Hypothesis

Shiro Takagi

2021/4/12

1 question

Q: How to make an intelligent system that learns to manipulate symbols?

2 survey

2.1 cognitive science

Chomsky says that the power to taming syntax is innate property of human brain (universal grammar) [1]. Ibbotson says that "the complexity of language emerges not as a result of a language-specific instinct but through the interaction of cognition and use" (usage-based theory) [7]. According to the usage-based theory, linguistic structure develops by 1. categorization, 2. chunking, 3. rich memory, 4. analogy, and 5. cross-modal association [2, 7]. Children use a limited number of reliable short frames.

" *Overall it seems there is good evidence to support the usage-based prediction that language structure emerges in ontogeny out of experience (viz. use) and when a child uses core usage-based cognitive processes – categorization, analogy, form-meaning mapping, chunking, exemplar/item-based representations – to find and use communicatively meaningful units.* [7] "

The meaning of symbols is established by convention [12, 14, 9].

2.2 transformers

Transformer learns syntactic information [11, 5, 3, 15].

2.3 comparative study

Watson et al. claims that "*nonadjacent dependency processing, a crucial cognitive facilitator of language, is an ancestral trait that evolved at least 40 million years before language itself*" [17]. Wilson et al. explains that sufficient cues play crucial role for human nonadjacent dependency learning [18]. Okanoya and Merker propose the hypothesis that human language is established through string-context mutual segmentation: "*song strings and behavioral contexts are mutually segmented during social interactions*" [10].

2.4 Neo-Vygotskian Theory

The key novelties in human evolution were all, in one way or another, adaptations for an especially cooperative, indeed hypercooperative, way of life [16]. Tomasello states that there are two human-unique capabilities: joint intentionality and collective intentionality. He says that Human developed preparation for these abilities and children first acquire joint intention and then collective intention in their developmental stages.

3 Discussion

3.1 transformers

If transformers really capture syntax, **how it develops the syntactic representation during the pre-training?** Previous studies seem to find that pre-trained transformer have syntactic representation but how to do that remains to be answered. If I can single out the cause of the syntax emergence, I may be able to model a guiding principle for an intelligent agent to learn syntax.

3.2 syntax

It might be plausible that infants first identify phrases in sentences. To identify phrases, it might be necessary that the phrase is used in multiple sentences. By observing the subset of sentences in multiple sentences repeatedly, infants could identify which subset is the the phrase. However, a phrase less likely to appear many times in multiple contexts. Thus, I could hypothesize that infants first understand too common and too often appearing phrase.

Then, they understand that theres is the concept “phrase” in their society. Finally, they could start to generalize their knowledge and to do try and error to manipulate phrase order. In sum, **I hypothesize that artificial intelligence should follow the following path for language acquisition: phrase identification - phrase order arrangement - phrase manipulation.** If this is the case, I should create an environment where identifying key phrase will give reword to the agent. This view is similar to that in [10] which I explained above.

3.3 keyword detection

Dual-coding memory may be a key because visual information is pseudo label there [6]. Even if no instructor exists, the agent can learns the concept him/herself.

3.4 collective intention

A distinct feature of human intelligence is its mastery of languages. On the other hand, Tomasello argues that humans differ from other animals in terms of collective intentionality. If this is the case, **a natural hypothesis from this is that human develop language skills thanks to their collective intentionality.**

3.5 Action and Symbolic Manipulation

Human understand a notion of “research” and know that it includes “science”, for example. Or, I can use a more abstract notion of “object” and apply an “operation” on it. Human beings seem to excel at these symbolic manipulation. I believe that memory is “used” for these operations.

I support the view that the semantics of neural representation is designed through the culture the agent is in [12]. Repeated activations of neurons constructs an abstract concept and the symbols are attached to these concepts and segments the neural representation space. Thus, the abstract concept is formed through energy minimization and semantics is grounded to these abstract through social interactions. I hypothesize that human attach the symbol because it improves the predictability of internal and external states. In other words, I do so because it is useful. I think that the relation is also formed through this process.

I think that the symbolic operation is a generalization of the physical action in the neural representation space. When I move a cup, the cup will change its position after the action. In the similar vein, I act on a symbol and produce another symbol. Planning, making a hypothesis, proving a proposition, all of these mental actions can be regarded as like this. I find a book that presents a similar idea a bit [4]. I also think that symbolic operation occurs just because it is "useful".

Therefore, I believe that the temporal characteristics of the environment humans live in matters for symbolic manipulation.

