Decision Transformer:

Reinforcement Learning via Sequence Modeling

Deep Learning Sessions Lisboa (DLSL)
Reading Group
14th June 2021

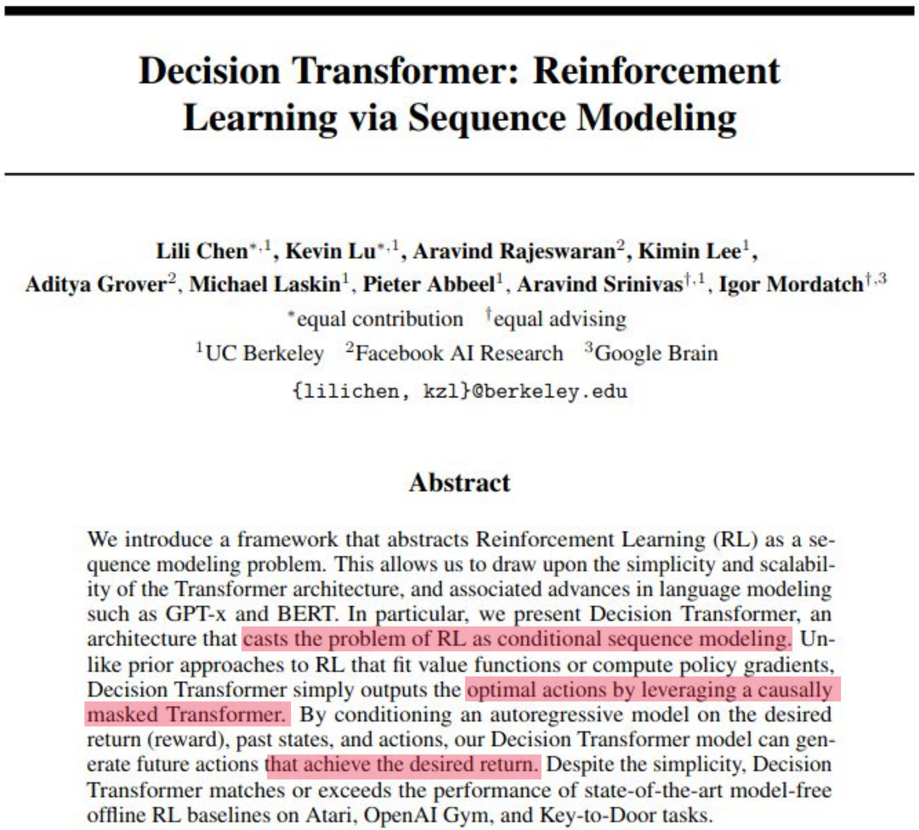
Meet & Greet!

(name, background, motivation/interest in paper)

The paper

(summary of the paper)

TLDR;



Decision Transformer: Reinforcement Learning via Sequence Modeling

Lili Chen[*,1], Kevin Lu[*,1], Aravind Rajeswaran[2], Kimin Lee[1],
Aditya Grover[2], Michael Laskin[1], Pieter Abbeel[1], Aravind Srinivas[†,1], Igor Mordatch[†,3]
[*]equal contribution [†]equal advising
[1]UC Berkeley [2]Facebook AI Research [3]Google Brain
{lilichen, kzl}@berkeley.edu

Abstract

We introduce a framework that abstracts Reinforcement Learning (RL) as a sequence modeling problem. This allows us to draw upon the simplicity and scalability of the Transformer architecture, and associated advances in language modeling such as GPT-x and BERT. In particular, we present Decision Transformer, an architecture that casts the problem of RL as conditional sequence modeling. Unlike prior approaches to RL that fit value functions or compute policy gradients, Decision Transformer simply outputs the optimal actions by leveraging a causally masked Transformer. By conditioning an autoregressive model on the desired return (reward), past states, and actions, our Decision Transformer model can generate future actions that achieve the desired return. Despite the simplicity, Decision Transformer matches or exceeds the performance of state-of-the-art model-free offline RL baselines on Atari, OpenAI Gym, and Key-to-Door tasks.

Keywords

- Reinforcement Learning (RL)
- Offline RL
- Model-free RL
- Conditional sequence modeling
- Autoregressive model
- Transformer

Research Question

Can we tackle RL as a generative trajectory modeling problem?

→ **Approach:**

‘(...) train transformer models on collected experience using a sequence modeling objective.

Research Question

Can we tackle RL as a generative trajectory modeling problem?

→ **Approach:**

‘(...) train transformer models on collected experience using a sequence modeling objective.

→ **Motivation:**

Overcome limitations of RL algorithms such as of temporal difference (TD) learning algos;

Research Question

Can we tackle RL as a generative trajectory modeling problem?

→ **Approach:**

‘(...) train transformer models on collected experience using a sequence modeling objective.

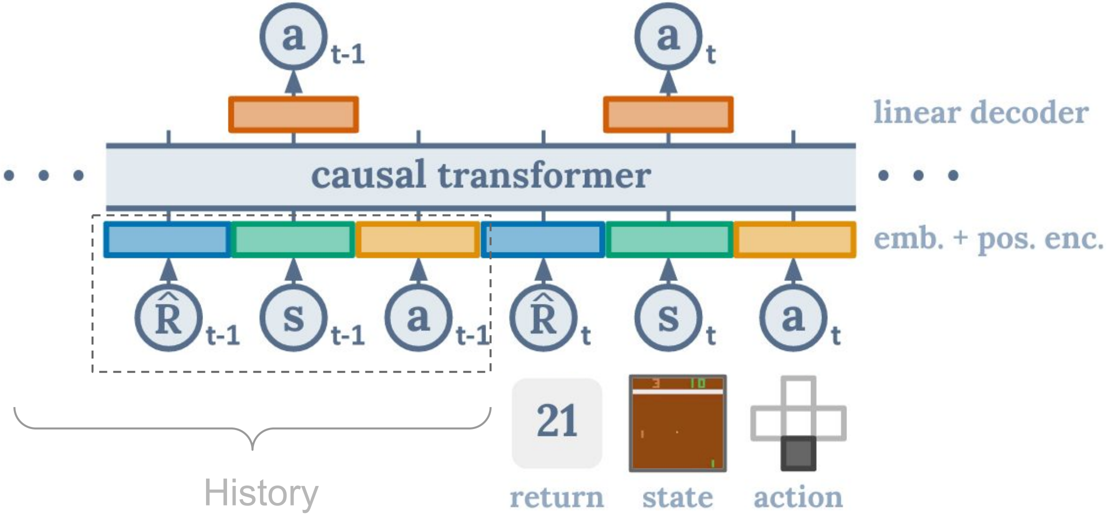
→ **Motivation:**

Overcome limitations of RL algorithms such as of temporal difference (TD) learning algos;

