
Error! Reference source not found. ***Enhancing Sequential Recommendation via Decoupled Knowledge Graphs***

Bingchao Wu (Institute of Software, Chinese Academy of Sciences), Chenglong Deng (Institute of Software, Chinese Academy of Sciences), Bei Guan (Institute of Software, Chinese Academy of Sciences), Yongji Wang (Institute of Software, Chinese Academy of Sciences) and Yuxuan Kangyang (Institute of Software, Chinese Academy of Sciences).

-- *Review 1* --

Time: Feb 11, 14:10

Originality: 2 (Excellent)

Technical quality: 2 (excellent)

Reproducibility and availability of resources: 1 (Excellent)

Related Work: 1 (Comprehensive)

Presentation quality: 1 (Excellent)

Relevance to ESWC: 1 (In Scope)

Strengths:

(1) Strong technical quality. (2) Extensive experiments and ablation study to investigate the performance and impact of the proposed model and its components. (3) Addresses relevant challenges in existing work and is very well-written.

Weaknesses:

(1) The lack of available implementation prevents the reproducibility of results, even though both the model's architecture, as well as the experimental and parameters settings are clearly described in the paper. (2) Knowledge-enhanced, GNN-based sequential recommendation models are not included in the related wok. (3) Some design choices and results can be further clarified (see points (3)-(4) below).

Overall evaluation: 3 (strong accept)

The paper addresses the problem of knowledge-enhanced sequential recommendation. It aims to tackle two main challenges identified in existing works: the problem of confounding heterogeneous information of items (attribute-level and behavior-level semantic information) by embedding-based methods which use a single vector representation for items, and the fact that traditional distance-based and semantic matching knowledge graph embedding models used in sequential recommendation ignore higher-order connectivity between items.

The paper proposes a novel method of decoupling the knowledge graph into two complementary subgraphs based on user behavior, and intrinsic attribute information, respectively. These are separately encoded using two GNN-based knowledge subextractors, before the heterogeneous semantic information is combined to obtain high-level semantic features that are incorporated in the sequential model. To this end, the paper clearly identifies gaps in the literature and addresses a relevant research topic.

The overall formulation is sound. The proposed model is compared against a wide range of representative and state-of-the-art models, and the results are extensively discussed. The quality of the experimental protocol is solid, and the obtained results seem very promising in demonstrating *the importance of separating heterogeneous KG information for sequential recommendation. I particularly appreciated the extensive ablation studies which further showcase the strengths of the proposed recommender.*

However, there are a few areas which could be improved to further increase the quality of the paper.

(1) I am concerned that although the model's architecture, and the experimental and parameter settings are described in detail, the lack of an available implementation prevents a faithful reproducibility of the results presented in the paper.

(2) The paper mentions that the "existing embedding-based methods integrated into sequential recommendation models are divided into two categories, i.e., traditional distance-based models [...] and traditional semantic matching models [...]", which ignore the high-order connectivity between items. However, GNNs are another type of embedding-based methods which are able to take into account high-order connections between items (e.g. KGCN [1], KGAT [2], [3]), and have already been integrated in sequential recommenders (e.g. [4, 5]). In this context, I am wondering why the authors did not also take into account existing knowledge-enhanced, GNN-based sequential recommenders in their discussions and comparisons.

(3) In Sections 3.1 and 3.2, the authors mention the item embeddings E_b used as input to the AKS and, respectively, E_a, used as input to the BKS. How were these item embeddings obtained?

(4) In Section 4.3, the authors claim that CrbiaNet-BKS outperforms all other three model variants. However, the results in Table 2 indicate that CrbiaNet-BKS outperforms only CrbiaNet-AKS and CrbiaNet-RANDOM, while obtaining slightly lower HR@10 and NDCG@10 than CrbiaNet-ADD. I believe that the authors should clarify this. Moreover, is the difference in results between CrbiaNet and CrbiNet-ADD statistically significant?

