WASP Software Engineering Assignment

Amir Mohammad Karimi Mamaghan

August 2024

1 Introduction

Existing neural networks are not yet close to human-level generalization. They need a vast amount of data, have difficulty adapting to new tasks, and are vulnerable to distribution shifts in data. Nevertheless, when the conditions are met, they can effectively learn and model complex and compositional patterns in real-world data. The main reason for their limited ability to generalize is that they often focus on associations rather than understanding deeper concepts such as the underlying system and causal mechanisms, which hinders systematic generalization. Despite significant efforts to tackle this issue, achieving human-level generalization is still an unresolved challenge. One possible approach to address this limitation is to mimic how human perception works by forming meaningful entities from unstructured sensory inputs (segregation), maintaining this separation of information at a representational level (representation), and using these entities to construct new inferences, predictions, and behaviors (composition). A practical way to achieve this is through Object-Centric (OC) Learning [1, 2] which has shown promising results in recent years. In OC learning, we assume that visual scenes are composed of multiple entities or objects that interact with each other, and exploit this compositional property as an inductive bias for neural networks. Relying on this inductive bias, object-centric representations are conjectured to be more robust than distributed representations, and to enable the systematic generalization typical of symbolic systems while retaining the expressiveness of other approaches. In my thesis, the goal is to better understand the essential requirements of learning generalizable and universal representations, potentially through inductive biases such as object-centric bias. We aim to dive deep and answer the following question: "What drives the representation learning capabilities?" To answer this, we would like to analyze how exactly each component of the models affects the quality of representations. We aim to investigate the impact of various design choices systematically and thus enable systematic progress towards better compositional representation learning approaches.

2 Two principles from the lectures

2.1 Verification and Robustness of the Models

With today's daily advancements in AI and Large Language Models (LLMs) and their worldwide accessibility, robustness has become an important and necessary aspect to consider. Every day, several new models are released for daily use, and it's important for the model to be robust to adversarial attacks and potential misuse, and relatively resistant to distribution shifts, for several reasons. For example, recently I attended the ICML conference, and a few papers introduced adversarial attacks to extract training samples from models, which might violate the privacy of the datasets or the company. My work on object-centric bias and representation learning in the image domain is related to Vision Language Models (VLMs) because visual models – such as the ones I'm working on – can be considered vision tokenizers for VLMs. Therefore, proposed models should be verified and heavily tested before being put into production and public use.

2.2 Group-level Behavioral Concepts

From the early days I joined our group at KTH as a PhD student, I've noticed that people in our group come from a diverse range of backgrounds, from economics to math and computer science. This results in having different perspectives on how to present the work, which direction to focus on, what the current hot topics are to dive into or avoid, etc., making decision-making for the whole group and for each individual a bit challenging. This is where the group-level behavioral concepts introduced in Robert's lectures will be useful. We can reach a (sub)optimal decision that benefits the group as a whole while avoiding potential conflicts by simply proper communication about different aspects such as people's opinions, needs, and project requirements. Nonetheless, I completely agree that this is a complicated matter and we might not always be able to reach a consensus.

3 Two Concepts from Guest Lectures

3.1 Safety of AI and AI Automation

In Per's lecture, the discussion on the safety of AI and AI automation was really interesting and relevant to me. It is closely related to the robustness of models. I have a friend at Stability AI (the creators of Stable Diffusion), and during ICML this year, I discussed this exact topic with him. One of the main challenges they faced when open-sourcing Stable Diffusion was ensuring the safety of the model in producing novel images for certain age groups of users. This challenge might not be relevant for me now, but it becomes an important aspect and a necessity to consider if we continue to develop vision models and decide to open-source them.

3.2 Clarity in Communications

During Per's lecture, he talked about clarity in communication under the topic of Behavioral Software Engineering. It is essential for a group to work properly and healthily. I also have personal experience with this. I'm currently involved in one main project that I'm leading and one side project in which I help another PhD student with tasks such as coding and running experiments. For both projects, we did not clearly communicate our roles and expectations, and thus, we had some minor misunderstandings and conflicts at the start, and people mainly started doing more than expected. After a couple of weeks, we found the issue and, in a clarification meeting, addressed all the misunderstandings and transparently mentioned our opinions. Since then, everything has been working out smoothly.