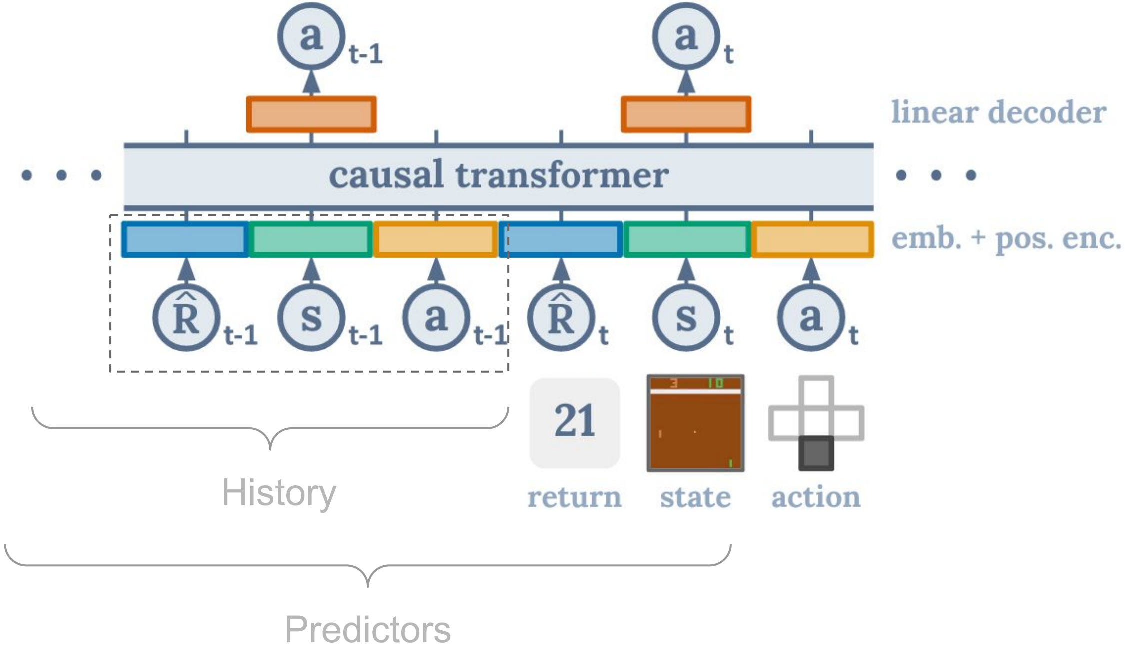
→ **Setting:**

Offline RL, Model-free RL.

Today's paper



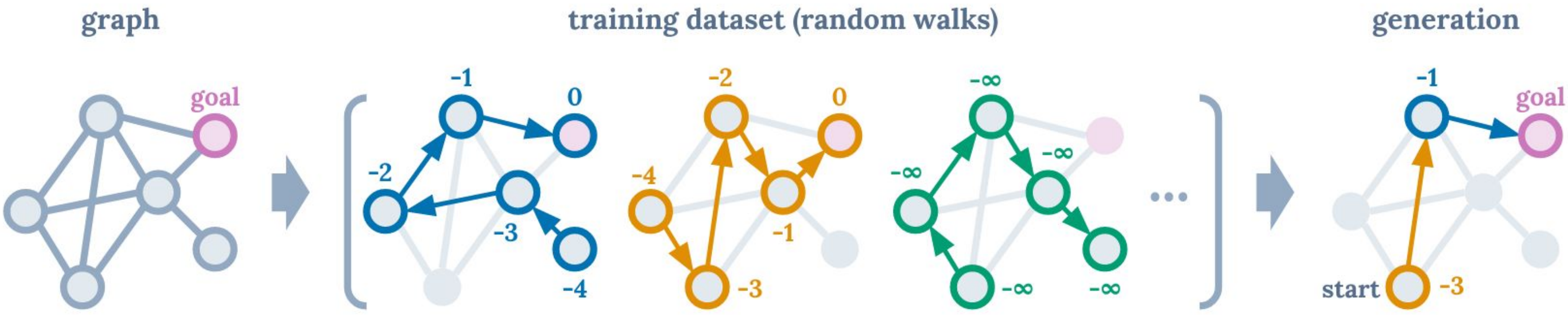
Today's paper



History

Predictors

In a nutshell, the end goal is...



Reflection

A few questions remain unanswered:

- ? Are the returns-to-go always the same? How are they computed? How do we use them at inference time?

- ? How are the states and actions represented? In Atari, we might need information about acceleration other than just the image. How is that represented?

- ? How do we train this model? What are the criteria for picking one action over another when we have multiple candidate actions?

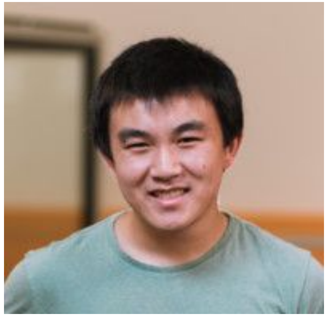
- ? Is this generalizable to the online RL setting?

Journalist

(check authors backgrounds, impact of the paper in the media, gossip, ...)

