Convolutional Knowledge Graph Link Prediction with Reshaped Embeddings

Yanhong Li*
yanhongli@uchicago.edu
University of Chicago
Chicago, Illinois, USA

Abstract

While the majority of current knowledge graphs remain incomplete, a plethora of models have been proposed with the intent of predicting the missing facts within them. In recent years, multi-layer models that leverage Convolutional Neural Networks (CNNs), such as Convolutional 2D Knowledge Graph Embeddings (ConvE), have exhibited superior performance over their shallower counterparts. The key to their success is largely due to the intricate interplay between entity and relation embeddings. In this study, we present ConvRE, a refined version of the ConvE model, and disclose our state-of-the-art results across multiple datasets. We further introduce a quantifiable measure of interaction, termed 'interaction length,' and propose an innovative approach to reshaping both relation and entity embeddings with the aim of intensifying their interaction. Upon employing these reshaping techniques, we successfully enhance the MRR score of the ConvE model by up to 3.28% (FB15K-237) and 3.72% (WN18RR). In addition, we conduct an in-depth analysis of the correlation between the degree of feature interaction and model performance. This exploration substantiates our assertion that augmenting the number of interactions between embeddings facilitates superior link prediction performance.

CCS Concepts: • **Information systems** → **Data mining**; • **Computing methodologies** → *Neural networks.*

Keywords: Knowledge Graph, Knowledge Graph Embedding, Link Prediction

1 Introduction

Knowledge graphs (KGs) are multi-relational graphs in which nodes symbolize entities and edges denote the relationships between these entities. Entities and relations are structured in the form of knowledge triples (s, r, o), signifying a relation 'r' between a subject entity 's' and an object entity 'o'. These knowledge graphs find extensive applications in real-world scenarios, such as recommender systems [7, 19], information retrieval [4], and information extraction [3, 8].

Several knowledge bases (KBs) exist today, including Freebase [1], WordNet [9], and YAGO [13]. However, a significant number of these knowledge graphs are marred by incompleteness [6]. Link Prediction (LP) is a promising approach that aims to predict missing facts among entities already present in a KG [12]. Many link prediction models leverage knowledge graph embedding [2, 17, 18]. By learning low-dimensional representations of entities and relations, these models can predict missing facts in incomplete knowledge graphs.

In recent years, multi-layer models employing Convolutional Neural Networks (CNNs) have seen an upsurge in popularity, thanks to their superior performance over their shallower counterparts [5, 16]. The application of multiple convolutional layers to the embeddings of both entities and relations enables these models to learn more expressive and rich features. The success of CNN-based models can largely be attributed to the enhanced interactions between the embeddings, which lead to more meaningful representations.

In this paper, we propose ConvRE (Convolutional Knowledge Graph Link Prediction with Reshaped Embeddings), a model that not only builds upon but also improves ConvE.

Our contributions are specific and distinct:

- We enhance the performance of the ConvE model by up to 3.28% (FB15K-237) and 3.72% (WN18RR) after applying reshaping techniques on embeddings.
- We suggest a novel approach for reshaping relation and entity embeddings to augment their interactions. The relation technique is demonstrated in Figure 1.
- We provide a quantitative definition of interaction – 'interaction length.' This metric measures the length of the contact surface between entity and relation embeddings. A greater interaction length signifies a higher degree of interaction.
- We further establish that increasing the number of interactions between embeddings enhances link prediction performance.

Please find the instructions for reproducing the results at https://github.com/yanhong-lbh/ConvRE.

2 Related Work

Current Knowledge Graph Embedding (KGE) models broadly fall into three categories: (i) translation-based models (geometric models), (ii) tensor decomposition models, and (iii) neural network models [20].

- **Translation-based models**: Over the past few decades, the advent of the TransE [2] model has spurred a wealth of research on neural link prediction models.

KDD-UC, August 7, 2023, Long Beach, CA
2023.

