Reasoning with Transformer-based Models: Deep Learning, but Shallow Reasoning

Chadi Helwe, Chloé Clavel and Fabian Suchanek

Télécom Paris, Institut Polytechnique de Paris

{firstname.lastname}@telecom-paris.fr

Motivation

Transformer-based models do great on many NLP tasks.

But, do they really understand natural language?

This survey paper discusses the performance of transformer-based models on different reasoning tasks.

Known pitfalls for BERT-based models

Negation and Mispriming

Positive Statement: *Marcel Oopa died in the city of [MASK]*
Target Answer: *Paris*
BERT's Top-3 Predictions: *Paris (-2.3)* 😀*, Lausanne (-3.3), Brussels (-3.3)*

Negation

Negative Statement: *Marcel Oopa did not die in the city of [MASK]*
Target Answer: *Any city that is not Paris*
BERT's Top-3 Predictions: *Paris (-2.4)* ☹*, Helsinki (-3.5),Warsaw (-3.5)*

Mispriming

Misprimed Statement: *Yokohama? Marcel Oopa died in the city of [MASK]*
Target Answer: *Paris*
BERT's Top-3 Predictions: *Yokohama (-1.0)* ☹*, Tokyo (-2.5), Paris (-3.0)*

Pattern Heuristics and Word Order

Pattern Heuristic

Statement: *The doctor was paid by the actor → The doctor paid the actor.*
Target Answer: *Non Entailment*
BERT's Prediction: *Entailment* ☹

Word Order

Statement: *Paul loves Real Madrid*
BERT's Prediction: *Yes* 😀

Modified Statement: *Real Madrid loves Paul*
BERT's Prediction: *Yes* ☹

Do transformer-based models have deep reasoning capabilities?

Short answer: NO

Reasoning with Transformer-based Models that Works

The strength of transformer-based models comes from two components: simple patterns in the training data, combined with background knowledge from the pretraining.

Thus, transformer-based models can perform well on tasks such as:

Horn Rule Reasoning

Context: Erin is young. Erin is not kind. If someone is young and not kind then they are big.
Question: Is Erin is big ?
Expected Answer: Yes

Simple Commonsense Reasoning

Context: Ravens can [MASK]
Expected Answer: fly

Simple Mathematical Reasoning

Context: Calculate -841880142.544 + 411127
Expected Answer: -841469015.544

Reasoning with Transformer-based Models that Fails

Transformer-based models fail on tasks where patterns and background-knowledge are absent, e.g.:

Implicit Reasoning

Context: David knows Mr. Zhang’s friend Jack, and Jack knows David’s friend Ms. Lin. Everyone of them who knows Jack has a master’s degree, and everyone of them who knows Ms. Lin is from Shanghai.
Question: Who is from Shanghai and has a master’s degree?
Options: (A) David (B) Jack (C) Mr. Zhang (D) Ms. Lin

Adversarial Commonsense Reasoning

Context: A prindag is smaller than a flurberg, so a flurberg is [MASK] likely to contain a prindag
Expected Answer: more

Mathematical Word Reasoning

Context: Jack had 8 pens and Mary had 5 pens. Mary gave 3 pens to Jack.
Question: How many pens does Jack have now?
Expected Answer: 8 + 3 = 11

Is there any task that transformer-based models cannot solve, even if they are trained on a large dataset?

Short answer: YES

Impossible Reasoning Tasks

Transformer-based models have theoretical limitations. The main limitations come from the fact that self-attention does not have the same level of expressiveness as recurrent models such as LSTMs. In particular, they cannot model two languages: Parity and Dyck-2.

To show the impact of these limitations on natural language we developed two tasks: Light Switch Task and Cake Task.

Even Parity😞

0011 → valid
010 → not valid

Dyck-2 😞

([])[]() → valid
([])[) → not valid

Light Switch Task 😞

Context: The light is off. I operate the light switch, and I eat a pizza, and I eat a pizza.
Is the light on?
Expected Answer: Yes

Cake Task 😞

Context: I make a cake. I add a peanut layer and I eat a chocolate layer.
Is the cake gone?
Expected Answer: No

To Learn More

Paper

Code

https://github.com/dig-team/FailBERT