

因果関係と事態分散表現を用いた 雑談対話応答のリランキングにおける 傾向分析

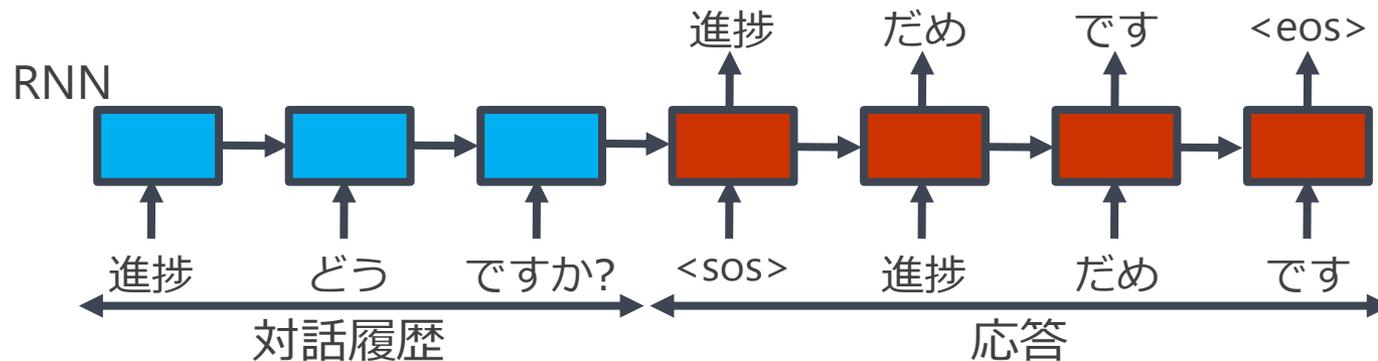
田中翔平¹, 吉野幸一郎¹², 須藤克仁¹, 中村哲¹

¹奈良先端科学技術大学院大学

²科学技術振興機構さきがけ

はじめに

Neural Conversational Model (NCM)



