A Criminal Detection of Mystery Novel Using the Principal Components Regression Analysis Considering Co-Occurrence Words

Shuhei Katsushima
Graduate School of Integrative Science and Engineering
Tokyo City University
Tokyo, Japan

National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan
g2081419@tcu.ac.jp

Hajime Anada
Graduate School of Integrative Science and Engineering
Tokyo City University
Tokyo, Japan
h-anada@tcu.ac.jp

Shusaku Egami
National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan
s-egami@aist.go.jp

Ken Fukuda
National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan
ken.fukuda@aist.go.jp

ABSTRACT

The interpretability problem, where even experts cannot explain the reasoning process of machine learning, has garnered considerable attention recently. Knowledge Graph Reasoning Challenge 2018, a contest concentrating on interpretability, was conducted in Tokyo. A previous study proposed a method based on word embedding to understand the meaning of the word in the novel. However, the method resulted in ignoring the flow of events. Therefore, we proposed a method that uses the principal components regression analysis of the feature vector of words considering word co-occurrence.

CCS CONCEPTS

• Computing methodologies→Artificial intelligence →Knowledge representation and reasoning; Ontology engineering;

• Computing methodologies→Artificial intelligence→Knowledge representation and reasoning; Reasoning about belief and knowledge

KEYWORDS

Reasoning on Knowledge Graph, Machine Learning, Explainable AI, Principal Components Analysis, Regression Analysis

ACM Reference format:

Shuhei Katsushima, Hajime Anada, Shusaku Egami and Ken Fukuda. 2021. A Criminal Detection of Mystery Novel Using the Principal Components Regression Analysis Considering Co-Occurrence Words. In *The 1st international Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence co-located with 10th International Joint Conference on Knowledge Graphs (IJCKG 2021). 4 pages.*

1 Introduction

With the recent development of artificial intelligence (AI), especially machine learning, social expectations for these technologies have increased. However, machine learning models such as deep learning, have an issue with the process of judgment and prediction being a black box, and even experts cannot give an interpretation to the results.

The Ministry of Internal Affairs and Communications created the Draft Guidelines for AI development in 2017 to promote AI and control the risks of their unexpected behaviors [1]; the guideline involves the principle of accountability of AI. Further, research on the interpretability of machine learning models has been getting a lot of attention.

The Knowledge Graph Reasoning Challenge (KGRC) was organized in Tokyo in 2018 [2] by the Special Interest Group on Semantic Web and Ontology Group of the Japanese Society for Artificial Intelligence (JSAI). The challenge aims to promote techniques for explaining AI using knowledge graphs. The contest provides a common task to identify criminals with a reasonable explanation based on open knowledge graphs from the Sherlock Holmes stories.

A previous study [3] converted novels represented by knowledge graphs into a structure called triples; then converted words into vectors by using an embedding method called TransE [4], and assigned correctness scores to the triple relations of the vectors to identify murderers. However, by converting the novel data into triples, information regarding the instantaneity of the target words like location and time, which should be learned simultaneously, was lost.

This study proposes a method for studying novels based on the Continuous Bag of Words model, which considers the instantaneity of words and an explanation approach based on principal component regression analysis.

1.1 Knowledge Graph

To consider the time course of scenes, KGRC assigned IDs to the contents, which are divided into the smallest units for each scene and the relationships between personages, and described them using a data structure called Knowledge Graph. Figure 1 shows an image of a knowledge graph.

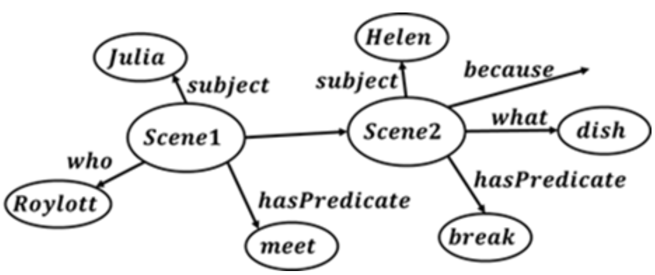

Figure 1: Image of Knowledge Graph

Examples of the relationship between scenes (transitions represented by arrows) in Figure 1 are listed as follows:

- Subject: the person who is the subject of the description between scenes
- hasPredicate: a predicate that describes the content between scenes
- Object: object that describes details between scenes; who, what
- Relations between scenes: then, because.

