Predicting Relations between RDF Entities by Multi-Order Interaction Neural Network

Qiang Gao, Fei Guo, Yunjie Wu, Xiaowang Zhang⋆, and Zhiyong Feng

College of Intelligence and Computing, Tianjin University, Tianjin, China
Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
⋆ Corresponding Author

Abstract. In this paper, we present a multi-order interaction neural network (MINN) for relation prediction, which can calculate the 2nd-order and 3rd-order feature interactions and automatically combine them within linear complexity. The proposed MINN contains three layers: embedding, multi-order interaction pooling (MI-Pooling), and Deep & Wide layer. In embedding layer, we convert sparse input features into dense representations to narrow the size of features. In MI-Pooling layer, the 2nd-order and 3rd-order feature interactions are calculated and combined in linear time. The results of MI-Pooling will be input into a deep neural network (DNN) for learning nonlinear feature in the last Deep & Wide layer, which will output the final prediction value. The experiments evaluating on two well-known datasets, WN18 and FB15k, show that MINN efficiently performs better than most of the state-of-the-art sparse models in relation prediction.

1 Introduction

In general, some relations among entities are uncertain in the semantic web due to information missing. It's meaningful and interesting to predict accurately if a particular and unknown relation exists between two entities. For example, we may want to kown whether two people are brothers.

An RDF triple in semantic web can be represented as (entity1, relation, entity2). When given such a triple, the goal of this paper is to predict if it is valid. This task can be considered essentially as sparse prediction when a few relations exist among a large number of entities. There are many researches about sparse prediction recently. Factorization Machine (FM) is a classical method by using the 2nd-order feature interactions [1]. The Wide&Deep Learning is presented to capture high-level nonlinear features [2]. DeepFM and Neural FM (NFM) are two excellent methods that combined 2nd-order feature interaction with DNN [3, 4]. InteractionNN is good at learning multilevel hidden features, which makes it perform very well on sparse prediction [5]. The higher-order feature interactions that contain more complex and crucial information is usually ignored since they would bring the cost of higher computational complexity. And it's difficult for DNN to learn more higher-order nonlinear features. In this paper, we propose MINN by combining 2nd-order and 3rd-order feature interaction while the complexity remains linear.

2 Muti-Order Interaction Neural Network

Each element of a triple is treated as a feature. We transform it into vector represent (x_1, x_2, x_3) as the input of MINN by one-hot encoder, where $x_i \in [1...size\ of\ feature_i]$. The MINN will output the existence probability of the triple finally. MINN contains three main modules: embedding layer, MI-pooling, Deep & Wide layer. The architecture of model is shown in Fig. 1.

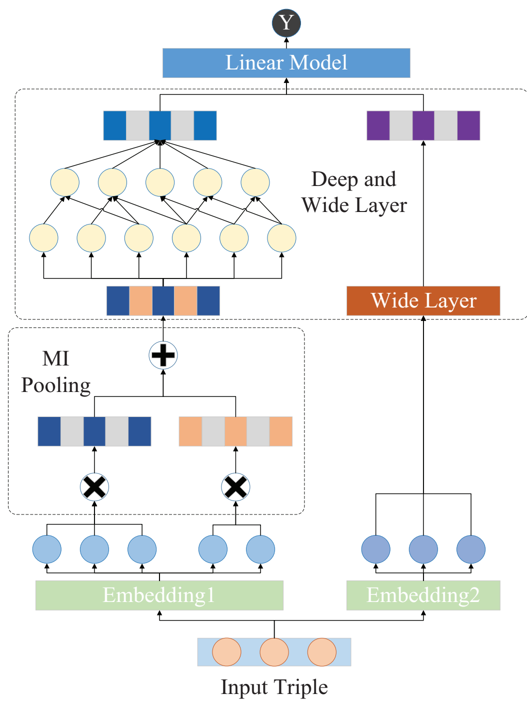

Fig. 1. The architecture of MINN

Embedding layer convert the sparse one-hot vector into dense representation, which can learn the initial information and reduce computation. Formally, let an input vector be $X_{in} = (x_1, x_2, x_3)$ and embedding matrix $vec_{em} \in R^{3\times K}$, where K is the embedding size. We can get dense embedding feature $x_{emb} \in R^K$:

$$x_{em} = X_{in} \cdot vec_{em} = (xv_1, xv_2, xv_3) \tag{1}$$

MI-pooling is employed for extracting the 2nd-order and 3rd-order feature interactions from embedding features firstly. The former can be calculated by formula (2).

$$f_{SO}(x_{em}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} xv_i \cdot xv_j = \frac{1}{2} \left[\left(\sum_{i=1}^{n} xv_i \right)^2 - \sum_{i=1}^{n} (xv_i)^2 \right] \tag{2}$$

Similarly, the calculation of the 3rd-order interaction can be simplified as (3).

$$f_{TO}(x_{em}) = -\frac{1}{3} \left[\left(\sum_{i=1}^{n} xv_i \right)^3 - \sum_{i=1}^{n} (xv_i)^3 - 3 \sum_{i=1}^{n} \cdot (f_{SO}(x_{em})) \right] \tag{3}$$

The complexities of both formulas are $O(n)$. The two interactions are combined by weighted summation to learn muti-order features. The weights can be learned by the model so that the MINN could automatically select more important information from 2nd-order and 3rd-order interactions.

