NLP+Robotics

***Abstract*—**
***Index Terms*—NLP, Robotics**

I. Problem Formulation

II. Related Work

Chet et al. [1] develop a Language-Model based Commonsense Reasoning (LCMR) method which enables robots to listen to incompletely specified instructions and uses environmental context and commonsense approach to fill in missing information. They first convert the raw instruction to 'verb frames'. A verb frame is set containing predicate (verb) and semantic roles. If the human says "pour me some water". The verb frame (pour, Theme: water, Destination: ?) is generated using Semantic Role Labeling (SRL) by He et al. [2]. The original instruction does not specify where to pour into, LCMR infers out destination on its own, i.e. a cup through environment observations. They use YouCook2 [3] and Now You're Cooking datasets as training corpora. They used motion planning toolkit Moveit! [4] to compute a motion plan for the robot to accomplish each task (based on ROS). System Model: (1) Speech Recognition using Google Cloud API, (2) Detect objects in environment using Mask R-CNN, (3) Predicate Argument Parsing using RNN to (a) conference resolution (b) SRL (c) Rule based mapping system using PropBank to convert to verb frame and do pruning. (3) Complete verb frame by plugging in ? in verb frames by objects in vicinity based on probability ranking (trained using croudsourcing Amazon Turks and used RNN model with GRUs). (4) Motion planning based on complete verb frames. Strength: they compared with other works like Co-occur, Word2Vec, ConceptNet and their results are much better. Weakness: Only restricted to knowledge from environment, can extend this to web.

Tenorth et al. [5] describe and discuss use of World Wide Web based information for autonomous service robots. They show web sources like: (1) ehow.com and wikihow.com which give step wise instructions for millions of everyday activities like making pancakes. (2) Lexical databases like wordnet.princeton.edu group verbs, adverbs and nouns semantically into sets of synonyms (synsets), which are linked to concepts in encyclopedic knowledge bases like opencyc.org (ontological relationships). Knowledge is represented as first order logic to parse say instuctions on how to make pancakes into usable forms. (3) Common sense knowledge bases like openmind.hri-us.com to tell like what are pancakes, tools, kitchen, stove, etc. (4) Object appearance like germandeli.com, images.google.com to show how stove looks like, or other things like pancake or pancake mix look like. (5) Object shape in 3D CAD models to identify objects in vicinity sketchup.google.com/3dwarehouse/ (6) Object properties like those extracted from shopping websites germandeli.com.

III. System Architecture

Identify applicable funding agency here. If none, delete this.

References

[1] H. Chen, H. Tan, A. Kuntz, M. Bansal, and R. Alterovitz, "Enabling robots to understand incomplete natural language instructions using commonsense reasoning," *arXiv preprint arXiv:1904.12907*, 2019.

[2] L. He, K. Lee, M. Lewis, and L. Zettlemoyer, "Deep semantic role labeling: What works and whats next," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 473–483.

[3] L. Zhou, C. Xu, and J. J. Corso, "Towards automatic learning of procedures from web instructional videos," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[4] S. Chitta, I. Sucan, and S. Cousins, "Moveit![ros topics]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.

[5] M. Tenorth, U. Klank, D. Pangercic, and M. Beetz, "Web-enabled robots-robots that use the web as an information resource," *IEEE Robotics and Automation Magazine*, vol. 18, no. 2, p. 58, 2011.