Grounded Language Processing in Different Domains

Abstract

Grounded language processing is itself a huge area of research. Considering the past 25 years, an immense amount of research has been done. Papers from various sources like Google scholar, IEEE Xplore, and ACL Anthology were chosen for reviewing. Also the different domains the grounded language processing to used to solve various problems from human-robot interaction to automatic speech recognition to data-to-text generation and so on, a wide variety of applications are considered. And different aspects of the field are discussed in this paper.

1 Introduction

Looking at all the sources available at the point like IEEE Xplore and Google scholar, the first thing was to know how much research has been done till now. As per the IEEE Xplore alone, 464 items related to grounded language processing were found. This included the conference paper, journals, and books. Then I filtered out the recent 25 years of published work which turned out to be 459 papers and books in total. This shows that there has been an immense amount of research work done in the past 25 years in the field of natural language processing. Coming to the grounded language processing in robotics, the number of papers came down to 62 in total as per the IEEE Xplore alone. Considering the wide range of application in which the Grounded language processing is used, papers from from different were selected for the review. Matuszek[1] in her paper discussed regarding the use of grounded language processing in the robotics. The interaction between human and robot in the real time considering the other factors affecting. Bisk et al[2] in their paper discussed the the datatset they worked on for the interaction between human and machine such that the interaction can become smoother and machine are able to understand the human. They built a database which consists of the pictures and frames. Chen et al[3] proposed knowledge grounded pre-training model. Data-to -text generation where the model can adapt the context easily. Pillai et al[4] in their paper discussed the human robot interaction and the learning part was done using grounded language processing. In the human environment the robot was trained using the description of the object provided by the user. Yang et al[5] in their paper discussed the embedding imputation. when the unseen or new words are encountered. Grounded information is used in the form of knowledge graph.

2 Previous works

As the robots are becoming part of the human society in various fields, it is important to understand each other commands and coordinate with each other to work efficiently. Thomason et al[13] in their paper discussed the interaction between the robot and human. At first they considered the ”Amazon Mechanical Turk” for the task like ”navigation, delivery, and relocation”. They developed the agent on the virtual platform first and then they implemented in the actual robot to study how agent works. The user provide the command to the robot which ”contains compositional language”. One of the important thing which they considered is ”end-to-end” pipeline. The robot was supposed to move from one room to another while ”picking and placing task”. Authors discussed various steps in the ”Clarification Dialog”, first step the providing ”commands to the agent” where the commands given by the user ”processed by semantic parsing”. Then the mapping is done which is considered as the next step ”Belief state”. The agent was capable to take input from the user and if the command was not understood then it asked for the command

again for clarification. Fleischman et al[6] in their paper discussed grounded language in "Automatic Speech Recognition of Sports Video". They mentioned that their model is better then that of the human commentary in the game. Here the considered three types of features as "visual context feature, camera motion features, and audio context features". In the video feature they converted the video in the "segments" and then each segment was examined individually. In the camera feature, they considered the the "motion, zoom-in, zoom-out" features which can cause the problem while evaluating anything the game. In the audio context feature, they considered the "background noise, cheers, music". This raw date was then "encoded" using a codebook and then matched with those present in the codebook. If the frame do not occur in the video, then the value is zero and for an particular instant the value is equal to the number of shots present in the video. Thus they described implemented the ways to increase the quality of speech recognition in the video. Pillai et al[4] in their paper discussed the grounded language processing in the robotics. Authors focused on the active learning while the robot encounter a new object. Here the robt get input from the user directly.They udsed 20 object for the descroption and asked 10 users to describe them. The experiment was conducted in two phase where in first phase the the user decribed the object completely and in the second phase they used "interactive labeling" where user and robot interacted with each othe and the robot labels an object considering the user's confidence.

References

[1] Matuszek, Cynthia.. Grounded Language Learning: Where Robotics and NLP Meet (invited talk). Proceedings of the International Joint Conference on Artificial Intelligence.

[2] Yonatan Bisk, Daniel Marcu and William Wong, "Towards a Dataset for Human Computer Communication via Grounded language Acquisition", 30th AAAI Conference on Artificial Intelligence(2016).

[3] Wenhu Chen,Yu Su, Xifeng Yan, William Yang Wang, "KGPT: Knowledge-Grounded Pre- Training for Data-to-Text Generation", 2020.

[4] Cynthia Matuszek, Nisha Pillai, Karan K Budhraja, 2016, "Improving grounded Language Acquisition Efficiency using Interactive Labeling", Robotics:Science and Systems(R:SS).

[5] Z Yang, C Zhu, V Sachidananda, E Darve, "Embedding imputation with grounded language information", 57th Annual Meeting of ACL(2019).

[6] Michael Fleischman, Deb Roy, "Grounded Language Modeling for Automatic Speech Recognition of Sports Video",2008.

[7] David L. Chen and Raymond J. Mooney, "Planning for Gold: Finding Relevant Semantic Content for Grounded Language Learning", MLSLP(2011).

[8] Bernd Huber, Daniel McDuff, Chris Brockett, Michel Galley, and Bill Dolan, "Emotional Dialogue Generation using Image-Grounded Language Models", CHI(2018).

[9] Andrea Vanzo, Danilo Croce, Emanuele B, Robert B, Daniele N, "Grounded Language interpretation of Robotic Command through Structured Learning", Artificial Intelligence(2020).

[10] Steve R Howell, Damian J, Suzanna B, "A model of grounded Language Acquisition: Sensorimotor Features improve lexical and grammatical Learning", Journal of Memory and Language(2005).

[11] Haonan Yu, Haichao Z, Wei Xu, "Interactive Grounded Language Acquisition and Generalization in a 2D World", ICLR(2018).

[12] Giovanni P, Lawrence W B, Angelo C, Martin H F, Ken M, Michael J S, "Computational Grounded Cognition: a new alliance between grounded cognition and computational modeling", Frontiers in Psychology(2013).

[13] J. Thomason et al., "Improving Grounded Natural Language Understanding through Human-Robot Dialog," 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 6934-6941, doi: 10.1109/ICRA.2019.8794287.