

Neural module networks

Shuang Sha

June 4, 2018

Abstract

This paper described a procedure for constructing and learning neural module networks, which compose collections of jointly-trained neural modules into deep networks for question answering. This method decomposes questions into linguistic substructures, and uses these structures to dynamically instantiate modular networks (with reusable components for recognizing things, classifying colors, etc.).

1. Introduction

This paper describes an approach to visual question answering based on a new model architecture that is called a neural module network (NMN). For example, as shown in Figure 1, given an image and an associated question, and this question is where the dog is. A corresponding answer predicted is that the dog is on the couch. Figure 1 is a schematic representation of this paper proposed model. The visual question answering task has significant applications to human-robot interaction, search, and accessibility, and has been the subject of a great deal of recent research attention [2, 3, 4]. The main contribution of this paper is that this paper described neural module networks, a general architecture for discretely composing heterogeneous, jointly-trained neural modules into deep networks.

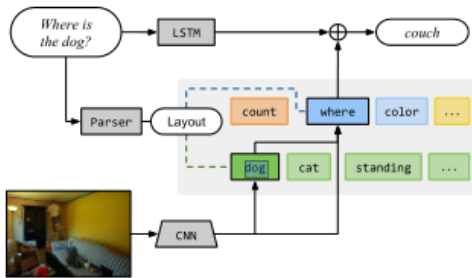


Figure 1. a schematic representation of this paper proposed model.

2. Formatting your paper

Thus the goal in this paper is to specify a framework for modular, compostable, jointly-trained neural networks. In this framework, researches first predict the structure of the computation needed to answer each question individually, then realize this structure by constructing an appropriately shaped neural network from an inventory of reusable modules [1]. These modules are learned jointly, rather than trained in isolation, and specialization to individual tasks (identifying properties, spatial relations, etc.) arises naturally from the training objective. In this paper, researchers established a module to solve each question by predicting. Then these modules are connected to each other to form whole. These modules can realize to train jointly. And this method can modularized the problem and improve the efficiency of the work.

References

- [1] J. Andreas, M. Rohrbach, T. Darrell, and D. Klein. Neural module networks. In *CVPR*, pages 39–48, 2016. 1
- [2] L. Ma, Z. Lu, and H. Li. Learning to answer questions from image using convolutional neural network. In *AAAI*, page 16, 2016. 1
- [3] M. Malinowski, M. Rohrbach, and M. Fritz. Ask your neurons: A neural-based approach to answering questions about images. In *ICCV*, pages 1–9, 2015. 1
- [4] L. Yu, E. Park, A. C. Berg, and T. L. Berg. Visual madlibs: Fill in the blank description generation and question answering. In *ICCV*, pages 2461–2469, 2016. 1