Dungeons and DQNs
Toward Reinforcement Learning Agents that Play Tabletop Roleplaying Games
LARA J. MARTIN, SRIJAN SOOD, MARK O. RIEDL
It’s an exciting time for AI
Sometimes AI needs to “roleplay” or improvise

Open the pod bay doors, HAL.

I’m sorry, Dave. I’m afraid I can’t do that.
Or in the real world...

AI: You’re doing a lot better today!

Patient: Give it to me straight, doc.

AI: You have 1 month to live.
Consider Dungeons & Dragons

- Players create characters to play & describe their character’s actions
- Characters exist in a shared imaginary world
- Game/Dungeon Master (GM/DM) mediates and sets up scenarios—or campaigns
Outline

- Why D&D?
- TRPGs compared to other media
- Our starting point
1) Unlimited Actions
2) Unexpected Consequences
3) Dynamic, Continuous World
4) Distributed Game World
5) Intrinsic Reward
6) Collaborative
Outline

- Why D&D?
- TRPGs compared to other media
- Our starting point
Using **Interactive Fiction** for Roleplaying

You find yourself in a dungeon. You see a **FLASK**. Obvious exits are NORTH and SOUTH. What do you do?

> Get flask

You can’t get the flask! I’m certainly not going to tell you why not.

- Constrains to a preexisting story, uncovered through puzzles

Example adapted from Homestar Runner’s “video games” Strong Bad Email
Using *Experience Management* for Roleplaying

> show proprietor the puzzle box

The proprietor takes the puzzle box and turns it over in his hands carefully. “Now, this is a tricky one,” he says. “Frightfully difficult, unless you know the catch of course.” His fingers flicker dexterously over the box, sliding a panel here, pressing a corner in there. Suddenly the lid pops open with a faint snick.

He places the box on top of the display case. “And there you have it,” he says. “A present for you.”

> x display case

Inside the display case are a deck of tarot cards, an amulet, and a geode.

> ask for the amulet

![Diagram of a puzzle box with connections to Drama manager and temp_deny_get_amulet nodes.](image)

- Intervenes in storyline to keep things “on track” for quality

Once Sally began to run, John pulled out the gun and directed it at the bank guard. John wore a stern stare as he pointed the gun at Sally. Sally saw the gun and instantly screamed before she could stop herself. John told her she had one minute to get the money and shook the gun at her. John gave Sally a bag to put the banks money in.

**Generates new stories for well-defined domains; usually not collaborative**

Outline

► Why D&D?
► TRPGs compared to other media
► Our starting point
“Simplified” D&D

- No dice rolling (i.e. no combat, etc.)
- Agent is always in character
- DMs aren’t refereeing
Our Approach: Deep Reinforcement Learning

- **Explore**
- **Exploit**
- **Reward**
- **Environment**
- **Policy Update**

or

- **Policy Update**
- **Explore**
- **Exploit**
- **Reward**
- **Environment**
Our Approach: Deep Reinforcement Learning

Atari Games - 2015
Go - 2016
Doom - 2016
DOTA 2 - 2018
But D&D isn’t your average game...

- We need a pre-defined goal for RL so it knows how to get reward
- How are sentences used as actions?
  - We need a way of breaking them down so we can update the game state
r 2 d 2 carrying some drinks on a tray strapped to his back passes yoda who uses his force powers to hog the drinks

**Expected:**
obi wan and anakin are drinking happily when chewbacca takes a polaroid picture of anakin and obi wan

**Predicted:**
can this block gives him the advantage to personally run around with a large stick of cheese
Event Representation

- From sentence, extract event representation (S, V, D, M, P)
- Use our linguistic knowledge to bootstrap the RL

For example...

**Original sentence:** They’re scanning for non-terrestrial technology.

**Event:** they scan EmptyParameter technology for

**Generalized Event:**
<PRP> assessment-34.1 EmptyParameter use.n.01 for
Eva delivered the message to Javier.

Eva swaggered away.

<Person>0 send-11.1 abstraction.n.01
<Person>1: has_possession(abstraction.n.01)
<Person>0: moved

<Person>0 run-51.3.2 Ø Ø
<Person>1: has_possession(abstraction.n.01)
The Proposed System (Training)

- **Policy Update**
  - **Seq2Seq (Genre Model)**
  - **Reward**
    - Updated State or <None>
  - **Exploit**
    - Event Selection & Current State
    - Updated State or <None>
  - **Explore**
    - Distribution of Next Events
    - Updated State or <None>

- **Environment (Rule Model & State Updater)**
World Model

1. Genre Expectation Model
   - Seq2Seq network generates next event in the story
   - Trained on relevant genre

2. Commonsense Rules Model
   - Things that aren’t mentioned in stories (see: Principle of Minimal Departure)
   - Temporal & physical rules
How could our model be improved?

- What about other forms of reward besides “quest completion”?
- It could learn when to break rules
Conclusion

- TRPGs are the next AlphaGo
Conclusion (In other words)

1. AI research is moving fast, but it needs the ability to adapt in the moment
2. We think testing on TRPGs can help with this
3. Our first step is to create a DQN that takes events from a genre and uses general rules of the world
Thank you!

LARA MARTIN

LJMARTIN@GATECH.EDU