Graph Representation And Process With Neural Networks

Cheng Guan

May 25, 2018

1 Graph representation of scenes and questions

The input data for each training or test instance is a question, and a parameterized description of contents of the scene. The question is processed with the Stanford dependency parser [1], which outputs the following.

- A set of N^Q words that constitute the nodes of the question graph.Each word is represented by its index in the input vocabulary, a token $x_i^Q \in \mathbb{Z}\left(i...N^Q\right)$.
- A set of pairwise relations between words, which constitute the edges of our graph.An edge between words i and j is represented by $e_{ij}^Q \in \mathbb{Z}$,an index among the possible types of dependencies.

The dataset provides the following information about the image .

- A set of pairwise relations between all objects.They form the edges of a fully-connected graph of the scene.The edge between objects i and j is represented by a vector $e_{ij}^S \in \mathbb{R}^D$ that encodes relative spatial relationships .
- A set of N^S objects that constitute the nodes of the scene graph. Each node is represented by a vector $x_i^S \in \mathbb{R}^C$ of visual features $\left(i \in 1...N^S\right)$. Please refer to the supplementary material for implementation details.

Our experiments are carried out on datasets of clip art scenes, in which descriptions of the scenes are provided in the form of lists of objects with their visual features. The method is equally applicable to real images, with the object list replaced by candidate object detections. Our experiments on clip art allows the effect of the proposed method to be isolated from the performance of the object detector. Please refer to the supplementary material for implementation details

$$x_i^{'Q} = W_1[x_i^Q]e_i^{'Q} = W_2[e_{ij}^Q] \tag{1}$$

$$x_i^{'S} = W_3[x_i^Q] + b_3 e_i^{'S} = W_4[e_{ij}^S] + b_4 \tag{2}$$

with W_1 the word embedding (usually pretrained, see supplementary material), W_2 the embedding of dependencies, $W_3 \in \mathbb{R}^{h\times c}$ and$W_4 \in \mathbb{R}^{h\times d}$ weight matrices, and $b_3 \in \mathbb{R}^c$ and $b_r \in \mathbb{R}^d$.

References

[1] M.-C de Marneffe. and C.D. Manning. The stanford typed dependencies representation. *In COLING Workshop on Cross-framework and Cross-domain Parser Evaluation*, 2008.