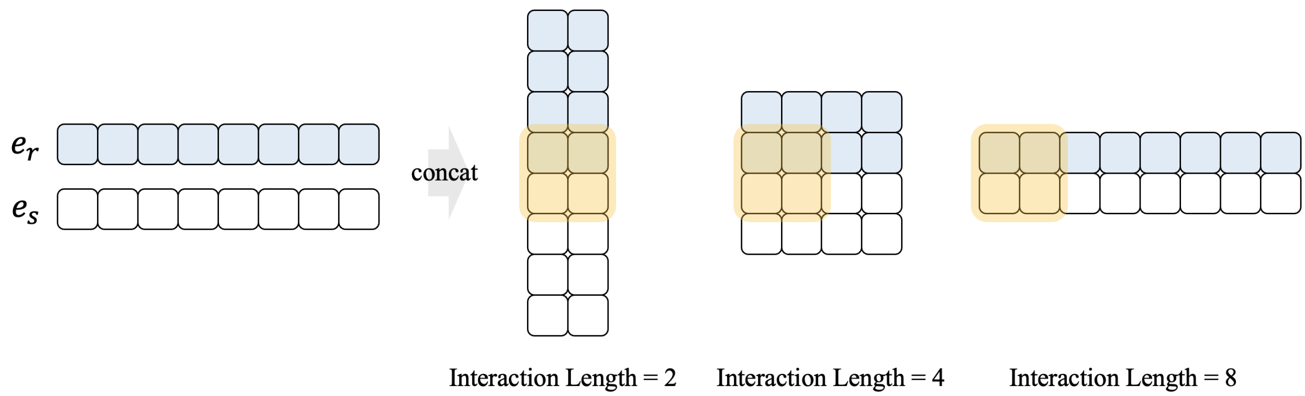

Figure 1. A 2D demonstration of reshaping embeddings by setting different interaction lengths. Here, we assume the shapes of the entity's and relation's embeddings are (1, 8) and that the convolutional filter's shape is (2,2). The embeddings are reshaped into shapes of (4, 2), (2, 4), and (1, 8). After concatenation, the interaction lengths are 2, 4, and 8, respectively, demonstrating an increasing degree of interaction.

Early models like TransE [2] and TransH [17] deploy dot products and matrix multiplications in their score function computations.

- **Tensor decomposition models**: These models posit that the score of a triple can be factorized in the form of tensors. Several tensor factorization-based models have been proposed, such as the Bilinear Diagonal model (DistMult) [18], and the ComplEx model, an extension of the DistMult model into the complex space [15].
- **Neural network models**: Recent studies have demonstrated that enhancing interactions between entity and relation embeddings can bolster a model's expressiveness [11, 18]. This insight has spurred many researchers to focus on feature interactions between embeddings. The Convolutional 2D Knowledge Graph Embeddings (ConvE) model, for instance, uses 2D convolutions over embeddings for link prediction in knowledge graphs. It comprises a single convolution layer followed by a projection layer to the embedding dimension and subsequently an inner product layer [5]. ConvKB [10] employs a convolution layer over the embeddings of the head entity, relation, and tail entity. Inspired by ConvE, InteractE further optimizes the performance of embeddings by intensifying feature interactions, utilizing three key concepts: Feature Permutation, Checkered Reshaping, and Circular Convolution.

Most existing KGE models predict missing links by optimizing a score function. In Table 1, we summarize the score functions of several KGE models. The vectors $\mathbf{e}_s$, $\mathbf{e}_r$, and $\mathbf{e}_o$ refer to the subject embedding, the relation embedding, and the object embedding, respectively. In the ComplEx model, $\mathbf{e}_s, \mathbf{e}_o \in \mathbb{C}^k$, whereas in all other models, $\mathbf{e}_s, \mathbf{e}_o \in \mathbb{R}^k$. Here, k denotes the embedding size, and $\mathbb{C}$ and $\mathbb{R}$ denote the complex and real-valued spaces, respectively. The symbol $*$ denotes circular-correlation, while $\star$ denotes convolution.

Table 1. Scoring Function of Various KGE Models

Model	Scoring Function $\psi(e_s, e_r, e_o)$
TransE	$\|e_s + e_r - e_o\|_p$
DistMult	$\langle e_s, e_r, e_o \rangle$
HolE	$\langle e_r, e_s * e_o \rangle$
ComplEx	$\mathrm{Re}\left(\langle e_s, e_r, e_o \rangle\right)$
ConvE	$f\left(\mathrm{vec}\left(f\left([\overline{e_s}; \overline{e_r}] \star w\right)\right)\mathbf{W}\right) e_o$
ConvRE	$f\left(\mathrm{vec}\left(f\left(g\left(\overline{e_s}; \overline{e_r}\right) \star w\right)\right)\mathbf{W}\right) e_o$

In this paper, ConvRE focuses on improving $g\left(\overline{e_s}; \overline{e_r}\right)$, the reshaping function, to achieve superior performance in link prediction.