4 Two CAIN Papers

Here I will talk about the following publications:

1. *Developer Experiences with a Contextualized AI Coding Assistant: Usability, Expectations, and Outcomes* [3]
2. *A Combinatorial Approach to Hyperparameter Optimization* [4]

4.1 Paper 1: AI Coding Assistant

The study is mainly about the importance of context-aware AI assistants in software development. It starts by depicting the rapid evolution of AI and its immensely growing applications in software development which makes software development much easier and faster. With all the recent developments of AI assistants and Large Language Models (LLMs), it has become more crucial for the assistants to give accurate and correct answers to the developers. However, general-purpose AI assistants fall short when employed in a specific domain or application and thus, the assistants need to have access to the *context*. In context-aware AI assistants, the model has access to the details of e.g. an e-commerce platform, with all the coarse-to-fine-grained information, from the general processes and procedures to the name of the files, objects, and classes. With this information, the assistant can provide more accurate and domain-specific help to the developers. The paper then talks about the Retrieval Augmented Generation (RAG) technique which is the general way to pass the context to the models, like in LLMs. Furthermore, the authors experiment with a group of software developers in which they are given access to a context-aware AI assistant, and based on their feedback, the authors analyze the assistant's benefits and challenges. Overall, the authors see significant potential in the use of context-aware AI assistants.

4.1.1 How the paper relates to my research

This paper does not directly relate to my research but as I am a Computer Science PhD student, I write code – mainly in Python – for the projects I'm involved in. AI assistants like Microsoft Copi-

lot and ChatGPT are two AI tools that I use every day to develop code faster and more efficiently, and sometimes they are not accurate and need more context which I provide manually which is not optimal. Thus, I believe using context-aware AI tools would be a better and more efficient option.

4.1.2 How my research and its results would fit into a larger AI-intensive software project where one of the core ideas from the paper would benefit the project if applied

Currently, I do not have any plans to work on a larger AI-intensive software project where this paper can help. Furthermore, Since I'm working on some Computer Vision models, I cannot see a project in which this paper can directly help, except for the general benefits AI assistants provide to develop the project's code base.

4.1.3 How my research could be potentially adapted/changed to make AI engineering in the project based on the idea of the paper even better/easier

The AI assistants I use generally do not work super well with images as context, or they do not necessarily generate relevant images when asked. One potential avenue for applying my research into these AI assistants is to improve them in this aspect by incorporating the state-of-the-art vision models' representations into the models and allowing them to have access to these rich representations to be able to better understand the context and improve image generation.

4.2 Paper 2: A Combinatorial Approach to Hyperparameter Optimization

This paper is about Hyperparameter Optimization (HPO), which is one of the significant steps in Machine Learning. In HPO, the goal is to obtain the ideal hyperparameters before starting to train the model in order to enhance its performance, generalization, and training efficiency, while minimizing the utilized computing resources. To address the limitations and reduce the computational needs of traditional methods like grid search and random search, a new HPO algorithm has been introduced by the authors of the paper. They introduce a novel method that restricts the hyperparameter search space to a smaller space by performing t-way testing. Finally, they compare the proposed method with traditional HPO methods and show that it is superior to traditional HPO methods in terms of efficiency, scalability to large dimensions, and optimization performance.

4.2.1 How the paper relates to my research

The proposed HPO method has the potential to be used in any kind of hyperparameter search, including the one I need for the models I work with. When I start training vision models, a considerable amount of compute (typically around 10% of the whole compute budget of the project) goes for hyperparameter search. Using this approach can reduce this computational cost and thus, it allows me to use the remaining compute budget for other purposes such as conducting more experiments.

4.2.2 How my research and its results would fit into a larger AI-intensive software project where one of the core ideas from the paper would benefit the project

As I mentioned in section 2.1, the models I'm working on can be considered vision tokenizers for VLMs, which are large-scale all-purpose AI models. Therefore, these large-scale models can benefit from an efficient hyperparameter selection method to reduce the overall computational costs and become more efficient in terms of sustainability and environmental impact.

4.2.3 How my research could be potentially adapted/changed to make AI engineering in the project based on the idea of the paper even better/easier

I cannot think of any direct ways for my research to improve the idea of the paper. However, one obvious way is to consider some of the object-centric models I work with as baselines and models in the paper and see how the proposed method performs on these types of models.

References

[1] Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

[2] Klaus Greff, Sjoerd Van Steenkiste, and Jürgen Schmidhuber. On the binding problem in artificial neural networks. *arXiv preprint arXiv:2012.05208*, 2020.

[3] Gustavo Pinto, Cleidson De Souza, Thayssa Rocha, Igor Steinmacher, Alberto Souza, and Edward Monteiro. Developer experiences with a contextualized ai coding assistant: Usability, expectations, and outcomes. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI*, pages 81–91, 2024.

[4] Krishna Khadka, Jaganmohan Chandrasekaran, Yu Lei, Raghu N Kacker, and D Richard Kuhn. A combinatorial approach to hyperparameter optimization. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI*, pages 140–149, 2024.