(5) Minor comments: - P. 1 "from massive items" -> "from a (massive) collection of items" - P. 3 "we construct experiments" -> "we conduct experiments" - P. 5 "it indicates: -> "this indicates" - P.6 "translated-based method" -> "translation-based method" - P. 9 "carried out" -> "constructed" - P. 13 "HKS" -> "HSK" - Table 1: the meaning of the underlined results and the stars should be explain in the table caption

[1] Wang, H., Zhao, M., Xie, X., Li, W. and Guo, M., 2019. Knowledge Graph Convolutional Networks for Recommender Systems. In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 3307–3313. [2] Wang, X., He, X., Cao, Y., Liu, M. and Chua, T.S., 2019. Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 950-958). [3] Wu, S., Sun, F., Zhang, W. and Cui, B., 2020. Graph neural networks in recommender systems: a survey. arXiv preprint arXiv:2011.02260. [4] Wang, B. and Cai, W., 2020. Knowledge-enhanced graph neural networks for sequential recommendation. Information, 11(8), p.388. [5] Liu, F., Liu, W., Li, X. and Ye, Y., 2020. Inter-sequence Enhanced Framework for Personalized Sequential Recommendation. arXiv preprint arXiv:2004.12118.

==================== POST-REBUTTAL UPDATE The authors' response and clarifications addressed my points. I particularly appreciate that they provided the source code in an anonymised format. Therefore, I raised my reproducibility and related work scores accordingly.

Reviewer's confidence: 4 (high)

Error! Reference source not found. ***Supervised Knowledge Aggregation for Knowledge Graph Completion***

Patrick Betz (University of Mannheim), Christian Meilicke (Universität Mannheim) and Heiner Stuckenschmidt (University of Mannheim).

--- *Review 1* ---

Time: Feb 02, 14:03

Originality: 1 (Creative)

Technical quality: 1 (good)

Reproducibility and availability of resources: 0 (Reasonable)

Related Work: -1 (Inadequate)

Presentation quality: 1 (Excellent)

Relevance to ESWC: 1 (In Scope)

Strengths:

S1. The paper is very clearly written. S2. The techniques are layered, each of which, e.g., aggregation learning and latent feature representation, seems reasonable.

Weaknesses:

W1. The related work is outdated, especially for the knowledge graph embedding. W2. The experimental results do not show a clear conclusion. No model, e.g., Dense or Sparse, consistently outperforms others. W3. The selected competitors are not state-of-the-art.

Overall evaluation: 1 (weak accept)

In this paper, the authors studied the knowledge graph completion problem and proposed supervised knowledge aggregation to tackle this problem. The main application focus is on drug repurposing.

The authors proposed learnable aggregator with interpretability. The experiments showed promising results compared with some conventional knowledge graph embedding models.

I have several concerns as follows:

1. Please compare with state-of-the-art knowledge graph embedding models. For example, M2GNN can achieve 0.275 of Hits@1 on FB15k-237, which is far better than all the models reported in the paper.

2. Please give a detailed comparison between Dense and Sparse. Based on the reported results, I don't see which one is better and why.

-- After rebuttal,

I appreciate the response from the authors. The explanation about Dense and Sparse may reduce some concerns about the reasoning complexity in practice.

Reviewer's confidence: 3 (medium)

-- *Review 2* --

Time: Feb 11, 15:22

Originality: 2 (Excellent)

Technical quality: 2 (excellent)

Reproducibility and availability of resources: 1 (Excellent)

Related Work: 1 (Comprehensive)

Presentation quality: 0 (Reasonable)

Relevance to ESWC: 1 (In Scope)

Strengths:

The paper is well written, the assumptions are clear and the design choices are well discussed and motivated.

There is a good mixture of symbolic and subsymbolic considerations and techniques. The motivations of the usage of subsymbolic techniques starting from symbolic knowledge is reasonable and well-motivated

The discussion about interpretability is interesting and useful. Interpretability is crucial in healthcare-related tasks; thus I appreciated the interpretability used as motivation about the usage of the rule-based approach and also the discussion about how to interpret the behavior of the model once trained.

Weaknesses:

I think that examples about real rules come very late in the paper. I am not an expert in this domain, and so only when I saw the rules in section 7 I fully understood which kind of rule they are and partially understood how they are mined starting from triples. I think that having a real example in section 3.2 about how a rule can be mined starting from triples and an example of a rule is crucial to correctly explain the paper content.

