

Abstraction \iff Long-Time Horizon Prediction

Aman Bhargava, Claude-3 Opus

April 1, 2024

1 Introduction

The ability to make accurate predictions over long time horizons is a fundamental challenge in artificial intelligence and computational neuroscience. Recent breakthroughs in language models, such as GPT-3 [?], have demonstrated impressive performance on a wide range of natural language tasks, including long-term dependency modeling. Similarly, protein language models like ESM-2 [?] have shown the ability to capture long-range structural information in protein sequences. In the domain of computer vision, deep neural networks like ResNet [?] and Transformer-based models [?] have achieved state-of-the-art performance on tasks requiring long-term spatial reasoning.

These advancements raise important questions about the role of abstraction in enabling long-time horizon prediction. We hypothesize that the ability to make accurate predictions over extended time periods is intimately linked to the capacity for abstraction. In this article, we formalize the concept of abstraction and investigate its relationship to long-time horizon prediction.

2 Thesis

Proposition 1. *Prediction over long time horizons \iff abstraction*

3 Definition of Abstraction

Definition 1. Given input information I in the original feature space F , let F^* be the set of all computable functions on F . An abstraction function $A : I \rightarrow I'$ maps the input information I to an abstract representation I' based on a subset of features $F' \subseteq F^*$, where F' is a set of computable functions on F . The abstraction function A should satisfy the following properties:

1. Information preservation: Mutual information $MI(I; I')$ is high.

$$MI(I; I') = \sum_{i \in I, i' \in I'} p(i, i') \log \frac{p(i, i')}{p(i)p(i')}$$

where $p(i, i')$ is the joint probability distribution of I and I' , and $p(i)$ and $p(i')$ are the marginal probability distributions of I and I' , respectively.

2. Complexity reduction: Kolmogorov complexity $K(I')$ is low.

$$K(I') = \min_{p \in P} \{|p| : U(p) = I'\}$$

where P is the set of all programs, $|p|$ is the length of program p , and U is a universal Turing machine that outputs I' when given p .

3. Generalization: Performance on related tasks T is high. Let $T = \{t_1, t_2, \dots, t_n\}$ be a set of related tasks, and let $P(A, t_i)$ be the performance of abstraction function A on task t_i . The generalization score $G(A)$ is defined as:

$$G(A) = \frac{1}{n} \sum_{i=1}^n P(A, t_i)$$

4. Task-relevance: Performance metric $P(A)$ for the intended task is high. Let t be the intended task and let $P(A, t)$ be the performance of abstraction function A on task t . The task-relevance score is simply $P(A, t)$.

The overall abstraction quality score is defined as:

$$Q(A) = w_1 \cdot MI(I; I') - w_2 \cdot K(I') + w_3 \cdot G(A) + w_4 \cdot P(A, t)$$

where w_1, w_2, w_3, w_4 are weighting factors that determine the relative importance of each property.

4 Argument for Abstraction \implies Long-Time Horizon Prediction

Consider a bounded computational system tasked with making predictions over a long time horizon. Let I_t be the input information at time t , and let \hat{I}_{t+h} be the predicted input information at time $t+h$, where h is the prediction horizon. The system's objective is to minimize the prediction error:

$$E(h) = \sum_t d(I_{t+h}, \hat{I}_{t+h})$$

where d is a distance function that measures the difference between the actual and predicted input information.

As h increases, the complexity of the input information grows exponentially, making it infeasible for the system to process and store all the detailed information. Therefore, to make accurate predictions, the system must employ an abstraction function A that maximizes the overall abstraction quality score $Q(A)$ while operating within the computational constraints.

Let A_1 and A_2 be two abstraction functions. If $Q(A_1) > Q(A_2)$, then A_1 is a better abstraction than A_2 for the given task and input. The system gravitates towards abstraction functions with high mutual information, low Kolmogorov complexity, good generalization performance, and high task-relevance, as they enable efficient compression and extraction of predictive features.

5 Argument for Long-Time Horizon Prediction \implies Abstraction

Let A be an abstraction function that satisfies the definition of abstraction. The low Kolmogorov complexity of I' implies the existence of a small program p that can generate I' :

$$|p| \leq K(I') + c$$

where c is a constant that depends on the choice of the universal Turing machine U .

This compact representation captures the essential features and patterns of the input, allowing for efficient storage and processing. The small program

p acts as a predictive model, able to generate future abstract representations based on the current state:

$$\hat{I}'_{t+h} = p(I'_t)$$

Moreover, the high generalization performance of the abstraction function ensures that the abstract representation I' captures the underlying regularities and structures that are relevant across related tasks. Let $T = \{t_1, t_2, \dots, t_n\}$ be a set of related tasks, and let $P(A, t_i)$ be the performance of abstraction function A on task t_i . The generalization score $G(A)$ satisfies:

$$G(A) = \frac{1}{n} \sum_{i=1}^n P(A, t_i) \geq 1 - \epsilon$$

where ϵ is a small constant that represents the desired level of generalization performance.

This generalization ability enables the system to make accurate predictions in novel situations by leveraging the learned abstract patterns:

$$E(h) = \sum_t d(I_{t+h}, \hat{I}_{t+h}) \leq \delta$$

where δ is a small constant that represents the desired level of prediction accuracy.

6 Conclusion

In conclusion, the need for long-term prediction drives the development of effective abstraction functions, and conversely, abstraction enables efficient and accurate prediction over extended time horizons. The key insights revolve around the compression, generalization, and task-relevance properties of abstraction. By finding compact and informative representations, abstracting away irrelevant details, and capturing the essential predictive features, abstraction enables bounded computational systems to make reliable predictions in the face of complexity and uncertainty.

The relationship between abstraction and long-time horizon prediction has significant implications for the design of intelligent systems, from language models to protein structure prediction and computer vision. By understanding and leveraging the principles of abstraction, we can develop more efficient and effective algorithms for a wide range of tasks requiring long-term reasoning and prediction.