Lec 18 - Future of NLP + Deep Learning

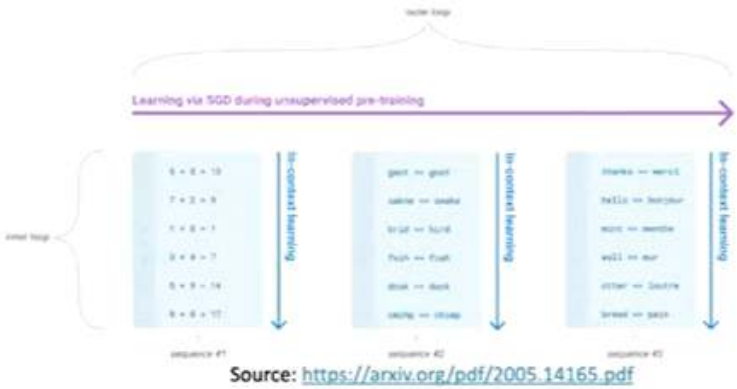
1. Extremely large models and GPT3
-> better than other models at language modeling and related tasks such as story completion
->commands to bash one-liners
->blend concepts i.e. find a word between 2 given words
->copycat analogy problems

GPT-3: Limitations and Open questions

- Seems to do poorly on more structured problems that involve decomposing into atomic / primitive skills:
  - RTE / arithmetic / word problems / analogy making
- Performing permanent knowledge updates interactively is not well studied.
- **Doesn't seem to exhibit human like generalization (systematicity).**
- **Language is situated and GPT-3 is merely learning from text without being exposed to other modalities.**

What's new about GPT-3: Flexible "in-context" learning

- GPT-3 demonstrates some level of fast adaptation to completely new tasks. This happens via "in-context" learning
- The language model training (outer loop) is learning how to learn from the context [inner loop]



24

Source: https://arxiv.org/pdf/2005.14165.pdf

2. Compositional Representations and Systematic Generalization

-1. Are neural representations compositional?

Compositional Representations and Systematic Generalization

Compositionality: closely related to the idea of systematicity is the principle of compositionality.

Rough Definition:

"The meaning of an expression is a function of the meaning of its parts"

More concrete definition (Montague):

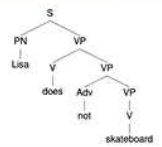
A homomorphism from syntax (structure) to semantics (meaning). That is, meaning of the whole is a function of immediate constituents (as determined by syntax)

-2. Do neural NLP models generalize systematically?

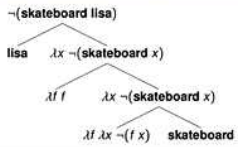
Systematicity and Compositional Generalization: Are neural representations compositional?

- According to Montague, Compositionality is about the existence of a homomorphism from syntax to semantics:

Lisa does not skateboard = ⟨Lisa, ⟨does, ⟨not, skateboard⟩⟩⟩



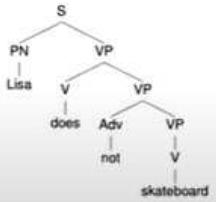
m(Lisa does not skateboard) = ⟨m(Lisa), ⟨m(does), ⟨m(not), m(skateboard)⟩⟩⟩



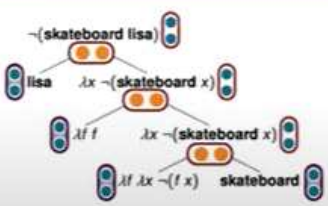
- Tree Reconstruction Error (TRE) [Andreas 2019]: Compositionality of **representations** is about how well the representation approximates an explicitly homomorphic function in *a learnt representation space*

- TRE [Andreas 2019]: Compositionality of representations is about how well the representation approximates an explicitly homomorphic function in a learnt representation space

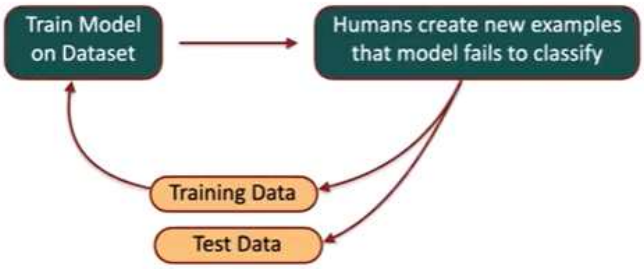
Lisa does not skateboard = ⟨Lisa, ⟨does, ⟨not, skateboard⟩⟩⟩



NN(Lisa does not skateboard) ≈ f(v(Lisa), f(v(does), f(v(not), v(skateboard))))



3. Improving how we evaluate models in NLP

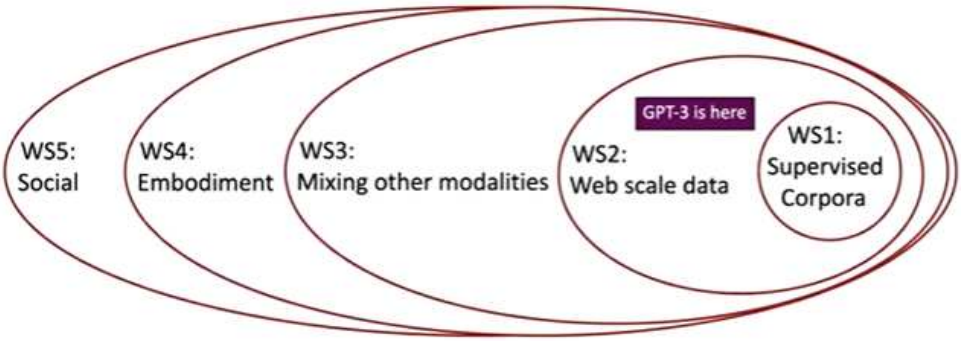


Overview of dynamic benchmarks

4. Grounding language to other modalities

Grounding Language to other modalities

- Many have articulated the need for using modalities other than text
- Bender and Koller [2020]: Impossible to acquire "meaning" (communicative intent of the speaker) from form (text / speech signal) alone
- Bisk et al [2020]: Training on only web-scale data limits the world scope of models.



5. Getting involved with NLP and Deep Learning Research