3 Model

In this paper, we introduce ConvRE, a model that proposes a novel feature reshaping method to enhance the model's expressiveness. We define the interaction length between entity and relation embeddings as the length of the contact line between the two. Our findings suggest that expanding the interaction length between the entity and relation embeddings prior to the concatenation process can increase a model's expressiveness.

The default embedding size for ConvE is 200, and the default value for the first dimension (number of rows) of the reshaped 2D embedding is 20. It can be inferred that the default shape of both the entity and relation embeddings is (20, 10), with 20 rows and 10 columns. The ConvE model opts to concatenate the two embeddings on the first dimension

Table 2. Nine Combinations of Shapes and Their Corresponding Interaction Length

Shape of Embeddings	Shape After Concatenation	Interaction Length
(50, 4)	(100, 4)	4
(40, 5)	(80, 5)	5
(25, 8)	(50, 8)	8
(20, 10)	(40, 10)	10
(10, 20)	(20, 20)	20
(8, 25)	(16, 25)	25
(5, 40)	(10, 40)	40
(4, 50)	(8, 50)	50
(2, 100)	(4, 100)	100

Table 3. Statistics of datasets

Dataset	$\|\mathcal{E}\|$	$\|\mathcal{R}\|$	# Triples		
			Train	Valid	Test
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

(rows) of the embedding's shape, resulting in a shape of (40, 10) after concatenation, and an interaction length of 10.

Building upon the ConvE model, ConvRE reshapes the embeddings before stacking them along their rows. ConvRE tests all possible combinations of the number of rows and columns of the embedding shape. In total, there are nine different embedding shapes (Table 2), with the interaction length ranging from 4 to 100.

Among the nine embedding shapes, (20, 10) is the default shape used by the ConvE model, comprising 20 rows and 10 columns. The following shapes of entity/relation embeddings are omitted: (200, 1), (100, 2), and (1, 200). This is because the shapes after concatenation for these reshaped embeddings are (400, 1), (200, 2), and (2, 200), all of which have a dimension that is smaller than the default kernel size (3, 3) used by the ConvE model.

4 Experiment Setup

- **Datasets**: We evaluate ConvRE primarily on two of the most commonly used datasets: FB15k-237 [14] and WN18RR [5]. More details about these datasets are shown in Table 3.
- **Baselines**: We compare ConvRE with DistMult [18] , ComplEx [15], and ConvE [5].
- **Evaluation protocol**: We follow the standard evaluation metrics: Mean Reciprocal Rank (MRR), Mean Rank (MR), Hits@10, Hits@3, and Hits@1.

5 Results

In this section, we attempt to answer the following questions:

- **Q1**: How does ConvRE perform compared with the baselines? (Section 5.1)
- **Q2**: What is the effect of increasing the interaction length on the model's performance? (Sections 5.2 and 5.3)
- **Q3**: For the FB15K-237 dataset, why does the shape (2, 100), the one with the highest degree of feature interactions, have worse performance compared to its (8, 25), (5, 40), and (4, 50) counterparts? (Section 5.2)
- **Q4**: Why do abnormal values only appear in results on FB15K-237, but not on the WN18RR dataset? (Section 5.3)

5.1 Performance Comparison

We compare ConvRE against DistMult [18], ComplEx [15], and ConvE [5] on the FB15k-237 and WN18RR datasets. The results are summarized in Tables 4 and 5. Since our model aims to improve ConvE, we specifically compare it against this model. The results show that ConvRE outperforms ConvE across four out of five metrics on FB15k-237 (MRR, MR, Hits@10, Hits@3) and across all metrics for WN18RR (MRR, MR, Hits@10, Hits@3, Hits@1). After applying reshaping techniques, ConvRE achieves an improvement of 3.28% and 3.72% on FB15k-237 and WN18RR on MRR over ConvE, respectively. When comparing with other baseline KGE methods, ConvRE outperforms DistMult and ComplEx across all metrics on both FB15k-237 and WN18RR datasets. These results validate the effectiveness of ConvRE.

