A Variational Approach to Knowledge Graph Node Embeddings

July 5, 2017

Abstract

We present a Variational approach to to the problem of embedding knowledge graph nodes.

1 Introduction

Knowledge graph node embeddings are vector representations of the nodes in a knowledge graph. Effective embeddings are an important first step in many pipelines for knowledge graphs visualisation, completion, classification and other tasks. Many embedding algorithms model the relations in the knowledge graph as maps in the embedding space: either translations, or more complex transformations. If for instance, we have the triple $(\texttt{John}, \texttt{parentof}, \texttt{Mary})$, then we learn embeddings $v_{\texttt{John}}, v_{\texttt{Mary}} \in \mathbb{R}^k$ (for some chosen embedding dimension k) and a map $f_{\texttt{parentof}} : \mathbb{R}^k \to \mathbb{R}^k$ for the relation $\texttt{parentof}$ such that $f_{\texttt{parentof}}(v_{\texttt{John}})$ should be as close to $v_{\texttt{Mary}}$ as possible. This scheme works well, but has trouble representing multiple relations. If John has more than one child, what should $f_{\texttt{parentof}}(v_{\texttt{John}})$ be? Many approaches have been introduced, often requiring an advanced mathematical structure for the embedding space, such as complex vectors [**?**], or holographic projections [**?**]. We introduce a variational approach. We model each knowledge graph node not as a point in the embedding space, but as a spherical Gaussian of fixed variance. Transforming this Gaussian by the relational map should then result in a new Gaussian which maximizes the likelihood of the nodes connected to the original by the given relation.[1] The variational approach has several advantages:

---

[1] This approach is inspired by the Coherent Point Drift algorithm [], which uses a similar approach for the problem of point set registration.

- The embedding space is a simple Euclidean space, making no special demands on downstream architectures.

- The multiple-relation issue is solved in an intuitive manner: mapping a parent node to its children creates a probability distribution that covers the children.

- For linear maps, we can decompose the loss function into factors which can be solved analytically. Thus, the problem can be solved using an iterative algorithm which is guaranteed to converge to a local optimum.

- Negative sampling is not required. The basic framework constrains the problem sufficiently to lead to models that generalize well.

We test our embeddings on the tasks of knowledge base completion and node classification. We outperform the state-of the art in each case. We also compare a general-purpose stochastic gradient descent approach for our model, to the variational search, showing that the latter converges faster, in fewer iterations, and to better solutions.

2 Model

3 Experiments

4 Conclusion