Who are the authors? (1)

Lili Chen
(UC Berkeley)
Lili (@lchen915) / Twitter
Lili Chen - Google Scholar



Kevin Lu
(UC Berkeley)
Kevin Lu (kzl.github.io)
Kevin Lu - Google Scholar

Aravind Rajeswaran
(FAIR)
Aravind Rajeswaran (washington.edu)
Aravind Rajeswaran - Google Scholar

Kimin Lee
(UC Berkeley)
Kimin Lee (google.com)
Kimin Lee - Google Scholar

Who are the authors? (2)

* advising

Aditya Grover

(FAIR)

Aditya Grover (aditya-grover.github.io)

Aditya Grover - Google Scholar

Michael Laskin

(UC Berkeley)

https://mishalaskin.github.io

Michael (misha) Laskin - Google Scholar

Pieter Abbeel

(UC Berkeley)

Pieter Abbeel--UC Berkeley--Covariant--Gradescope

Pieter Abbeel - Google Scholar

Aravind Srinivas*

(UC Berkeley)

Aravind Srinivas (berkeley.edu)

Aravind Srinivas - Google Scholar

Igor Mordatch*

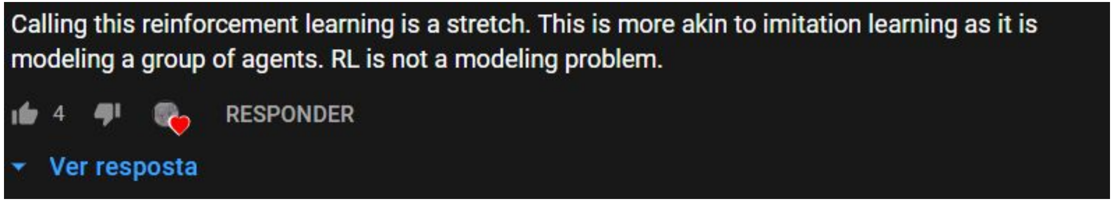
(FAIR)

Igor Mordatch (@IMordatch) / Twitter

Igor Mordatch - Google Scholar

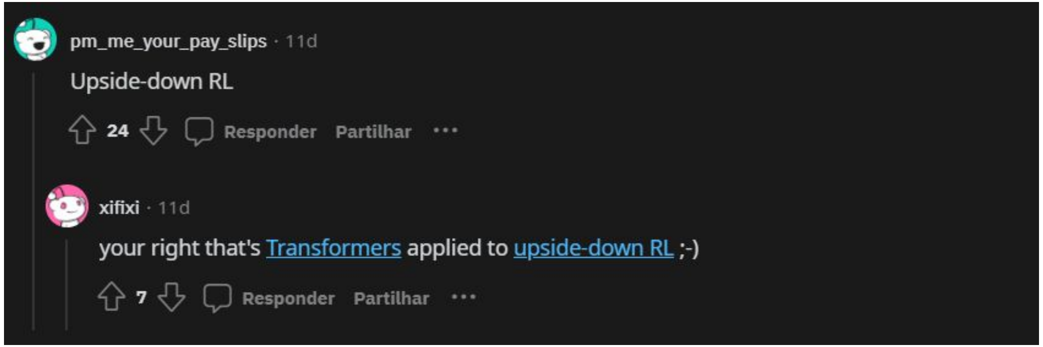
Hot Discussions on Social Media

Is this really Reinforcement Learning?



Calling this reinforcement learning is a stretch. This is more akin to imitation learning as it is modeling a group of agents. RL is not a modeling problem.

<source>



pm_me_your_pay_slips · 11d

Upside-down RL

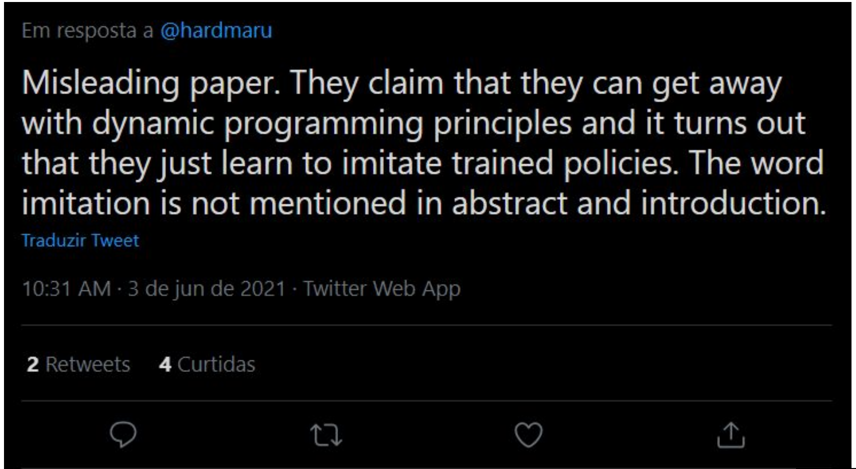
xifixi · 11d

your right that's Transformers applied to upside-down RL ;-)

<source>

Isn't this just Schmidhuber's paper on "Upside-Down RL"[1] with a Transformer?

Hot Discussions on Social Media



Em resposta a @hardmaru

Misleading paper. They claim that they can get away with dynamic programming principles and it turns out that they just learn to imitate trained policies. The word imitation is not mentioned in abstract and introduction.

Traduzir Tweet

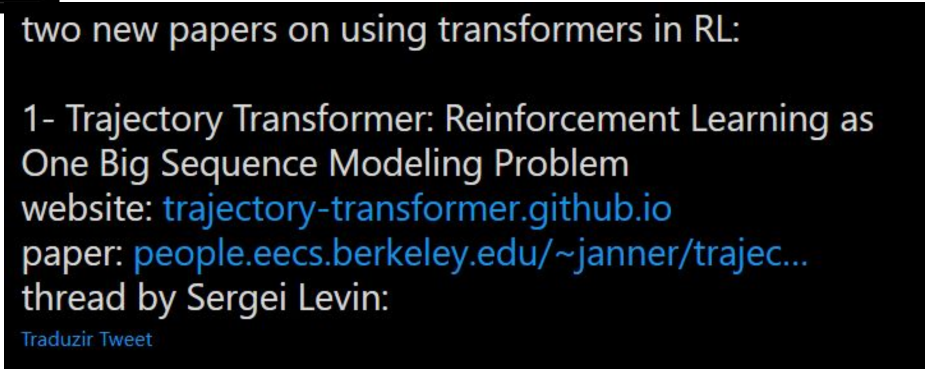
10:31 AM · 3 de jun de 2021 · Twitter Web App

2 Retweets **4** Curtidas

<source>

Isn't this just Imitation Learning?