In the introduction, it is explained that long-range dependencies are important in the biological domain but challenging to manage. I do not understand how the rules described in section 3.2 can exploit long-range dependencies

The results do not have a p-value nor a standard deviation term; both in the paper and in supplementary material I did not find the information of how many model instances for each setup you trained on each dataset, thus I think that you trained only one instance for each setup on each dataset. I think it is extremely important to have information about the stability of a model when the model is claimed to outperform other state-of-the-art models.

Overall evaluation: 1 (weak accept)

The article is well written and mixes both symbolic and subsymbolic notions and techniques with good results in terms of performance and interpretability. Thus I think that it is innovative and has to be considered for the main conference.

The justifications for the scores are:

Originality 2: an interesting mixture of symboling and subsymbolic techniques are used, both focused on resolving domain-related problems like the number of rules to consider, the meaning of the rules, and the importance of interpretability

Technical Quality 2: the technical sections are well explained and motivated; the proposed solutions are clear and reasonable

Reproducibility and availability of resources 1: the information in the article and the comprehensive supplementary material let interested researchers reproduce the approach explained in this article

Related Work 1: the literature review is complete and well described

Presentation quality 0: the presentation quality can be improved by adding clearer examples as I argued in the weaknesses

____ after rebuttal: I'd like to thank the authors for their responses. I am satisfied with the answers to my questions.

Reviewer's confidence: 2 (low)

-- *Review 5* --

Time: Feb 11, 15:42

Originality: 1 (Creative)

Technical quality: 1 (good)

Reproducibility and availability of resources: -1 (Not to standard)

Related Work: 0 (Reasonable)

Presentation quality: 1 (Excellent)

Relevance to ESWC: 1 (In Scope)

Strengths:

** Well and clearly written * Comprehensive evaluation*

Weaknesses:

** No implementation code available*

Overall evaluation: 2 (accept)

The paper describes an approach for learning knowledge graph completion rules, mainly in the biomedical domain. It is well written and evaluated in a comprehensive set of graphs, including non-biomedical ones. Only thing I am missing is links to the code/data for the reproduction of the experiments.

[After the rebuttal] Thanks for the authors' response, I maintain my evaluation.

Reviewer's confidence: 3 (medium)

Error! Reference source not found. ***Expressive Scene Graph Generation using Commonsense Knowledge Integration for Visual Understanding and Reasoning***

Muhammad Jaleed Khan (National University of Ireland, Galway), Edward Curry (National University of Ireland, Galway) and John G. Breslin (National University of Ireland, Galway).

-- *Review 1* --

Time: Feb 14, 14:58

Originality: 1 (Creative)

Technical quality: 1 (good)

Reproducibility and availability of resources: 0 (Reasonable)

Related Work: 1 (Comprehensive)

Presentation quality: 1 (Excellent)

Relevance to ESWC: 1 (In Scope)

Strengths:

+ novelty + shows potential of combining ML + (CS)KGs + sound evaluation + clear paper

Weaknesses:

- only minors, ie long sentences or typos

Overall evaluation: 2 (accept)

The paper proposes to integrate the commonsense knowledge from KGs in the task of scene graph generation. The Common Sense KG is used as it includes 7 different types of knowledge. The scene graphs of a given image is enriched with facts from the CSKG using graph embedding-based similarity. The proposed method is compared to existing approaches on a subset of the Visual Genome dataset, achieving higher results, and on the task of image generation.

I really liked the paper as it is. The topic is very timely, i.e. more and more deep learning methods for image understanding are appearing but they often lack of integration of common sense background knowledge at scale (to the best of my knowledge only ConceptNet/WordNet were used before). The related work are extensive (maybe half of the 60+ references would be enough). The method itself is a simple but effective combination of existing techniques, which results in a solid method and sound results. The evaluation is thoroughly designed and the paper reads very well.

A few things that I wonder : - the CSKG is not only heterogenous in knowledge, but also quite noisy given the way the (sub)grahps have been generated. The comparative analysis mentions some improvements wrt the literature, but no discussion on how much noise is introduced is given. I actualy think that the little improvement might be due to such noise. - I would be interested in knowing is how much each of the 7 graphs contributed to the improvement in performance. I bet not everything contributed the same. - Finally, there is a number of steps in the method. It might be useful to know how much error is collected throughout the steps, i.e. which step is affected the most? These are points that should be addressed.

