Effective Deep Learning Through Bidirectional Reading on Masked Language Model

Hiroyuki Nishimoto[1],

[1] *Kochi university*

Nankoku, Kochi 783-8505, Japan

ABSTRACT

Google BERT is a neural network that is good at natural language processing. It has two major strategies. One is “Masked Language Model” to clear the word-level relationships, and the other is “Next Sentence Prediction” to clear sentence-level relationships. In the masked language model, with the task of masking some words in sentences, BERT learns to predict the original word from context. Some questions come to mind. Why BERT achieves effective learning by reading in two ways from fore and back? What is the difference between bidirectional reading? BERT learns to predict the original word using the surrounding words as context and to make two-way predictions by forward and backward readings in order to increase the precision. Besides, the bidirectional reading technique can be applied to scenario planning especially using back-casting from the future. This paper clarifies these mechanisms.

Keywords: Masked language model, Bidirectional reading, Future thinking, Back-casting from the future

AI IS CREATED BY IMITATING THE HUMAN BRAIN

Google BERT

Google BERT (Jacob Devlin et al 2019), Bidirectional Encoder Representations from Transformers, specializes in Natural Language Processing. Human-speaking languages are called natural languages as opposed to computer languages.

BERT has two major strategies. One is “Masked Language Model” to clear the word-level relationships, and the other is “Next Sentence Prediction” to clear sentence-level relationships. In the Masked Language Model with the task of masking some words in sentences, BERT learns to predict the original word from context.

Some questions come to mind. How does the context work? Why BERT achieves effective learning by reading in two ways from fore and back? What is the difference between bidirectional reading from fore and back? Each answer lies in human thinking mechanism because AI is created by imitating the human brain.

How does the context work?

In the case of humans, we understand meaning in context. How does the context work? This mechanism is illustrated in Figure 1. For example, what does “study” mean? In the context of high school as shown in the left side, “study” means “learning”. On the other hand, in the context of healthcare as shown in the right side, “study” means “clinical trials”. The context determines its meaning, but not the word (Hiroyuki Nishimoto. et al. 2019).

Neural networks use the surrounding words as context?

Instead of the context, neural networks use the surrounding words as context. How do the surrounding words work with the central word? For example, what is the meaning of the central word of “diabetes”? If the surrounding words are “he developed”, it indicates own medical history. If the surrounding words are “his mother”, it means family medical history. AI, such as BERT, uses the surrounding words as context, to guess the meaning of the central word.

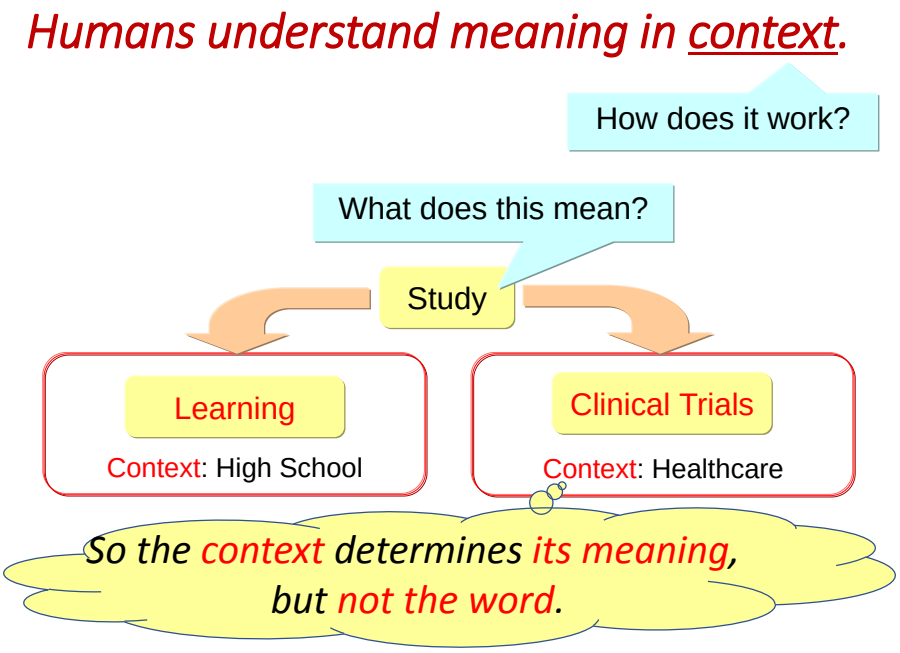

Figure 1. Semantic illustration of human context

Effective deep learning through bidirectional reading

Google BERT has two major strategies. One is “Masked Language Model” In the masked language model with the task of masking some words in sentences, BERT learns to predict the original word from context. In addition, masked language model provides two-way reading from fore and back. And that raises the question of what is the difference between two-way reading. What does BERT learn from the masked language model? This mechanism is illustrated in Figure 2. For example, the middle sentence is “I ate [mask] every morning”. The candidates are an apple and beef steak.