2 Related work

A previous study used an embedding method called TransE to learn the meanings of words in the given knowledge graph. In TransE, a vector is allocated to each word of the unit called triples, which consists of <subject, predicate, object>, and the text is learned by judging the correctness of the triple.

The score function for evaluating the triples in eq. (1) are given as follows:

$$f(h,r,t) = \|\mathbf{V}_h + \mathbf{V}_r - \mathbf{V}_t\| \tag{1}$$

Where $\mathbf{V}_h, \mathbf{V}_r, \mathbf{V}_t$ are the vectors of the subject, predicate, and object, respectively.

Eq. (1) returns a small value for correct triples and a large value for incorrect triples. The objective function is defined in eq. (2) as follows:

$$\mathcal{L} = \sum_i ([\tau + f(h_i, r_i, t_i) - f(h_i', r_i, t_i')]_+) \tag{2}$$

Where τ is the margin, $[x]_+ = \max(0, x)$; (h_i, r_i, t_i) denotes positive examples, whereas (h_i', r_i, t_i') denotes negative examples. This function is minimized in training.

The knowledge graph of the novel was converted into units called triples in existing research to employ TransE for word learning. However, in generating triples from a single sentence, each of two or more objects (place, time, or object) is divided and added to the subject and predicate, respectively, to form a triple. However, because the temporal relationship between triples is not considered, the simultaneity of words is missing.

3 Proposed method

The detection process is divided into two parts. Firstly, use the Continuous Bags of Words (CBOW) to learn the sequence of words in each context; this method allows for increased similarity of vectors between the co-occurrence words [5]. Secondly, the construction of a regression model that explains words that co-occur with other words, which is required for identifying criminals by extracting and interpreting the principal components with large regression coefficients.

3.1 CBoW

The CBOW model is a basic word embedding technique of natural language processing. The model predicts target words using surrounding words as input. Input words through a hidden layer, then output the probability distribution of words by activating the vector from the fully connected hidden layer. Figure 2 shows the image of CBOW.

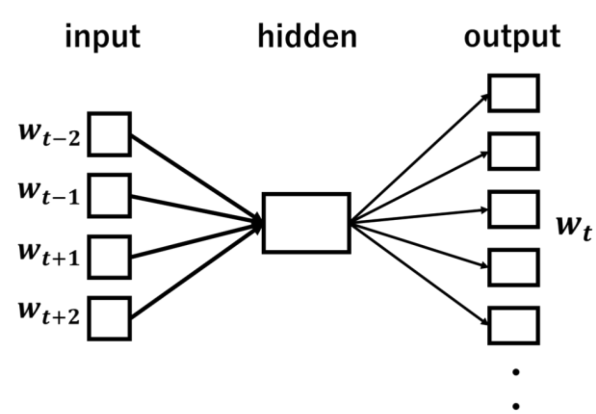

Figure 2: Continuous Bag of Words

As input words, $w_{t-i}, \cdots, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, \cdots, w_{t+i}$, which consist of i words surrounding w_t, the model outputs a score of w_t.

3.2 Principal Components Regression Analysis

This study uses the principal components regression analysis for feature vector of words to identify the murderer in the novel. Each scene in the novel is represented by several principal components in the principal component regression analysis, and the regression analysis is performed using the principal components as the explanatory variable; co-occurrence words necessary for the murderer's identification are the objective variables. This method has the following three key merits:

1 Multicollinearity in regression analysis is avoided because principal component analysis generated variables uncorrelated with each other.

2 The linear regression model is expressed as a linear combination of weights, and the effect of independent variables on the target variable is explicit.

3 Evaluating the factor loadings of the highly influential principle components makes interpreting the principal components simple.

3.2.1 Input of Principal Components Analysis

Principal Components Analysis (PCA) is a technique for reducing the dimension of a dataset to minimize information loss.