Is this somehow related to this other paper on "Reinforcement Learning as a One Big Sequence Modeling Problem"[2] *?*



two new papers on using transformers in RL:

1- Trajectory Transformer: Reinforcement Learning as One Big Sequence Modeling Problem
website: trajectory-transformer.github.io
paper: people.eecs.berkeley.edu/~janner/trajec...
thread by Sergei Levin:

Traduzir Tweet

<source>

Other Hot Discussions on Social Media

Does this work for large state spaces where it is unlikely that the current state has been observed before?

How important is the context length for the learning problem?

How great it is to see the cooperation of several companies?

Pointers

Paper’s Website

- https://sites.google.com/berkeley.edu/decision-transformer

Youtube

- AI Weekly Update - June 9th, 2021 (#34!) [Henry AI Labs]- https://youtu.be/z9mDGLKKqo0
- Decision Transformer [Yannic Kilcher] - https://youtu.be/-buULmf7dec

Synced

- https://syncedreview.com/2021/06/09/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-37/

Medium

- https://medium.com/syncedreview/pieter-abbeel-teams-decision-transformer-abstracts-rl-as-sequence-modelling-b8f4cf58ed5e

Twitter

- https://twitter.com/IMordatch/status/1400113795196809227

Reddit

- https://www.reddit.com/r/deeplearning/comments/nxfvgo/paper_explained_decision_transformer/
- https://www.reddit.com/r/MachineLearning/comments/nqqle6/r_decision_transformer_reinforcement_learning_via/

Archeologist

(extract knowledge out of the references...)

Digging the references

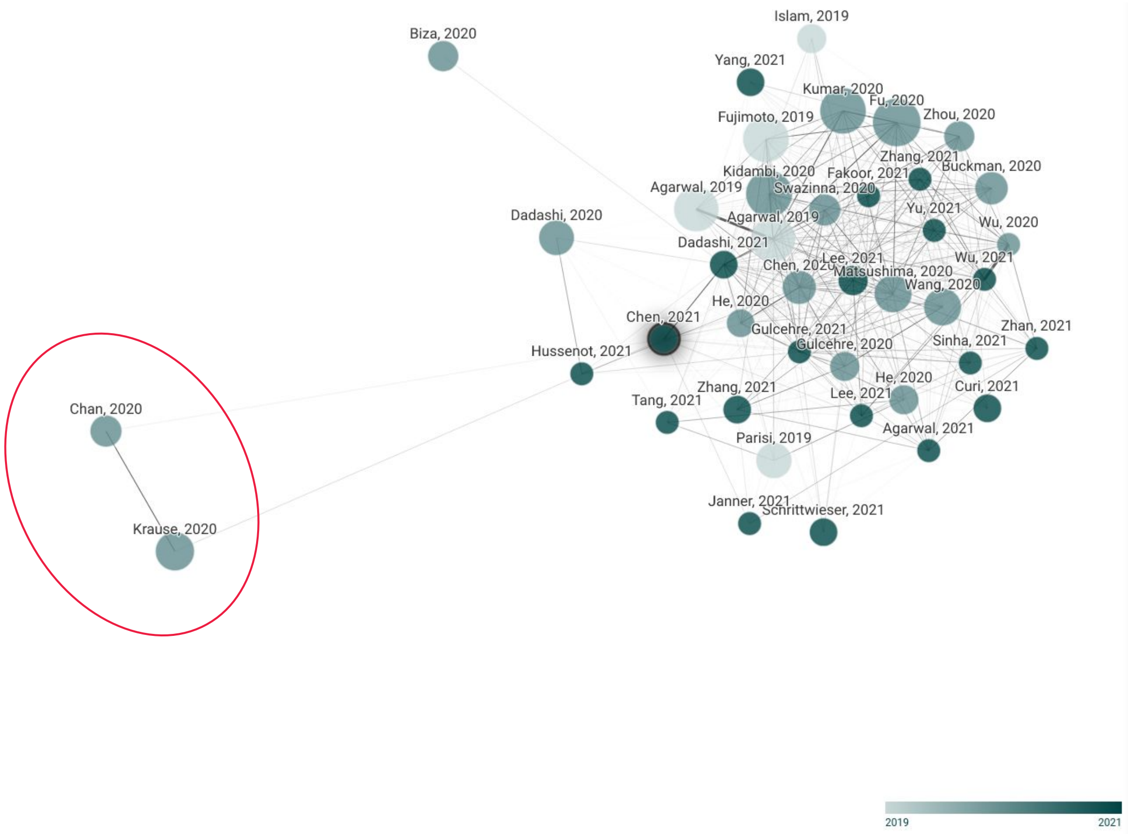
- Recent technical paper
  (dates back to 2nd of June of 2021);

- Comprises 66 references and 1 citation.

- References spread mostly across 2 different topics:

