Learning Action Preconditions from Step-by-step Intructions in Planning Domains

Maurício Steinert
Graduate Program in Computer Science - School of Technology
Pontifical Catholic University of Rio Grande do Sul - PUCRS
Porto Alegre, Brazil
mauricio.steinert@edu.pucrs.br

Abstract— **Automated Planning devises plans to achieve a specific goal. A planning domain must be provided in a formal language like Planning Domain Definition Language (PDDL) to be processed into a computer. Using other types of domain description is desirable. In this work, we propose using Natural Language sentences to describe a planning domain, given a sequence of step-by-step instructions to accomplish a task. Specifically, we want to investigate the possibility of extracting verbs within sentences based on word embeddings similarity with predefined keywords, and also if a neural network is capable of learning relationship between ordered step-by-step instructions in natural language format.**

I. Introduction

Automated Planning is a field of research in Artificial Intelligence that aim to devise a set of actions to achieve a specific goal. In Classical Planning, a domain problem is formally described as a 5-tuple $\Pi =< P, O, I, G, A >$, where P defines a set of first-order predicates, O is a set of symbols called objects, I is the initial state, G is a set of goal states to achive, and A is a set of actions that defines transition between states. A planning problem must also be formally described to be processed into a computer, what is usually done using description languages like Planning Domain Definition Language (PDDL).

Language modeling inside a computer is a challenging task, where approaches vary from using indexed dictionaries (Bag-Of-Words) to vector representation (embeddings) capable of capturing semantic relationship between words based on vicinity [6]. Many tasks like question-answering and text summarization attained significative results using word and sentence embeddings [3].

Using natural language to define a problem domain is an active area of research. Framer [5] and StoryFramer [4] are methods that accept text as input and translate them into PDDL. Actions and objects are acquired based on Named Entity Recognition (NER) systems that perform syntactic analysis for each word within sentences and assigns a label to them for further processing: verbs are included in actions set; nouns are included in objects set. Users must manually eliminate redundancies and assign preconditions for actions. By default, these systems consider that each action is performed in some location by someone, requiring that users provide this information to be used as preconditions.

Instead of natural language input, other data formats are used to describe problem domains. Asai et al [1] developed LatPlan, which accepts images of states transition as input and, using State Auto-Encoders, generate a latent space representation of states, transitions and actions that can be computed to devise a plan. LatPlan returns a set of images that represents the order of actions that must be performed to achieve a goal.

In this work, we propose using natural language to describe planning domain problems. We are going to use datasets composed by ordered step-by-step instructions to perform a task. Our main objectives are *(1)* identify actions in sentences using word embeddings by computing similarity between predefined keywords and *(2)* given a set of actions A provided as input in an ordered manner, evaluate if a Neural Network can learn relationship between sentences to use as preconditions.

II. Technical Approach

In this section we detail our intended approach.

A. Word Embeddings

Instead of using Named Entity Recognition systems, we propose using word embedding dictionaries [6] with clustering algorithms to classify words based on similarity to predefined keywords for actions set, consequently populating this set. For example, common verbs used in planning domains include *take*, *give*, *put*, *drop*, *walk*, to name a few. Word embedding dictionaries are trained in an unsupervised manner over huge text datasets like Wikipedia, overcoming the need for manual annotation of words.

B. Identify action preconditions based on ordered sentences

We want to investigate the possibility of identifying preconditions given step-by-step ordered instructions. Be A is a set of ordered actions, where $a_n \in A$, we propose that $pre(a_n) = \{a_1, a_2, ..., a_{n-1}\}$ is the set of preconditions of a_n, i.e. all predecessor actions of an action are their preconditions.

We will implement a Neural Network using A and $pre(A)$ as input that evaluates if a sequence of instructions sounds coherent. We suppose that exists a relationship between ordered sentences that a Neural Network is capable of learning.

C. Dataset

We will use WikiHow recipes and Text2HBM [2] datasets for training and test our model. Both datasets are composed by ordered instructions and use imperative voice in sentences.

III. Project Management

Week 1:

- Train a word embedding dictionary over Wikipedia dumps.
- Develop a program that receives as input a sequence of sentences, convert them to embedding and identify actions based on similarity to predefined similar words.
- Preprocess data from WikiHow to generate the dataset.

Week 2-4:

- Develop and train a Neural Network to evaluate coherence between ordered step-by-step instructions, in terms of preconditions for each action, as explained earlier.

Week 5:

- Evaluate obtained results and write final report.

IV. Conclusion

In this work, we aim to investigate the possibility of using word embeddings to identify verbs within sentences that describe actions. We also want to investigate if exist any kind of relationship between sentences that describe step-by-step instruction to accomplish a task that can be learned by a neural network. We believe that our proposed methods can open new possibilities to further research in translating natural language into planning domains. We expect to achieve satisfying results and get a better understanding of the problem at hand.

References

[1] Masataro Asai and Alex Fukunaga. Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary, 2017.

[2] Debajyoti Paul Chowdhury, Arghya Biswas, Tomasz Sosnowski, and Kristina Yordanova. Towards evaluating plan generation approaches with instructional texts, 2020.

[3] Mahak Gambhir and Vishal Gupta. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1):1–66, 2017.

[4] Thomas Hayton, Julie Porteous, Joao Ferreira, Alan Lindsay, and Jonathon Read. Storyframer: From input stories to output planning models. pages –, June 2017. Workshop on Knowledge Engineering for Planning and Scheduling (KEPS). The 27th International Conference on Automated Planning and Scheduling (ICAPS), KEPS 2017 ; Conference date: 18-06-2017 Through 23-06-2017.

[5] Alan Lindsay, Jonathon Read, Joao Ferreira, Thomas Hayton, Julie Porteous, and Peter Gregory. Framer: Planning models from natural language action descriptions. pages –, June 2017. 27th International Conference on Automated Planning and Scheduling, ICAPS 2017 ; Conference date: 18-06-2017 Through 23-06-2017.

[6] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.