Do Response Selection Models Really Know What's Next? Utterance Manipulation Strategies for Multi-turn Response Selection

Taesun Whang*, Dongyub Lee*, Dongsuk Oh, Chanhee Lee,

Kijong Han, Dong-hun Lee, and Saebyeok Lee

WISEnut kakao KOREA UNIVERSITY kakaoenterprise

*Equal Contribution.

Motivation

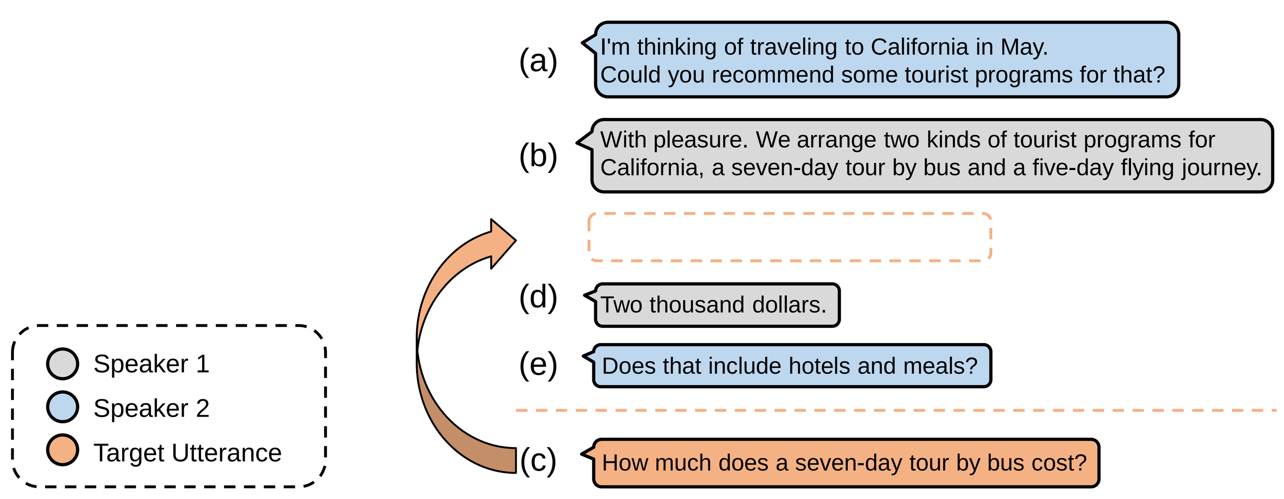
- Domain adaptation based on an additional training on a target corpus is extremely time-consuming and computationally costly.
- Formulating response selection as a dialog-response binary classification task is insufficient to represent intra- and inter-utterance interactions as the dialog context is formed by concatenating all utterances.
- The models tend to select the optimal response depending on how semantically similar it is to a given dialog.

Contribution

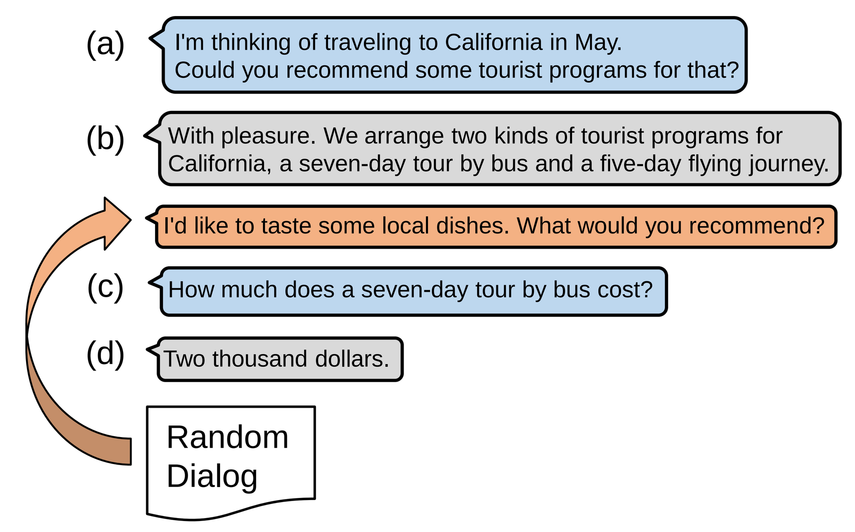
- We show that existing response selection models are more likely to predict a semantically relevant response with its dialog rather than the next utterance (Adversarial Experiment).
- We propose highly effective self-supervised learning methods, **utterance manipulation strategies (UMS)**, which aid the model towards maintaining dialog coherence.
- We obtain **state-of-the-art** performance on multiple public benchmarks (i.e., Ubuntu, Douban, and E-commerce).

