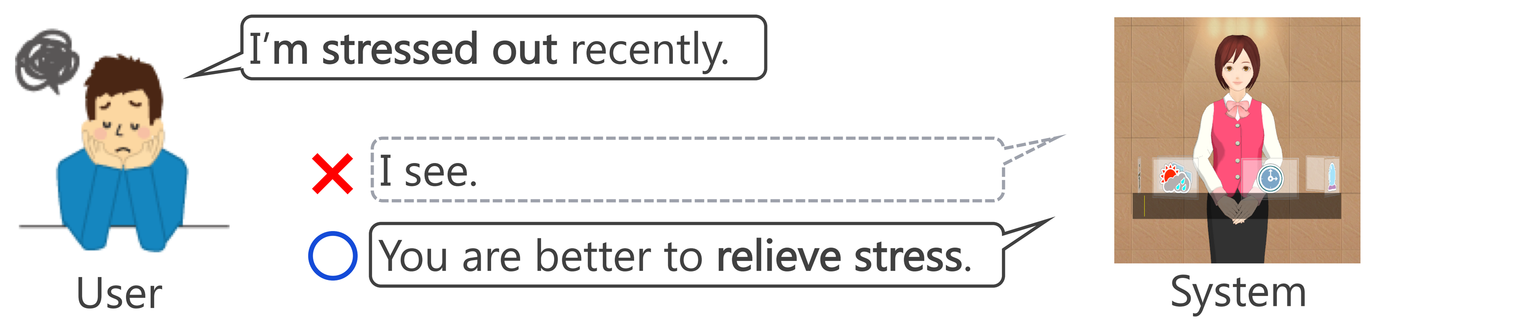
Dull Response Problem of NCM

NCM [Vinyals et al., 2015] often generates simple and dull responses due to the limitation of its ability.

I see. I don't know.

