A Method to Constract a Masked Knowlege Graph Model using Transformer for Knowledge Graph Reasoning

1st Ryoya Kaneda
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
kaneda@ss.cs.osakafu-u.ac.jp

2nd Makoto Okada
Graduate School of Informatics
Osaka Metropolitan University
Osaka, Japan
okada@omu.ac.jp

3rd Naoki Mori
Graduate School of Informatics
Osaka Metropolitan University
Osaka, Japan
mnao@omu.ac.jp

***Abstract*—Most of the previous methods using machine learning for this challenge generate a new knowledge graph from the original one, and some information is lost in the process of creating a new knowledge graph. Therefore, we proposed a new model to estimate the criminal without changing the original knowledge graph. The proposed model uses a Transformer and allows the estimation of unknown criminals in nonexistent scenes by learning similar to Masked Language Modeling in BERT. This model, which uses the original knowledge graph, is expected to infer information about the crime scene at the same time as predicting the criminal. We confirmed by experiments that the model had gained the ability to estimate the hidden story parts by considering the surrounding stories.**

***Index Terms*—Knowledge Graph, Transformer, Masked learning**

I. Introduction

In this challenge, many previous methods using machine learning created a new knowledge graph from the original for predicting. However, we considered that the generated knowledge graphs lose some information. Therefore, we propose a new model, its training method, and a method to estimate the criminal, which can be input without processing the original knowledge graph.

II. Previous methods

A. Processing of knowledge graphs in previous methods

Tables I and II show examples of the original knowledge graphs and the processed knowledge graph by previous methods [1], [2].

As shown in Table I, most of the original data is a set of triples whose head is a scene entity. A processed knowledge graph is generated as below.

1) Separate the triple set for each scene entity in the head.
2) Extract the tail of the triple whose relation is “subject” and “hasPredect”.
3) Create a new triple set with head as the subject, relation as the predicate, and other tails as the tails.

This work was supported by JSPS KAKENHI Grant, Grant-in-Aid for Scientific Research(B) JP19H04184. This work was supported by JSPS KAKENHI Grant, Grant-in-Aid for Scientific Research(C) JP20K11958.

TABLE I: Original data

head	relation	tail
Scene1	subject	Holmes
Scene1	predicate	stand
Scene1	when	morning
Scene1	then	Scene2

TABLE II: Processed data

head	relation	tail
Holmes	stand	moning
Holmes	stand	Scene2

By creating the data in this way, we can represent the triple at the murder scene as “<criminal> kill sufferer”, which leads us to the problem of predicting the person who may be in this “<criminal>”. The task of predicting the missing entity of the triple is called the “Link Prediction Task”. It is one of the general tasks, so many approaches have been proposed.

B. Problems with previous methods and the original data

The problem with processing data by previous methods is that two major types of information in the original data is lost.

First, the relevance of the information to the scene and other information in the same scene is lost. Existing methods using co-occurrence information are available as an approach to the latter, but in any case, they still do not solve the problem of removing the scene itself. Second, due to the loss of the relation of the original data by processing, the information on what the information represents in six Ws is lost.

On the other hand, it is not easy to treat the original data as it is. It is because there is no scene that matches the situation we wish to predict, that is, the situation in which the criminal committed the crime. Due to this major problem, we cannot solve the problem as a “Link Prediction Task” to predict one of the triple missing entities.

III. Proposed Method

In this study, we propose a new inference method to solve the problems of previous methods and the major problems of original data. This method allows us to predict the criminal by utilizing the scene entities and relations in the original data.

The proposed methods are similar to the existing methods in that each entity and relation should be embedded. The difference from the existing methods is that they can handle the scene entities as they are in the original knowledge graph.

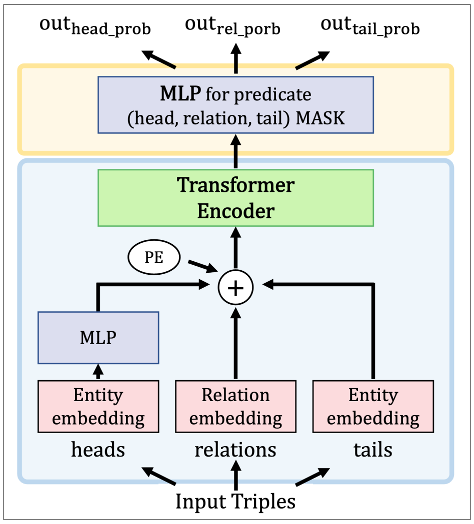

Fig. 1: Proposed Model. "PE" means "Position Encoding".