Proposed Method

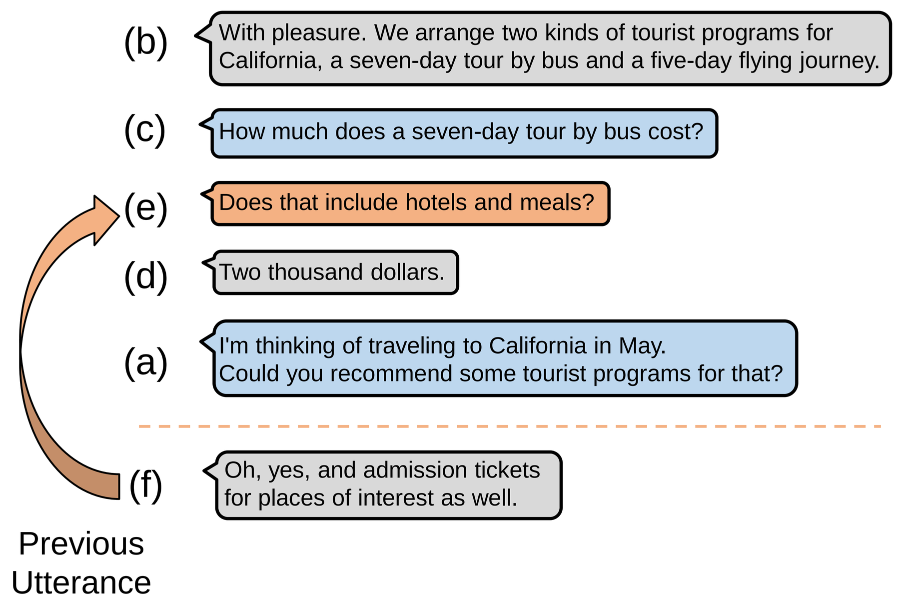
[Utterance Insertion]



[Utterance Deletion]



[Utterance Search]



Multi-task Learning

- The output representations of special are used to classify whether each toke is in a correct position to be inserted, deleted, and searched.
- Target tokens for each task ($[\textbf{INS}]_t$, $[\textbf{DEL}]_t$, and $[\textbf{SRCH}]_t$) are labeled as 1, otherwise 0.
- Binary cross-entropy loss for all auxiliary tasks to optimize the model.

$$p(y_{\text{TASK}} = 1|\mathbf{X}_{\text{TASK}}) = \sigma(\mathbf{w}^\top \mathbf{x}_{\text{TASK}} + b)$$

Utterance Insertion

- Find where the selected utterance should be inserted.
- [INS] tokens are positioned before each utterance and after the last utterance.
- $\boldsymbol{u_t}$ is the target utterance and $[\textbf{INS}]_{\boldsymbol{t}}$ is the target insertion token.

$$\mathbf{X}_{\text{INS}} = [[\text{CLS}]\ [\text{INS}]_1\ u_1 [\text{INS}]_2\ u_2 \ldots u_{t-1}\ [\text{INS}]_t\ u_{t+1} \ldots u_k\ [\text{INS}]_k\ [\text{SEP}]\ u_t\ [\text{SEP}]]$$

Utterance Deletion

- Find an unrelated utterance to the dialog.
- The unrelated utterance is sampled from the random dialog.
- [DEL] tokens are positioned before each utterance.
- $\boldsymbol{u^{rand}}$ is the utterance from the random dialog and $[\textbf{DEL}]_{\boldsymbol{t}}$ is the target deletion token.