Finally, we employ a linear model to obtain the output in deep & wide layer, where the deep part can learn the high-level non-linear features and the wide can learn the linear and original features.

3 Experiments and Evaluation

We evaluate the performance of MINN on two well-known datasets, WN18 and FB15k. Since negative samples are required for training models, we need to generate negative datasets. We generate a new sample by random replacing one entity of a triple, and if the sample is not in the datasets, it is a valid negative sample. The ratio of positive and negative samples is 1:1. We randomly divide the datasets into three parts: 70% (training), 20% (validation), and 10% (test).

We compare MINN with the state-of-the-art sparse prediction models: FM [1], DeepFM [3], NFM [4] and InteractionNN [5]. RMSE(Root Mean Square Error) and AUC(Area Under ROC) are employed as the evaluation metrics. We set the batch size and embedding size to 256 for MINN on both dataset, and the initial learing is 0.05. In addition, we employed early stopping to avoid overfitting. Table 1 shows the performance of different models.

According to the experimental results, the following conclusions can be drawn. Firstly, the FM learns only 2nd-order features interaction, which performs worst. This demonstrates that 2nd-order features interaction is not enough to express the information of input features. Secondly, the DeepFM combines FM and DNN in parallel, and the performance of DeepFM is bettter than FM but inferior to the others. This shows that high-level non-linear features are necessary and low-order features play an important role when learning the high-level nonlinear feature. Finally, comparing to NFM and InteractionNN, the MINN combines the 2nd-order and 3rd-order feature interactions for learning nonlinear feature. The better performance demonstrates that the higher-order features have improved

Table 1. The RMSE and AUC of different models.

Model	WN18		FB15k	
	RMSE	AUC	RMSE	AUC
FM	0.5589	0.9786	0.4457	0.9865
DeepFM	0.5301	0.9805	0.4401	0.9914
NFM	0.4194	0.9827	0.3086	0.9926
InteractionNN	0.4175	0.9833	0.3105	0.9930
MINN	**0.4101**	**0.9832**	**0.2989**	**0.9932**

the performance of the model. However, the 3rd-order interactions of the RDF triples are not enough, that is, there is only one 3rd-order combination $xv_1xv_2xv_3$ that contains 3rd-order interaction information. We will take more auxiliary semantic information into considering to exhibit the better performance.

4 Conclusion

In this paper, we present a novel neural network MINN for relation prediction by reducing to sparse prediction with higher-order feature interactions to extract more semantics. Our proposed MINN provides a linear computation of 2nd-order and 3rd-order feature interactions. The linear interaction method exactly make higher-order feature interactions feasiable in relation predication. We believe that our proposal is also helpful to other sparse predication. In the future work, we will consider more datasets with more features in our experiments.

5 Acknowledgments

This work is supported by the National Key Research and Development Program of China (2017YFC0908401) and the National Natural Science Foundation of China (61972455, 61672377). Xiaowang Zhang is supported by the Peiyang Young Scholars in Tianjin University (2019XRX-0032).

References

1. Rendle S.: Factorization machines with LIBFM libfm. *ACM TIST*, 3(3), 1–22 (2012).
2. Cheng H. T., Koc L., Harmsen J., Shaked T., Chandra T., Aradhye H., Anderson G., Corrado G., Chai W., Ispir M., Anil R., Haque Z., Hong L., Jain V., Liu X., Shah H.: Wide & deep learning for recommender systems. In: *Proc. of RecSys 2016*, pp. 7–10.
3. Guo H., Tang R., Ye Y., Li Z., He X.: DeepFM: A factorization-machine based neural network for CTR prediction. In: *Proc. of IJCAI 2017*, pp. 1725–1731.
4. He X., Chua T. S.: Neural factorization machines for sparse predictive analytics. In: *Proc. of SIGIR 2017*, pp. 355–364.
5. Zhang X., Guo Q., Feng, Z.: InteractionNN: A neural network for learning hidden features in sparse prediction. In: *Proc. of IJCAI 2019*, pp. 4334–4340.