Review 5 - 1

The More You Know: Using Knowledge Graphs for Image Classification

Summary

This paper presents a method to use structured prior knowledge to classify and reason on image dataset. This paper uses the basic idea from Graph Gated Neural Network which is used to learn structured data. However the biggest problem of applying GGNN onto images is computational scalability. To tackle this problem, this paper introduces a new method called Graph Search Neural Network. The idea is rather than performing recurrent update over all of the nodes of the graph at once, only choosing to expand "useful" nodes. The model first initializes the nodes based on the likelihood of the concepts. To determine which notes to be expanded, this paper proposes an importance network to learn pre-node score. The model only expands top scoring nodes that have never been expanded. Besides, to adapt graph networks for vision problem, this paper comes up an idea to map the learned graph into a high-dimensional space and concatenate the feature vector with the image feature vector. To test the performance of the model, this paper conducts two experiments on multi-label classification and low-shot recognition respectively.

Pros:

1. The introduction of the GSNN as a way of incorporating potentially large knowledge graphs into an end-to-end learning system that is computationally feasible for large graphs.
2. Present a framework for using noisy knowledge graphs for image classification.
3. Introduce a new subset of Visual Genome designed to test 1-shot and few-shot learning without any overlap with classes from previous image dataset.
4. Outperform the performance of multi-label classification task both in full-data and low-data setting

Cons:

1. If the model predict wrong on the concept, it could cause a series of failures.

Review 5 - 2

A Diagram Is Worth A Dozen Images

Summary

This paper proposes a Diagram Parse Graphs (DPG) to tackle the diagram interpretation and reasoning problem. These two tasks are defined as Syntactic Parsing and Semantic Interpretation. Syntactic parsing involves detecting and recognizing constituents and their syntactic relationships in a diagram. Semantic interpretation is the task of mapping constituents and their relationships to semantic entities and events (real-world concepts). In DPG, nodes correspond to constituents and edges correspond to relationships between constituents. This paper models four types of constituents: Blobs, Text Boxes, Arrows and Arrow Head. To learn the relationships between different constituents, the authors of this paper introduces a Deep Sequential Diagram Parser which uses a LSTM RNN with full-connected layers used prior to. As for the semantic interpretation, they designs a neural network call DQA-NET to answer diagrammatic questions. This neural network first embeds the question and the DPG into a d-dimensional space, and then uses an attention module to learn to attend to the relevant diagram relations by selecting the best statement choice that has a high similarity with the relevant diagram relations. Besides to evaluate the performance of the model and provide a standard baseline, this paper also builds a new dataset comprised of more than 5000 diagrams from many different topics. The experiments show this model provides latest state-of-the-art performance both on synaptic parsing and diagram question answering tasks.

Pros:

1. Present two new tasks of diagram interpretation and reasoning.
2. Introduce the DPG representation to encode diagram and introduce a model that learns to map diagrams into DPGs.
3. Introduce a model for diagram question answering that learns the attention of questions into DPGs.
4. Present a new dataset to evaluate the above models with baseline.

Cons:

1. Doesn't consider to use external knowledge as the prior, which could be very helpful in this task.