The Transformers for Polystores - The Next Frontier for Polystore Research

Edmon Begoli$^{(\boxtimes)}$, Sudarshan Srinivasan, and Maria Mahbub

Oak Ridge National Laboratory, Oak Rdige, TN 37831, USA
{begolie,srinivasans,mahbubm}@ornl.gov

Abstract. What if we could solve one of the most complex challenges of polystore research by applying a technique originating in a completely different domain, and originally developed to solve a completely different set of problems? What if we could replace many of the components that make today's polystore with components that only understand query languages and data in terms of matrices and vectors? This is the vision that we propose as the next frontier for polystore research, and as the opportunity to explore attention-based transformer deep learning architecture as the means for automated source-target query and data translation, with no or minimal hand-coding required, and only through training and transfer learning.

Keywords: Polystore · Transformers attention-based neural networks · Deep learning

1 Introduction

In 2005, Stonebraker and Çetintemel [17] posited that the time of "One Size Fits All" database management systems is over. The era of "Big Data" brought the challenge of a variety of formats, large volumes, and specialized systems (relational, document, graph, etc.) required to manage different data domains.[1] As a result of this need for diversification, we have seen a rise of new data management systems and styles, successful new special-purpose technologies, and the research efforts, such as polystores, that bring back the idea of federated style databases [8]. Polystore research has inspired numerous new directions and solutions, which we survey in the next section.

This idea, however, faces similar issues that the original federated database idea faced. There is a challenge of creating a uniform interface against all sources

---

[1] The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

(islands) covered under the polystores while simultaneously maintaining the independence of individual sources' access and manageability. The engineering effort involved in developing and maintaining a unifying layer is significant. This engineering effort usually requires a significant amount of programming. Then, there is a need to perform a source-to-source translation between database query languages (e.g., from SQL to SQL-like dialects), and there is sometimes a need to do a translation between the data sources to produce useful and meaningful results (e.g., translate long narratives to computable summaries).

However, in the field of artificial intelligence, we observed state-of-the-art techniques that enable human-like performance in machine translations, language generation, and other useful transformations. Our experience with the development and use of polystores architectures and related approached, and our exposure to these AI techniques inspire us to propose a new research approach that combines the AI research with polystores research, with a promise that some of these AI techniques could help simplify some of the engineering challenges related to the implementation, use, and maintenance of polystores.

2   Challenges with Current Approaches

As we have discussed already, a canonical polystore architecture uses *shims* for language translation between the native, *island* datastore and a polystores. While this is a convenient and necessary feature to create a user-friendly experience, it is also a complex one, and it comes with some significant challenges. For example, polystore system requires multiple shims to carryout translation from one language to another which requires all the supporting engineering mechanisms to translate one database language to another. While this is a sound, albeit labor-intensive approach, the ideal solution would be to automate the translation between the two languages (or two dialects of the same database language), and hence remove the need for significant software engineering effort. In the next sections, we discuss what this solution could be.

3   Natural Language Processing with Transformers

In recent years, deep learning has revolutionized the field of natural language processing (NLP). While many deep learning architectures have been used for processing natural language (such as convolutional neural networks (CNN) [9] long short-term memory (LSTM) networks [12], Temporal Convolutional Networks (TCN) [11]), attention based networks [7] have been at the forefront of deep learning based NLP models. The attention mechanism is a part of a neural architecture that enables to dynamically highlight relevant features of a sequence of textual elements. The transformer architecture [18] effectively uses attention for long sequences dispensing recurrence and convolutions entirely. Transformers have been been very successful in various NLP applications [21].

The transformer architecture along with the notion of transfer learning via pre-trained language models was effectively used to create the Bidirectional

Encoder Representations from Transformers (BERT) [4]. At the time of its publication, BERT broke several state-of-art results on quintessential NLP tasks on widely used datasets. Since then, there have many derivatives of BERT, many of them catering to specific scenarios. Examples include GPT [15,16], GPT-3 [1], Transformer-XL [3], XLNet [22], RoBERTa [13], and many others. For more information please see [20].

GPT-3 [1] is the most recently released model (July 2020). This is an autoregressive language model that was trained with large number of parameters (175 billion). Its abilities include translation, question-answering, Cloze tasks [14], and several tasks that require on-the-fly reasoning or domain adaption. Furthermore, due to large amount of training and parameters, this model is capable of few-shot and one-shot training [19] on certain tasks. Few-shot training refers to fine-tuning the language model on a few task specific example, while one-shot refers to fine-tuning with one example. Interestingly, it was shown that transformer architectures (e.g., Transformer) can be used to generate source code for programming languages [10]. We believe this aspect of pre-trained and fine-tuned transformers in general could be the subject of the next stage of Polystore research.

4   The Role for Transformers in Polystore Research

We propose to use this attention-based transformer architecture as a means of augmenting shim functionality. We believe that a neural machine translation system that uses transformer architecture can be trained to *translate* polystore language queries into the underlying island language query. In particular, we believe that the advanced models such as GPT-3 transformer with very minimal fine-tuning could be used to either augment the translation made by or completely replace the shim component of the polystore framework (Fig. 1).