NCM [Vinyals et al., 2015] **柔軟**に応答を生成可能

単純でつまらない応答を生成する傾向

ユーザのシステムに対する興味の消失

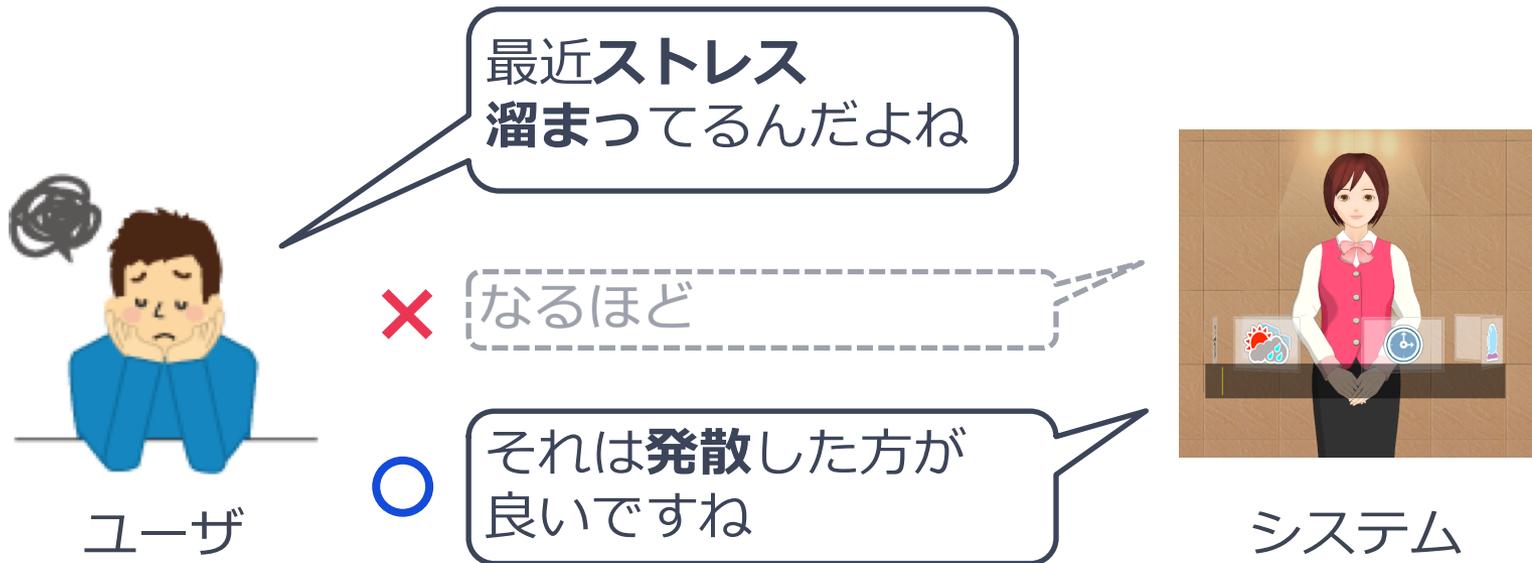
分かりません

なるほど

応答の一貫性や多様性を向上させることで
対話を継続する必要がある

因果関係に基づく応答選択

因果関係に基づいて NCM から生成された応答候補を
リランキング



「ストレスが溜まる」→「発散」という
因果関係を用いて応答を選択

因果関係とは？

2つの事態間に原因と結果の関係が成立する関係

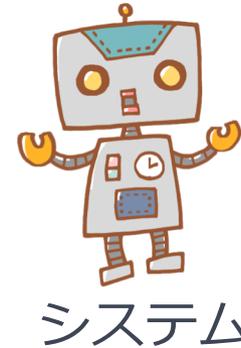
