Generating Text-based Adventure Games

Anna Orosz
A Thesis in Data Science
for the Degree of Master of Science in Engineering at the
University of Pennsylvania

December 20th, 2021

What is a text-based adventure game?

```
On the table is an elongated brown sack, smelling of hot peppers.
A clear glass bottle is here.
The glass bottle contains:
  A quantity of water.
>w
You are in the living room.  There is a door to the east.  To the west
is a wooden door with strange gothic lettering, which appears to be
nailed shut.
In the center of the room is a large oriental rug.
There is a trophy case here.
On hooks above the mantlepiece hangs an elvish sword of great antiquity.
A battery-powered brass lantern is on the trophy case.
There is an issue of US NEWS & DUNGEON REPORT dated 28-JUL-80 here.
>get sword
Taken.
>break egg with sword
You rather indelicate handling of the egg has caused it some damage.
The egg is now open.
There is a golden clockwork canary nestled in the egg.  It seems to
have recently had a bad experience.  The mountings for its jewel-like
eyes are empty, and its silver beak is crumpled.  Through a cracked
crystal window below its left wing you can see the remains of
intricate machinery.  It is not clear what result winding it would
have, as the mainspring appears sprung.
>_
```

A conversation from 30 years ago...

From: goetz@acsu.buffalo.edu (Phil Goetz)
Subject: Re: Adventure generators

> Date: 29 Oct 92 04:40:05 GMT
> From: morpheus@sage.cc.purdue.edu (Morpheus Nosferatu)
> Subject: Adventure generators
>
> Has anyone ever worked on, or even heard of, an adventure generator? I'm not talking about an adventure design language like TADS or Alan, but rather a stand-alone adventure generator that produces complete adventures, where the user need only give a minimal degree of input, such as the level of complexity, type of adventure (mystery, treasure hunt, etc.), size of adventure, and so forth? ... But as anyone ever heard of someone trying to come up with a generator whigh would produce infocom-style text adventures? I can just imagine what kind of limitations it would have, but I'm curious to know if anyone has tried this, and if so what degree of success they've had.

No. ... The generator you speak of is not written, not being written, and not anywhere on the horizon. In 50 years, maybe. In 20, definitely not. The problem of writing interesting stories, which adhere to someone's definition of a plot (with goal explanations, conflict, resolution, comlication, climax, etc., all occuring at appropriate intervals) is very hard, and I don't expect a solution soon. But the problem of writing clever puzzles involves much greater creativity, and I have seen NO evidence that ANYBODY has a clue in these creativity issues; the most you will find in the field are a few vague theories of creativity. This problem is what Stuart Shapiro calls "AI-complete": Solving it would be equivalent to solving all the other problems of AI.

Phil

Goals

- What makes a good text-adventure game?
- How to generate game-appropriate text?
- What is the most cost- and time-efficient method to implement for such auto-generation of Text-adventure games?
- How to evaluate the generated text and attributes?

The data - Facebook LIGHT data

The paper

- Intended for generating conversational data

- Provides great background for fictional text generation

Data used

- Extensive library that included dictionaries of:

  - Objects

  - Characters

  - Rooms/categories

- Wide range of attributes

The model

GPT-3

- Most advanced & largest language model yet

- 175 **billion** parameters

- Built by OpenAI

Fine-tuning

- re-training a pre-trained LM using custom data

- weights of the LM are updated for the characteristics of the data and the specific task
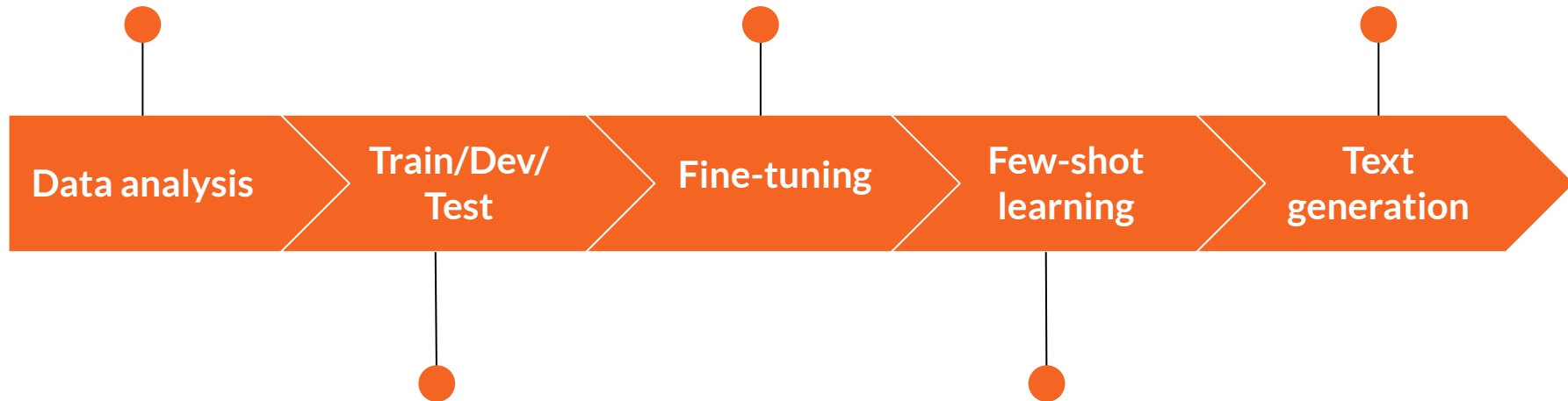
Few-shot learning

