

rgpt3: Making requests from R to the GPT-3 API

23 October 2022

Summary

The past decade has seen leap advancements in the field of Natural Language Processing (NLP, i.e., using computational methods to study human language). Of particular importance are generative language models which - among standard NLP tasks such as text classification - are able to produce text data that are often indistinguishable from human-written text. The most prominent language model is GPT-3 (short for: Generative Pre-trained Transformer 3) developed by OpenAI and released to the public in 2021 (Brown et al. 2020). While these models offer an exciting potential for the study of human language at scale, models such as GPT-3 were also met with controversy (Bender et al. 2021). Part of the criticism stems from the opaque nature of the model and the potential biases it may hence propagate in generated text data. As a consequence, there is a need to understand the model and its limitations so researchers can use it in a responsible and informed manner. This package makes it possible to use the GPT-3 model from the R programming language, thereby opening access to this tool to the R community and enabling more researchers to use, test and study the powerful GPT-3 model.

Statement of need

The GPT-3 model has pushed the boundaries the language abilities of artificially intelligent systems. Many tasks that were deemed unrealistic or too difficult for computational models are now deemed solvable (Maas, Snoek, and Stevenson 2021). Especially the performances of the model on tasks originating from Psychology show the enormous potential of the GPT-3 model. For example, when asked to formulate creative use cases of everyday objects (e.g., a fork), the GPT-3 model produced alternative uses of the objects that were rated of higher utility (but lower originality and surprise) than creative use cases produced by human participants (Stevenson et al. 2022). Others found that the GPT-3 model shows verbal behaviour similar to humans on cognitive tasks so much so that the model made the same intuitive mistakes that are observed in humans (Binz and Schulz 2022). Aside from these efforts to understand *how the model thinks*, others started to understand the personality that may be represented by the model. Asked to fill-in a standard personality questionnaire and a human values survey, the GPT-3 model showed a response pattern comparable with human samples and showed evidence of favouring specific values over others (e.g., self-direction > conformity) (Miotto, Rossberg, and Kleinberg 2022).

There is also ample evidence that the GPT-3 model produces biased responses (e.g., assigning attributes of brilliance more often to men than to women) (Shihadeh et al. 2022). Both the promises and challenges with the GPT-3 model require that we start to understand the system better. Of particular relevance in the ambition to study such a black box language model is the “machine behaviour” (Rahwan et al. 2019) approach, which harnesses research designs from psychological and social science research to map out the behaviour and processes of algorithms (e.g. GPT-3).

Since a large part of the behavioural and social science community who may be best placed to conduct such research is using the R environment, this package - as the first R access point to the GPT-3 model - could help break down barriers and increase the adoption of GPT-3 research in that community.

Examples

The `rgpt3` package allows users to interact via R with the GPT-3 API to perform the two core functionalities: i) requesting **text completions** and ii) obtaining embeddings representations from text input.

Completions

The core idea of text completions is to provide the GPT-3 model with prompts which it uses as context to generate a sequence of arbitrary length. For example, prompts may come in the form of questions (e.g., ‘How does the US election work?’), tasks (e.g., ‘Write a diary entry of a professional athlete:’), or open sequences that the model should finish (e.g., ‘Maria has started a job as a’).

This package handles completions in the most efficient manner from a `data.table` or `data.frame` object with the `gpt3_completions()` function. In the example, we provide the prompts from a `data.frame` and ask the function to produce 5 completions (via the `param_n` parameter) with a maximum token length each of 50 (`param_max_tokens`) with a sampling temperature of 0.8 (`param_temperature`). Full detail on all available function parameters is provided in the help files (e.g., `?gpt3_completions`)

The `output` object contains a list with two `data.tables`: the text generations and the meta information about the request made.

```
““{r eval=F, echo=T} prompt_data = data.frame(prompts = c('How does the US election work?' , 'Write a
diary entry of a professional athlete:' , 'Maria has started a job as a') , prompt_id = 1:3)
```

```
output = gpt3_completions(prompt_var = prompt_data$prompts , param_max_tokens = 50 , param_n =
5 , param_temperature = 0.8)
```

Embeddings

The second (albeit less relevant for computational social science work) functionality concerns text embeddings. An embedding representation of a document can help, for example, to calculate the similarity between two pieces of text. Embeddings can be derived as follows (using the package-provided `mini `travel_blog_data` dataset`):

```
```{r eval=F, echo=T}
data("travel_blog_data")
```

```
example_data = travel_blog_data[1:5,]
```

```
embeddings = gpt3_embeddings(input_var = example_data$gpt3
 , id_var = 1:nrow(example_data))
```

## A note on API access

When the GPT-3 model was announced, it was controversial whether the model should be made available to the public. OpenAI decided to make it available through an API that users can access with their own API key that they receive when creating an account. Using the GPT-3 API is not free (although at the time of writing, each user is provided with a small amount to get started, which is sufficient for most basic research ideas).

## References

- Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23.
- Binz, Marcel, and Eric Schulz. 2022. “Using Cognitive Psychology to Understand GPT-3.” *arXiv Preprint arXiv:2206.14576*.
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. “Language Models Are Few-Shot Learners.” *Advances in Neural Information Processing Systems* 33: 1877–1901.
- Maas, Han L. J. van der, Lukas Snoek, and Claire E. Stevenson. 2021. “How Much Intelligence Is There in Artificial Intelligence? A 2020 Update.” *Intelligence* 87 (July): 101548. <https://doi.org/10.1016/j.intell.2021.101548>.
- Miotto, Marilù, Nicola Rossberg, and Bennett Kleinberg. 2022. “Who Is GPT-3? An Exploration of Personality, Values and Demographics,” September. <https://doi.org/10.48550/arXiv.2209.14338>.
- Rahwan, Iyad, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, et al. 2019. “Machine Behaviour.” *Nature* 568 (7753): 477–86.
- Shihadeh, Juliana, Margareta Ackerman, Ashley Troske, Nicole Lawson, and Edith Gonzalez. 2022. “Brilliance Bias in GPT-3.”
- Stevenson, Claire, Iris Smal, Matthijs Baas, Raoul Grasman, and Han van der Maas. 2022. “Putting GPT-3’s Creativity to the (Alternative Uses) Test.” *arXiv Preprint arXiv:2206.08932*.