Response Re-ranking Based on Event Causality*

*cause-effect relations between two events

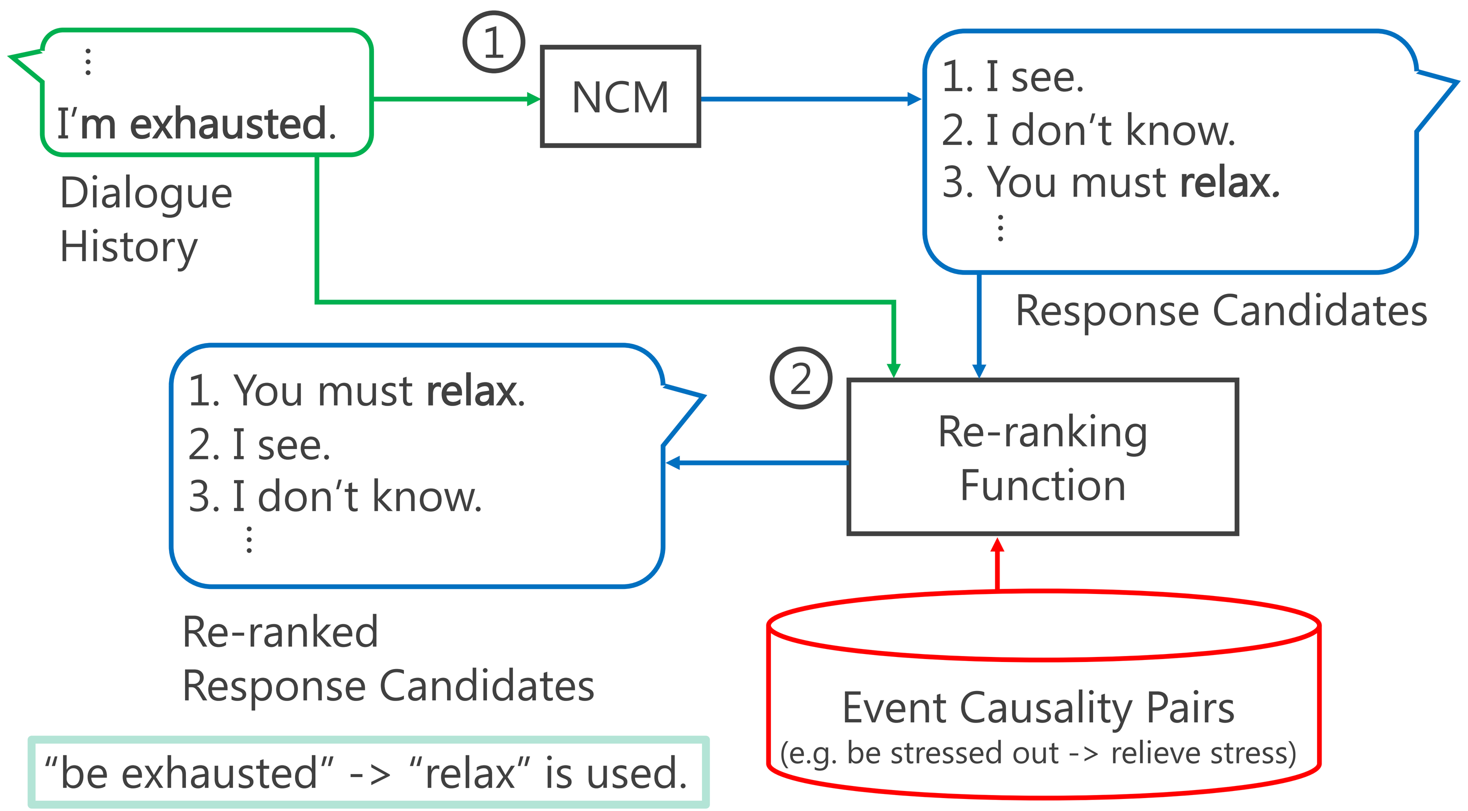


Response selection using a causality ("be stressed out" precedes "relieve stress")

We expected that our method improves response coherency and diversity to continue dialogues.

2 Method | Response Re-ranking Using Event Causality

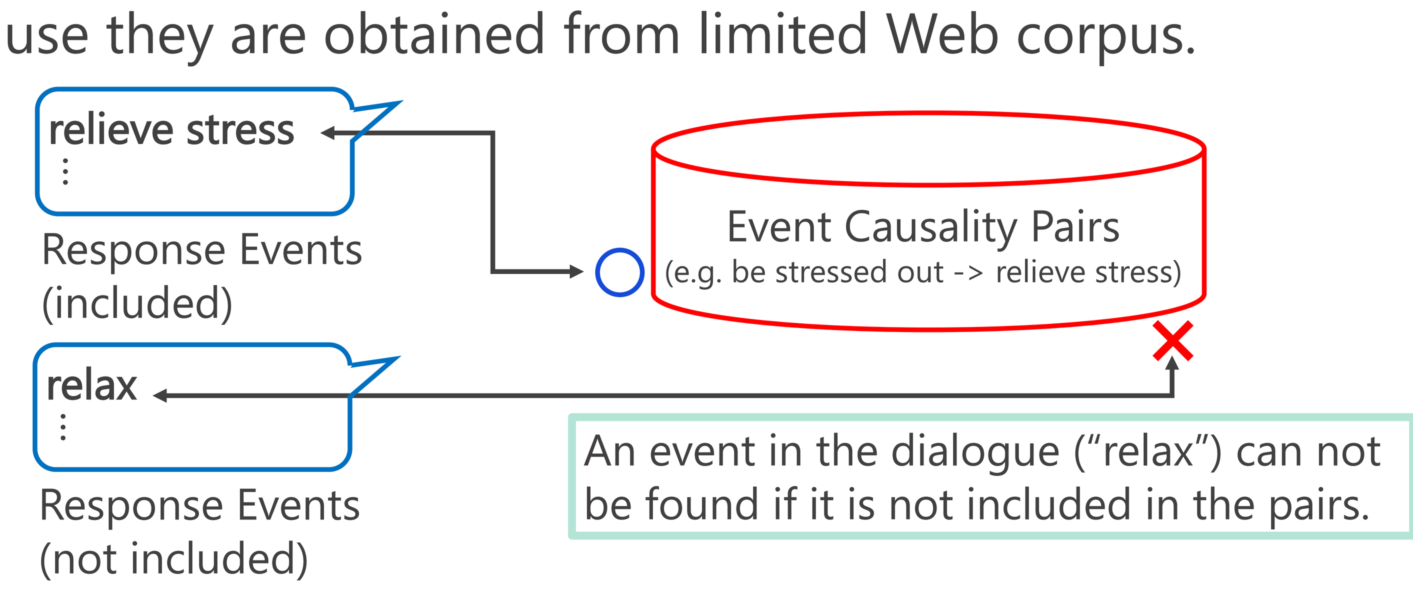
Overview of Re-ranking Based on Event Causality



