Building a Product Recommender System Using Knowledge Graph Embedding and Graph Completion

*Anjali Dileepkumar, *Anjana Dileepkumar, *Karthik Ramanan , **Dr. Anuraj Mohan
*UG Scholar, **Assistant Professor, Department of Computer Science and Engineering, NSS College of Engineering, Palakkad

Product Recommendation

- In real-world scenarios like online shopping systems, knowledge about user preferences is likely to be incomplete.
- Traditional information filtering systems fall short in discovering user preferences and other relevant data.

Problem Statement

Building an effective product recommender system using knowledge graph embedding method that reflects richer data inter-relationships and an efficient learning pipeline that can predict the user preferences with good precision.

Objective

- Our work aim to construct and design a knowledge representation that encodes complex data inter-relationships using an unstructured raw dataset on user, product and category relationships.
- From the knowledge representation as defined above, we experiment to learn good representations that suit for passing to a machine learning pipeline.
- Using the learned representations, this work aims to predict the user preferences with good precision and evaluate the overall system performance.

Amazon Review Dataset

- The dataset comprises of data such as review data and product metadata.
- The dataset contains a total of 233.1 million reviews and data collected is from May 1996 to October 2018.
- The data set contains information on 29 different categories of products. The subset chosen for this work is "Video games".

Tools

- Python
- Neo4j
- Scipy
- PyTorch
- Pykg2vec
- Py2neo
- sklearn
- PyKEEN
- Pandas

Evaluation Measures

- Area under the ROC curve
- Average Precision (AP)