e.g. ストレスが溜まる（原因） → 発散（結果）

Why-QA システム [Oh et al., 2013] で使用



なぜ津波が発生するの？

地震が地震波を起こす
からです



「地震が地震波を起こす」 → 「津波が発生する」
という因果関係を用いて質問に関連する回答を生成

なぜ因果関係が雑談応答選択に役立つのか？

雑談応答を因果関係に基づいて選択

応答中の事態は対話履歴に関連

→ 一貫性の向上

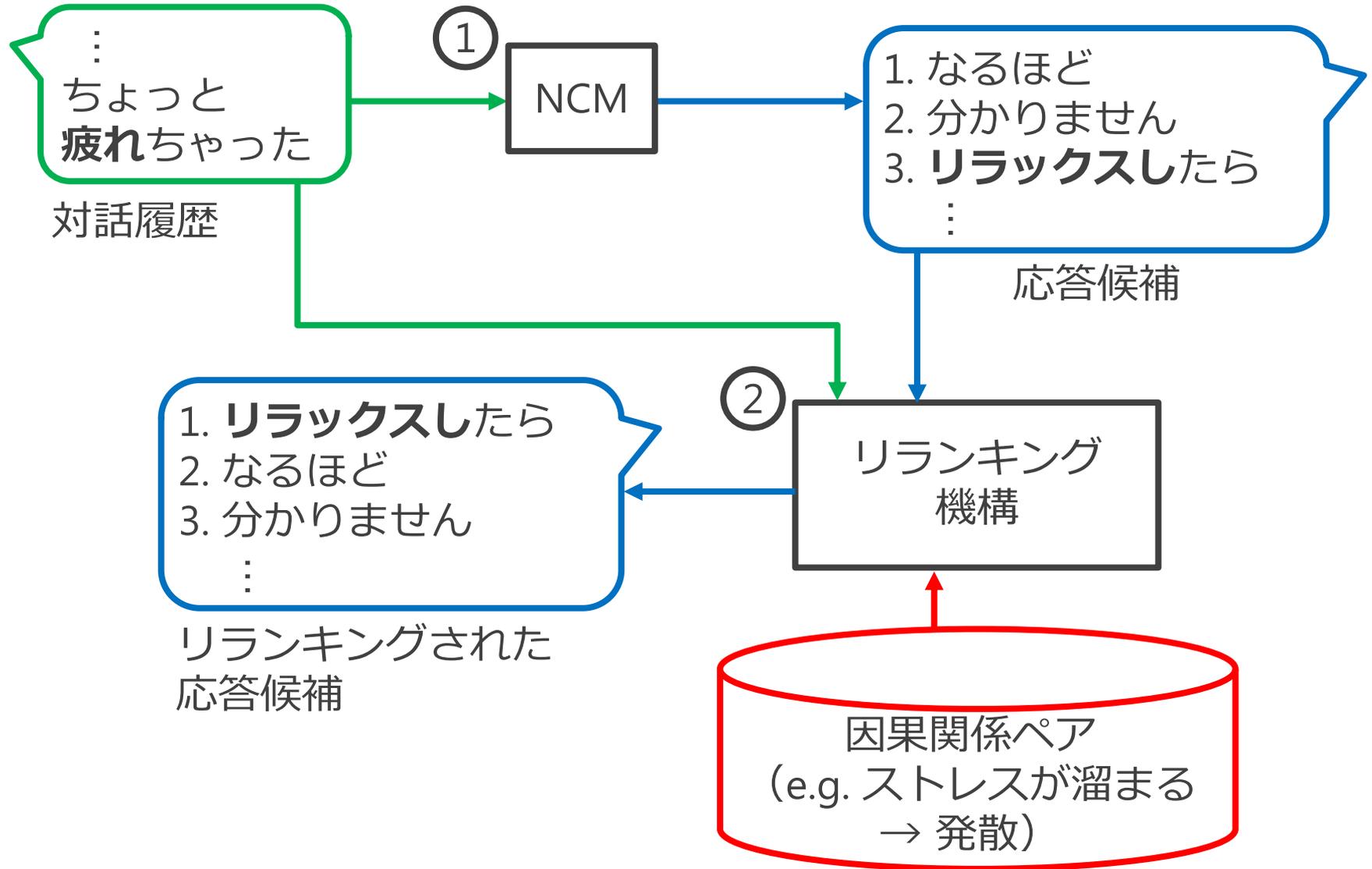
応答は対話履歴に対して高い相互情報量

→ 多様性の向上

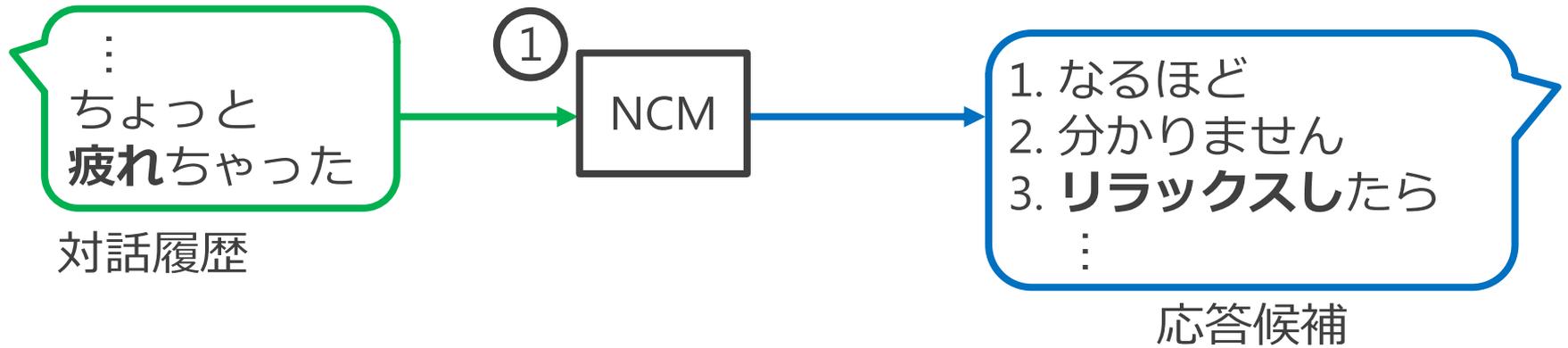
対話継続性の向上

因果関係を用いた応答の リランキング