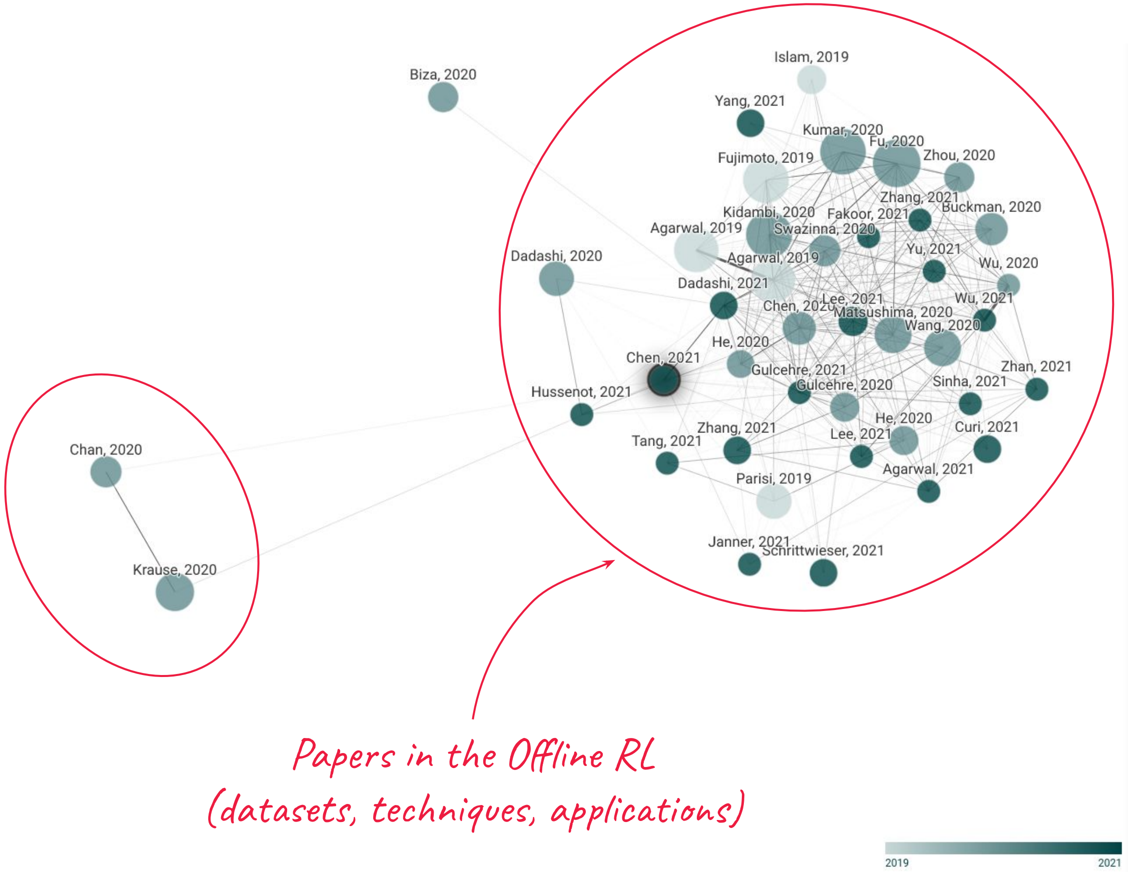
Transformers	Reinforcement Learning (RL)
Seminal paper on Transformers (Vaswani *et al.*, 2017)	Standard RL & Deep RL methods Offline RL (or Batch RL)
Applications in NLP domain (text generation, language models, language understanding)	Datasets Offline RL
	Credit Assignment

ConnectedPapers - Decision Transformer [1]



Papers in the NLP domain
(involving transformers)

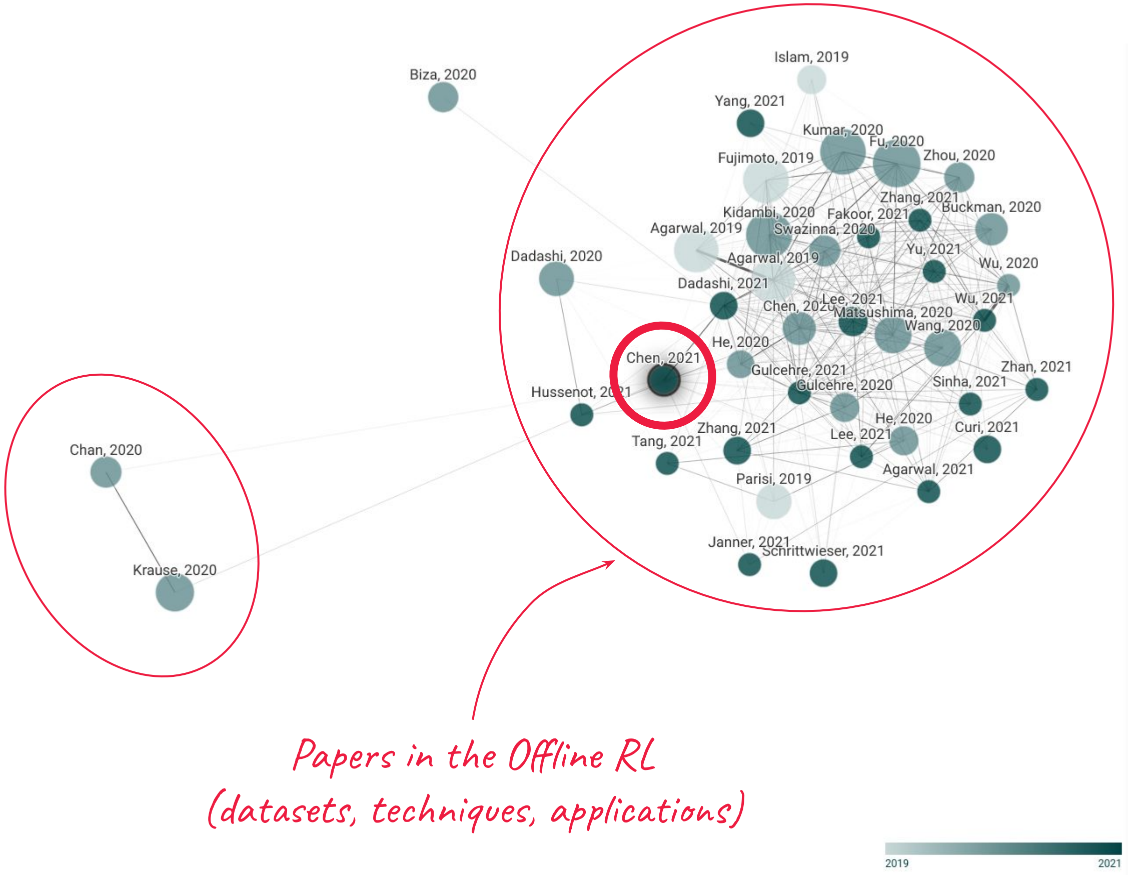
ConnectedPapers - Decision Transformer [1]



Papers in the NLP domain
(involving transformers)

Papers in the Offline RL
(datasets, techniques, applications)

ConnectedPapers - Decision Transformer [1]