For example, BigDAWG's *shim* can translate the *SELECT * FROM table* SQL query to *SCAN(table)* in SciDB, or it can translate one dialect of SQL to another. In fact, we posit that translation of one declarative language such as SQL to another, is a lesser challenge than translating English to Finnish, or French to Mandarin, where the grammatical differences, and possible variations are far greater than between the declarative database languages. Yet, we have seen transformers achieve the near-human, state-of-the-art results [2].

We further hypothesize that GPT-3-like transformer could be employed for polystore query translation with minimal or without any fine-tuning, and we base this hypothesis on above mentioned results in related work [2,10].
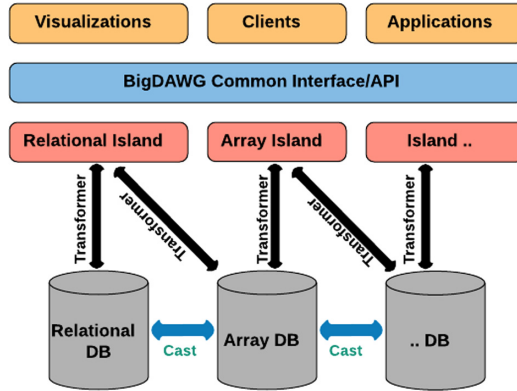
Fig. 1. Our Concept of BigDAWG [5,6] with Transformers

5   Future Work

An incorporation of transformer architectures into polystore research is an exciting and challenging idea. Given its novelty it is hard to exactly pin down all the possible directions that this research can take. For that reason, we discuss here areas that we observe as the ones that we plan to undertake, as well as the ones that are the most obvious – at least to us.

We expect that transformer architectures will play a role in two areas of polystore research, namely:

– use of transformers as source-to-source translators, and
– use of transformers as data translators.

While the current work has already shown that transformers can do a source-to-source translation between programming languages (transpilation), we expect that other transformer functions, such as auto-summarization, and other forms of transformations, could play a role in polystore research. For example, we could see in the future transformers used to summarize text into sentences, which could be served as columnar results, transform semi-structured data into structured tabular form, translate data in one native language to another, and many others.

There are perhaps too many ideas and future directions, and we see that as a good state. We hope that this paper will serve as an inspiration for durable and broad research into how one breakthrough technology can benefit the other.

References

1. Brown, T.B., et al.: Language models are few-shot learners. arXiv e-prints arXiv:2005.14165, May 2020
2. Brown, T.B., et al.: Language models are few-shot learners. arXiv preprint arXiv:2005.14165 (2020)
3. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q.V., Salakhutdinov, R.: Transformer-XL: attentive language models beyond a fixed-length context. arXiv e-prints arXiv:1901.02860, January 2019
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv e-prints arXiv:1810.04805, October 2018
5. Duggan, J., et al.: The BigDAWG polystore system. ACM SIGMOD Rec. **44**(2), 11–16 (2015)
6. Gadepally, V., et al.: The BigDAWG polystore system and architecture. In: 2016 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1–6. IEEE (2016)
7. Galassi, A., Lippi, M., Torroni, P.: Attention in natural language processing. arXiv e-prints arXiv:1902.02181, February 2019
8. Hsiao, D.K.: Federated databases and systems: part I–a tutorial on their data sharing. VLDB J. **1**(1), 127–179 (1992)
9. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751 (2014)
10. Lachaux, M.A., Roziere, B., Chanussot, L., Lample, G.: Unsupervised translation of programming languages. arXiv preprint arXiv:2006.03511 (2020)
11. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 156–165 (2017)
12. Liu, G., Guo, J.: Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing **337**, 325–338 (2019)
13. Liu, Y., et al.: RoBERTa: a robustly optimized BERT pretraining approach. arXiv e-prints arXiv:1907.11692, July 2019
14. Neville, M.H., Pugh, A.: Context in reading and listening: variations in approach to cloze tasks. Read. Res. Q. 13–31 (1976)
15. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I.: Improving language understanding by generative pre-training (2018)
16. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
17. Stonebraker, M., Cetintemel, U.: "One size fits all": an idea whose time has come and gone. In: 21st International Conference on Data Engineering (ICDE 2005), pp. 2–11. IEEE (2005)
18. Vaswani, A., et al.: Attention Is All You Need. arXiv e-prints arXiv:1706.03762, June 2017
19. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems, pp. 3630–3638 (2016)
20. Wolf, T., et al.: HuggingFace's transformers: state-of-the-art natural language processing. arXiv e-prints arXiv:1910.03771, October 2019

21. Wolf, T., et al.: Transformers: state-of-the-art natural language processing. arXiv preprint arXiv:1910.03771 (2019)
22. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., Le, Q.V.: XLNet: generalized autoregressive pretraining for language understanding. In: Advances in Neural Information Processing Systems, pp. 5753–5763 (2019)