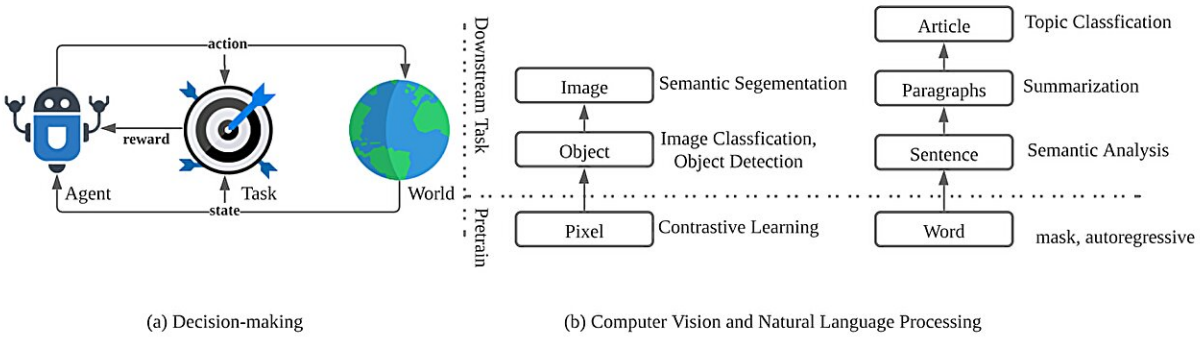
Large sequence models for sequential decision-making

December 15 2023



The difference between sequential decision-making tasks and prediction tasks, such as CV and NLP. (a) A sequential decision-making task is a cycle of agent, task, and world, connected by interactions. (b) In prediction tasks, tasks form a hierarchical structure. Credit: Muning Wen, Runji Lin, Hanjing Wang, Yaodong Yang, Ying Wen, Luo Mai, Jun Wang, Haifeng Zhang, Weinan Zhang

Transformer architectures have facilitated the development of large-scale and general-purpose sequence models for prediction tasks in natural language processing and computer vision, for example, GPT-3 and Swin Transformer.

Although originally designed for prediction problems, it is natural to inquire about their suitability in another important field, sequential decision-making and reinforcement learning problems, which are

typically beset by long-standing issues involving sample efficiency, credit assignment, and partial observability, etc.

In recent years, sequence models, especially the Transformer, have attracted increasing interest in the RL communities, spawning numerous approaches with notable effectiveness and generalizability.

To inspire more investigation into this trending topic and empower more real-world applications, e.g., robotics, automatic vehicles, and the automated industry, a research team led by Muning Wen published their survey in *Frontiers of Computer Science*.

Method	Sequence	Prediction	Discretized Tokens	Benefit	Notes
UPDeT [52]	s	a	No	Multi-task; Few-shot learning; Interpretability	Model-free; Online; Multi-agent
PIT [53]	s	Q values	No	Multi-task; Few-shot learning; Credit assignment	Model-free; Online; Multi-agent
DT [44]	rtg-s-a	a	No	Long sequence; POMDP; Credit assignment	Model-free; Offline
TT [45]	s-a-r(-rtg)	s-a-r	Yes	Long sequence; POMDP; Sparse-reward	Model-based; Offline
GDT [59]	$\psi(s,a)$-s-a	a	No	HIM problems	Model-free; Offline
PDT [46]	s-a	a	No	Few-shot learning	Model-free; Pre-train
MADT [50]	s-a	a	No	Multi-task; Long Sequence	Model-free; Offline; Multi-agent
ODT [49]	rtg-s-a	a	No	Few-shot learning	Model-free; Online
MAT [54]	s	a	No	Monotonic improvement; Multi-Task; Few-shot learning	Model-free; Online; Multi-agent
MGDT [55]	s-a-r-rtg	a-r-rtg	Yes	Multi-task; Few-shot learning	Model-free; Offline
TrMRL [60]	s	a	No	Multi-task; Few-shot learning	Model-free; Online; Meta-learning
PG-AR [61]	s	a	No	Monotonic improvement	Model-free; Online; Multi-agent
Prompt-DT [56]	rtg-s-a	a	No	Multi-task; Few-shot learning	Model-free; Offline
BooT [62]	s-a-r-rtg	s-a-r-rtg	Yes	Data Augmentation	Model-based; Offline

Detailed comparison between different Transformer-based methods for sequential decision-making. Credit: Muning Wen, Runji Lin, Hanjing Wang,

The survey presents a comprehensive overview of recent works aimed at solving sequential decision-making tasks with sequence models such as the Transformer, by discussing the connection between sequential decision-making and sequence modeling, and categorizing them based on the way they utilize the Transformer.

These works suggest the potential for constructing a large decision model for general purposes, that is, a large sequence model that can harness a vast number of parameters to perform hundreds or more sequential decision-making tasks, analogous to the way in which large sequence models have been leveraged for NLP and CV.

To examine the development of the Transformer in the field of sequential decision-making, the authors summarized recent works that convert the reinforcement learning problem into sequential form, to leverage sequence models for specific reinforcement learning settings.

Methods	Knowledge Domain	Downstream Task Indicator	What to Pre-Train	How to Pre-Train	How to Use Pre-Trained Model
Xland [91]	Online tasks	Predicates	Policy	RL	Zero-shot; Finetune
MIA [42]	Offline human demo	Text	Policy	BC	Zero-shot; Finetune
Gato [8]	Offline expert demo; Multi-modal data	Prompt	Policy	BC	Zero-shot; Finetune
SayCan [92]	Pre-trained LM	Text	Perception	SL; RL	zero-shot
Minedojo [51]	Internet video; Pre-trained LVM;	Text	Reward	SL	Online RL
VPT [9]	Internet video; Manual annotation	-	Policy; World Model	BC	Finetune
LM-Nav [93]	Pre-trained LVM; Pre-trained LM	Text	Perception	SL	Search method
Inner Mono. [94]	Pre-trained LM; Pre-trained VM	Text	Perception	SL; BC	Zero-shot

Analysis of what kind of data is used by these models (knowledge domain), how to understand the zero-shot generalization task (task indicator), what kind of component the sequence model is deployed as (what to pre-train), how to pre-train the model, and how to use the pre-trained model. Below is an explanation of the abbreviations in the table: Language model (LM), language and vision model (LVM), and behavior cloning (BC). Credit: Muning Wen, Runji Lin, Hanjing Wang, Yaodong Yang, Ying Wen, Luo MAI, Jun Wang, Haifeng Zhang, Weinan Zhang.

The authors' summary methods leverage diverse data to pre-train a large-scale sequence model for various downstream sequential decision-making tasks, inspired by the tremendous success of NLP and CV.

The team puts forth various potential avenues for future research to improve the effectiveness of large sequence models for sequential

decision-making, encompassing theoretical foundations, network architectures, algorithms, and efficient training systems. They hope this survey could inspire more investigation into this trending topic.

More information: Muning Wen et al, Large sequence models for sequential decision-making: a survey, *Frontiers of Computer Science* (2023). DOI: 10.1007/s11704-023-2689-5