I have annotated a few typos. If the paper is accepted I will send the authors a version for the camera-ready.

****

AR : Thank you for your answers regarding my questions.

Reviewer's confidence: 4 (high)

-- *Review 3* --

Time: Feb 10, 09:37

Originality: 1 (Creative)

Technical quality: 1 (good)

Reproducibility and availability of resources: 1 (Excellent)

Related Work: 1 (Comprehensive)

Presentation quality: 1 (Excellent)

Relevance to ESWC: 1 (In Scope)

Strengths:

** application-oriented approach * focus on commonsense knowledge * more aspects of evaluation*

Weaknesses:

** background knowledge includes the dataset (possibly bias) * a bit off-topic for Semantic web * several compared approaches were not mentioned in related work*

Overall evaluation: 2 (accept)

The paper deals with an application of commonsense heterogeneous knowledge source on scene graph generation. This has an promising effect on recall results in visual understanding domain. The paper first introduces the topic and provides illustrative motivating example where scene graph is enriched using commonsense knowledge with background information and related facts. Section 2 provides an overview related to scene graph generation from traditional ones to the recent works (a few) where commonsense knowledge graphs is explicitly leveraged for visual understanding and reasoning. There are introduced several commonsense knowledge resources of which seven are included in CommonSense Knowledge Graph (CSKG) heterogeneous resource. It is the first time CSKG has been applied for visual understanding and reasoning. The proposed method is described in Section 3. First Deep Neural Network components are briefly explained. The result is generated scene graph. This scene graph is subsequently refined (triples deduplicated) and enriched using CSKG via Knowledge Graph Toolkit. Evaluation is provided in Section 4 which includes comparison against state-of-the-art in terms of top K confident relationship predictions. Further, the evaluation also contains qualitative results in terms of illustrative results of generated scene graphs refined and enriched. Finally, there is a downstream task of scene graph generation based on scene graphs before and after integrating commonsense knowledge. In all the results of three evaluation aspects are promising.

In my opinion, the paper provides an interesting and original work with quite decent technical quality. The code is provided which ensures its reproducibility quality. The body of related work is quite extensive and it is quite easy to follow the paper even for non-experts in visual-understanding domain. Visual understanding is off-topic for the conference but we can see it as an intersection of domains: commonsense knowledge graph and visual understanding.

*Further remarks: * the related work is quite extensive but I am surprised that Cyc has not been mentioned there. Is there any reason for that? * Ad Figure 1: it seems to me there is a mistake. holding relation is in green but it should be blue, shouldn't be? * My main concern deals with the question whether there is not bias in the evaluation due to the fact that CSKG also includes Visual*

*Genome dataset and at the same time Visual Genome dataset is used for evaluation too. Does this cause bias? * Ad Figure 2, I am curious wheather there should two paralel BiLSTM boxes going into one pairwise concat box instead of sequence of two BiLSTM boxes? * Ad Algorithm 1, there is b2 on line 5. I would expect b1 on line 5. * p. 9, "CSKG and extract triplets from CSKG that include an object node in the predicted scene graph." - Why not subject node too? * p. 10, Could you provide the size of the dataset? There is information about the size of validation but not the whole dataset. * p. 11, In Section 4.2 there is first stated training of Faster RCNN and then SGG model. Does this mean that FasterRCNN is the first part and SGG model is means as the whole? * p. 12, It seems to me that not all approaches, used for comparison, are mentioned in related work. * In all, it would be interesting to know what are the computational costs of the approach (with and without integrating knowledge). * Do authors have some detail analysis (e.g. on the level of triples) in which commonsense knowledge graph help with regard to the the evaluation of the approach on given dataset? * Is there any reasoning with triples added based on commonsense knowledge? What authors mean with "higher-level reasoning"? Reasoning with higher triples?*

After the rebuttal

Thank you for your rebuttal. I am glad that you perfectly answered my questions and you will update the paper accordingly. Based on this I increase the overall score to 'Accept'.

Reviewer's confidence: 3 (medium)