Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 1
AIEA 2023

Intelligent Sentencing Assistance System based on BERT Model

Yangli Zhu and Peimeng Luo *

Shenzhen University Law School, Shenzhen, China

* Corresponding author Email: peimengluo@outlook.com

Abstract. This paper takes sentencing prediction as the research object, uses deep learning model, and predicts sentencing recommendations combined with specific cases. BERT+DNN model is used to quantify the text of the case information, endow the sentence with certain meaning, and then send data to DNN network to train the classification model so that it can automatically identify the sentencing circumstances in the text of different cases, and put forward sentencing recommendations based on various circumstances.

Keywords: Intelligent Sentencing Assistance System; BERT Model; Sentencing Prediction; BERT+DNN Model.

1. Introduction

In the context of the rapid development of information technology, the use of artificial intelligence to assist sentencing is an inevitable requirement to further promote the standardization of sentencing.

The author uses the BERT model to learn the case text, making full use of the multi-level multi-headed self-attention mechanism to serve the specific case and complete the intelligent sentencing prediction function. [1]

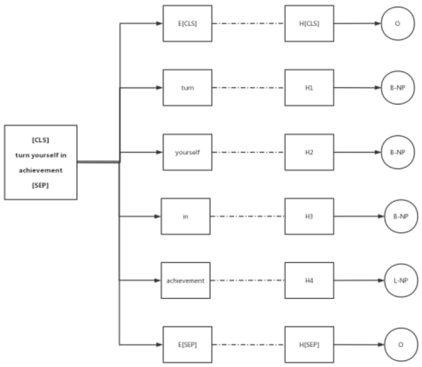

Figure 1. BERT model

2. Analysis of Applying the BERT Model to Sentencing Prediction

2.1. Inputs and Outputs

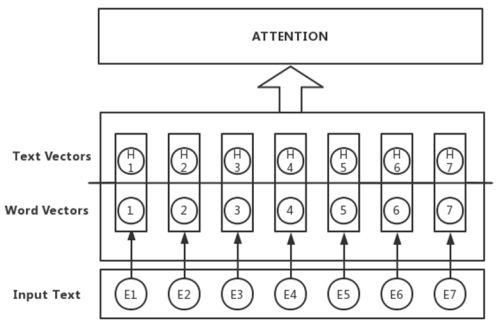

Figure 2. Inputs and outputs of BERT

The input of the BERT model is the original one-dimensional vector of each word or phrase; the output is the vector representation of each word or phrase in the text. In the process of machine learning, a large number of sentencing circumstances or elements in the case need to be marked, so that the machine can correctly identify the relevant judgment factors during the training process, and then continuously generate and continuously modify the algorithm for sentencing.

2.2. Transformer Coding Layer

The Attention mechanism obtains the Q vector representation of the target word, the K vector representation of each word of the context, and the original V representation of the target word and each word of the context by linear transformation, then calculates the similarity between the Q vector and each K vector as weights, and weighted fusion of the V vector of the target word and the V vector of each word of the context as Attention's output. The calculation formula is shown in Fig. [2]

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{SoftMax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}$$

Figure 3. Attention span calculation formula

To enhance the diversity of Attention, different Self-Attention modules can be further used to obtain the enhanced semantic vectors of each word in the text in different semantic spaces, and multiple enhanced semantic vectors of each word can be linearly combined to obtain a final enhanced semantic vector of the same length as the original word vector, i.e., Multi-head Self-Attention. [3]

On the basis of Multi-headSelf-Attention, Residual Connection is added to directly add the input and output of the module as the final output, and make 0 mean 1 variance normalization of a layer of neural network nodes and then make linear transformation. [4]

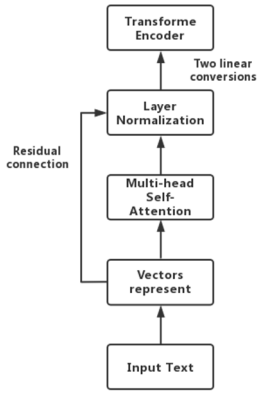

Figure 4. The structure of Transformer

2.3. DNN Classification

A DNN is a deep neural network, which is a neural network that contains an input layer, a hidden layer and an output layer. The network performs complex operations through its complex operations by its own neurons to integrate the relationship between input data and output data. New data is entered into the model, and calculations based on the parameters already stored can be automatically performed and down parameters for calculation can be automatically classified. [5] In order to detect the correctness of AI-assisted sentencing judgments, the "degree of deviation" is often used as an evaluation criterion, and the results of AI-assisted sentencing judgments are compared with the "correct" judgments to verify the deviation from the "correct answer" in terms of sentencing range. The degree of deviation from the "correct answer".[6]

In order to elucidate the applicability of DNN to the legal domain, a randomly generated dataset of the number of occurrences of five different crimes in five different cities over the same period of time is now generated through a Gaussian distribution.