PCA uses a multi-hot vector of the novel data, which is a binary matrix that indicates the presence of each sentence's word label. Figure 3 shows an example of a multi-hot vector.

$$\text{"Helen come Holmes\_house"} \Rightarrow \begin{array}{cccccc} \text{Helen} & \text{meet} & \text{has} & \text{come} & & \text{Holmes\_house} \\ [1 & 0 & 0 & 1 & \cdots & 1] \end{array}$$

Figure 3 Example of multi-hot vector

The number of different words that appear in the novel is represented by elements in the vector.

3.2.2 Regression Analysis

Regression Analysis is a statistical technique for evaluating the relationships between a dependent variable and one or more explanatory variables [6]. We defined $w_{keyward}$ as the key-ward for detecting murderers and examining the influence of the principal component on the $w_{keyward}$.

$$\mathcal{L} = softmax(w_t)_{w_{keyward}} \tag{3}$$

Where $\mathcal{L}$ outputs the score against input data (a sentence). The higher the co-occurrence between input data (a sentence) and $w_{keyward}$, the higher the value returned by this function.

4 Experiments

We carried out experiments using the following novels of the Sherlock Holmes series: "Speckled Band", "Devil's Foot," and "Abbey Grange." Here, the criminal estimation of "Speckled Band" and "Abbey Grange" are performed by combining the data of "Devil's Foot".

The CBOW model learns the sequence structure of words in each novel. The criminals of each novel, Roylott and Jack, are solved by CBOW as the fill-in-the-blank questions shown in Table 2, and the candidates appearing in each novel are ranked. If the output result for a personage in the fill-in-the-blank question output by the CBOW model is high, that personage is considered to be the criminal.

Table1 shows the experimental parameters.

Table 1 Experimental parameters

Embedding dim	100
Window size	2
Learning rate	0.01
Iteration	10
Optimize	Adam
$w_{keyward}$	kill

Table 2 shows fill-in-the-blank questions of "Speckled Band" and "Abbey Grange," which were used in the previous work.

Table 2 Fill-in-the-blank questions of each novels

Novel\data	*Speckled Band*	*Abbey Grange*
Question 1	(?) kill Julia.	(?) kill Brackenstall.
Question 2	On_death_day_of Julia, (?) kill Julia in_bedroom_of_Julia.	(?) kill Brackenstall, Lady_Brackenstall had been tied with a_red_string

In KGRC, the novels with the last 10% and 25% cuts are used.

5 Results

For each data, tables 3 and 4 show the rank of the real murderer for the fill-in-the-blank questions when an additional novel ("+Devil's Foot") is added to the target novel.

Table 3 Suspect ranking of each novel's personages (10% cut)

Novel (10% cut)	**Question 1**	**Question 2**
Speckled Band	+Devils Foot	
	6/9	6/9
Abbey Grange	+Devils Foot	
	5/7	5/7

Table 4 Suspect ranking of each novel's personages (25% cut)

Novel (25% cut)	**Question 1**	**Question 2**
Speckled Band	+Devils Foot	
	6/9	6/9
Abbey Grange	+Devils Foot	
	2/7	2/7

The number of input principal components to the regression analysis was limited to 170 and 260 principal components in the cases of the “Speckled Band” and “Abbey Grange” principal components, respectively, assuming 80% of the data can be explained.

Figure 4 describes the top five partial regression coefficients of the regression model constructed with the 10% cut novel.

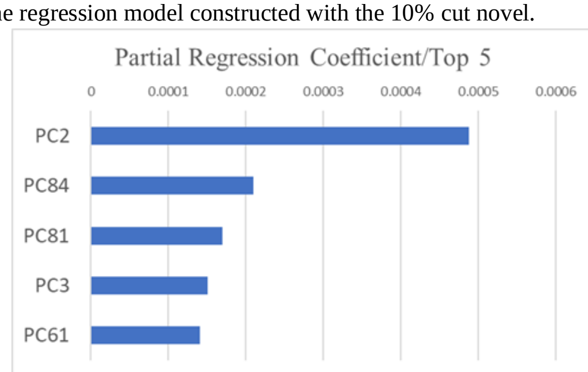

Figure 4 Partial regression coefficient / top 5

In Figure 4, the 2nd, 84th, 81st, 3rd, and 61st principal components were constructed as principal components with high partial regression coefficients.