3.6 Language and Localization

A research proposes an interesting view on localization of human language function in our brain [8]. They say that "*our hypothesis holds that the left hemisphere of the vertebrate brain was originally specialized for the control of well-established patterns of behavior under ordinary and familiar circumstances. In contrast, the right hemisphere, the primary seat of emotional arousal, was at first specialized for detecting and responding to unexpected stimuli in the environment ... In other words, the left hemisphere became the seat of self-motivated behavior, sometimes called top-down control. (We stress that self-motivated behavior need not be innate; in fact, it is often learned.) The right hemisphere became the seat of environmentally motivated behavior, or bottom-up control.*" [8]. This view is consistent with [10], where authors emphasize the importance of "song" and repeated phrases for the emergence of human language. An important implication from this literature is the connections among repeated action, top-down control, and language. **An intelligent agent may come to manipulate symbols after gaining top-down control through repeated actions.**

They present another interesting observation. They note that "*In humans the right hemisphere "takes in the whole scene," attending to the global aspects of its environment rather than focusing on a limited number of features. That capacity gives it substantial advantages in analyzing spatial relations. Memories stored by the right hemisphere tend to be organized and recalled as overall patterns rather than as a series of single items. In contrast, the left hemisphere tends to focus on local aspects of its environment.*" **If this is the case, I may assume that locality preference bias may be a good starting point for nurturing symbolic manipulations.**

They also hypothesize that "*to assess an incoming stimulus, an organism* *must carry out two kinds of analyses simultaneously. It must estimate the overall novelty of the stimulus and take decisive emergency action if needed (right hemisphere). And it must determine whether the stimulus fits some familiar category, so as to make whatever well-established response, if any, is called for (left hemisphere).*" **If I adopt this view, I may say that I can assume two kind of intrinsic motivations. One is the preference for the novelty, which is studied for a long time [13]. The other is the preference for categorization, which is not explored but can be important for higher cognitive functions.** They also say that "*Perhaps, then, those hemispheric specializations initially evolved because collectively they do a more efficient job of processing both kinds of information at the same time than a brain without such specialized systems.*"

References

[1] Noam Chomsky. *Syntactic structures*. Walter de Gruyter, 2002.

[2] Maria Angélica Furtado da Cunha. Language, usage and cognition. cambridge: Cambridge university press. 252 págs. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, 29(1), 2010.

[3] Yoav Goldberg. Assessing bert's syntactic abilities. *arXiv preprint arXiv:1901.05287*, 2019.

[4] Jeff Hawkins. *A Thousand Brains: A New Theory of Intelligence*. Basic Books, 2021.

[5] John Hewitt and Christopher D Manning. A structural probe for finding syntax in word representations. In *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)*, pages 4129–4138, 2019.

[6] Felix Hill, Olivier Tieleman, Tamara von Glehn, Nathaniel Wong, Hamza Merzic, and Stephen Clark. Grounded language learning fast and slow. In *International Conference on Learning Representations*, 2021.

[7] Paul Ibbotson. The scope of usage-based theory. *Frontiers in psychology*, 4:255, 2013.

[8] Peter F MacNeilage, Lesley J Rogers, and Giorgio Vallortigara. Origins of the left & right brain. *Scientific American*, 301(1):60–67, 2009.

[9] James L McClelland, Felix Hill, Maja Rudolph, Jason Baldridge, and Hinrich Schütze. Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences*, 117(42):25966–25974, 2020.

[10] Kazuo Okanoya and Bjorn Merker. Neural substrates for string-context mutual segmentation: A path to human language. In *Emergence of communication and language*, pages 421–434. Springer, 2007.

[11] Emily Reif, Ann Yuan, Martin Wattenberg, Fernanda B Viegas, Andy Coenen, Adam Pearce, and Been Kim. Visualizing and measuring the geometry of bert. *Advances in Neural Information Processing Systems*, 32:8594–8603, 2019.

[12] Adam Santoro, Andrew Lampinen, Kory Mathewson, Timothy Lillicrap, and David Raposo. Symbolic Behaviour in Artificial Intelligence. *arXiv preprint arXiv:2102.03406*, 2021.

[13] Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010.

[14] Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater, and Florentin Wörgötter 1. Symbol Emergence in Cognitive Developmental Systems: a Survey. *IEEE transactions on Cognitive and Developmental Systems*, 11(4):494 – 516, 2018.

[15] Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das, et al. What do you learn from context? probing for sentence structure in contextualized word representations. *arXiv preprint arXiv:1905.06316*, 2019.

[16] Michael Tomasello. *Becoming human: A theory of ontogeny*. Belknap Press, 2019.

[17] Stuart K Watson, Judith M Burkart, Steven J Schapiro, Susan P Lambeth, Jutta L Mueller, and Simon W Townsend. Nonadjacent dependency processing in monkeys, apes, and humans. *Science advances*, 6(43):eabb0725, 2020.

[18] Benjamin Wilson, Michelle Spierings, Andrea Ravignani, Jutta L Mueller, Toben H Mintz, Frank Wijnen, Anne Van der Kant, Kenny Smith, and Arnaud Rey. Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, 12(3):843–858, 2020.