5.2 Effects of Feature Interaction on FB15k-237

We find that, except for the last data point, which represents an interaction length of 100, increasing the number of interactions between embeddings improves the performance of link prediction. According to Figure 2, up until the interaction length reaches 50, there is a roughly positive linear relationship between the interaction length of the embeddings and MRR. The reason why an interaction length of 100 performs worse than previous feature interaction variations is explained as follows:

Table 4. Link prediction results for FB15K-237

Model	Shape	MRR	MR	Hits@10	Hits@3	Hits@1
DistMult [18]	–	.241	254	.419	.263	.155
ComplEx [15]	–	.247	339	.428	.275	.158
ConvE [5]	–	.316	246	.491	.350	.239
ConvRE (Proposed Method)						
	(50, 4)	.319	245	.496	.351	.230
	(40, 5)	.320	228	.497	.349	.232
	(25, 8)	.320	245	.499	.351	.232
	(20, 10)	.321	242	.500	.352	.232
	(10, 20)	.323	241	.502	.355	.235
	(8, 25)	.325	243	.507	.359	.234
	(5, 40)	.327	248	.508	.359	.237
	(4, 50)	.327	246	.507	.359	.237
	(2, 100)	.323	265	.508	.356	.232

Table 5. Link prediction results for WN18RR

Model	Shape	MRR	MR	Hits@10	Hits@3	Hits@1
DistMult [18]	–	.43	5110	.49	.44	.39
ComplEx [15]	–	.44	5261	.51	.46	.41
ConvE [5]	–	.43	4766	.51	.44	.39
ConvRE (Proposed Method)						
	(50, 4)	.427	4969	.502	.438	.393
	(40, 5)	.427	4791	.499	.440	.393
	(25, 8)	.428	4883	.498	.440	.394
	(20, 10)	.430	4971	.501	.442	.397
	(10, 20)	.431	5083	.501	.444	.397
	(8, 25)	.431	4953	.506	.444	.397
	(5, 40)	.433	5139	.505	.444	.399
	(4, 50)	.433	4754	.508	.443	.399
	(2, 100)	.446	4712	.516	.460	.412

Both the size of the hidden layers (trainable parameters in the hidden layers) in the ConvRE model (i.e., hidden size) and the degree of feature interaction may affect the effectiveness of the model for link prediction. A larger hidden size and more feature interaction benefit the model's performance. The abnormal results for shape (2, 100) are mainly due to the decrease in hidden size, which severely harms the performance of the model and offsets the positive effect of feature interaction on the model's expressiveness.

The default hidden size in the ConvE model is 9728, associated with the embedding shape (20, 10). The hidden size is 10368 for the embedding shape (10, 20), 10304 for (8, 25), 9728 for (5, 40), 9216 for (4, 50), and 6272 for (2, 100). All the hidden sizes listed above, except the last one, are within 6.6% of the default hidden size. The differences between these hidden sizes are relatively small and thus have only a limited impact on the model's performance. This makes the model's performance mostly dependent on the interaction length. Our analysis aligns with Figure 2, where a positive linear relationship appears when the hidden size stays roughly constant.

However, in the case of (2, 100), the hidden size is 35.5% smaller than the default. This significant drop in hidden size severely reduces the number of parameters that could be trained, harms the performance of the model, and diminishes the effects of feature interactions. Therefore, we can see a sudden decrease in MRR at an interaction length of 100 in Figure 2.

5.3 Effects of Feature Interaction on WN18RR

Similarly, we find that increasing the interaction length improves performance. According to Figure 3, there is a roughly

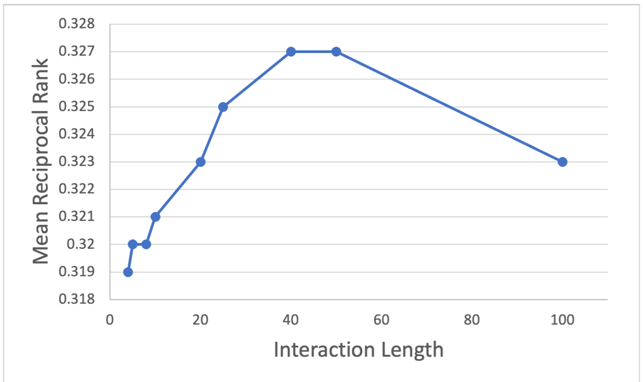

Figure 2. Performance on the data of FB15k-237 with different interaction lengths

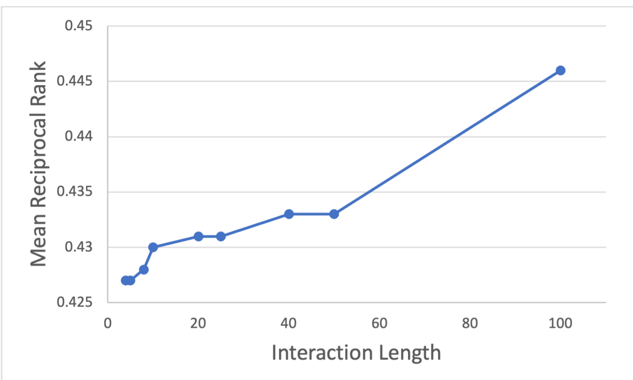

Figure 3. Performance on the data of WN18RR with different interaction lengths

positive linear relationship between the interaction length of the embeddings and MRR.

