Report for Relation Networks

Yixiao FEI

I Introduction

The paper I have chosen is named "A simple neural network module for relational reasoning", published in 2017 on *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. This is a study done in a British artificial intelligence company in London called **DeepMind Technologies**, founded in September 2010, currently owned by **Alphabet Inc**. [i] The full reference of the paper could be found at the end of this report. [ii]

Relational reasoning is a central component of generally intelligent behavior, but has proven difficult for neural networks to learn. The paper describes how to use Relation Networks (RNs) as a simple plug-and-play module to solve problems that fundamentally hinge on relational reasoning. The research has tested the RN-augmented networks on visual question answering, text-based question answering and complex reasoning about dynamic physical systems.

II Summary

II.1 Existing Approaches

The ability to reason about the relations between entities and their properties is central to generally intelligent behavior. Consider a reader piecing together evidence to predict the culprit in a murder-mystery novel: each clue must be considered in its broader context to build a plausible narrative and solve the mystery.

Symbolic approaches to artificial intelligence are inherently relational, where the relations between symbols are defined by using the language of logic and mathematics, and then relations are reasoned by using a multitude of powerful methods, including deduction, arithmetic, and algebra. But symbolic approaches suffer from the symbol grounding problem and are not robust to small task and input variations.

II.2 Relation Networks

An RN is a neural network module with a structure primed for relational reasoning. In its simplest form the RN is a composite function:

$$\mathrm{RN}(O) = f_{\emptyset}\left(\sum_{i,j} g_{\theta}(o_i, o_j)\right)$$

where the input is a set of "objects" $O = \{o_1, o_2, \ldots, o_n\}, o_i \in \mathbb{R}^m$ is the i^{th} object, and $f_{\emptyset}$ and g_{θ} are functions with parameters $\emptyset$ and θ, respectively. For our purposes, $f_{\emptyset}$ and g_{θ} are MLPs (multi-layer perceptrons), and the parameters are learnable synaptic weights, making RNs end-to-end differentiable. The study claims that RNs have three notable strengths: they learn to infer relations, they are data efficient, and they operate on a set of objects – a particularly general and versatile input format – in a manner that is order invariant.

II.3 Model

In their simplest form RNs operate on objects, and hence do not explicitly operate on images or natural language. A central contribution of this work is to demonstrate the flexibility with which relatively unstructured inputs, such as

CNN or LSTM embeddings, can be considered as a set of objects for an RN. As we describe below, we require minimal oversight in factorizing the RN's input into a set of objects.

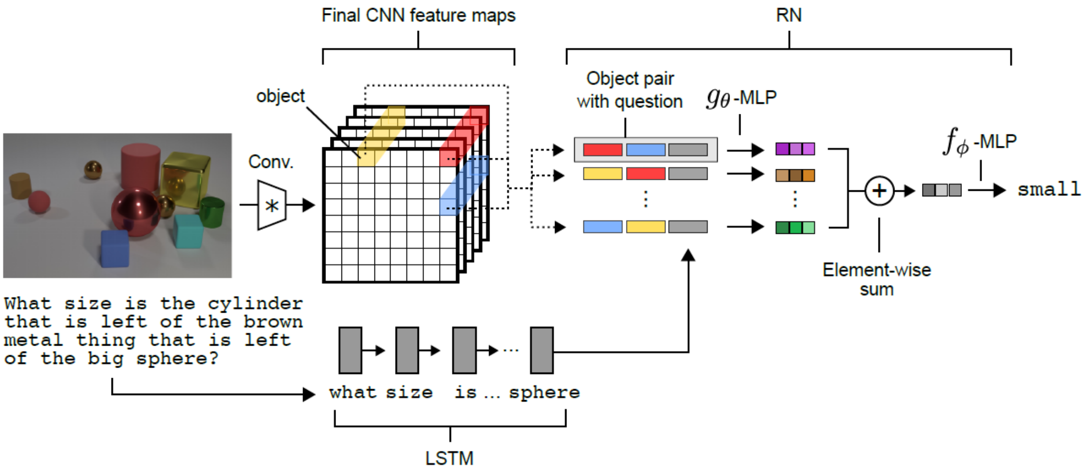

Figure 1: **Visual QA architecture**. Questions are processed with an LSTM to produce a question embedding, and images are processed with a CNN to produce a set of objects for the RN. Objects (three examples illustrated here in yellow, red, and blue) are constructed using feature-map vectors from the convolved image. The RN considers relations across all pairs of objects, conditioned on the question embedding, and integrates all these relations to answer the question.

II.4 Performance

The model achieved state-of-the-art performance on CLEVR at 95:5%, exceeding the best model trained only on the pixel images and questions at the time of the dataset's publication by 27%, and surpassing human performance in the task

Model	**Overall**	Count	Exist	Compare Numbers	Query Attribute	Compare Attribute
Human	92.6	86.7	96.6	86.5	95.0	96.0
Q-type baseline	41.8	34.6	50.2	51.0	36.0	51.3
LSTM	46.8	41.7	61.1	69.8	36.8	51.8
CNN+LSTM	52.3	43.7	65.2	67.1	49.3	53.0
CNN+LSTM+SA	68.5	52.2	71.1	73.5	85.3	52.3
CNN+LSTM+SA*	76.6	64.4	82.7	77.4	82.6	75.4
CNN+LSTM+RN	**95.5**	**90.1**	**97.8**	**93.6**	**97.9**	**97.1**

Table 1: **Results on CLEVR from pixels.** Performances of the model (RN) and previously reported models, measured as accuracy on the test set and broken down by question category.

III Comment

The research proposes RNs (Relation Networks), powerful, versatile, and simple neural network modules with the capacity for relational reasoning. From the table presented above, we can find that the performance of RN-augmented networks on CLEVR is especially notable, which surpasses that of humans. The study offers a neural module for simple use, but if we need to fully understand the mechanism behind the model, hard works need to be done to reconstruct the model by own and verify the performance.

IV Reference

[i] DeepMind - Wikipedia https://en.wikipedia.org/wiki/DeepMind

[ii] Santoro, A., Raposo, D., Barrett, D. G., Malinowski, M., Pascanu, R., Battaglia, P. & Lillicrap, T. (2017). A simple neural network module for relational reasoning. Advances in Neural Information Processing Systems 30 (NIPS 2017), 4967-4976. https://arxiv.org/abs/1706.01427