First, predict the masked word by forward reading, focusing on the middle sentence and the previous sentence “Considering my health, I decided to change the breakfast menu”. What is asked in general? We usually think about which is more realistic and reach “an apple”. The answer is “feasibility”.

Next, predict the masked word by backward reading, focusing on the middle sentence and the post sentence “A month later, I lost 3 kg and became healthy.” What is asked in general? In such a situation, we are looking for success factors. We usually think about which is more relevant and reach “an apple”. The answer is "success factors".

Therefore, BERT is learning to predict the feasibility by forward reading and the success factors by backward reading.

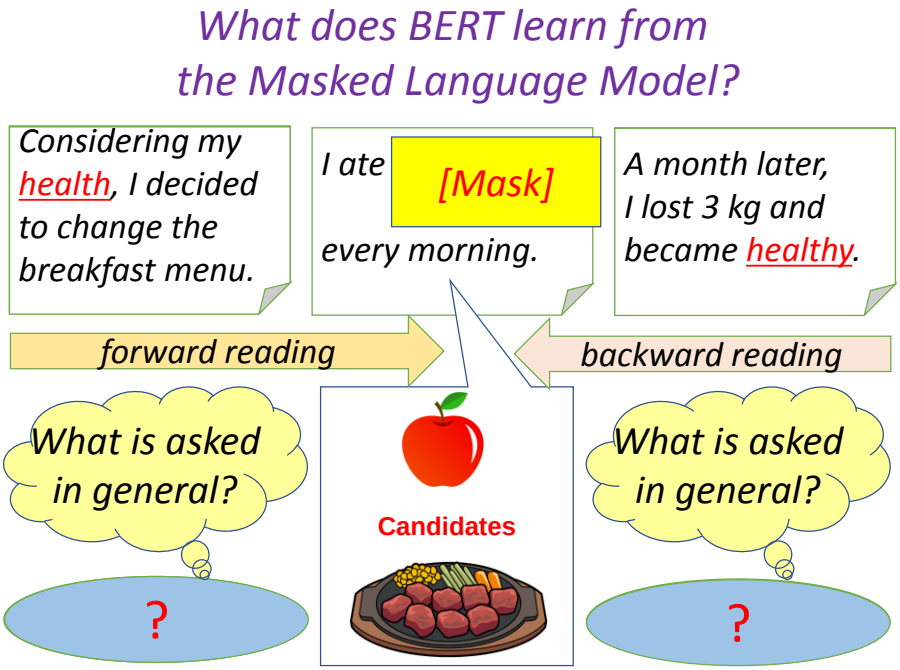

Figure 2. Semantic illustration of Masked Language Model

Bidirectional reading technique can be applied to scenario planning

Scenario planning using back-casting from the future

As described above, the bidirectional reading technique is excellent. It can be applied to others. One is scenario planning that is making assumptions on what the future is going to be. Since a scenario can be described in two ways, one is fore-casting and the other is back-casting.

As shown in Figure 3, fore-casting means viewing from the present to the future. In general, back-casting means viewing from the present to the past. But in this paper, it means viewing from the future to the present (Raja R. Timilsina, Yoshinori Nakagawa. 2020). Just as there are two different predictions for bidirectional reading, there is a big difference between fore-casting into the future and back-casting from the future.

