Multi-level Context Response Matching in Retrieval-Based Dialog Systems

Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin and Emmanuel Morin

LS2N, UMR CNRS 6004, Université de Nantes, France

firstname.lastname@ls2n.fr

1. Task Description

We participated in the *sentence selection* task of the 7^{th} edition of DSTC challenges. This task addresses the following points.

- The currently built systems are evaluated on non realistic scenarios.
- The number of candidate responses is small.
- Possibility of having more than one correct response.
- Sometimes none of the candidate responses is correct.

2. Subtasks

Our participation concerns the following three subtasks of sentence selection out of five.

- Subtask 1: One correct response among 100 candidate responses.
- Subtask 3: Between one and five correct responses (paraphrases) are correct among 100.
- Subtask 4: The 100 candidate responses may not include the correct response.

3. Datasets & Metrics

- Ubuntu Dialogue Corpus: Ubuntu related chat.
- Advising Corpus: Teacher-student conversations.
- We used the Recall@k, MRR and MAP evaluation metrics.

4. Approach

1. Encode the context and the response with a **shared** LSTM and compute their cross product: sequence-level similarity.
2. In parallel, compute a dot product between the embedding matrices and encode it with another LSTM: word-level similarity.
3. Concatenate both vectors and transform them into a probability using a FFNN: response ranking score.

5. Multi-Level Retrieval-Based Dialog System

- Our system is inspired by the dual encoder [1] and the Sequential Matching Network (SMN) [2].

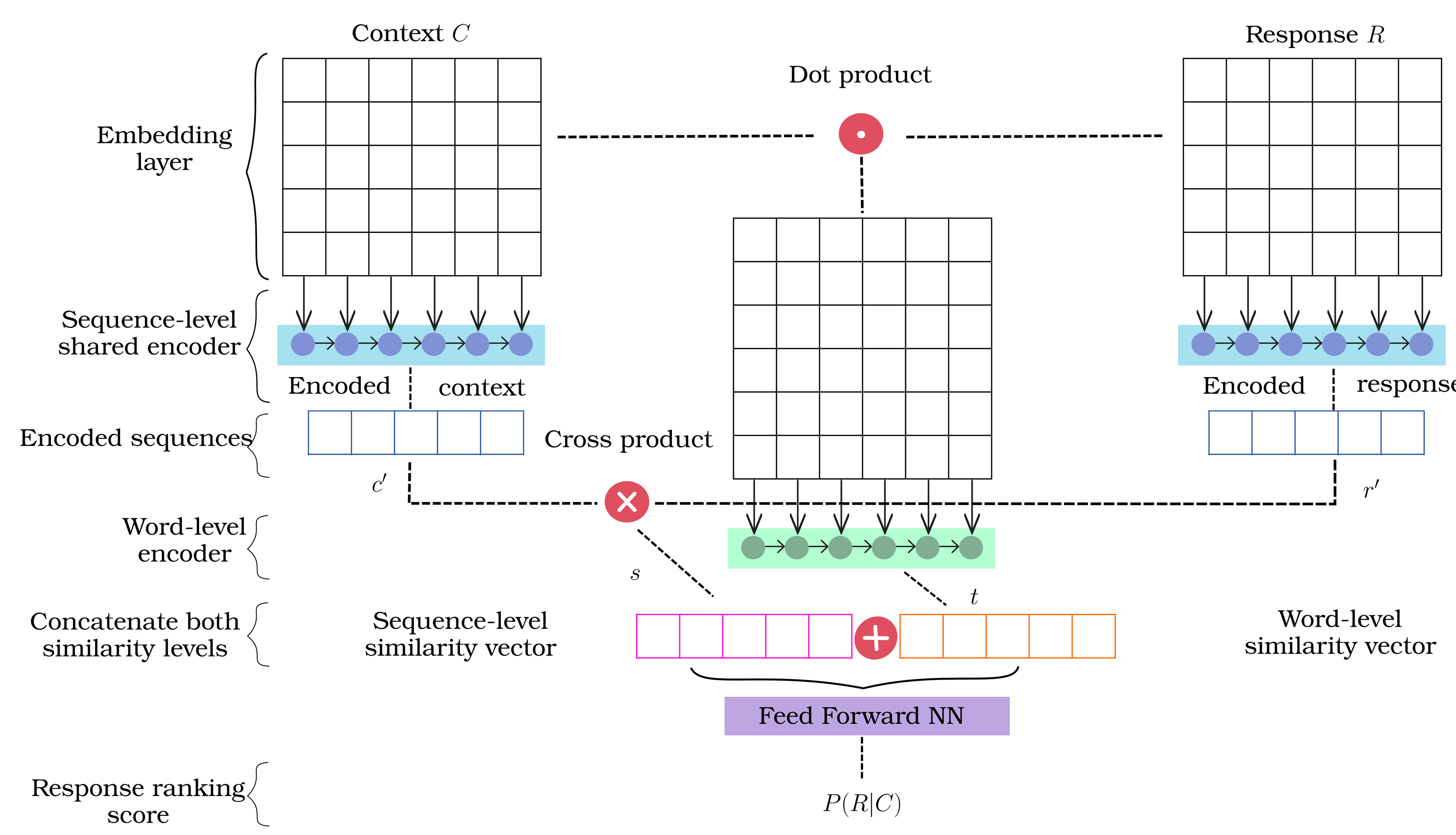

Figure 1: Architecture of our multi-level context response matching dialog system.