リランキングの概要



■ 応答候補の生成



対話履歴から応答候補を生成

因果関係に基づくリランキング

対話履歴との間に因果関係を持つ
応答候補を高い順位にリランキング

1. リラックスしたら
2. なるほど
3. 分かりません
- ⋮

リランキングされた
応答候補

「疲れる」→「リラックスする」
という因果関係を使用

②

リランキング
機構

因果関係ペア
(e.g. ストレスが溜まる
→ 発散)

因果関係ペア

各事態は述語項構造を用いて表現

述語：必須，格要素：持たない場合も

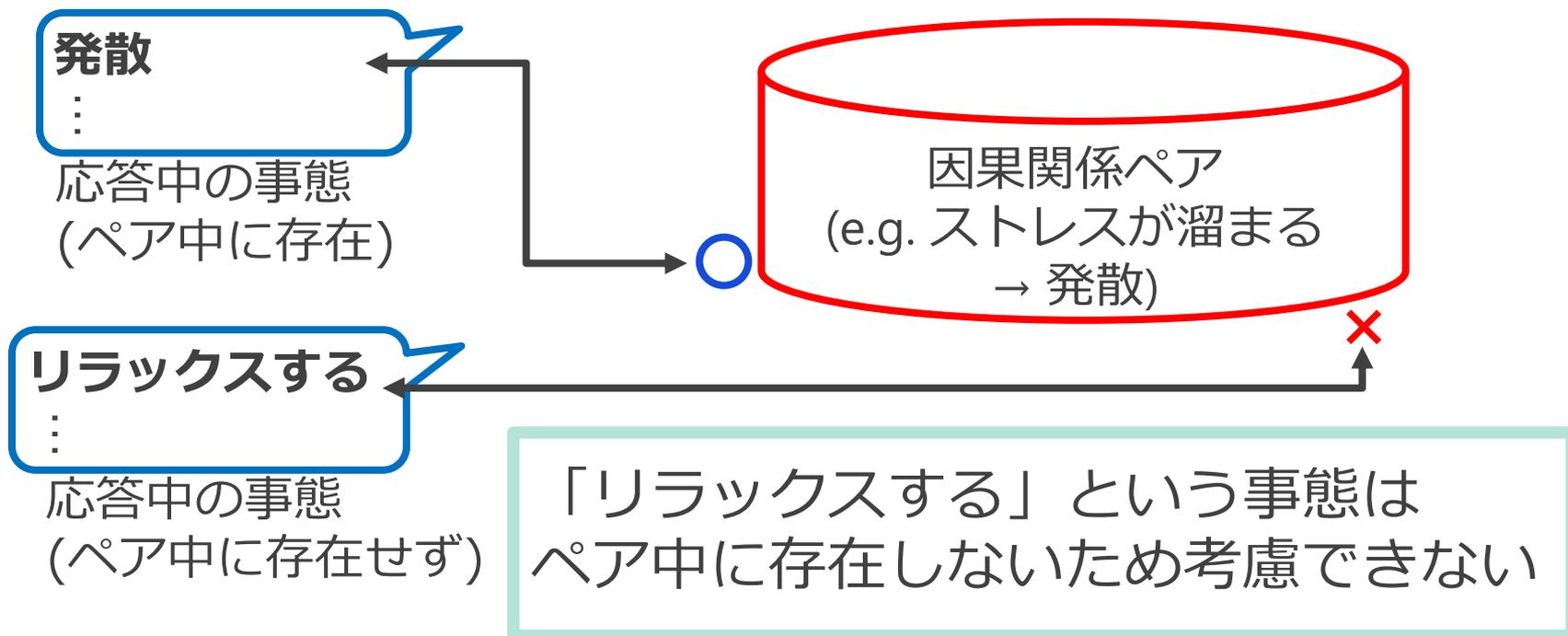
因果関係			
原因		結果	
述語	格要素	述語	格要素
溜まる	ストレス	発散	-