The values of the factor loadings of the 2nd principal component at this time are described in table 5.

Table 5 Factor loading / Top 5 and bottom 10

canSee	0.005605
buy	0.00203
by_falling_price_of_agricultural_products	0.001208
support	5.06E−05
intense_fear	4.99E−05

Crown_Inn	−0.01231
Leopard	−0.01252
beDivided	−0.01277
30_years_old	−0.01334
Indian_cigarettes	−0.01386
India	−0.01464
Crane_Water	−0.01488
Holmes	−0.01498
Roma	−0.01879
Stoke_Moran	−0.01974

Table 5 shows the five highest and ten lowest factor loadings of the second principal component; table 6 shows the suspected murderer ranking of personages.

Table 6 Suspect ranking of SpeckledBand's personages

Rnanking	Name list of SpeckledBand
1	Mother_of Helen
2	Percy_Armitage
3	Roma
4	Stoke_Moran
5	Sister
6	Roylott
7	Helen
8	Watson
9	Holmes

Tables 5 and 6 show that the person listed in the lower part of the factor loading is also ranked low in the suspected murderer ranking. Further, Roylott, the real murderer, also has low factor loadings outside table 5.

6 Discussion

The person listed in the lower part of the factor loading is also ranked low in the suspected murderer ranking as shown in the 2nd principal component. Therefore, this principal component shows the co-occurrence structure of the novel's words that influenced the murderer's estimation.

In the novel, the real murderer Roylott sends a poisonous snake to Julia's bedroom and it kills her. However, Highly ranked words in Table 5 are different from the reasoning process of Holmes in the novel. Table 6 shows that “Mother_of_Helen” is the murderer who dies early in the plot. Thus, it is necessary to add proper data such as ConceptNet [7] and crime method ontology ~~is~~ to correctly identify the murderer. In addition, this model has many principal components, but only a small number of principal components have been investigated. It is necessary to select the variable by Lasso Regression for interpretability improvement.

7 Conclusion

We proposed a method to explore the process from additional knowledge to identify the murderer by using the co-occurrence of words. This model shows the important data for explaining the inference result by analyzing the bias of the sequence structure of words and explains the inference result.

However, the form of additional knowledge has a strong influence in identifying the murderer. Further, this model cannot consider the lies that appeared in the novel because it assumes the content in the novel to be true.

We would like to study a model that can trace the reasoning process that is close to the actual reasoning by humans.

ACKNOWLEDGMENTS

This paper is based on results obtained from projects, JPNP20006 and JPNP180013, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] Ministery of Internal Affairs and Communications. 2017. Draft AI R&D GUIDELINES for international Discussions. The Conference toward AI Network Society. https://www.soumu.go.jp/main_content/000507517.pdf

[2] Kawamura T. et al. 2020. Report on the First Knowledge Graph Reasoning Challenge 2018. In: Wang X., Lisi F., Xial G., Botoeva E. (eds) Semantic Technology. JIST 2019. Lecture Notes in Computer Science, vol12032. Springer, Cham. https;//doi.org/10.1007/978-3-030-41407-8_2.

[3] Mori Kurokawa. 2019. Narejji gurafu umekomi ni motozuku hannin suitei (in Japanese). Criminal detection based on knowledge graph embedding. https://challenge.knowledge-graph.jp/submissions/2019/kurokawa/kddi_kurokawa_1226.pdf.

[4] Antoine Bordes. 2013. Translating Embeddings for Modeling Multi-relational Data, In: Usunier, N., Garcia-Duran, A., Weston, J., &Yakhnenko, O. In Proceedings of the 26th International Conference on Neural Information Processing Systems(NIPS"13), pp.2787-2795.

[5] Tomas Mikolov. 2013. Efficient Estimation of Word Representations in Vector Space, https://arxiv.org/abs/1301.3781.

[6] Jean-Francois Dupuy, 2018. A Brief Overview of Linear Models In Statistical methods for Overdispersed Count Data, pp.1-32.

[7] Robyn S. et al. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/viewPaper/14972