- we train a model on some classes and predict for a new class - model has only seen a handful examples

Penn Engineering

Generating Descriptions

Statistics about # of examples in objects, rooms, characters

Update model parameters to fit the task of generating descriptions

Generating object/room/character descriptions with fine-tuned model

Data analysis › **Train/Dev/ Test** › **Fine-tuning** › **Few-shot learning** › **Text generation**

Split data:
- 80% train (fine-tuning)
- 10% dev (few-shot learning)
- 10% test (generating descriptions)

Give model a handful of examples to generate text in desired format

Penn Engineering

	Original	Davinci	Curie	Curie w/ few-shot learning	Babbage
The Old Queen Farm	Once a fairly profitable farm, the Old Queen Farm is now a deteriorated old shack surrounded by uninhabitable fields with soil incapable of growing even weeds. The farm house shows signs that is being occupied by wild animals, and the roof is beginning to cave inwards.	The Old Queen Farm is a dilapidated farmhouse with a tiny little garden. It has a barn and stables for animals. It has a long driveway and meadow behind it. The farm has been vacant for over a year, but the grass hasn't grown in the meadow or by the house or by the barn or stables, because of some magic that keeps it from growing. The house is empty but perfectly clean, because it's being maintained by magic.	The old queen farm is a large, grassy field that is surrounded by several oak trees. There are two large, open barns on opposite sides of the field. The old wooden fence is beginning to show signs of age and wear.	This farm has been in the family for many generations and is very well known. The farm is enormous, stretching as far as the eye can see. It's barns are huge and full of every imaginable animal, and there are wagons with supplies coming and going all day. The fields are green with growing crops and cows grazing. There is a huge castle in the distance, but it seems to be out of place here.	This farm is the oldest one in the kingdom. It's run by the queen and her butler, a former soldier who has been transformed into a servant. The queen and her new assistant, the butler, are constantly arguing about what to do with it.

rating	1 - 5	4.4	4.6	4	4.4	3.8

Does length of text predict rating?

Generating Item Attributes

A closer look at 'object' data

```
{'base_form': ['sword', 'Sword'],
 'desc_entries': 2,
 'descriptions': ['The sword is very
old, you would assume it had once
belonged to a legendary warrior.',
  "The sword's legend is known by
everyone, it is famous throughout the
land."],
 'ex_room_ids': [],
 'holding_character_ids': [],
 'in_room_ids': [12],
 'is_container': 0.0,
 'is_drink': 0.0,
 'is_food': 0.0,
 'is_gettable': 1.0,
 'is_plural': 1.0,
 'is_surface': 0.0,
 'is_weapon': 1.0,
 'is_wearable': 0.0,
 'link_entries': 1,
 'name': 'Legendary swords',
 'object_id': 1188}
```

- 'Objects' dataset has a myriad of attributes:
  - mostly binary attributes

- Used two different engines (Curie+Davinci) to test performance

- Compared natural vs non-natural language during training

Penn Engineering

Let's look at the results

Question: Does Davinci or Curie predict object attributes better?

Predicting 'food' attribute, in %			
metric	Curie	Davinci	Baseline
precision	3.25	2.63	0.0
recall	80	80	N/A
accuracy	48.28	26.96	41.6

Answer: cost-to-value is better for Curie overall

Question: How important is it that the prompt is in natural language?

Predicting 'weapon' attribute, in %			
metric	natural	non-natural	baseline
precision	3.25	2.63	0.0
recall	80	80	N/A
accuracy	48.28	26.96	94.83

Answer: much better performance w/ natural language

Penn Engineering

Conclusion and Future work

1. Fine-tuning is a powerful tool that allows less-powerful models (Curie) to be as effective as any (Davinci)

2. Few-shot learning performs well in case of small trainingset -> need to be careful about class imbalance!

3. GPT-3 performs best with natural language

4. Future work: analyse performance using more metrics

Special thanks to my advisors, without whom this work would not have been possible:

Professor Chris Callison-Burch

Lara Martin, PhD

Q+A