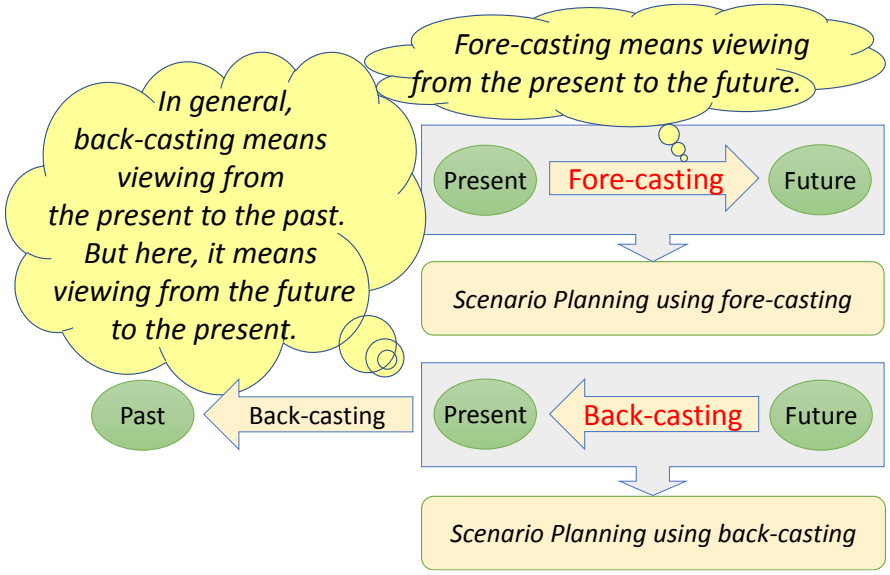

Figure 3. Disposition of bidirectional viewpoints on scenario planning

Scenarios are better used to discuss how to deal with structural changes

The purpose of scenario planning is making assumptions on what the future is going to be. Scenarios are better used to discuss how to deal with structural changes than to evaluate how to deal with this scenario of structural changes, what to do now, and so on. As shown in Figure 4, destructive innovations have been changing the social structure. Structural changes have been witnessed with major technological innovations, such as the internet technologies, and the deep neural networks.

Combination of EMR and PHR makes a structural change

The future is unpredictable. But we can find signs of change at least. For ex-ample, in the healthcare development, one sign is telemedicine, which is accelerating the demand for home medical devices. This will be a game changer in healthcare development because the device development is less risky and costly than new drug development. As shown in Figure 5, Electronic Medical Record (EMR) records patient visit da-ta in hospitals when the patient is ill. On the other hand, home medical devices generate data called PHR (Personal Health Record). In essence, the PHR can cover daily data collected by home medical or wearable devices. Ideally, the combination of EMR and PHR enables lifelong data analysis and preventive analysis, that covers healthy data.

