Evaluating Explanations of Relational Graph Convolutional Network Link Predictions on Knowledge Graphs

Nicholas Halliwell

▶ **To cite this version:**

Nicholas Halliwell. Evaluating Explanations of Relational Graph Convolutional Network Link Predictions on Knowledge Graphs. AAAI 2022 - 36th AAAI Conference on Artificial Intelligence, Feb 2022, Vancouver, Canada. hal-03454121

HAL Id: hal-03454121
https://hal.science/hal-03454121

Submitted on 29 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluating Explanations of Relational Graph Convolutional Network Link Predictions on Knowledge Graphs

Nicholas Halliwell

Inria, Université Côte d'Azur, CNRS, I3S, France
nicholas.halliwell@inria.fr

Abstract

Recently, explanation methods have been proposed to evaluate the predictions of Graph Neural Networks on the task of link prediction. Evaluating explanation quality is difficult without ground truth explanations. This thesis is focused on providing a method, including datasets and scoring metrics, to quantitatively evaluate explanation methods on link prediction on Knowledge Graphs.

Introduction

Knowledge Graphs represent facts as triples in the form *(subject, predicate, object)*, where a *subject* and *object* represent a real-world entity, linked by some *predicate*. Knowledge Graphs often do not explicitly contain every available fact. Link prediction on Knowledge Graphs is used to identify unknown facts from existing ones. Relational Graph Convolutional Networks (RGCN) (Schlichtkrull et al. 2018) extends Graph Convolutional Networks (Kipf and Welling 2017) for applications to link prediction on Knowledge Graphs, using the scoring function from DistMult (Yang et al. 2015) as an output layer, returning a probability of the input triple being a fact.

Several algorithms have been proposed to explain the predictions of an RGCN used for link prediction, in particular: ExplaiNE (Kang, Lijffijt, and Bie 2019) quantifies how the predicted probability of a link changes when weakening or removing a link with a neighboring node, while GNNExplainer (Ying et al. 2019) explains the predictions of any Graph Neural Network, learning a mask over the adjacency matrix to identify the most informative subgraph.

Contributions

Defining Ground Truth Explanations

The weakness of these papers is their evaluation of explanation quality due to the lack of available datasets with ground truth explanations. Evaluating the quality of explanation is essential to determining when to prefer one explanation method over another. Without ground truth explanations, researchers will have difficulties determining if their newly created algorithm is generating high quality explanations. This thesis addresses this issue by defining a method, including datasets and scoring metrics, to quantitatively evaluate these explanation methods. For this thesis, ground truth explanations are defined as a single justification for an entailment. We use an open-source semantic reasoner with rule-tracing capabilities (Corby et al. 2012) to generate ground truth explanations for each rule we choose to define. For this thesis, we focus on explaining family relationships, as no prior domain knowledge is needed.

Benchmarking Non-Ambiguous Explanations

The first step to quantitatively evaluating explanation methods starts with defining explanations where one and only one possible explanation can exist for some input triple. We define two datasets, Royalty-20k and Royalty-30k (Halliwell, Gandon, and Lecue 2021a), where each triple in the training and test set has exactly one set of triples determining why a link exists between the two given entities. Additionally, we propose the use of a scoring metric for non-ambiguous explanations, which involves computing the Jaccard similarity between the predicted and ground truth explanation. Lastly, in the first work of this thesis, we benchmark two state-of-the-art explanation methods, ExplaiNE and GNNExplainer, using the proposed dataset and scoring metric. In practice, there is often more than one way to explain a prediction. The beginning of this thesis focuses on the simplified case where there is only one way to explain each triple.

Benchmarking Ambiguous Explanations

When evaluating explanation quality, one must consider that there can be multiple ways to explain why a link could exist between two nodes. In other words, explanations can be non-unique. After defining datasets with only one ground truth explanation, we further expand on this idea of using justifications of an entailment as explanations to address the issue of benchmarking explanation methods with non-unique explanations.

The second work in this thesis involves designing a method, including a dataset and performance metrics, for evaluating explanations with non-unique explanations. This work focuses on 6 family relationships: *hasSpouse*, *hasBrother*, *hasSister*, *hasGrandparent*, *hasChild*, and *hasParent*. We construct a dataset including all possible explanations for each triples using these 6 family relationships. Furthermore, several scoring metrics are adapted to this task based on the generalized precision and recall (Kekäläinen and Järvelin 2002). Indeed the binary precision and recall

could be used for this task, however, these metrics fail to account for the fact that some explanations can be more intuitive than others to users. Both metrics would give a score of 1 when a predicted explanation exactly matches a ground truth explanation. However, an explanation method could predict an unintuitive explanation, and receive the highest possible evaluation score, potentially misleading practitioners into thinking the predicted explanation is of high quality. Therefore, scoring metrics used for this task must compare a predicted explanation to all possible explanations, and account for the fact that explanations have different degrees of relevance.