Methodology

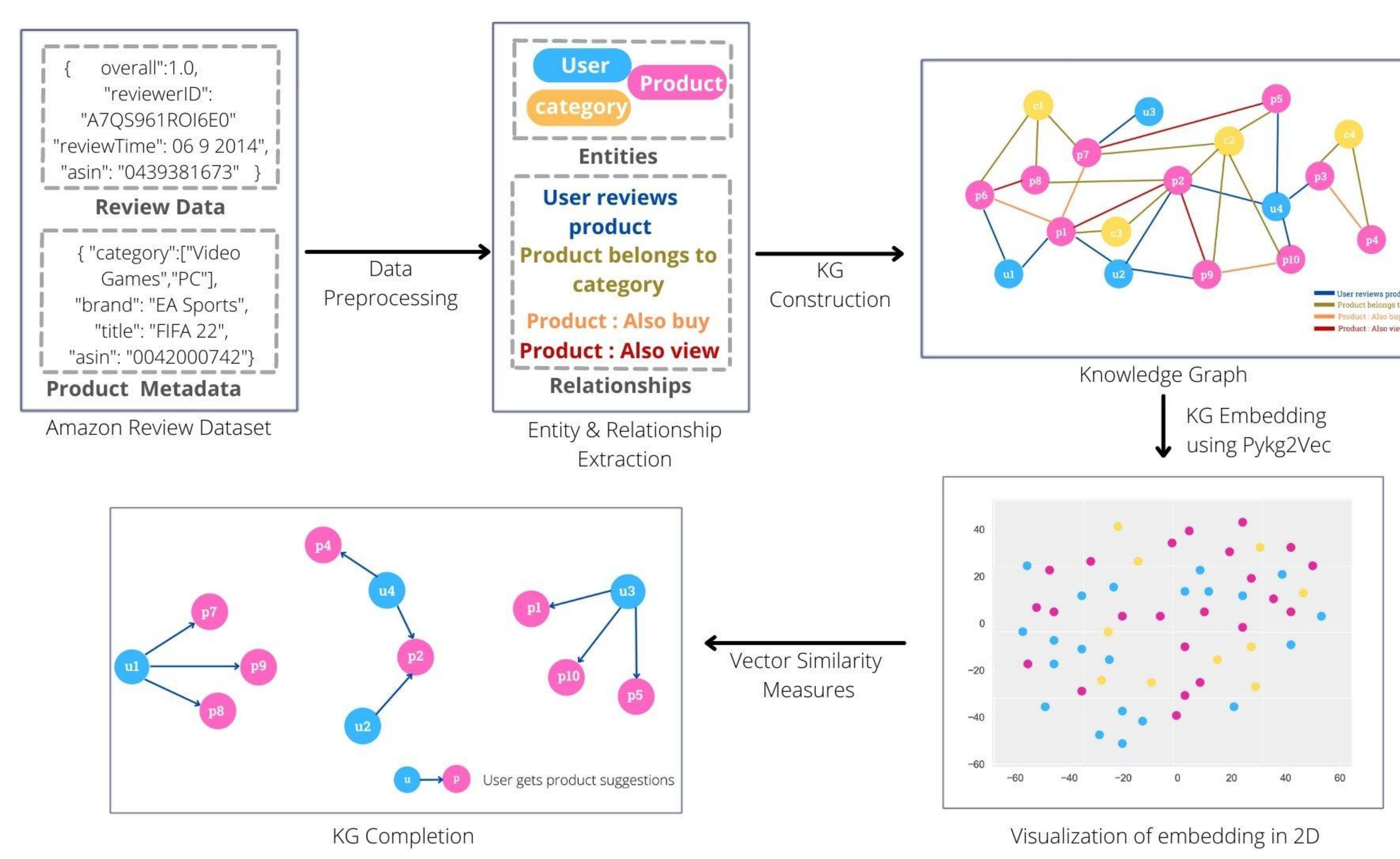

Fig. 1. System architecture

Various steps involved in the proposed system are:

- Preprocessing the dataset to extract entities and relationships.
- Knowledge graph construction.
- Perform knowledge graph embedding using TransE, TransR and transH.
- Multiclass link prediction and vector similarity measure to implement knowledge graph completion.

Results & Analysis

Classifier	AP	ROC
LR	**0.85310**	0.78466
RF	0.79607	**0.82206**
SVM	0.77841	0.79760
KNN	0.78714	0.79710

Table 1: AP and ROC of the proposed system w.r.t. various classifiers

Embedding Algorithm	AP Score	ROC Score
TransE	0.85310	**0.78466**
TransR	**0.85686**	0.78372
TransH	0.85010	0.77908

Table 2: Comparison w.r.t. Embedding Algorithms

Embedding Dimension	AP Score	ROC Score
25	0.84885	0.77707
50	0.85310	**0.78466**
75	0.84598	0.77352
100	0.85015	0.77856
150	**0.86258**	0.78219

Table 3: Comparison w.r.t. Embedding Dimension

Size of Test Set	Size of Validation Set	AP Score	ROC Score
10%	20%	0.82206	**0.79607**
10%	30%	0.85076	0.77768
20%	20%	**0.85310**	0.78466
20%	30%	0.84354	0.77050

