Title	Prototypical Contrastive Transfer Learning for multimodal language understanding
Sub Title	
Author	小槻, 誠太郎(Otsuki, Seitaro)
Publisher	慶應義塾大学AI・高度プログラミングコンソーシアム
Publication year	2023
Jtitle	AICカンファレンス予稿集 (2023.) ,p.25- 25
JaLC DOI	
Abstract	We focus on the task of identifying target objects in domestic environments according to free-form natural language instructions. In this work, we propose a novel transfer learning approach for multimodal language understanding, Prototypical Contrastive Transfer Learning (PCTL) which uses a new contrastive loss, Dual ProtoNCE. We introduce PCTL to the target task. To validate PCTL, we built new real-world and simulation datasets. Our experiment demonstrated that PCTL outperformed existing methods. Specifically, PCTL achieved an accuracy of 78.1%, while simple fine-tuning achieved an accuracy of 73.4%.
Notes	会議名 : AICカンファレンス2023 開催地 : 慶應義塾大学日吉キャンパス 日時 : 2023年3月4日 第2章ポスター発表要旨 ポスター要旨-2
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0025

Prototypical Contrastive Transfer Learning for Multimodal Language Understanding

Seitaro Otsuki[1]

[1]Department of Information and Computer Science, Keio University

Abstract: We focus on the task of identifying target objects in domestic environments according to free-form natural language instructions. In this work, we propose a novel transfer learning approach for multimodal language understanding, Prototypical Contrastive Transfer Learning (PCTL) which uses a new contrastive loss, Dual ProtoNCE. We introduce PCTL to the target task. To validate PCTL, we built new real-world and simulation datasets. Our experiment demonstrated that PCTL outperformed existing methods. Specifically, PCTL achieved an accuracy of 78.1%, while simple fine-tuning achieved an accuracy of 73.4%.

Keywords: Multimodal Language Understanding, Prototypical Contrastive Transfer Learning, Dual ProtoNCE

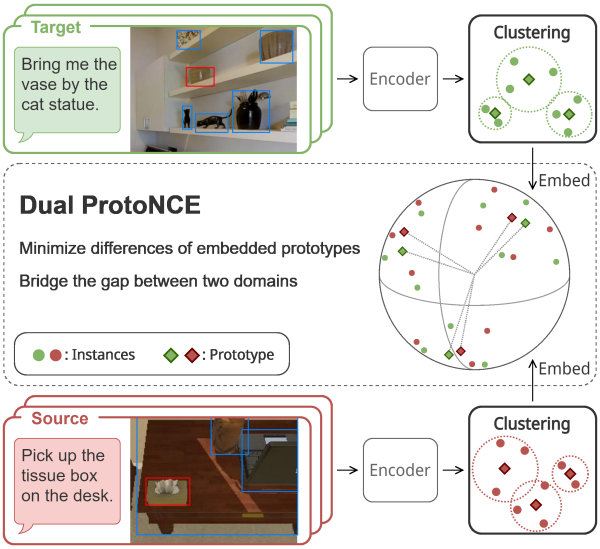

Fig. 1 A summary of our method, PCTL.

1. Introduction

Domestic Service Robots (DSRs) are gaining popularity as a solution for increasing demand for daily care and support, due to their ability to physically assist individuals. However, the DSRs currently lack the capability for smooth interaction with people through natural language. For training their language comprehension models, it is desirable to use data collected in real-world environments. However, collecting and annotating such real-world data can be labor-intensive. In contrast, collecting training data with a simulator is much more cost-effective than collecting data in real-world environments. Hence, it is advantageous to leverage simulation data through a transfer learning framework.

2. Proposed Method

In this paper, we propose Prototypical Contrastive Transfer Learning (PCTL) - a novel transfer learning approach for the multimodal language understanding task. PCTL performs contrastive learning between the source and target domain data using our new contrastive loss, Dual ProtoNCE. PCTL is expected to alleviate the influence of the gap between the two domains by minimizing Dual ProtoNCE. A summary of our method is depicted in Fig. 1.

Our key contributions are as follows:

- We introduce transfer learning to the task of identifying target objects in domestic environments according to free-form natural language instructions.
- We propose PCTL, which is a novel transfer learning approach for the multimodal language understanding task.
- Within PCTL, we develop Dual ProtoNCE - a novel contrastive loss generalized for transfer learning.

3. Experiments

We conducted comparison experiments to compare the proposed and baseline methods. Table 1 shows the accuracy on the REVERIE-fetch dataset. The right column shows the means and standard deviations over five trials.

4. Conclusions

In this study, we proposed PCTL, which is a novel transfer learning approach for the multimodal language understanding task. Specifically, we developed Dual ProtoNCE - a new contrastive loss generalized for transfer learning.

References

[1] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, pages 3723–3732, 2018.

Table 1
Quantitative results on the REVERIE-fetch dataset.

Method	Accuracy [%]
Target domain only	73.0 ± 1.87
Fine-tuning	73.4 ± 11.8
MCDDA+ [1]	74.9 ± 3.94
Ours	$\mathbf{78.1 \pm 2.49}$