$$\mathbf{X}_{\text{DEL}} = [[\text{CLS}]\ [\text{DEL}]_1\ u_1\ [\text{DEL}]_2\ u_2 \ldots [\text{DEL}]_t\ u^{rand}\ [\text{DEL}]_{t+1}\ u_t \ldots [\text{DEL}]_{k+1}\ u_k\ [\text{SEP}]]$$

Utterance Search

- Find the previous utterance of the last utterance from the jumbled utterances.
- Shuffle utterances except for the last utterance.
- [SRCH] tokens are positioned before each utterance.
- $\boldsymbol{u'_t}$ $(\boldsymbol{u_{k-1}})$ is the previous utterance of the last utterance $\boldsymbol{u_k}$ and $[\textbf{SRCH}]_{\boldsymbol{t}}$ is the target search token.

$$\mathbf{X}_{\text{SRCH}} = [[\text{CLS}]\ [\text{SRCH}]_1\ u'_1 [\text{SRCH}]_2\ u'_2 \ldots [\text{SRCH}]_t\ u'_t \ldots u'_{k-1} [\text{SEP}]\ u_k\ [\text{SEP}]]$$

Results and Discussion

Quantitative Results

Models	Ubuntu			Douban						E-commerce		
	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP	MRR	P@1	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
BERT (ours)	0.820	0.906	0.978	0.597	0.634	0.448	0.279	0.489	0.823	0.641	0.824	0.973
ELECTRA	0.826	0.908	0.978	0.602	0.642	0.465	0.287	0.483	0.839	0.609	0.804	0.965
UMS_{BERT}	0.843	0.920	0.982	0.597	0.639	0.466	0.285	0.471	0.829	0.674	0.861	0.980
$\text{UMS}_{\text{ELECTRA}}$	0.854	0.929	0.984	0.608	0.650	0.472	0.291	0.488	0.845	0.648	0.831	0.974
BERT+	0.862	0.935	0.987	0.609	0.645	0.463	0.290	**0.505**	0.838	0.725	0.890	0.984
ELECTRA+	0.861	0.932	0.985	0.612	0.655	0.480	0.301	0.499	0.836	0.673	0.835	0.974
$\text{UMS}_{\text{BERT+}}$	**0.875**†	**0.942**†	**0.988**†	**0.625**	**0.664**	**0.499**	**0.318**	0.482	**0.858**	**0.762**	**0.905**	**0.986**
$\text{UMS}_{\text{ELECTRA+}}$	**0.875**	0.941	**0.988**	0.623	0.663	0.492	0.307	0.501	0.851	0.707	0.853	0.974