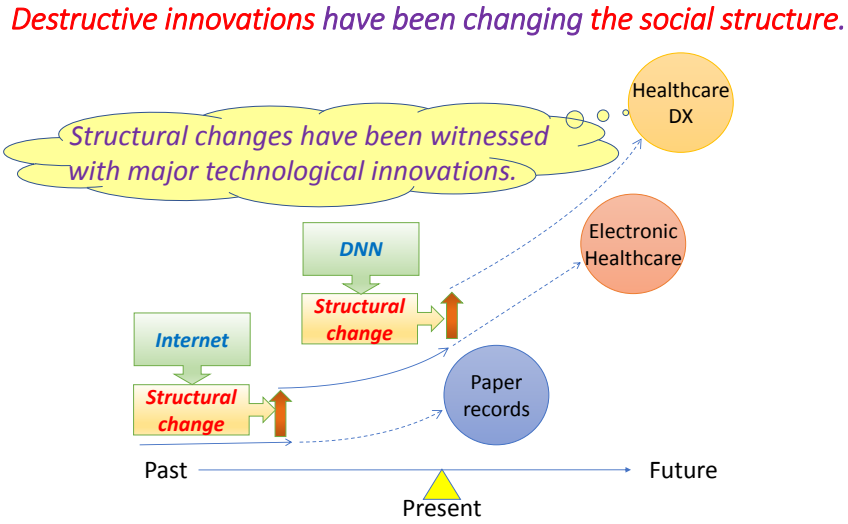

Figure 4. Destructive innovations changing the social structure

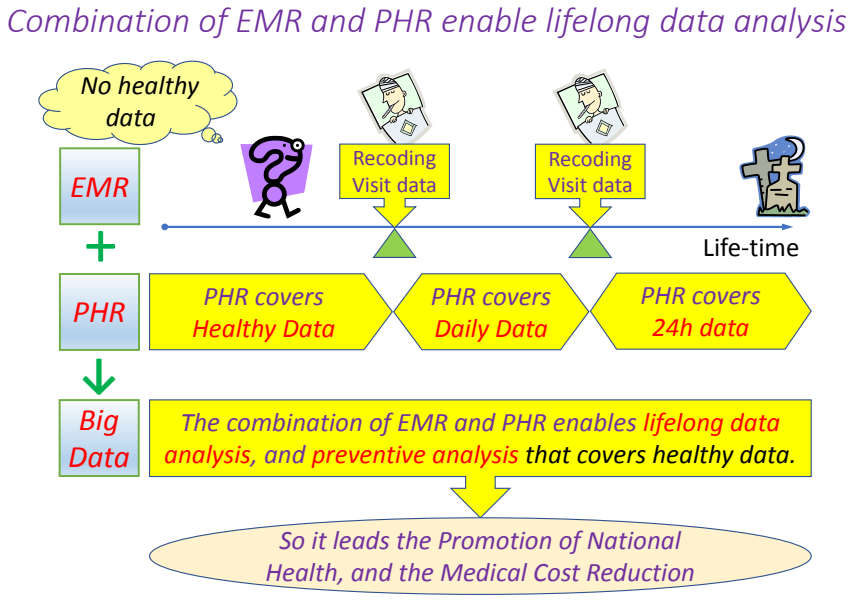

Figure 5. Lifelong data analysis combining EMR and PHR

Scenario Planning in two ways

The above-mentioned scenario in terms of the combination of EMR and PHR can be described in two ways using fore-casting and back-casting as follows:

1. Scenario Planning using fore-casting from the present to the future

Rapid aging is a fundamental problem in Japan. In addition, the COVID-19 pandemic is causing further confusion. The need for telemedicine to resolve them may be increasing. With the promotion of home medical care, there is a possibility that the development of home medical devices is also progressing. It has the potential to be more advanced than drug development due to its low development costs and health risks. And more home medical devices could be used for lifelong data and disease analysis. The device may store many personal health records (PHRs), which are expected to be integrated into electronic medical records (EMRs) and big data. This data might be used in precision medicine and contribute to extending the healthy lifespan of people. Finally, the promotion of National Health Insurance might be expected to achieve continued reductions in medical costs.

2. Scenario Planning using back-casting from the future to the present

Rapid aging was a fundamental problem in Japan. In addition, the COVID-19 pandemic caused a lot of confusion. There was a growing need for telemedicine to resolve them. With the promotion of home medical care, the development of home medical devices has progressed. The development of home medical devices is ahead of drug development due to lower development costs and reduced health risks. And more home medical devices were used for lifelong data and disease analysis. Many personal health records (PHRs) are stored on the device and merged with electronic medical records (EMRs) into big data. This data is used in precision medicine and contributes to extending the healthy life expectancy of people. Finally, the promotion of National Health Insurance has continuously reduced medical costs.

Did you notice the difference? How do you feel about the first scenario using fore-casting? You will find that you tend to focus on feasibility. It starts a long debate about the feasibility, such as “Is it possible?”, “Is it difficult?”, and “How to achieve?”.

On the other hand, how do you feel about the second scenario using back-casting from the future to the present? This scenario has to be written in past mode because of back-casting. If it is written in past mode, you feel it has done and someone has resolved all the problems by that time. The surrounding words in past mode change your feeling from prediction to event context. You will find that you tend to focus on success factors. You can escape from the long debate.

CONCLUSIONS

Google BERT is a neural network that specializes in natural language processing. It has two major strategies. One is “Masked Language Model” to clear the word-level relationships, and the other is “Next Sentence Prediction” to clear sentence-level relationships. In the masked language model, with the task of masking some words in sentences, BERT learns to predict the original word from context with bidirectional reading. It has a big difference between forward and backward readings. BERT is learning to predict the feasibility by forward reading and the success factors by backward reading. Besides, the bidirectional reading technique can be applied to scenario planning using back-casting from the future to the present. This scenario has to be written in past mode because of back-casting. If it is written in past mode, you feel it has done and someone has resolved all the problems by that time. The surrounding words in past mode change your feeling from prediction to event context. You tend to focus on causal factors of success. You can escape from the long debate.

Scenario planning using back-casting from the future to the present makes a good proposition, avoiding long discussions. Besides, in terms of the mystery of deep learning, each answer lies in human thinking mechanism because AI is created by imitating the human brain.

REFERENCES

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). “BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding.” Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).

Hiroyuki Nishimoto, Tomoyoshi Koyanagi, Makoto Sarata, Ayae Kinoshita and Mitsukazu Okuda. (2019). "Memes" UX-Design methodology based on cognitive science re-garding Instrumental Activities of Daily Living, Human Computer Interaction 2019, LNCS 11582.

Raja R. Timilsina, Yoshinori Nakagawa, and Koji Kotani. (2020). Exploring the Possibility of Linking and Incorporating Future Design in Backcasting and Scenario Planning, Sus-tainability, 12(23), 9907.