Unlike FB15k-237, the results for WN18RR do not contain any abnormal values. This is because, according to the statistics of the datasets (see Table 3), the training set for WN18RR contains 86,835 triples, which is 68.1% smaller than the size of the FB15K-237 dataset's training set. This means the number of trainable parameters required for WN18RR should be smaller than that needed for FB15K-237, i.e., the hidden layer size of 6272 under the (2, 100) reshaping setting is sufficient for the smaller dataset, WN18RR. Therefore, when the number of parameters associated with hidden layers diminishes, although the performance of the model on FB15K-237 worsens, the results on WN18RR remain relatively unaffected.

6 Conclusion

In this paper, we propose ConvRE, a novel KGE model that improves the performance of ConvE by reshaping the embeddings of entities and relations. We offer a quantitative definition of feature interaction, named interaction length. Through experiments, we provide empirical evidence that by increasing feature interactions, the performance of link prediction on various datasets can be consistently improved. We suggest that future work could further explore novel approaches for increasing feature interactions to complete the existing knowledge graphs.

References

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. 2008. Freebase. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data - SIGMOD '08*. ACM, New York, NY, USA. https://doi.org/10.1145/1376616.1376746

[2] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*. 1–9.

[3] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes. 2013. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th International Conference on Semantic Systems - I-SEMANTICS '13*. ACM, New York, NY, USA. https://doi.org/10.1145/2506182.2506198

[4] J. Dalton, L. Dietz, and J. Allan. 2014. Entity query feature expansion using knowledge base links. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, New York, NY, USA. https://doi.org/10.1145/2600428.2609628

[5] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. 2018. Convolutional 2d Knowledge Graph Embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. AAAI. https://doi.org/10.1609/aaai.v32i1.11573

[6] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA. https://doi.org/10.1145/2623330.2623623

[7] Y. Hao, Y. Zhang, K. Liu, S. He, Z. Liu, H. Wu, and J. Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Stroudsburg, PA, USA. https://doi.org/10.18653/v1/p17-1021

[8] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld. 2011. Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Stroudsburg, PA, USA, 541–550.

[9] G. A. Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41. https://doi.org/10.1145/219717.219748

[10] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung. 2018. A novel embedding model for knowledge base completion based on Convolutional Neural Network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. ACL. https://doi.org/10.18653/v1/n18-2053

[11] M. Nickel, L. Rosasco, and T. Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30. AAAI. https://doi.org/10.1609/aaai.v30i1.10314

[12] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo. 2021. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis. *ACM Transactions on Knowledge Discovery from Data* 15, 2 (2021), 1–49. https://doi.org/10.1145/3424672

[13] F. M. Suchanek, G. Kasneci, and G. Weikum. 2007. Yago: A Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web - WWW '07*. ACM, New York, NY, USA. https://doi.org/10.1145/1242572.1242667

[14] K. Toutanova and D. Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality*. ACL. https://doi.org/10.18653/v1/w15-4007

[15] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *ICML 2016*. https://doi.org/10.48550/arXiv.1606.06357

[16] S. Vashishth, S. Sanyal, V. Nitin, N. Agrawal, and P. Talukdar. 2020. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. AAAI, 3009–3016. https://doi.org/10.1609/aaai.v34i03.5694

[17] Z. Wang, J. Zhang, J. Feng, and Z. Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 28. AAAI. https://doi.org/10.1609/aaai.v28i1.8870

[18] B. Yang, W.-tau Yih, X. He, J. Gao, and L. Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *ICLR 2015*. https://doi.org/10.48550/arXiv.1412.6575

[19] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma. 2016. Collaborative knowledge base embedding for Recommender Systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA. https://doi.org/10.1145/2939672.2939673

[20] Z. Zhang, F. Zhuang, H. Zhu, Z. Shi, H. Xiong, and Q. He. 2020. Relational graph neural network with hierarchical attention for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. AAAI, 9612–9619. https://doi.org/10.1609/aaai.v34i05.6508