因果関係ペアを用いて対話履歴と応答候補の間に存在する因果関係を探索

因果関係ペアのカバレッジ問題

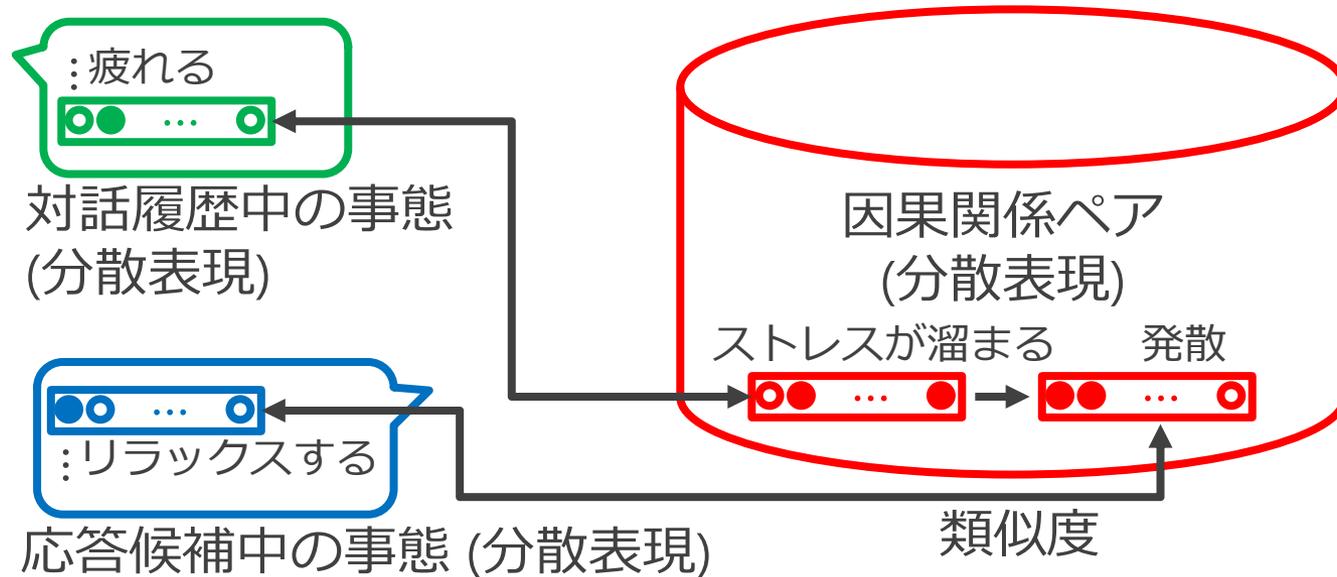
限られた Web コーパスから構築されているため

**因果関係ペアは対話中の全ての
因果関係を網羅していない**



事態分散表現に基づくマッチング

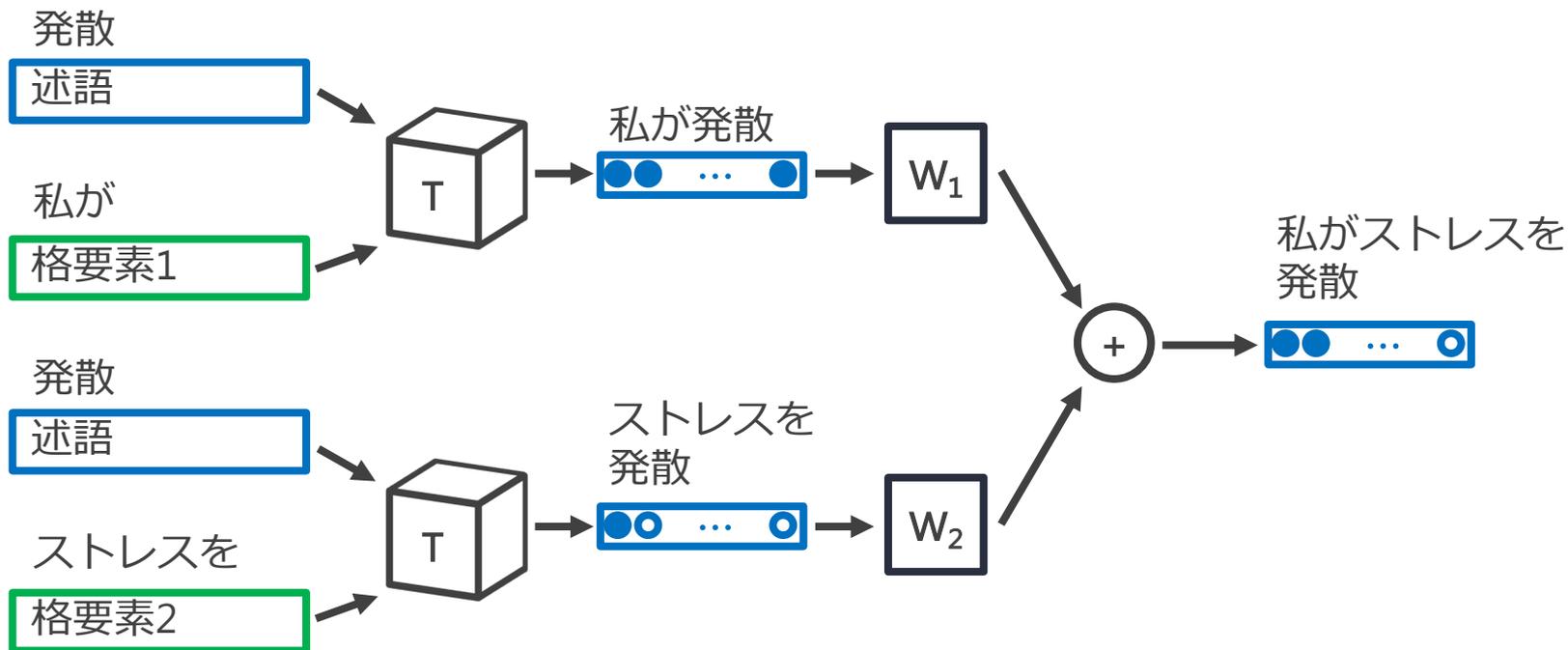
類似した因果関係をベクトル空間上で探索



対話中の「疲れる」→「リラックスする」という因果関係は「ストレスが溜まる」→「発散」という類似した因果関係がペア中に含まれるため考慮可能

Role Factored Tensor Model (RFTM) [Weber et al., 2018]

事態を述語と付随する格要素の関係に基づいて
分散表現に変換



述語の詳細な意味を捉えることが可能

実験

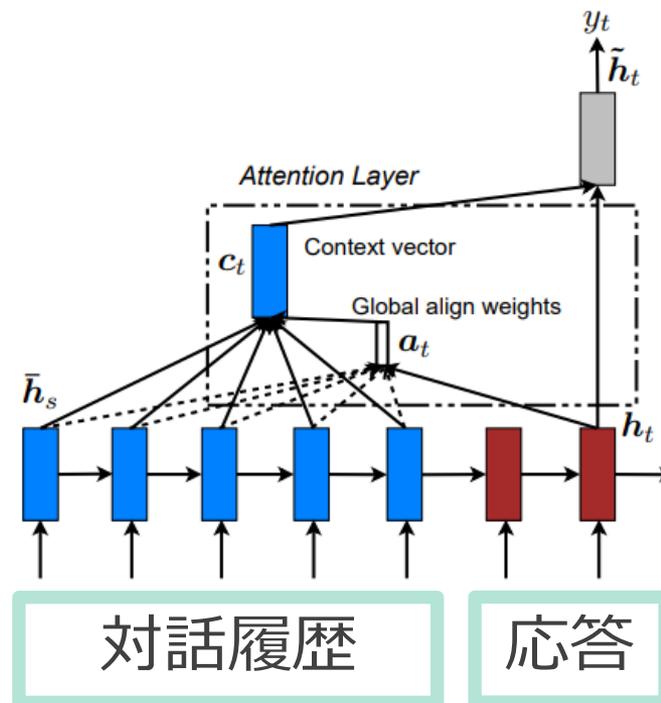
■ 実験設定

	設定
NCM	EncDec, HRED
リランキング	1-best (リランキングしない), 分散表現なし/あり
データ	260万対話の Twitter データ (テストデータは6万)

Encoder-Decoder with Attention [Luong et al., 2015]

