

Attention is all you need

Ashish Vaswani Noam Shazeer Niki Parmar

<https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>

Abstract

- Attention is a breakthrough in natural language processing and is essential to understanding cutting-edge technologies such as GPT.
- Attention is a mechanism that determines which words in a sentence are important.
- This solve the problem of the conventional model's lower accuracy with long sentences.

Example of natural language processing

Please see the text below.

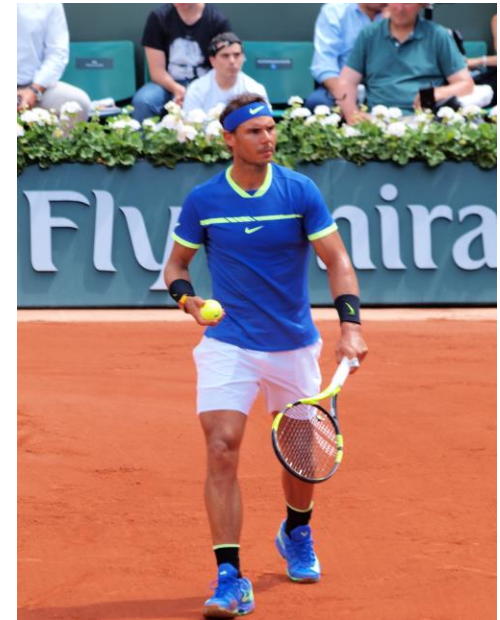
"My parents make somen, so I used to eat somen every day, but now I live in Aizu, so I can't eat somen."

When we see a sentence like this, we focus on the important words and understand them.

"My parents make **somen**, so I used to eat somen **every day**, but now I live in **Aizu**, so I **can't eat** somen."

Example of image processing

- Humans unconsciously perform this kind of processing, and **attention is what applies this to AI**.
- It can also be applied to image processing, such as extracting important objects.
- For example, in the photo on the right, we can focus on the man, the racket, and the ball.



Conventional vs Attention

- Conventional methods process entire sentences.
- It could process short sentences correctly, but **the accuracy decreased for long sentences.**
- Specifically, the accuracy decreases rapidly when the number of words exceeds 30.
- On the other hand, models using Attention can highlight important words and accurately translate even long sentences.

Achievements of this paper

- The reason Attention is said to be a breakthrough is because it is an important mechanism that connects to libraries such as GPT-3.
- Attention was announced in 2015, but natural language processing has advanced rapidly since then.
- Two years later, Transformer was released, based on Attention.

Are attention mechanism enough?

- A new architecture called Transformer successfully developed high-quality models without using any recursion or convolutional neural networks.
- Furthermore, it can be parallelized and **training time is significantly reduced.**

Conclusion

- Attention has a positive effect on speed and quality.
- The latest model uses a model with improved attention.
- Attention can be used in many fields such as natural language and image processing.