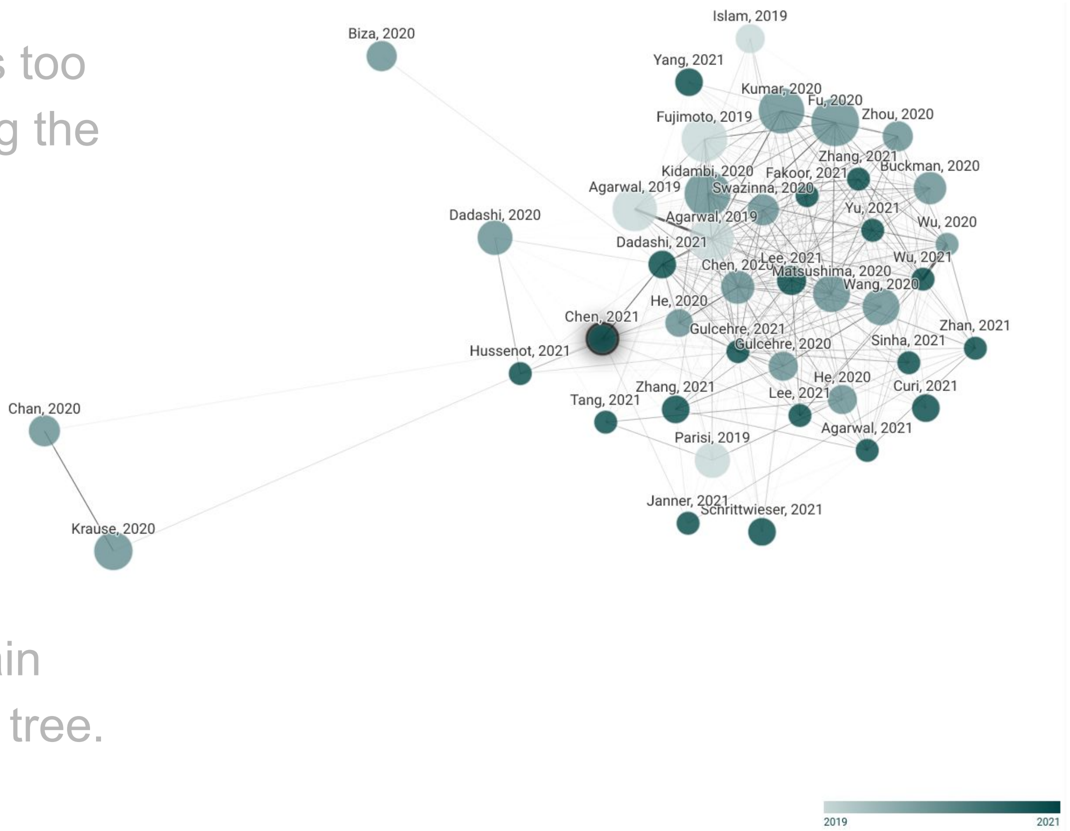


Papers in the NLP domain
(involving transformers)

Papers in the Offline RL
(datasets, techniques, applications)

ConnectedPapers - Decision Transformer [1]

Every work in the similarity tree is too specific, only tangentially covering the mentioned topics.



This paper stands as a paradigm changing paper, which may explain the heterogeneity in the similarity tree.

Digging the paper’s references - Most influential

- ***Human-level control through deep reinforcement learning*** ***(2015)***
  Deep Q-network (DQN)’s seminal paper. It learns successful policies directly from high-dimensional sensory inputs using end-to-end RL. It represents a turning point in the RL community.

- **Attention is all you need** ***(2017)***
  Transformers’ seminal paper. Despite not being fundamental for reading today’s paper, reading it will provide you with the details on the architecture and the overall mechanisms associated.

- **D4RL: Datasets for Deep Data-Driven Reinforcement Learning** ***(2020)***
  a work merely focused on the creation of datasets for the offline RL setting.

Digging the paper's citations - Trajectory Transformer

The paper *Reinforcement Learning as One Big Sequence Modeling Problem* **(2021),** published 1 day after this session's paper, proposes the ***Trajectory Transformer***.

- Trajectory Transformer concerns the use of transformers to handle the RL problem as a sequence modeling task as well. They mention their method work both in the online and offline setting.
- Through empirical experiments in different datasets their model is able to more reliably make predictions over long-horizons than conventional dynamics models.
- Overall, they also mention that they are able to achieve competitive SotA performance in the offline setting when coupling Trajectory Transformers w/ beam search.

Looking into the field - Recommendations

The main recommendations in terms of resources to look for when trying to get a better understanding of this paper are:

- **Richard S Sutton and Andrew G Barto .Reinforcement learning: An introduction. MIT Press** (2018)
  (it seems to be the best introduction to the field of reinforcement learning. Introduces the fundamental concepts).
- **Imitation Learning - Tutorial** (2018)
  (broad overview of imitation learning techniques and recent applications).
- **Reinforcement Learning Upside Down: Don’t predict rewards -- Just Map them to Actions** (2019)
  (it seems to be the best introduction to the field of reinforcement learning. Introduces the fundamental concepts).

Other resources

Companies working with RL: DeepMind, Maluuba, Amazon, Facebook AI Research, Microsoft, ...

RL-based groups:

- Microsoft - Reinforcement Learning Group
- Reddit post with several groups in RL
- Meetup list of Largest Deep RL groups around the world

