Extracting Persona Triples in Utterances

Yoonna Jang[1], Kinam Park[2*]
[1]Korea University, [2]Human-inspired AI Research
{morelychee, spknn}@korea.ac.kr

Abstract

In this paper, we study the extraction of persona triples from utterances in conversations. Utilizing utterance sentences and their corresponding triple pairs, a model is trained using a multi-task learning approach to extract persona triples from utterance sentences or persona sentences. The model utilizes a pre-trained language model BART with an encoder-decoder structure and learns three tasks: head, relation, and tail extraction. Head and relation extraction is classified, and tail extraction is generated to finally extract persona triples with the structure of head, relation, and tail. We found that the accuracy of each task is higher than 90%.

Keywords: Persona Dialogue, Persona Extraction, Open-domain Dialogue

1. Introduction

As research on persona dialogues [1, 2] has evolved, machines have been trained to generate consistent utterances that do not contradict what they have previously said, just like humans. For consistency, dialogue natural language inference (DNLI) [3] assigns triples to utterances and persona sentences and labels them for entailment, contradiction, and neutral relation between utterance-utterance and persona-utterance sentences. A number of studies have utilized this dataset to generate utterances for persona-based conversations that are consistent with the persona's sentences or previous utterances. [4, 5, 6]

In this study, we utilize utterance sentences and their corresponding triples to extract persona triples from utterance sentences. We utilize a pre-trained language model BART [7] with an encoder-decoder structure to train the task of extracting triples from sentences. Through experiments, we show that the model achieves high scores of over 90% for head and relation prediction accuracy and tail token accuracy.

2. Method

We utilize the DNLI [3] dataset to extract personas from utterances by utilizing sentence-triple pairs. The training consists of three subtasks: (1) head, (2) relation, and (3) tail extraction. First, in the head extraction, the model learns a classification task that judges whether a given sentence is the target of a persona triple extraction. Then, in the relation extraction, the model learns to select one of 61 predefined relation labels [1]. Finally, the tail extraction directly generates the target tail. The entire training is performed using multi-task learning, which means that it learns about all three tasks simultaneously, and each task is given equal weight.

3. Experiments

The dataset we used for our experiments is dialogue natural language inference (DNLI) [3]. The total data con-

*Corresponding author.

[1]Pre-defined relation labels are as follows: 'none', 'have_sibling', 'own', 'like_watching', 'favorite_hobby', 'teach', 'have_family', 'live_in_general', 'place_origin', 'like_food', 'favorite_place', 'has_degree', 'employed_by_general', 'employed_by_company', 'dislike', 'favorite_music', 'favorite_activity', 'gender', 'like_goto', 'have_children', 'member_of', 'like_movie', 'job_status', 'favorite_sport', 'like_activity', 'like_animal', 'marital_status', 'favorite_color', 'attend_school', 'favorite_animal', 'school_status', 'physical_attribute', 'favorite_food', 'nationality', 'like_general', 'like_music', 'favorite_book', 'favorite_music_artist', 'have_vehicle', 'want', 'has_hobby', 'like_drink', 'misc_attribute', 'have', 'other', 'has_profession', 'favorite_season', 'like_read', 'want_job', 'favorite_movie', 'previous_profession', 'favorite_show', 'live_in_citystatecountry', 'have_pet', 'not_have', 'has_ability', 'like_sports', 'want_do', 'favorite_drink', 'has_age', 'favorite'

Table 1. Experimental results. Head and relation indicate accuracy, and tail indicates token accuracy. Results are rounded to three decimal places.

Model	Head (%)	Relation (%)	Tail (%)
BART	94.538	91.213	99.429

sists of 310,110, 16,500, and 16,500 sentence pairs and labels in the train, validation, and test sets, respectively. Pytorch Lightning [8] and HuggingFace's Transformers [9] are used for the experiments. The model used for training is BART [7] base model, a pre-trained language model with an encoder-decoder structure. Training is performed using the AdamW [10] optimizer with a learning rate of 5e-05, and the model is trained for 2 epochs.

As shown in Table 1, BART shows the head, relation and tail accuracy over 90%. Especially, head and relation classification is 94.538 and 91.213, respectively. Tail token accuracy is about 99%, which indicates the model generates the most target tokens in tails.

4. Conclusion

In this paper, we studied a method to extract persona triples from sentences by utilizing triple pairs with utterance sentences or persona sentences. We trained a model with an encoder-decoder structure and added layers for extracting head, relation, and tail, and demonstrated that it performed over 90% accuracy.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425). Also, this research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2022R1A2C1007616)

Reference

[1] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?" *arXiv preprint arXiv:1801.07243*, 2018.

[2] J. Urbanek, A. Fan, S. Karamcheti, S. Jain, S. Humeau, E. Dinan, T. Rocktäschel, D. Kiela, A. Szlam, and J. Weston, "Learning to speak and act in a fantasy text adventure game," *arXiv preprint arXiv:1903.03094*, 2019.

[3] S. Welleck, J. Weston, A. Szlam, and K. Cho, "Dialogue natural language inference," *arXiv preprint arXiv:1811.00671*, 2018.

[4] Y. Cao, W. Bi, M. Fang, S. Shi, and D. Tao, "A model-agnostic data manipulation method for persona-based dialogue generation," *arXiv preprint arXiv:2204.09867*, 2022.

[5] E. Mitchell, J. J. Noh, S. Li, W. S. Armstrong, A. Agarwal, P. Liu, C. Finn, and C. D. Manning, "Enhancing self-consistency and performance of pre-trained language models through natural language inference," *arXiv preprint arXiv:2211.11875*, 2022.

[6] Y. Nie, M. Williamson, M. Bansal, D. Kiela, and J. Weston, "I like fish, especially dolphins: Addressing contradictions in dialogue modeling," *arXiv preprint arXiv:2012.13391*, 2020.

[7] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[8] W. Falcon and The PyTorch Lightning team, "PyTorch Lightning," Mar. 2019. [Online]. Available: https://github.com/Lightning-AI/lightning

[9] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, "Transformers: State-of-the-art natural language processing," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Oct. 2020. [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-demos.6

[10] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.