	crime A	crime B	crime C	crime D	crime E
city NO. 1	73	83	83	71	92
city NO. 2	94	75	97	71	93
city NO. 3	75	93	100	94	88
city NO. 4	80	78	73	74	70
city NO. 5	71	80	87	70	70

Figure 5. a Randomly Generated Dataset

First we start by putting 3 neurons as input layer as shown below:

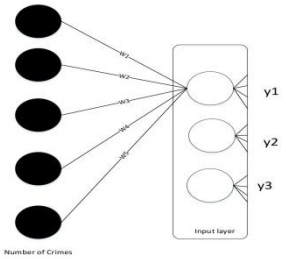

Figure 6. Input Layer

Then all the inputs of number of crimes are connected to the first neuron to get y1. To summarize, y1=w1*x1+w2*x2+w3*x3+w4*x4+w5*x5,using the same calculation we can get y1,y2,y3.

By multiplying and adding the above calculations we can get a linear calculation, the next step is to make this model nonlinear by using an excitation function, which is used to solve nonlinear problems in conjunction with the legal field.

After the nonlinear process, y can basically be taken as the understanding of this neuron for the input counseling, each neuron in the input layer receives the number of crimes in different cities in the same period of time, plus the weighted neurons are shown in the figure below.

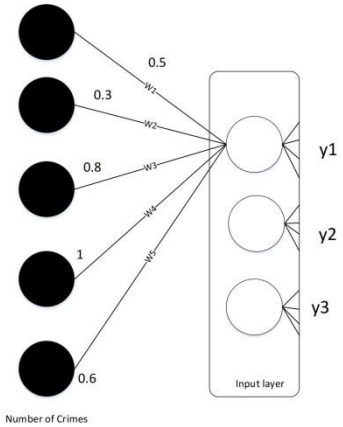

Figure 7. Input Layer(Weighted Calculations)

This represents how much importance neurons attach to the number of crimes in different cities, each neuron gets different perceptions through different weighting matrices, each of which is completely randomized, which allows each neuron to have a different weighting tendency in the beginning. DNN classification analyzes the crime rates in different cities through these neurons with different tendencies, and then comes up with the results.

Because the neurons sum up all the inputs, the neurons in the next layer also sum up the outputs of the previous layer, meaning that each neuron is connected to all the neurons in the previous layer, this form of neural network is also known as a fully-connected neural network, as shown in the figure below.

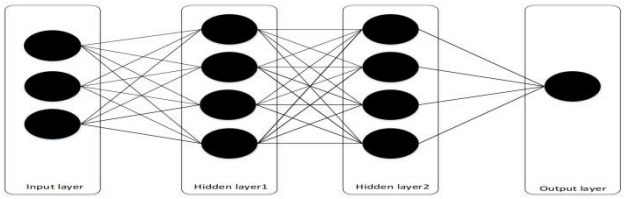

Figure 8. Input Layer and Output Layer

3. Analysis and Conclusion

Sentencing is a mathematical question about legal rules and discretion, the intelligent sentencing assistance system is to map the legal rules in the code and algorithm for artificial intelligence technology to operate according to the legal rules and get accurate results based on the legal circumstances entered, through fine code construction, the traditional repetitive manual calculation process is simplified, and accurate results are obtained in a short time to improve judicial work efficiency and promote the development of judicial intelligence.

Acknowledgments

This thesis is funded by the following projects:

1. Research on the Application of Competency Model Theory in Legal Education for Digital Society (Guangdong Provincial Education Science Planning Leading Group Office Project No. 2021 GXJK 231) Instructor Zhang Xianwei.

2. Research on the integration of intellectual property law teaching with ideological elements of the curriculum (Department of Academic Affairs, Shenzhen University, Project No. 000030110214) Instructor Zhang Xianwei.

References

[1] Hong Wenxing, Hu Zhiqiang, Weng Yang. et al. Automatic construction of case knowledge graphs for judicial cases[J]. Journal of Chinese Information,2020,34(01):34-44.

[2] Qiu Liwei, Guan Weili, Zhang Weijin. An exploration of the BERT language model [J]. Computer Programming Skills and Maintenance,2021, No.427(01):21-23.

[3] Illustrated BERT Model: Building BERT from Scratch - Tencent Cloud Developer Community - Tencent Cloud (tencent.com).

[4] Fan Aman, Wang Yanchuan. An intelligent adjudication method for multitasking legal cases based on BERT model[J]. Microelectronics and Computers,2022,39(09):107-114.

[5] Jiayi Gu. Research on policy condition identification based on BERT model[J]. Science and Technology Perspectives,2020, No.301(07):251-252.

[6] Cheng Long. The problem and way out of artificial intelligence-assisted sentencing[J]. Journal of Northwestern University (Philosophy and Social Science Edition),2021,51(06):163-174.