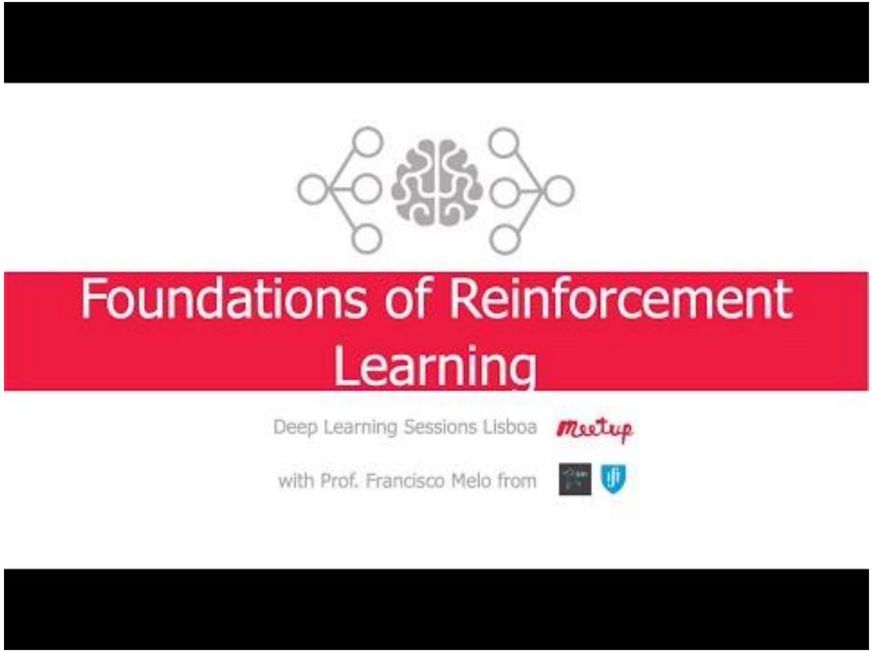
Courses

- UC Berkeley Courses
- RL Course by David Silver
- RL specialization @ coursera

Misc Resources:

- Github with cool resources for researchers on RL

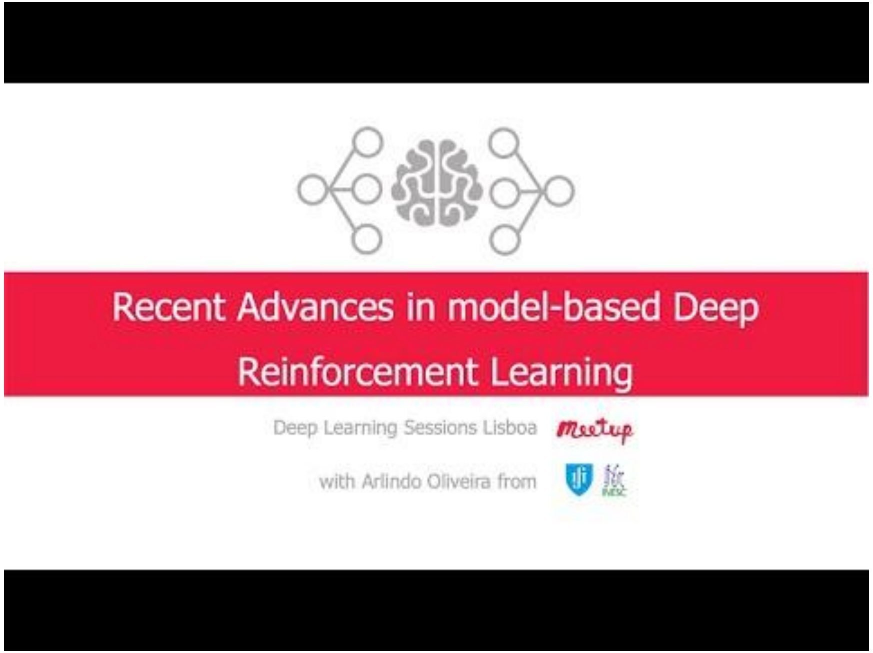
Check out our meetups on RL!



Foundations of Reinforcement Learning

Deep Learning Sessions Lisboa Meetup

with Prof. Francisco Melo from



Recent Advances in model-based Deep Reinforcement Learning

Deep Learning Sessions Lisboa Meetup

with Arlindo Oliveira from

Developer

(check for code, tutorials, implementations, ...)

Related Code

- Paper repo: https://github.com/kzl/decision-transformer
  (notes: only contains Atari and Gym, no implementation of the door experiment nor other techniques)

- Implementation of Schmidhuber’s ⅂□: https://github.com/BY571/Upside-Down-Reinforcement-Learning
  (notes: also has experiments on Gym, a direct comparison may be possible)

- TD implementation: https://github.com/dennybritz/reinforcement-learning
- BC implementation: https://github.com/omerbsezer/Reinforcement_learning_tutorial_with_demo
  (notes: both repos with old school TF </3)

Good Peer Reviewer

(find the strengths of the paper, the most valuable aspects, ...)

Positive Aspects of the Paper ↑

- ✔ Authors were able to pose RL as a sequence modeling problem, making use of Transformers, that have the following advantages:
  - ✔ selecting other tokens in the sequence to make predictions (e.g. a state at the very beginning can be crucial to decide the current action)
  - ✔ attention weights could be used to understand why the model choose a given action.

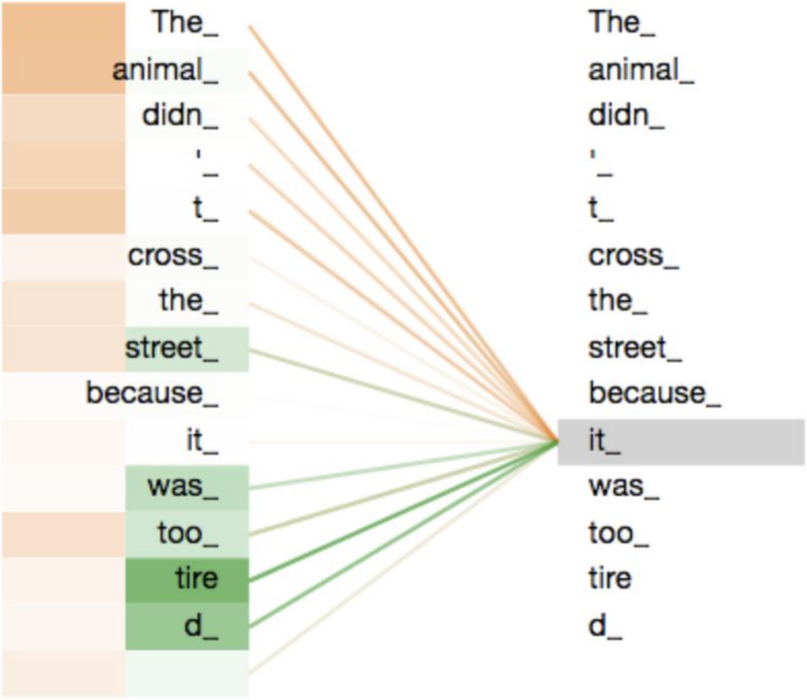


Image from Jay Alammar.

