

Large Language Models

Google is touting its Large Language Model, PaLM, with a claimed 540 billion nodes.

It learns by reading text.

A few problems.

Firstly, complex text – a piece of legislation, a high value contract or a Defence specification – has its own definitions, in a glossary or declared somewhere in the document. Words or phrases defined this way are meant to override the common meanings.

state of mind of a person includes:

- (a) the knowledge, intention, opinion, suspicion, belief or purpose of the person; and
- (b) the person's reasons for the intention, opinion, belief or purpose.

Secondly, complex text relies heavily on bullets, with references to a particular bullet capturing the phrase or clause it links to.

- (b) a regular premium policy to which paragraph (a) does not apply

corporate group has the meaning given by subsection 123(12).

In other words, a complex text document is its own thing, and no amount of reading other texts will help to understand the meanings in it.

LLM doesn't seem aimed at complex text. What sort of text is it aimed at?

“...the idea is that we will try to attack this problem very directly, this problem of few-shot learning, which is this problem of generalizing from little amounts of data. (that is what a dictionary is – might be a good place to start)”

...the main idea in what I'll present is that instead of trying to define what that learning algorithm is by N and use our intuition as to what is the right algorithm for doing few-shot learning, but actually try to learn that algorithm in an end-to-end way.

And that's why we call it learning to learn or I like to call it, meta learning.”

The goal with the few-shot approach is to approximate how humans learn different things and can apply the different bits of knowledge together in order to solve new problems that have never before been encountered.

The advantage then is a machine that can leverage all of the knowledge that it has to solve new problems.

As soon as someone mentions “algorithm” in a discussion of language, it should raise a red flag. The meaning is in the words, and how they connect together.

The obvious question is – Instead of reading random text to learn things, why not read a dictionary? A dictionary encapsulates hundreds of man-years of effort.

A dictionary:

May use other words to help define a word.

Hilarious -> extremely amusing

May use grammatical objects to give a broader definition of a word

Of Sub Sense 1

Definition: followed by a noun expressing the object of the verb underlying the first noun – the owner of the boat

May indicate the purpose of the word is not to add meaning, but emphasis.

(used for emphasis) only to a small extent; not much or often

May indicate the purpose of the word is figurative or hyperbolic.

Barnacle Sense 1

Figurative: Sense 0

Definition: a tenacious person or thing

Freezing

Sub Sense 0

(used hyperbolically) very cold

The other aspect of a dictionary is the sheer number of definitions – “set” has 74, “run” has 48, “on” has 21 as preposition and 7 as adverb.

Another aspect is the number of Parts of Speech a word can have – 313 words have four POS, 38 words have five POS, 2 have six. Very hard to learn these from random text. Without knowing the forms of verb a word can take (in total about 47), extremely difficult to know (in general) whether a word is being used as a noun or verb.

In other words, a dictionary would appear to be the obvious thing to read to gain initial knowledge of a language.

Currently, we have 20,000 files containing definitions from the OED, 70,000 definitions, and 700,000 words in those definitions, supporting a vocabulary of 45,000 words and 15,000 wordgroups (combinations of words – “bank account”). As a rough rule, each word will use ten network elements, so seven million network elements, leaving thirty million elements to handle a particular problem. The system has the ability to look up a word it doesn’t know, and merge the new definitions into its structure. It also has to fabricate definitions for comparative and superlative adjectives and gerunds “on the run”.

Dictionaries have a bad rap about circularity. The machine can carry out instructions in English to clean itself in different ways.

Google has had great success with a huge database using key word searching. Knowing that, it is not surprising that they thought a huge structure would be useful for language. Why would anyone else think that?

ORION Design Note