応答を生成する各ステップにおいて、対話履歴のどの部分に注視するか (attention を掛けるか) を計算

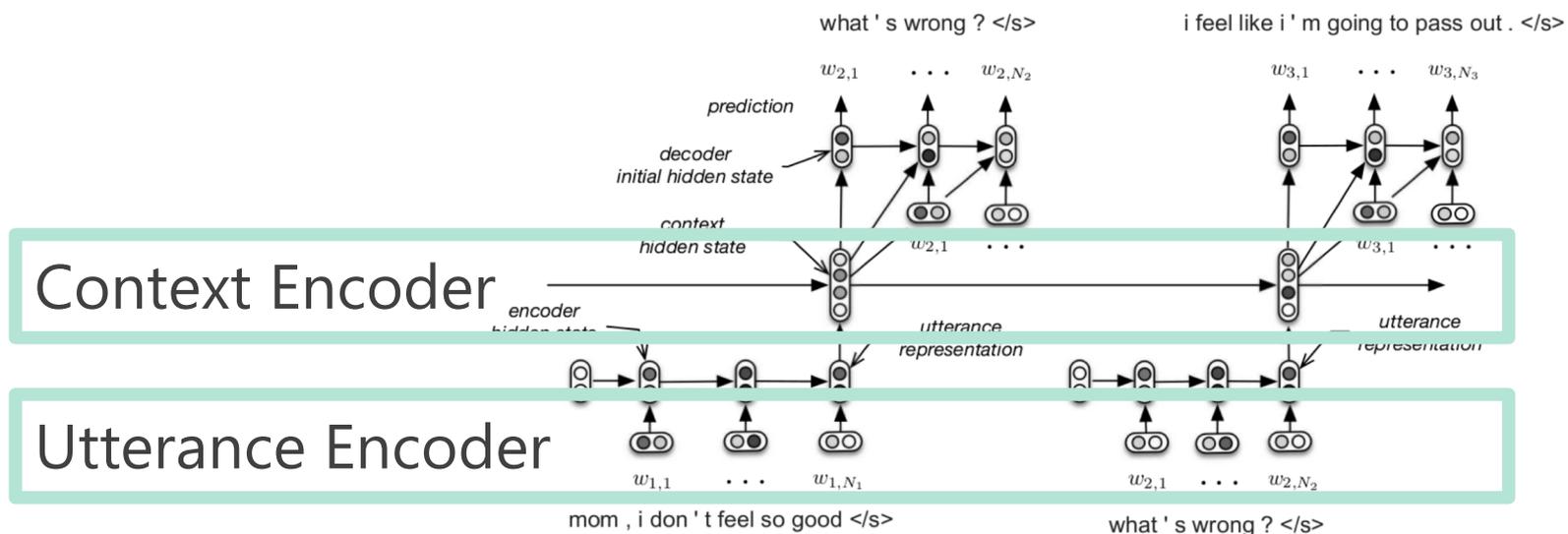
Attention の計算には Encoder, Decoder それぞれの hidden state vector の内積などを使用



Hierarchical Recurrent Encoder-Decoder [Serban et al., 2016]

対話履歴を階層的 (hierarchical) にエンコード

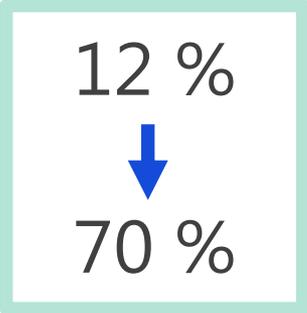
各発話を utterance encoder でエンコードした後、対話履歴全体を context encoder でエンコード



■ リランキングされた応募候補の割合

どの程度リランキングが適用可能かを表す

リランキング	NCM	割合
分散表現なし	EncDec	6,469 (12.72)
	HRED	6,231 (12.25)
分散表現あり	EncDec	35,284 (69.39)
	HRED	36,373 (71.53)



事態分散表現を用いることでリランキングされる割合が劇的に上昇

■ 正解応答に対する類似度スコア

- BLEU

実際の応答との N-gram の一致率

実際の応答は文脈を考慮

システム応答が文脈を考慮しているかをある程度表現

- NIST

BLEU に似ているが、低頻出語に重み付けしている点で異なる

正解応答に対する類似度スコア (続き)

- Vector Extrema

正解応答と生成された応答のコサイン類似度

各文ベクトル e_s はそれぞれの応答中に含まれる単語の単語ベクトル e_w より各次元 d において極値を取ることで取得

$$e_{sd} = \begin{cases} \max_{w \in s} e_{wd} & \text{if } e_{wd} > |\min_{w' \in s} e_{w'd}| \\ \min_{w \in s} e_{wd} & \text{otherwise} \end{cases}$$

BLEU, NIST, and extrema

NCM	リランキング	BLEU	NIST	extrema
EncDec	1-best	1.12	1.19	0.42
	分散表現なし	1.09	1.17	0.42
	分散表現あり	1.00	1.04	0.39
HRED	1-best	1.34	2.74	0.42
	分散表現なし	1.33	2.73	0.42
	分散表現あり	1.28	2.74	0.41



リランキングにより類似度スコアが低下

NCM は正解応答に類似した応答を生成

1-best 応答が最も高いスコアを持つ

多様性および一貫性の評価

- Dist-1, 2

全応答中の N-gram の異なり数の割合

システム応答が多様であることを意味

- Pointwise Mutual Information (PMI)

対話履歴中の単語

$$\text{PMI} = \frac{1}{|\text{response}|} \sum_{wr}^{|\text{response}|} \max_{wh} \text{PMI}(wr, wh)$$

応答の一貫性を意味

応答中の単語

Dist, PMI は正解応答とは無関係

Dist and Pointwise Mutual Information (PMI)

NCM	リランキング	dist-1	dist-2	PMI
EncDec	1-best	0.06	0.18	1.77
	分散表現なし	0.06	0.19	1.78
	分散表現あり	0.07	0.21	1.77
HRED	1-best	0.07	0.20	1.84
	分散表現なし	0.06	0.20	1.84
	分散表現あり	0.06	0.20	1.86



事態分散表現を用いたモデルが最も高いスコア

多様性と一貫性が**向上**

■ 人手評価に用いる NCM

ベースライン：HRED

V.S.

