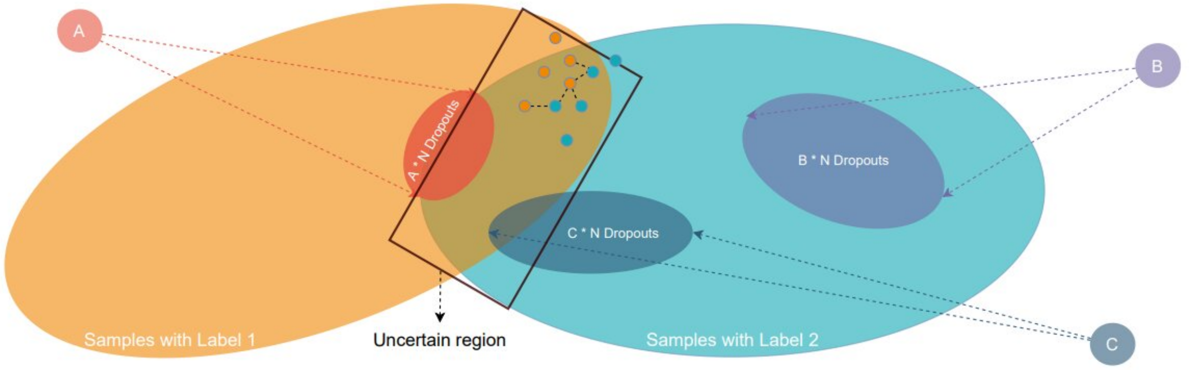
Researchers make language models scalable self-learners

June 5 2023, by Rachel Gordon



Visualization of the SimPLE method. The figure shows the embedding space of natural sentences, and different colors represent different predicted labels. Each data sample is labeled with multiple random dropouts, and we use the SETRED algorithm to detect the uncertain pseudo-labels. The final label is voted by confident inferences. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2305.17197

Socrates once said, "It is not the size of a thing, but the quality that truly matters. For it is in the nature of substance, not its volume, that true value is found."

Tell that to large language models. But does size always matter? A thought-provoking query. In a technological landscape bedazzled by large language models taking center stage, MIT CSAIL researchers think the smaller models shouldn't be overlooked, especially for natural language standing products widely deployed in the industry.

To that end, they cooked up an approach to long-standing problems of inefficiency and privacy associated with big, text-based AI models. A logic-aware model that outperforms 500x bigger counterparts on some language understanding tasks without human-generated annotations, while preserving privacy and robustness with high performance. Their study is published on the *arXiv* preprint server.

Large language models, which have shown some promising skills in generating language, art, and code, are computationally expensive, and their data requirements can risk privacy leaks when using APIs for data upload. Smaller models have been historically less capable, particularly in multitasking and weakly supervised tasks compared to their larger counterparts.

So what's helping these smaller models act so mighty then? Something called "textual entailment," a way to help these models understand a variety of language tasks, where if one sentence (the premise) is true, then the other sentence (the hypothesis) is likely to be true as well. For example, if the premise is, "all cats have tails" then the hypothesis "a tabby cat has a tail" would be entailed by the premise.

This concept is used to train an "entailment model" which proved to be less biased than other language models, from the team's previous research. They then created "prompts" that the models can use to figure out if certain information is entailed by a given sentence or phrase according to different tasks. This method improved the model's ability to adapt to different tasks without any additional training, known as zero-shot adaptation.

In the realm of "Natural Language Understanding," there are various applications that hinge on determining the relationship between two pieces of text. For example, in sentiment classification, a statement like "I think the movie is good" can be inferred or entailed from a movie review that says, "I like the story and the acting is great," indicating a positive sentiment.

Another is news classification, where the topic of a news article can be inferred from its content. For example, a statement like "The news article is about Sports" can be entailed by an article if the main content of the article reports on an NBA game. The key insight was that many existing natural language understanding tasks could be recast as an entailment (i.e., logical inference in natural language) task.

"Our research is about improving the ability of computer programs to understand and process natural language—the way humans speak and write. Our self-trained, 350M-parameter entailment models, without human-generated labels outperform supervised language models with 137 to 175 billion parameters," says MIT CSAIL Postdoctoral associate Hongyin Luo, lead author.

"This has potential to reshape the landscape of AI and machine learning, providing a more scalable, trustworthy, and cost-effective solution to language modeling," says Luo. "By proving that smaller models can perform at the same level as larger ones for language understanding, this work paves the way for more sustainable and privacy-preserving AI technologies."

The team discovered that they could improve the model's performance even more by using a technique called "self-training," where the model uses its own predictions to teach itself, effectively learning without human supervision and additional annotated training data. The self-training method significantly improved performance on a bunch of downstream tasks, including sentiment analysis, question answering, and news classification. It outperformed both Google's LaMDA and FLAN in zero-shot capabilities, GPT models, and other supervised algorithms.

However, one challenge with self-training is that the model can sometimes generate incorrect or noisy labels that harm performance. To overcome this, they developed a new algorithm called "SimPLE" (Simple Pseudo-Label Editing), a process to review and modify the pseudo-labels made in initial rounds of learning. By correcting any mislabeled instances, it improved the overall quality of the self-generated labels. This not only made the models more effective at understanding language, but more robust when faced with adversarial data.

As with most research, there are some limitations. The self-training on multi-class classification tasks didn't perform as well as on binary NLU tasks, indicating the challenge of applying entailment models to multi-choice tasks.

"This research presents an efficient and effective way to train large language models (LLMs) by formulating natural language understanding tasks as contextual entailment problems and employing a pseudo-labeling self-training mechanism to incorporate large quantities of unlabelled text data in the training process," adds MIT Professor and CSAIL Principal Investigator James Glass, who is also an author on the paper.

"While the field of LLMs is undergoing rapid and dramatic changes, this research shows that it is possible to produce relatively compact language models that perform very well on benchmark understanding tasks compared to their peers of roughly the same size, or even much larger language models."

"Entailment task is a popular proxy to evaluate 'understanding' of a given context by an AI model," says Leonid Karlinsky, research staff member at the MIT-IBM Watson AI Lab. "It is used in many areas analyzing models with unimodal, like LLM's, and and multi-modal, like VLMs inputs, simplifying the task of question answering about a given input context to a binary classification problem—does this context entail a certain (e.g., text) conclusion or not? This paper makes two contributions in this space. First, it proposes a way to improve the zero-shot (without additional tuning) NLU performance and robustness to adversarial attacks via tuning with synthesized (specialized) entailment tasks generated for the primal NLU task. Second, it offers a self-supervised SimPLE method including pseudo-labeling and confidence-based filtering to further improve large LLMs NLU performance."

"NLU is a crucial module for effective industrial AI systems," says Facebook AI Research Manager Daniel Li. "Traditional NLU models are task dependent and trained with a significant amount of human annotated data. This work shows exciting and promising results for a computation-efficient, self-learning, and robust model that are versatile among a wide range of NLU tasks."

Luo and Glass wrote the paper with CSAIL member and assistant professor in MIT's Department of Electrical Engineering and Computer Science Yoon Kim. Their work will be presented at the Meeting of the Association for Computational Linguistics in Toronto, Canada this July.

More information: Jiaxin Ge et al, Entailment as Robust Self-Learner, *arXiv* (2023). DOI: 10.48550/arxiv.2305.17197

Provided by MIT Computer Science & Artificial Intelligence Lab

Citation: Researchers make language models scalable self-learners (2023, June 5) retrieved 10 December 2023 from https://techxplore.com/news/2023-06-language-scalable-self-learners.html