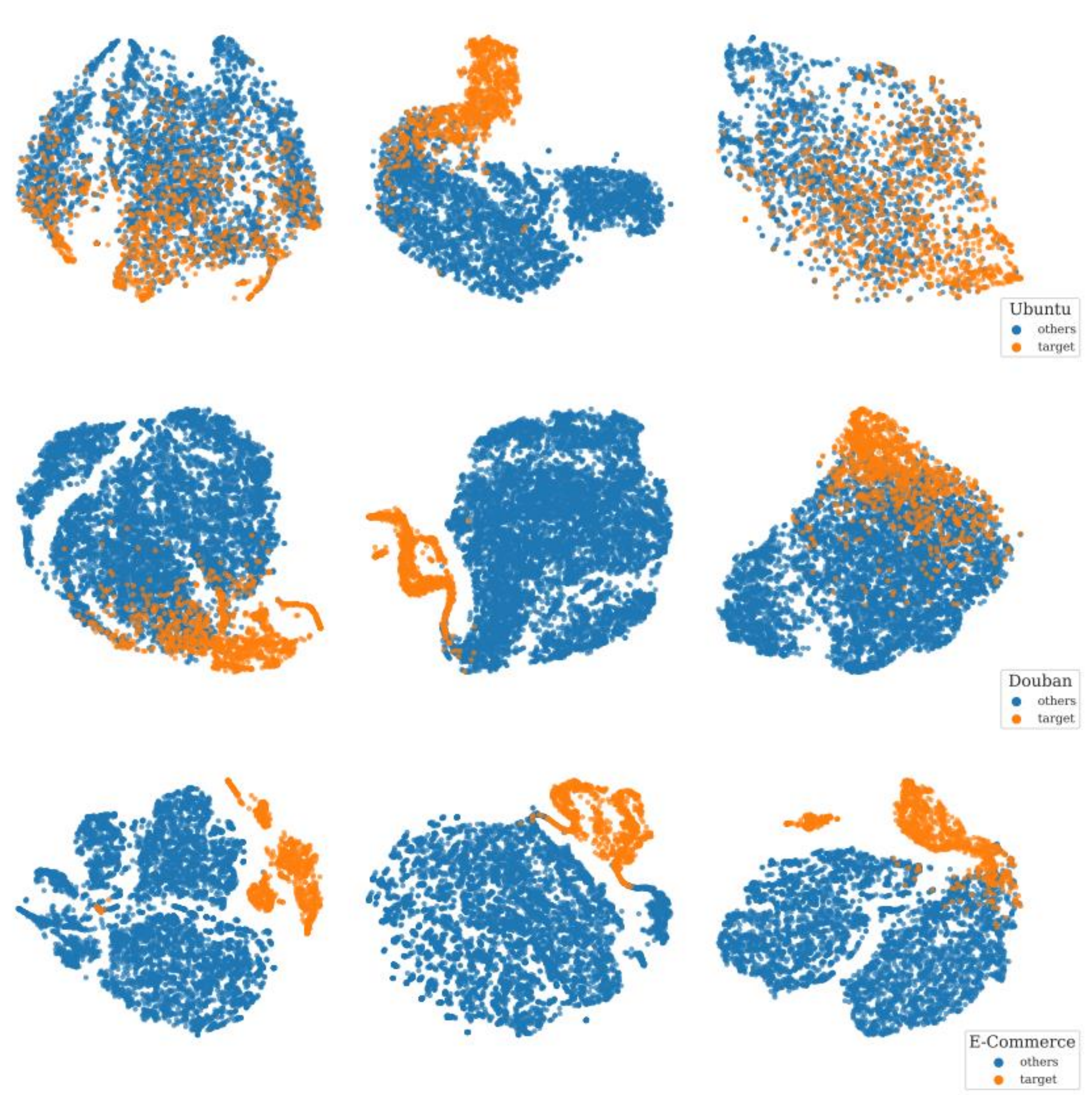
Adversarial Experiment

Approach	Model	Original		Adversarial	
		$R_{10}@1$	MRR	$R_{10}@1$	MRR
Baselines	BERT	0.820	0.887	0.199	0.561
	BERT+	**0.862**	**0.915**	0.203	0.573
	ELECTRA	0.826	0.890	0.304	0.614
	ELECTRA+	0.861	0.914	**0.329**	**0.636**
	Avg	0.842	0.902	0.259	0.596
UMS	BERT	0.843	0.902	0.310	0.622
	BERT+	**0.875**	**0.923**	0.363	0.656
	ELECTRA	0.854	0.910	0.397	0.668
	ELECTRA+	**0.875**	0.922	**0.437**	**0.692**
	Avg	0.862	0.914	0.377	0.660

Ablation Study

	Auxiliary Tasks	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MRR
1	None	0.826	0.908	0.978	0.890
2	INS	0.836	0.917	0.980	0.897
3	DEL	0.848	0.924	0.983	0.905
4	SRCH	0.834	0.915	0.981	0.896
5	INS + DEL	0.853	0.927	0.984	0.909
6	INS + SRCH	0.841	0.920	0.982	0.901
7	DEL + SRCH	0.852	0.927	0.983	0.908
8	INS + DEL + SRCH	**0.854**	**0.929**	**0.984**	**0.910**

Visualization



(a) Insertion (b) Deletion (c) Search