提案手法：
HRED + 分散表現なし/ありでリランキング

■ 人手評価

10人の評価者が100個の応答を以下の2つの基準で比較

- **Word coherency**

どちらの応答に含まれる単語がより対話履歴に関連しているか

- **Dialogue continuity**

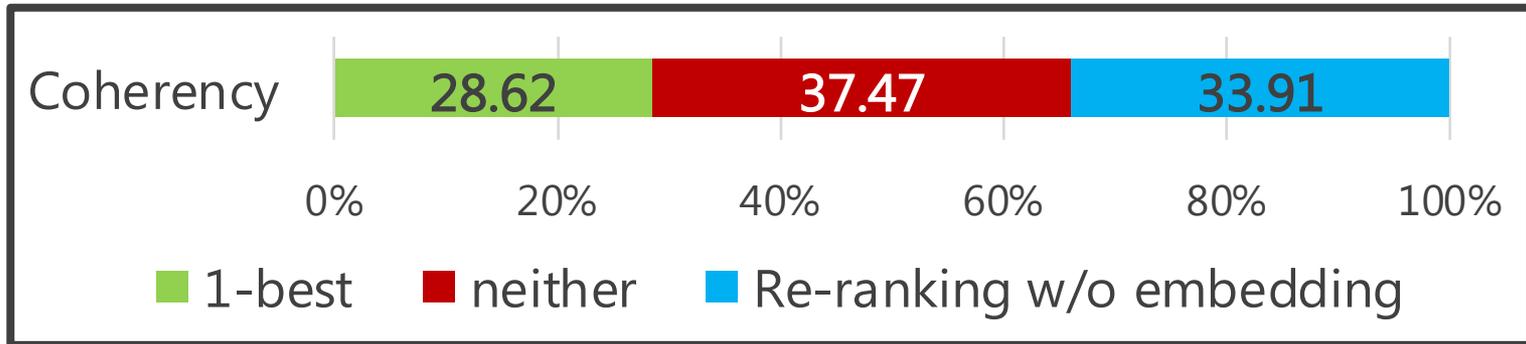
どちらの応答により返答したいと思うか

評価者の負担を軽減するため、以下の対話を除外

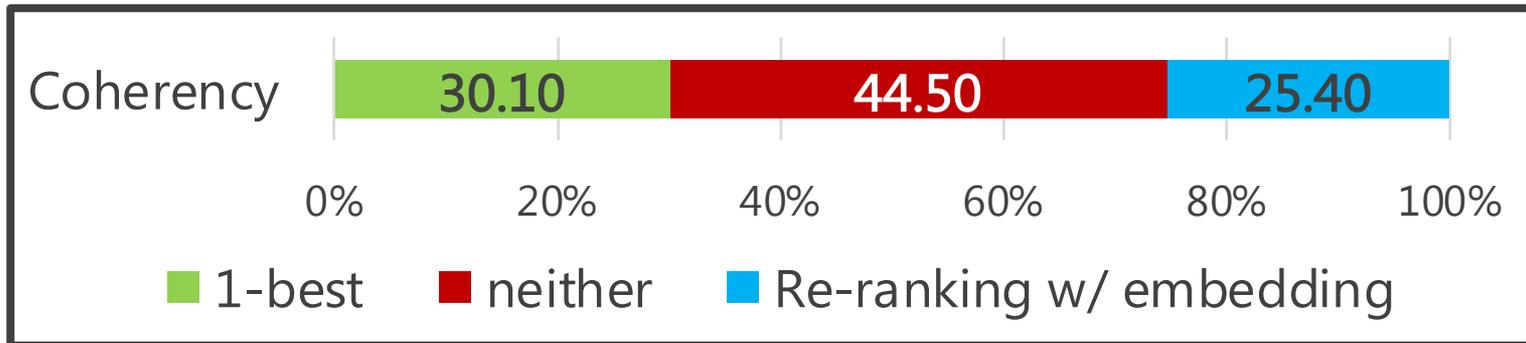
- ユーザの発話数が3以上
- 評価に外部知識が必要

Word Coherency

分散表現なし



分散表現あり

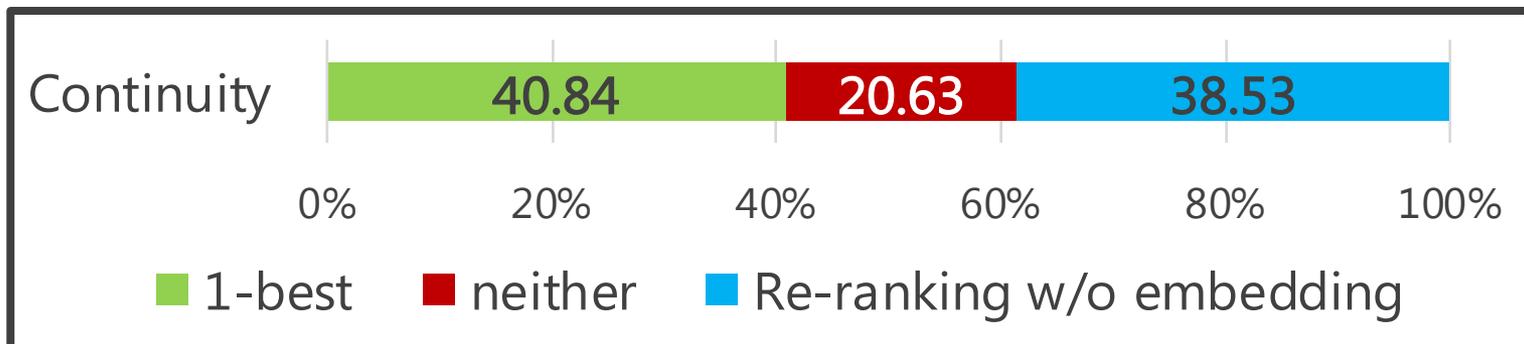


分散表現なしのリランキングでは**向上**するが、分散表現ありのリランキングでは**低下**

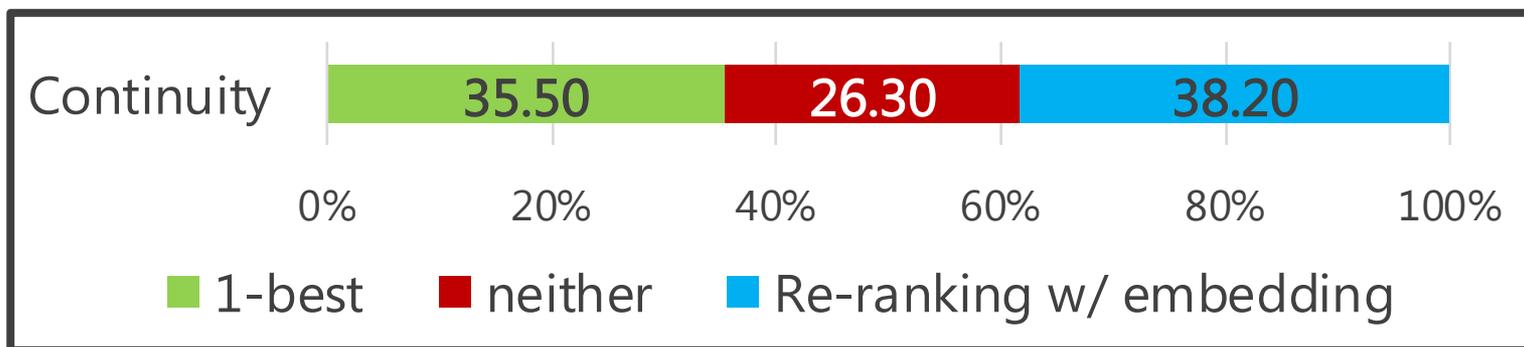
分散表現が事態を**過汎化**

Dialogue Continuity

分散表現なし



分散表現あり



単純な応答の割合が減少したために分散表現ありのリランキングで向上

■ 実験結果のまとめ

人手評価において

- Word coherency が向上
- Dialogue continuity が向上

自動評価においても多様性 (dist) と一貫性 (PMI) が向上することを確認

リランキング結果の事例分析

■ リランキング結果の分類 (分散表現なし)

リランキング結果およびリランキングに用いられた因果関係の妥当性を分類

分散表現なし

リランキング /因果関係	良い	悪い (ペア)	計
良い	14	12	26
悪い	14	10	24
両方良い	12	7	19
両方悪い	11	20	31
計	51	49	100

$$26 + 19 = 45$$

良い+両方良いを妥当なリランキングとすると、分散表現を用いない場合50%前後のリランキングが妥当