Table 4: Comparison w.r.t. test-validation set splitting

Real Time Recommendations

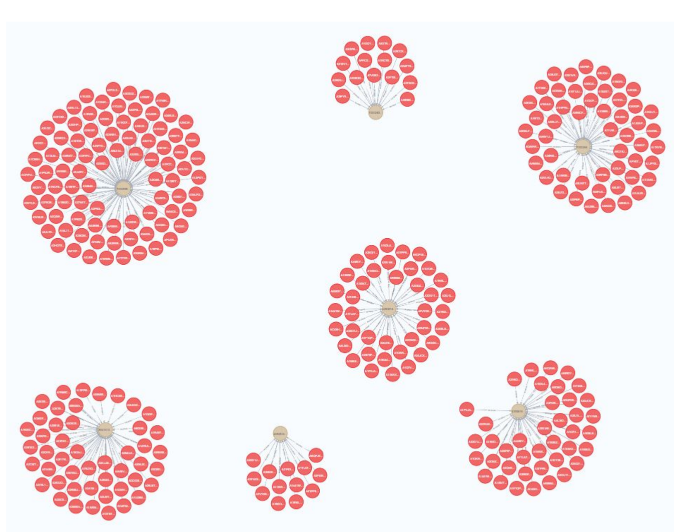

Fig. 2. User-product sub knowledge graph using relation reviews

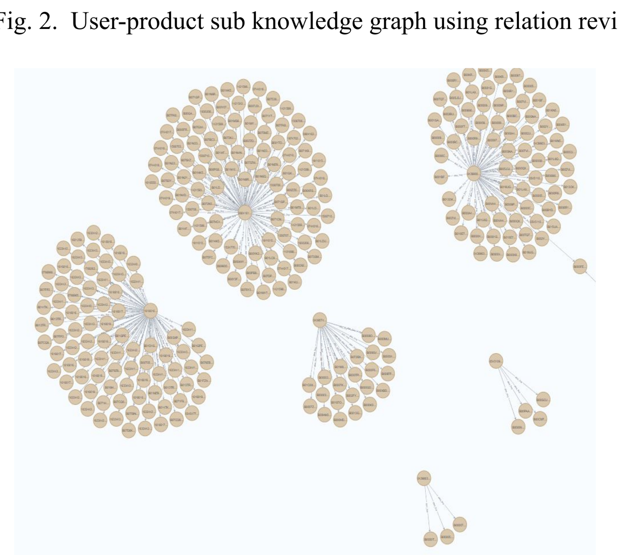

Fig. 3. Product-product sub knowledge graph

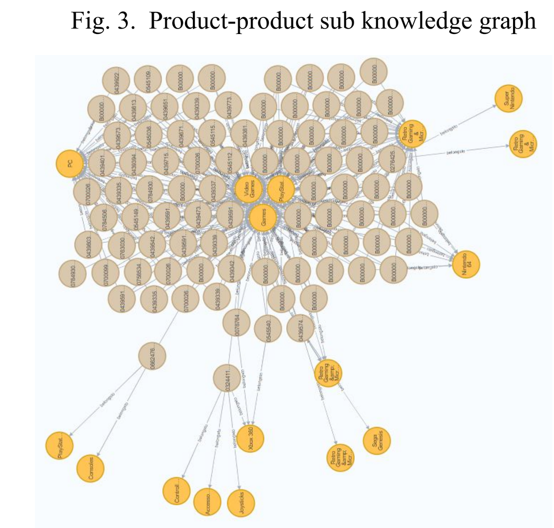

Fig. 4. Product-category sub knowledge graph

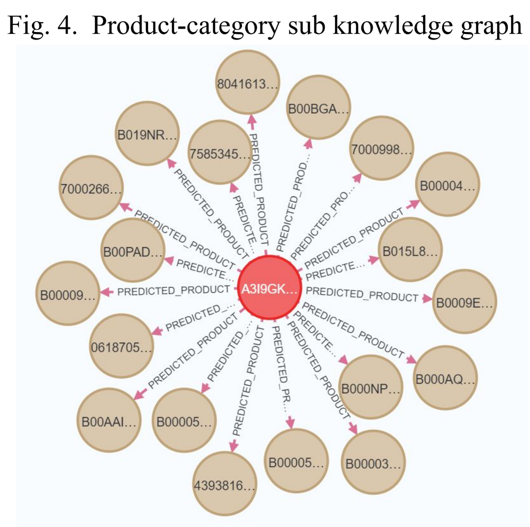

Fig. 5. Predictions for user A3I9GK5OO42B0I

Conclusion

We have implemented a novel pipeline by combining knowledge graph modeling and network representation learning to enhance the performance of product recommendation systems. The components are :

- Process the raw data available from amazon product reviews and construct a knowledge graph.
- Use knowledge graph embedding methods to learn graph representations.
- Perform knowledge graph completion to predict user preferences.

Future Work

- Advanced embedding algorithms can be used to generate representations, and this is likely to result in an improvement in performance.
- Taking a large amount of data for preprocessing and following a big data approach will lead to exponential increase in performance.
- In future, we would like to work with the big graph analytic pipeline that uses distributed graph processing engines to process large knowledge graphs.

References

[1] McAuley, Julian, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. "Image-based recommendations on styles and substitutes." In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval (2015) : 43-52.

[2] Yu, Shih-Yuan, Sujit Rokka Chhetri, Arquimedes Canedo, Palash Goyal, and Mohammad Abdullah Al Faruque. "Pykg2vec: A Python Library for Knowledge Graph Embedding." J. Mach. Learn. Res. 22 (2021): 1-16.

[3] Lucas Hu, Thomas Kipf, & Gökçen Eraslan. lucashu1/link-prediction: v0.1: FB and Twitter Networks. Zenodo.(2018): 1-14.

[4] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," In Twenty-ninth AAAI conference on artificial intelligence, (2015), pp. 2181–2187.

[5] Chen, Zhe, Yuehan Wang, Bin Zhao, Jing Cheng, Xin Zhao, and Zongtao Duan. "Knowledge graph completion: A review." IEEE Access 8 (2020): 192435-192456.