Positive Aspects of the Paper ↑

- ✔ Decision Transformers eliminate the need for classical RL concepts such as temporal-differences, value- and Q-functions.

- ✔ Section 5.7 states a big disadvantage of TD methods:

  TD-learning optimizes an already approximated function, which can exploit any inaccuracies in the value function approximation, causing failures in policy improvement.

- ✔ This paper connects the concepts of RL and Supervised Learning.

Mean Peer Reviewer

(find the cons/weaknesses of the paper)

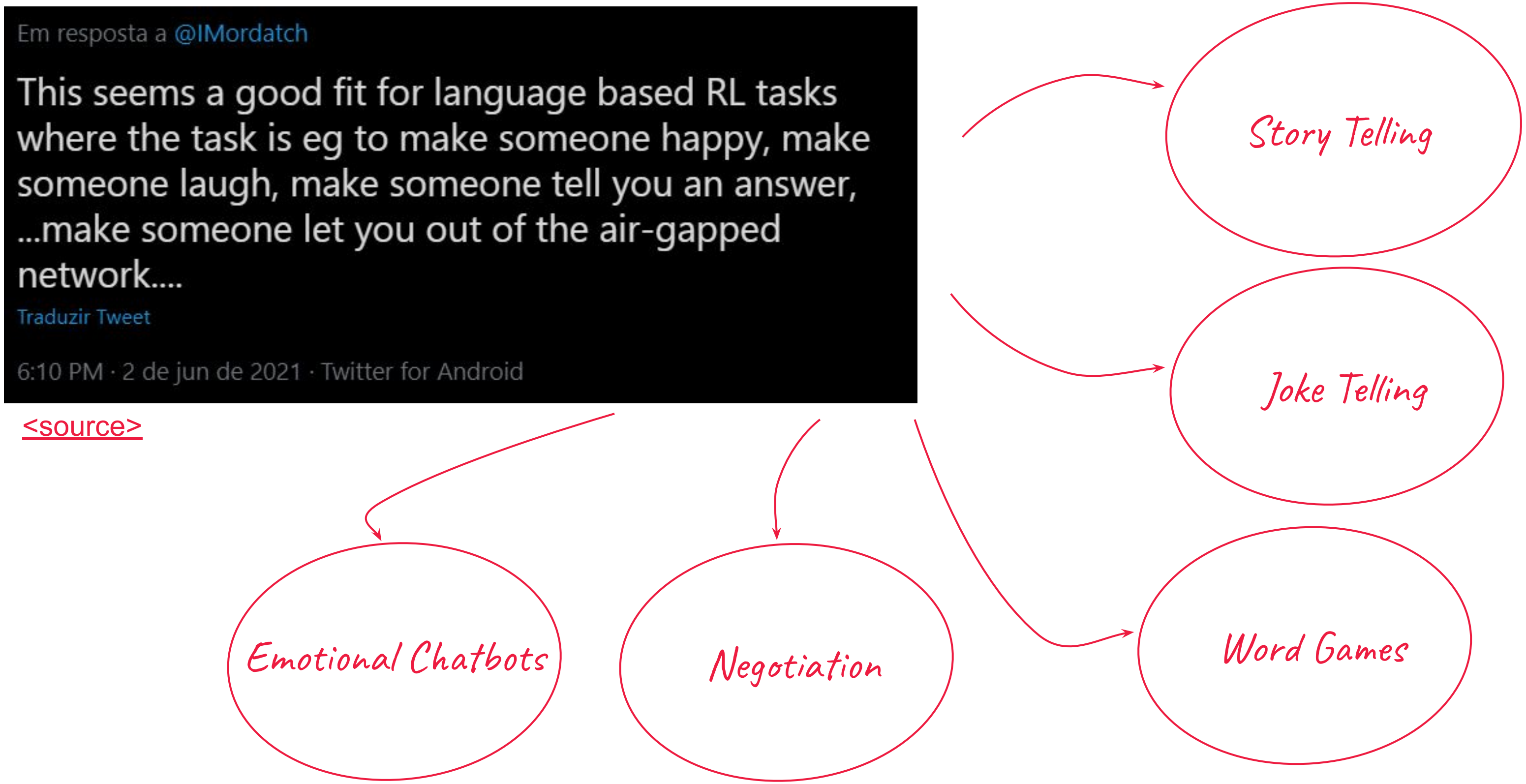
Lowpoints of the Paper ↓

- × Some implementation details are unclear: how are desired rewards determined? How are these updated at training time? What about inference time?
- × What is the computational cost of training this method? How does it compare with others in the literature? What if the context is not enough to model the whole trajectory?
- × Can we adapt it to the online setting?
- × Several questions remain about the evaluation of the paper. What was the methodology employed? How were the parameters selected? Why aren’t there standard deviations for the baseline methods?

Entrepreneur

(come up w/ ideas for applications/products and pitch them)

Application to Language based RL



Em resposta a @IMordatch

This seems a good fit for language based RL tasks where the task is eg to make someone happy, make someone laugh, make someone tell you an answer, ...make someone let you out of the air-gapped network....

Traduzir Tweet

6:10 PM · 2 de jun de 2021 · Twitter for Android

<source>

Story Telling

Joke Telling

Emotional Chatbots

Negotiation

Word Games

Wrap-up

(discussion; come up w/ ideas for titles)

Open Questions

Transformer models in NLP work well with Self-Supervised Learning. Can that also be applied to the Decision Transformer?

?

?

It's not clear how returns-to-go are defined. If these are always the maximal value, isn't it redundant to use the same tokens at every time step?

Summary

Decision Transformer casts offline RL as a sequence modelling problem.

Uses the transformer as an autoregressive model and provides as inputs the desired reward to achieve, the previous states, and actions.

The evaluation compares model-free offline RL baselines in game-related tasks including Atari and OpenAI Gym. Obtained results show Decision Transformer achieves competitive SotA results.

Some resources...

Paper; Code;

Paper review (by Yannick's Kilcher);

Thank you!

Deep Learning Sessions Lisboa (deep.learning.lx@gmail.com)

Vote for the next paper @ List Suggested Papers