因果関係も50%前後が妥当

■ リランキング結果の分類 (分散表現あり)

分散表現あり

リランキング /因果関係	良い	悪い (ペア)	悪い (過汎化)	計
良い	2	0	6	8
悪い	2	2	5	9
両方良い	0	0	14	14
両方悪い	0	4	65	69
計	4	6	90	100

$$8 + 14 = 22$$

分散表現を用いる場合20%前後のリランキングが妥当

因果関係は5%前後が妥当

過汎化が大きな問題

■ 妥当なリランキング

対話1：

ユーザ：もう不安なことが多すぎて**ストレスが溜まって**く

システム (1-best)：大丈夫ですか

システム (Re-ranked)：大丈夫ですか**無理しない**でくださいね

「**無理をする**」→「**ストレスが溜まる**」という妥当な因果関係により**一貫した**応答を選択

過汎化

対話2：

ユーザ：お酒を飲めない社長がビールをくれる...

システム (1-best)：わかる

システム (Re-ranked)：お酒飲む

「お酒を飲む」 → 「お酒を飲める」 は因果関係ペア中に含まれる
「レストランに入る」 → 「ビールを頼む」という因果関係を汎化

事態分散表現が事態を過汎化

事態分散表現を改善することで過汎化を防ぐ必要

■ 応答の自然性の低下

対話3：

ユーザ：お、起き上がれない...会社**休む**...熱...

システム (1-best)：お大事に

システム (Re-ranked)：お**疲れ**様です

提案手法は述語項構造のみに着目
(「疲れ」→「休む」)

応答の自然性が低下

リランキング前後で自然性を保つ必要

まとめ

まとめ

因果関係に基づき一貫した多様な応答を選択する手法を提案

リランキングにより応答の一貫性、多様性、および対話継続性が向上



もう不安なことが多すぎて
ストレスが溜まってく



大丈夫ですか



大丈夫ですか無理しないでくださいね



システム

今後の課題

- 事態分散表現の改善
- 因果関係の妥当性や応答の自然性の考慮