Evaluation d'explications pour la prédiction de liens dans les graphes de connaissances par des réseaux convolutifs

Fabien Gandon, Nicholas Halliwell, Freddy Lecue

Evaluation d'explications pour la prédiction de liens dans les graphes de connaissances par des réseaux convolutifs

Fabien Gandon[1], Nicholas Halliwell[1], Freddy Lecue[1,2]

[1] Inria de l'Université Côte d'Azur, I3S, CNRS
[2] CortAIx, Thales

prénom.nom@inria.fr

Résumé

Nous résumons ici l'article [1] dont l'approche permet de générer un jeu de test comparatif et fournit des métriques pour évaluer les explications de prédictions de liens dans les graphes de connaissances par des réseaux convolutifs pour les graphes relationnels et ceci en présence de plusieurs explications possibles.

Mots-clés

prédiction de liens, IA explicable, graphes de connaissances, réseaux de neurones, évaluation d'explication.

Abstract

We summarize the paper [1] which the approach allows to generate a benchmarks and provides metrics to evaluate explanations of link predictions in knowledge graphs by relational graphs convolutional networks when several possible explanations do exist.

Keywords

link prediction, explainable AI, knowledge graphs, graph neural networks, explanation evaluation.

1 Introduction

Les réseaux convolutifs dédiés aux graphes relationnels (RGCN) [2, 3] sont utilisés sur des graphes de connaissances [4] notamment pour prédire des liens manquants mais malheureusement comme des boîtes noires. Plusieurs méthodes de génération d'explications ont été proposées pour expliquer leurs prédictions. Cependant la performance des méthodes d'explication est difficile à évaluer en absence d'une vérité terrain. De plus, il peut y avoir plusieurs explications pour une même prédiction dans un graphe de connaissances. Jusqu'à présent, il n'existait aucun jeu de données où les observations avaient de multiples explications avec lesquelles se comparer. De plus, il n'existait pas de métrique standard pour comparer les explications prédites par rapport à de multiples explications possible. Dans l'article [1], nous avons proposé une méthode et un jeu de données (FrenchRoyalty-200k), pour évaluer les performances de systèmes d'explication des RGCN sur la tâche de prédiction de liens dans un graphe de connaissances en présence de plusieurs explications possibles. De plus nous avons mené une expérience où des utilisateurs ont évalué chaque type d'explication possible de la vérité terrain en fonction de leur compréhension de l'explication et ceci afin d'affiner l'évaluation de la qualité des explications choisies par un système. A partir de cette vérité terrain, nous proposons l'utilisation de plusieurs métriques utilisant des poids agrégeant les scores des utilisateurs pour chaque explication prédite. Pour valider notre approche, nous nous avons évalué sur ce jeu de données des méthodes d'explication récentes pour la prédiction de liens en utilisant les mesures proposées.

2 De la nécessité d'évaluer les explications des prédictions de liens

Peu d'algorithmes existent pour aider à comprendre les prédictions des RGCN sur un graphe de connaissances. ExplaiNE [2] mesure le changement du score de sa méthode dû à une petite perturbation dans la matrice d'adjacence du graphe pour évaluer l'importance de la présence ou l'absence d'un lien dans la prédiction d'un autre lien. ExplaiNE repose sur l'hypothèse qu'une explication peut être fournie en sélectionnant l'un des voisins de la prédiction. GNNExplainer [3] explique les prédictions en apprenant un masque sur la matrice d'adjacence d'entrée pour y identifier le sous-graphe le plus impactant pour la prédiction. Une faiblesse de ces méthodes est l'évaluation de la qualité de l'explication. Les auteurs d'ExplaiNE reconnaissent eux-mêmes qu'il est difficile de mesurer la qualité de l'explication en l'absence de vérité terrain [2]. ExplaiNE mesure la qualité de ses explications en utilisant la similarité de Jaccard moyenne entre les genres pour un film recommandé donné, et l'ensemble des genres des 5 premières explications sélectionnées. Il s'agit d'une évaluation très limitée qui ne se généralise pas à d'autres tâche ou graphes facilement. De même, GNNExplainer n'a pas été évalué sur la tâche de prédiction de liens explicables sur les graphes de connaissances. Les évaluations sont donc limitées et, a fortiori, ne permettent pas non plus de comparer les méthodes d'explication entre elles.

Dans l'article [5] nous avions commencé par proposer une méthode et des ressources pour évaluer quantitativement et qualitativement les méthodes d'explication sur la tâche de

prédiction de liens dans un graphe de connaissances à partir de données du Web sémantique et dans le cas d'explications uniques pour chaque prédiction. Dans l'article [1] nous avons consolidé et étendu ces résultats en fournissant un autre jeu de données incluant de multiples explications possibles pour une même prédiction et en introduisant des métriques permettant l'évaluation et la comparaison des méthodes de prédiction.