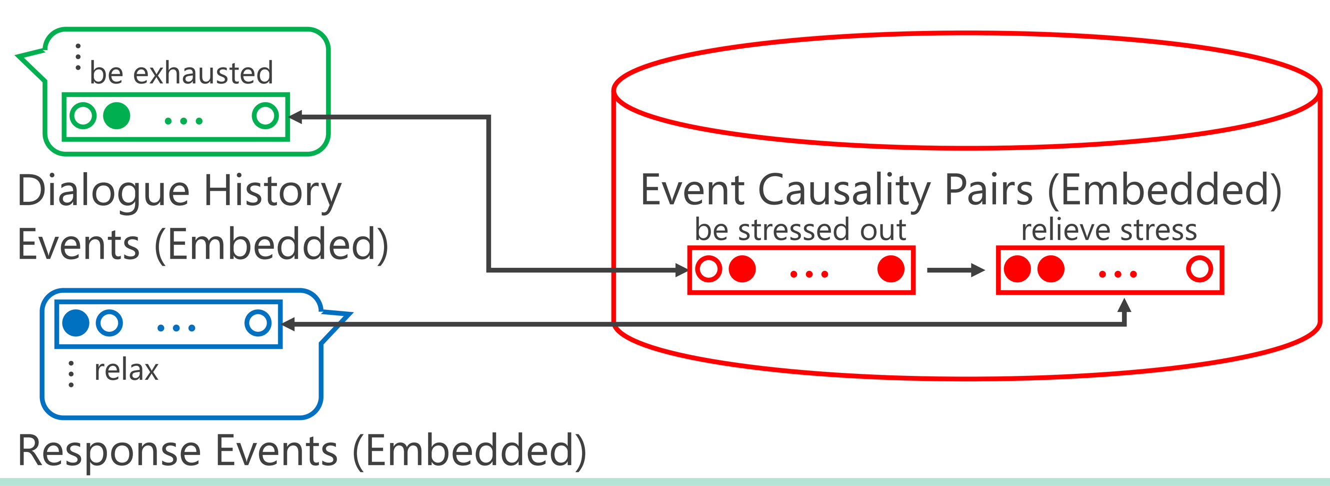
① NCM generates response candidates from a dialogue history.
② Re-ranking function selects response candidates with event causality related to the dialogue history.
Proposed method uses event causality pairs [Shibata et al., 2014] to find causalities between the dialogue history and response candidates.

Causality Matching Based on Event Embedding

Event causality pairs do not include all causalities in dialogue because they are obtained from limited Web corpus.



Solve this problem by using **event embedding** [Weber et al., 2018].
Find a similar event causality pair on vector space.



A causality in the dialogue ("be exhausted" -> "relax") is found if a similar causality ("be stressed out" -> "relieve stress") is included in the pairs.