6. Experiments

System	Subtask	Measure	Ubuntu Dialogue Corpus	Advising Corpus case 1	Advising Corpus case 2
Baseline	Subtask 1	R@1	0.083	0.008	0.008
		R@10	0.359	0.102	0.094
		R@50	0.794	0.542	0.498
		MRR	0.175	0.053	0.048
Our system	Subtask 1	R@1	**0.446**	**0.114**	**0.1**
		R@10	**0.732**	**0.398**	**0.42**
		R@50	**0.937**	**0.782**	**0.802**
		MRR	**0.551**	**0.205**	**0.200**
	Subtask 3	R@1	-	**0.212**	**0.176**
		R@10	-	**0.586**	**0.57**
		R@50	-	**0.906**	**0.926**
		MRR	-	**0.338**	**0.297**
		MAP	-	**0.37**	**0.343**
	Subtask 4	R@1	**0.388**	**0.088**	**0.066**
		R@10	**0.592**	**0.31**	**0.316**
		R@50	**0.751**	**0.618**	**0.686**
		MRR	**0.462**	**0.163**	**0.15**

Table 1: Experimental results on test sets of Subtasks 1, 3 and 4.

7. Discussion

- Our system outperforms the baseline system on both datasets and on all metrics.
- Retrieving paraphrases was easier compared to retrieving only one response.

	Train	Dev	Test Case 1	Test Case 2
Ubuntu	20%	20%	20.20%	-
Advising	20.05%	18.80%	23.40%	18.40%

Table 2: Percentage of cases where no correct response is available (Subtask 4).

- Only 20% of training samples are cases where no correct response is available.

8. System Ablation

- Both similarity levels are important.
- With only sequence-level similarity, our system outperforms the baseline.

			Ubuntu	Advising
	Baseline	R@1	0.083	0.062
		R@10	0.359	0.296
		R@50	0.800	0.728
		MRR	-	-
Our system	Only seq sim	R@1	0.290	0.080
		R@10	0.575	0.364
		R@50	0.910	0.800
		MRR	0.389	0.176
	Word + seq sim	R@1	**0.399**	**0.116**
		R@10	**0.693**	**0.444**
		R@50	**0.944**	**0.848**
		MRR	**0.501**	**0.219**

Table 3: Ablation results on *valid* of Subtask 1.

Code and data

Available at `https://github.com/basma-b/multi_level_chatbot`

9. System Extension

Subtask 4 requires the model to recognize cases where no candidate response is correct.

- We added the following classifier on top of our system.
- The candidate scores are fed into a SVM classifier.
- It predicts the presence of a correct response.

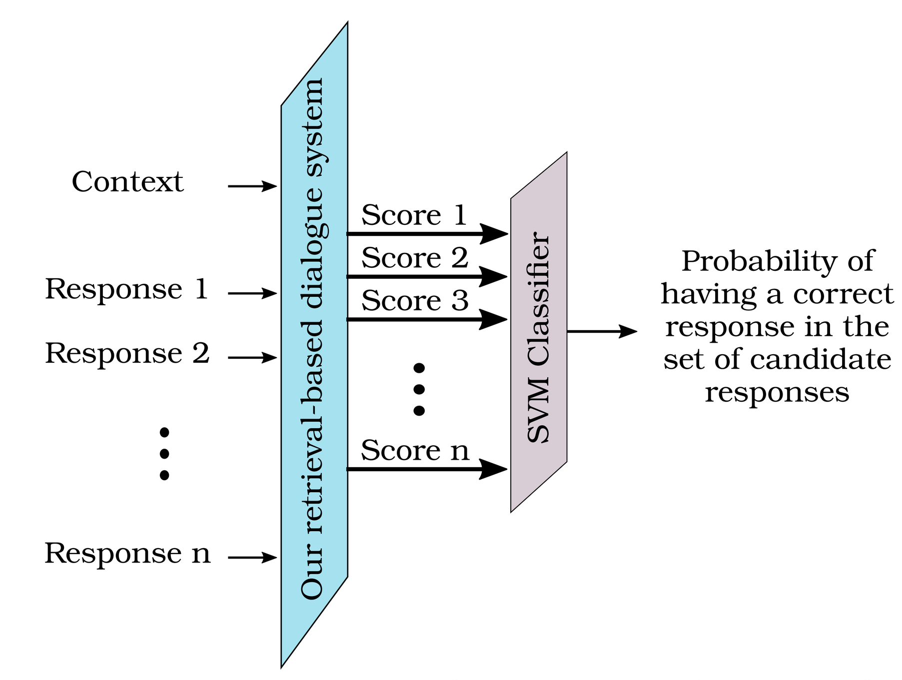

Figure 2: Extension of our proposed system for subtask 4.

10. Conclusion

- We proposed an end-to-end retrieval-based dialog system that matches the context with the correct response on **two levels**.
- Performance improvement compared to the baseline system.
- One simple system for the three subtasks.

11. References

[1] Ryan L., Nissan P., Iulian S., and Joelle P.
The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems.
In SIGDIAL 2015.

[2] Yu W., Wei W., Chen X., Ming Z., and Zhoujun L.
Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots.
In ACL 2017.

Contact Information

- Web: https://basma-b.github.io/
- Email: basma.boussaha@univ-nantes.fr