3 Des traces d'inférences notées par les utilisateurs comme explications

Dans un graphe de connaissances, la sémantique formelle disponible nous permet de proposer comme vérité terrain pour des explications de prédiction de liens la justification implication logique de ce lien. Grâce à un raisonneur sémantique open-source avec des capacités de traçage de règles [6] nous avons généré automatiquement des explications pour des règles dérivant de nouveaux liens. Ce traçage identifie les liens qui ont causé la génération d'un autre lien que nous pouvons soumettre à des méthodes qui tenteront à leur tour de le prédire et de s'en expliquer. Cette approche générique de génération d'explications de la vérité du terrain peut être appliquée à de nombreux graphes de connaissances et à de nombreux ensembles de règles. De plus en multipliant les règles nous pouvons générer plusieurs explications possibles pour un lien. Certaines explications peuvent être plus faciles à comprendre que d'autres et l'évaluation d'une méthode devrait prendre cela en compte. Nous avons mené une expérience auprès d'utilisateurs pour noter chaque type d'explication possible. Cela nous permet de distinguer les explications qui sont intuitives de celles qui ne le sont pas, sans nous appuyer sur des hypothèses préalables. Au total, 42 utilisateurs ont répondu, de 11 nationalités différentes, issus de milieux informatiques et non informatiques. Nous avons normalisé les scores moyens entre 0 et 1 pour chaque explication possible, et les avons arrondis au dixième le plus proche pour en faire des poids utilisables dans l'évaluation des explications choisies par un système. Toutes les ressources utilisées et produites sont disponibles en ligne, y compris le lien de téléchargement du raisonneur, le code et les jeux de données [1].

4 Métriques, évaluation et résultats

Nous avons proposé d'évaluer les méthodes d'explication en adaptant la précision et le rappel généralisés [6] proposés à l'origine pour la recherche de documents. Nous définissons aussi la mesure F_1 généralisée, comme la moyenne harmonique entre précision et rappel généralisés et nous proposons la métrique max-Jaccard pour identifier quelle explication a le plus de points communs avec une explication prédite. La métrique max-Jaccard mesure à quel point une méthode d'explication est capable de prédire avec précision une des explications possibles. La précision et le rappel généralisés intègrent à quel point l'explication prédite a reçu un score élevé de la part les utilisateurs. Ces deux mesures empêchent une méthode d'explication de prédire uniquement des explications peu intuitives tout en recevant un score élevé. La mesure F_1 généralisée fournit une vue d'ensemble de la performance.

A partir des traces de règles appliquées à un jeu de données issu de DBpedia nous avons construit un graphe de connaissances de plus de 200 000 triplets avec différentes explications possibles et leurs scores. En utilisant les mesures introduites, nous obtenons un jeu de test permettant d'évaluer et comparer quantitativement différentes méthodes d'explication. Nous montrons que les méthodes d'explication n'essaient pas toujours de prédire la meilleure explication possible (i.e. celle avec le meilleur score des utilisateurs) et qu'elles tentent parfois de prédire des explications avec un score inférieur et n'y réussissent que partiellement. Le graphe de connaissance et son schéma nous permettent aussi d'effectuer une analyse d'erreur sur les prédictions les plus fréquentes et une comparaison des comportements en fonction des caractéristiques des liens (ex. symétriques, inverses, etc.).

Au final, la méthode introduite dans l'article, son jeu de données et ses métriques permettent aux chercheurs de développer de nouvelles méthodes d'explication et d'évaluer quantitativement et qualitativement leurs résultats d'une manière qui leur était auparavant impossible.

1. https ://github.com/halliwelln/multiple-explanations/

Remerciements

Ce travail est soutenu par le 3IA Côte d'Azur - ANR-19-P3IA-0002

Références

[1] N. Halliwell, F. Gandon, F. Lecue, *User Scored Evaluation of Non-Unique Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs*, InProceedings of the 11th on Knowledge Capture Conference 2021 Dec 2 (pp. 57-64).

[2] B. Kang, J. Lijffijt, T. De Bie, *ExplaiNE : An Approach for Explaining Network Embedding-based Link Predictions*, CoRR abs/1904.12694 (2019).

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

[4] F. Gandon, *Dessine-moi un graphe de connaissances !*, Binaire, 5 Oct. 2021.

[5] N. Halliwell, F. Gandon, F. Lecue, *Linked Data Ground Truth for Quantitative and Qualitative Evaluation of Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs*, WI-IAT 2021 - 20th IEEE/WIC/ACM Int. Conference on Web Intelligence and Intelligent Agent Technology, 2021.

[6] O Corby, A Gaignard, C Faron Zucker, J Montagnat. *KGRAM Versatile Inference and Query Engine for the Web of Linked Data*, In IEEE/WIC/ACM Int. Conference on Web Intelligence, 2012.