3 Experiments | Improvement of diversity and coherency

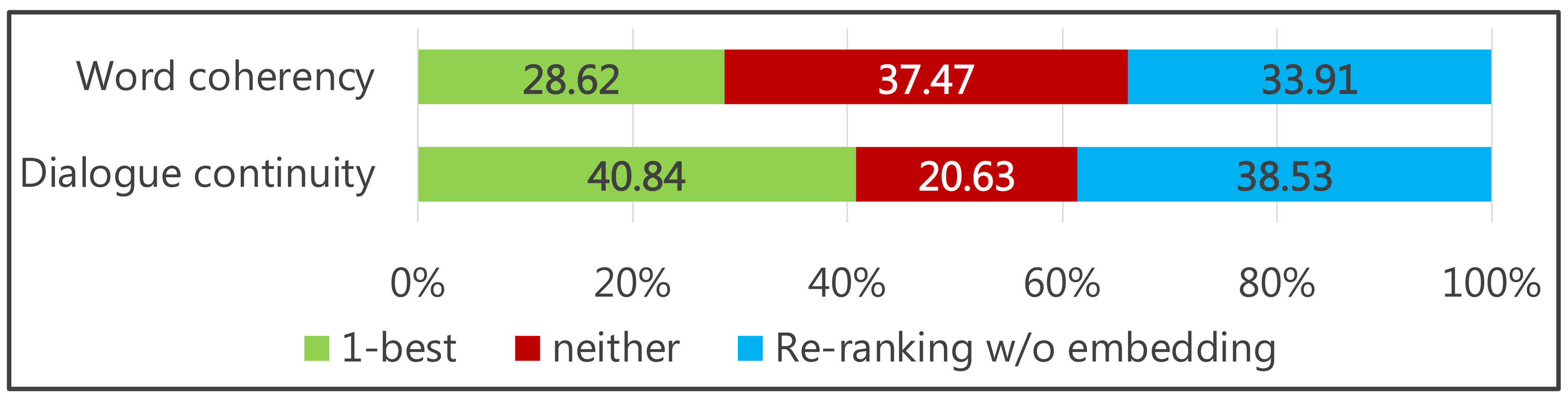
Automatic Evaluation

Comparison in Automatic Metrics

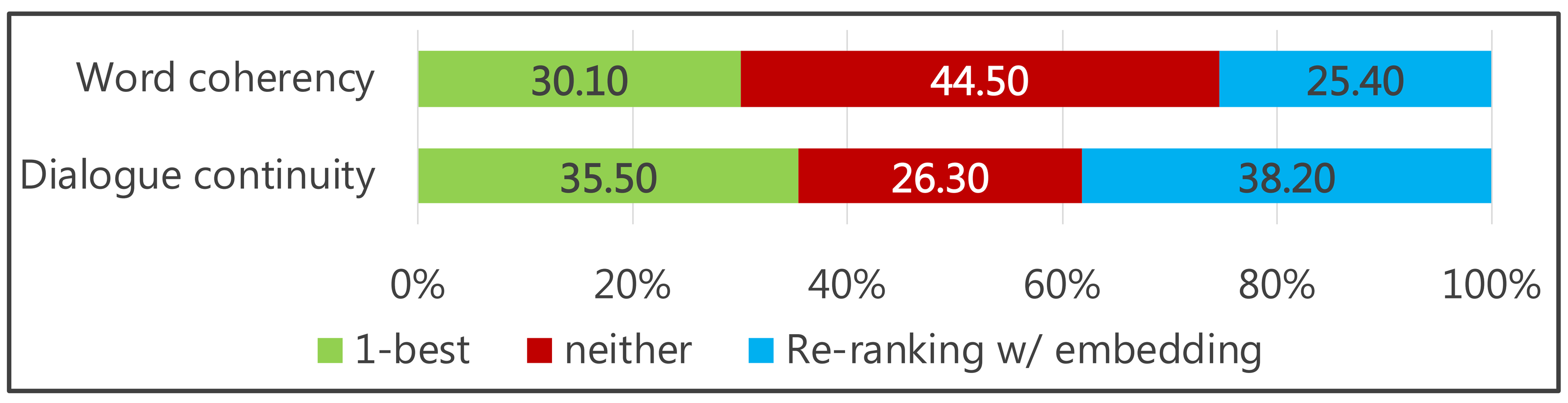
NCM	Re-ranking	Re-ranked (%)	dist-1	dist-2	PMI
EncDec	1-best	-	0.06	0.18	1.77
	w/o embedding	12.72	0.06	0.19	1.78
	w/ embedding	69.39	0.07	0.21	1.77
HRED	1-best	-	0.07	0.20	1.84
	w/o embedding	12.25	0.06	0.20	1.84
	w/ embedding	71.53	0.06	0.20	1.86

- Re-ranked ratios were increased drastically by the embedding.
- Dist is a number of distinct words in responses.
  -> Diversity was improved.
- PMI is a mutual information between a history and a response.
  -> Coherency was improved.

Human Evaluation



Preference Test; 1-best v.s. Re-ranking w/o embedding; # evaluators: 10; # dialogues: 100



Preference Test; 1-best v.s. Re-ranking w/ embedding; # evaluators: 10; # dialogues: 100

- Coherency was improved by the model without the embedding but worsened with the embedding.
  -> Evaluators did not acknowledge causalities because the event embedding over-generalized events.
- Continuity was improved by the model with the embedding.
  -> Number of dull responses was reduced.

Proposed method improved coherency, diversity, and continuity.

4 Case Study | Re-ranking examples

Appropriate Re-ranking

User: Because of my fears, I have **been stressed out**.
System (1-best): Are you OK?
System (Re-ranked): Are you OK? Don't **work too hard**.

("work too hard" -> "be stressed out") was used appropriately.
Examples such as e.g. 1 are rarely observed.

Over-generalization

User: Our president gives me beer, but he **can't drink**.
System (1-best): I see.
System (Re-ranked): I **drink**.

("drink" -> "can drink") was obtained by generalizing ("enter restaurant" -> "order beer"), which is included in the pairs.

Coherency and naturalness were worsened by re-ranking.

Updating the event embedding and maintaining the response naturalness are necessary.