We conduct a user experiment, where for each predicate, users are shown all possible explanations and asked to assign a score based on how intuitive the explanation is. These user scores are used as relevance scores in the generalized precision and recall adapted for this task. We benchmark ExplaiNE and GNNExplainer on this dataset, using these performance metrics. This work was published as (Halliwell, Gandon, and Lecue 2021b).

Remaining Work

The remainder of this thesis involves two remaining tasks; understanding the role graph embeddings play on the quality of explanation generated by an explanation method, and comparing these state-of-the-art explanation methods against rule based algorithms.

Understanding the Role of Graph Embeddings on Explanation Generation. Indeed the graph embeddings learned by the RGCN plays some role in the quality of explanation generated by a post-hoc explanation method such as ExplaiNE or GNNExplainer. For all of our previous experiments, the embeddings are kept fixed, i.e., the exact same embeddings are used to benchmark both explanation methods. The extent to which the graph embedding influences the explanation is unknown. One natural question that stems from this idea is if the accuracy of the RGCN is increased or decreased slightly, how does the learned embedding influence the explanations generated by ExplaiNE or GNNExplainer? In other words, is there a relationship between perturbations in a given entity embedding and the performance metrics of the quality of explanation generated? When an explanation method generates an inaccurate explanation, is the explanation method flawed, or is this due to a bad embedding that is not capturing enough information.

In order to investigate this further, one first step would be to understand what properties of the graph the graph embedding has learned. Indeed this is no trivial task, however, one would want to ensure that a graph embedding captured the relationship between each triple and all its possible sets of explanations. I hope to be able to identify when an explanation method is generating a bad explanation from when the graph embedding model is learning a bad latent representation of the entities. This work may involve adapting the loss function and weight matrices of the RGCN to provide more insights as to what information the embedding is capturing.

Rule Based Link Prediction. Using an RGCN along with ExplaiNE or GNNExplainer is not the only way to perform explainable link prediction. Rule based methods can also be applied to this task. An interesting experiment would be to compare these rule based approaches to more recent explanation methods such as ExplaiNE and GNNExplainer. As of Fall 2021, rule based methods have not been benchmarked on any datasets constructed during this thesis. All datasets and metrics designed in this thesis are readily available to be used on other link prediction models and explanation methods.

Conclusion

This thesis allows researchers and practitioners to quantitatively evaluate explanation methods on the task of link prediction on Knowledge Graphs in ways they were previously unable to.

References

Corby, O.; Gaignard, A.; Faron Zucker, C.; and Montagnat, J. 2012. KGRAM Versatile Inference and Query Engine for the Web of Linked Data. In *IEEE/WIC/ACM Int. Conference on Web Intelligence*, 1–8. Macao, China.

Halliwell, N.; Gandon, F.; and Lecue, F. 2021a. Linked Data Ground Truth for Quantitative and Qualitative Evaluation of Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs. In *International Conference on Web Intelligence and Intelligent Agent Technology*. Melbourne, Australia.

Halliwell, N.; Gandon, F.; and Lecue, F. 2021b. User Scored Evaluation of Non-Unique Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs. In *International Conference on Knowledge Capture*. Virtual Event, United States.

Kang, B.; Lijffijt, J.; and Bie, T. D. 2019. ExplaiNE: An Approach for Explaining Network Embedding-based Link Predictions. *CoRR*, abs/1904.12694.

Kekäläinen, J.; and Järvelin, K. 2002. Using graded relevance assessments in IR evaluation. *J. Assoc. Inf. Sci. Technol.*, 53(13): 1120–1129.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Int. Conf. on Learning Representations, ICLR*.

Schlichtkrull, M. S.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling Relational Data with Graph Convolutional Networks. In *European Semantic Web Conference, ESWC*.

Yang, B.; Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *3rd International Conference on Learning Representations, ICLR*.

Ying, Z.; Bourgeois, D.; You, J.; Zitnik, M.; and Leskovec, J. 2019. GNNExplainer: Generating Explanations for Graph Neural Networks. In *Advances in Neural Information Processing Systems*.