Unsupervised/NLP Project

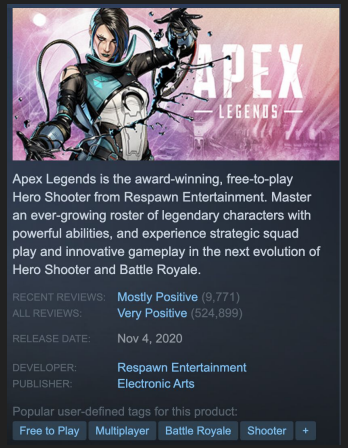
Colin Gallagher

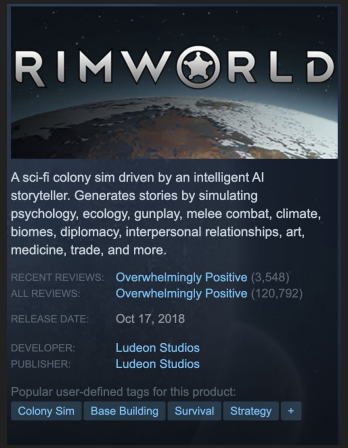
Introduction

- The goal of this project was to create a recommendation system
- Use NLP and unsupervised learning techniques to recommend games to consumers on the Steam online game marketplace
- Clients that would benefit would be Steam and their consumers
- Consumers would be recommended more games that they are interested in and buy more
- Steam sells more
- Both parties benefit

Methodology

- Scraped game descriptions from the Steam store website
- Each game has own webpage on the store
- Title, Description, Genre, Tags, Publisher, and Developer
- Only made use of Title and Description
- A total of 1,915 games' information was scraped (tried for 2,000)





Methodology Continued

ABOUT THIS GAME

Counter-Strike: Global Offensive (CS: GO) expands upon the team-based action gameplay that it pioneered when it was launched 19 years ago.

CS: GO features new maps, characters, weapons, and game modes, and delivers updated versions of the classic CS content (de_dust2, etc.).

"Counter-Strike took the gaming industry by surprise when the unlikely MOD became the most played online PC action game in the world almost immediately after its release in August 1999," said Doug Lombardi at Valve. "For the past 12 years, it has continued to be one of the most-played games in the world, headline competitive gaming tournaments and selling over 25 million units worldwide across the franchise. CS: GO promises to expand on CS' award-winning gameplay and deliver it to gamers on the PC as well as the next gen consoles and the Mac."

MATURE CONTENT DESCRIPTION

The developers describe the content like this:

Includes intense violence and blood.

Methodology Continued

- Created a corpus of game descriptions
- Each game description is a document
- CountVectorizer and TF-IDF to build document term matrices
- Latent Semantic Analysis, Non-Negative Matrix Factorization, and Latent Dirichlet Allocation for topic modeling
- Iterative process for cleaning the corpus and making the topic modeling better
    - View top words for each topic
    - Remove general, unnecessary, junk words (game, players, games, gamemode, etc.)
    - Repeat until satisfied

Methodology Continued

- With topic modeling finalized, the document-topic matrix used for the data points for recommendation system
- Final model:
  - Doc-term: TF-IDF
  - Topic Modeling: Non-Negative Matrix Factorization
  - 10 topics
- 10 column matrix reduced to 3-D using Principal Component Analysis
- Simple Recommendation System: recommended games are the closest games in euclidean distance

Results (Topics)

Hero Adventure	heroes, monsters, characters, character, adventure
Real Time Strategy	war, units, strategy, command, campaign
Cars/Racing	car, cars, racing, race, simulator
Survival	build, resources, island, craft, building
Action Story-Based	mature, weapons, combat, violence, story
Simulation/Management/Tycoon	support, company, products, guidance, warriors
Space/Futuristic	space, ship, galaxy, star, ships
Comic/Movie Heroes	city, batman, lego, marvel, arkham
Sports Simulation/Management	football, manager, team, club, soccer
Multiplayer Shooter	duty, multiplayer, warfare, modern, wwii

Results (Example)

If user likes “DayZ” recommends:

Dead Maze

Night of the Dead

Kenshi

Green Hell

Project Zomboid

The Isle

Graveyard Keeper

Stranded: Alien Dawn

The Wild Eight

The Forest

Results (Example)

If user likes “iRacing” recommends:

Project CARS 2

DiRT 4

Asphalt 9: Legends

BeamNG.drive

My Summer Car

DiRT Rally 2.0

RaceRoom Racing Experience

Automobilista 2

American Truck Simulator

Results (Example)

If user likes “PUBG” recommends:

Requiem: Desiderium Mortis

A Plague Tale: Innocence

Dishonored®: Death of the Outsider™

Katana ZERO

The Bridge Curse Road to Salvation

Detroit: Become Human

Borderlands 3

NARAKA: BLADEPOINT

DEATHLOOP

Bayonetta

Conclusions

- A recommendation system was made to recommend games using written descriptions on the store pages
  - Used NLP and unsupervised learning techniques
- Topic modeling and dimensionality reduction was used to give the games coordinates for distance calculations
- System performs well for most games, but sometimes recommendations can be questionable

Future Work

- Make use of the other scraped columns
- Improvement on preprocessing the text data
- Simple web app
- Add more games
- Incorporate other marketplaces together
- Further investigate poor recommendations

Questions?