A. Approach

As an approach to the absence of scenes, we introduce a virtual scene in which the criminal commits the crime. It allows us to create a triple set with the virtual scene entity as the head for the original data.

However, it is difficult to learn unknown scenes that do not exist originally and to predict unknown criminals on them. Therefore, we propose a model that predicts the hidden elements from the neighborhood information. This model allows us to predict the hidden parts of a scene, including unknown scenes and unknown criminals.

B. Overview of the proposed model

Figure 1 shows an overview of the proposed model. As shown in the figure, this model consists of the blue layer that creates embeddings for each triple in the knowledge graph using a transformer and the yellow layer that predicts unknown scenes, relations, and tails by MLP using the output.

By training the model, it learns an embedded representation of each entity and relation in the knowledge graph, considering the surrounding information, and predicts unknown scene entities and unknown criminals.

C. Training Model

Figure 2 shows the training by the proposed model together with examples of data. The method of learning embedded expressions using this model is similar to that of the Masked Language Model of BERT. Specifically, some elements of the triple are replaced to "<mask>" and these are used as input to the Transformer Encoder. The model and embedding are learned by predicting "<mask>" parts by MLP.

D. How to predict the criminal

To infer the criminal, we create a virtual scene. Since it is a virtual scene, the scene entity is "<mask>". This scene is a scene where the subject is "<mask>", the predicate is "kill",

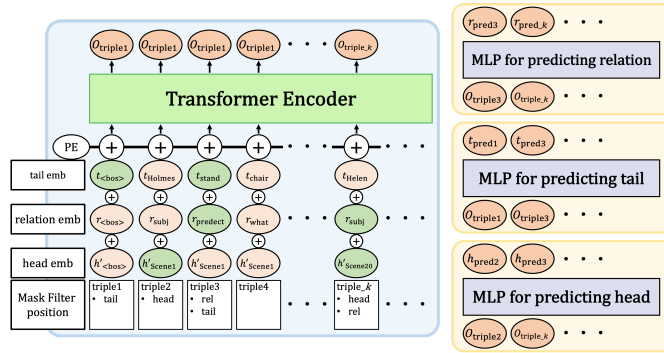

Fig. 2: Training with the proposed model

and the triple tail with the relation of "whom" is the "kill". The triple set of the virtual scene is inserted before or after the triple set of existing scene, and then input to the model, and the model predicts the masked part of the virtual scene.

In addition, the model can predict the object used in the crime by inputting a triple with the relation of "what" representing the object information. Similarly, the model can predict the reason for the crime by inputting a triple with the relation "why".

IV. Experiment

To confirm the validity of the model, we experimented with the case of no missing scenes. All titles were used for training the model. When considering missing scenes, the triple containing the relevant scene was excluded. After training the model, we compared the accuracy of guessing the criminal by placing virtual scenes as below.

1) Inserts only a virtual scene as input
2) Insert a virtual scene after scenes behind the same title
3) Insert a virtual scene before scenes behind the same title

The experimental results showed that the prediction of the criminal was least accurate when only the virtual scene was used as input, and most accurate when the virtual scene was inserted after the existing scene.

In the case of using a good model, the model can make a complete prediction for the problem of guessing the criminal on the condition of eliminating the victim himself or herself or non-humans as potential criminals. These results considered that the model derives the criminal from the flow of the story.

In addition, it is interesting to note that most of the possibilities for "<mask>" are the objects or persons that appear in the title. This result also shows that the results can be used for estimation.

References

[1] M. Kurokawa, "Searching for criminals based on knowledge graph embedding," *Second Knowledge Graph Reasoning Challenge*, 2020.

[2] S. Katsushima and H. Anada, "Predicting criminals in mystery novels using principal component regression analysis considering co-occurrence," *The 83rd National Convention of IPSJ*, vol. 5, p. 08, 2021.