Computer Science > Machine Learning

[Submitted on 24 Jul 2023 (v1), last revised 3 Oct 2023 (this version, v3)]

A Real-World WebAgent with Planning, Long Context Understanding, and Program Synthesis

Izzeddin Gur, Hiroki Furuta, Austin Huang, Mustafa Safdari, Yutaka Matsuo, Douglas Eck, Aleksandra Faust

Pre-trained large language models (LLMs) have recently achieved better generalization and sample efficiency in autonomous web automation. However, the performance on real-world websites has still suffered from (1) open domainness, (2) limited context length, and (3) lack of inductive bias on HTML. We introduce WebAgent, an LLM-driven agent that learns from self-experience to complete tasks on real websites following natural language instructions. WebAgent plans ahead by decomposing instructions into canonical sub-instructions, summarizes long HTML documents into task-relevant snippets, and acts on websites via Python programs generated from those. We design WebAgent with Flan-U-PaLM, for grounded code generation, and HTML-T5, new pre-trained LLMs for long HTML documents using local and global attention mechanisms and a mixture of long-span denoising objectives, for planning and summarization. We empirically demonstrate that our modular recipe improves the success on real websites by over 50%, and that HTML-T5 is the best model to solve various HTML understanding tasks; achieving 18.7% higher success rate than the prior method on MiniWoB web automation benchmark, and SoTA performance on Mind2Web, an offline task planning evaluation.

Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

Cite as: arXiv:2307.12856 **[cs.LG]**
(or arXiv:2307.12856v3 **[cs.LG]** for this version)
https://doi.org/10.48550/arXiv.2307.12856

Submission history

From: Hiroki Furuta [view email]
[v1] Mon, 24 Jul 2023 14:56:30 UTC (571 KB)
[v2] Sun, 1 Oct 2023 10:30:27 UTC (1,445 KB)
[v3] Tue, 3 Oct 2023 03:51:14 UTC (1,445 KB)