IEEE BigData 2021 Cup: Predicting Victories in Video Games
Andrzej Janusz, Maciej Świechowski, Rafał Tyl
AGENDA

- Tactical Troops: Anthracite Shift
  - The game
  - Tactical Troops Analytics portal
  - The Fogs of War project
- The new challenge at KnowledgePit.ml
  - The scope and research aims
  - What have we learned so far?
  - Our plans for the following months
Tactical Troops: Anthracite Shift
Tactical Troops: Anthracite Shift - general information

1. Features:
   a. Turn-based, tactical game
   b. Gridless, physics-based gameplay
   c. Teleportation
   d. Singleplayer, 30 hours of gameplay
   e. Multiplayer, 24 different maps
   f. Rich selection of tools of destruction: 34 weapons, 7 gadget types
   g. Advanced, diverse opponent AI

2. Official website: https://tacticaltroops.net/
3. Twitter QED Games Team: https://twitter.com/QEDGamesTeam
4. Steam: https://store.steampowered.com/app/1266890/Tactical_Troops_Anthracite_Shift/
5. Youtube - Devastator mode explained: https://www.youtube.com/watch?v=r7CIG9sVLgo
6. Youtube - Domination mode explained: https://www.youtube.com/watch?v=WY6_nnG3DuF
Gridless, physics-based gameplay with destructible environment

- Turn-based
- Movement, combat and all environment interactions are based on 2D physics
- Many obstacles can be pushed around and serve many different purposes, like building temporary covers
The goal of the game (multi-player)

- Eliminate all opponent's units
- Only one in a given match
  - Domination mode
  - Devastator mode
Domination game mode

In this game mode, the goal is to secure two out of three control points and hold them for two full rounds.
Devastator game mode

In this game mode, the goal is to destroy all generators belonging to your enemy
Teleportation

- Units, grenades, bullets and crates can be teleported
- There’s a special weapon class (plasma weapons) that bypasses the teleportation mechanic
- Teleportation makes map navigation and target selection much more complicated
Rich equipment selection

- **Weapon types:** SMG, machine guns, rifles, plasma rifles, sniper rifles, rocket launchers
- **Gadget types:** grenades, shields, mortars, stimulants, mines
Many diverse map types
Bots in Tactical Troops

Strategy:
- balancing victory conditions
- general goals assignment to units
- forming “ad-hoc” attack squads

Tactics:
- choosing actions (e.g. attack vs throw vs movement)
- movement paths
- coordination

Heuristics:
- possible movement paths
- choosing the right weapon
Bots in Tactical Troops

- Hierarchical combination of Utility AI and Monte Carlo Tree Search
Tactical Troops Analytics
Users of Tactical Troops Analytics (TTA)

- Part of Sensei R&D project
- Analytics for games
- Tactical Troops is one of proof-of-concepts
- One of many facets of AI/ML in games
Logs as the main source of data

- Incremental logs directly from the game
- “Logs are the food”
- Analysis outside from the game
Feature example 1: win chances

Goal: predict chances for the win → [0,1] based on starting parameters

Method: neural networks, embeddings
Feature example 2: loadout recommender / completer

**Goal:** predict strong and fun loadout for the player (with constraints)

**Method:** XGBoost; a special way of training
Feature example 3: formations analysis

Goal: observe how players group their units; which go together; centers
Method: DBSCAN + cluster analysis + heuristics
Feature example 4: playstyle & players clustering

**Goal:** individual player characteristic, comparing players

**Method:** mRMR, feature selection, smooth normalization, GUI widgets
Feature example 5: heatmaps

Goal: show ``where” and ``how much” something happens on a map
Method: heatmaps
Feature example 6: direct statistics

- Player efficiency & ranking
- Weapon efficiency
- Popular choices
- Basic game information
Selected other aspects

- Data processing pipeline
- Motivation for players e.g. forum with gamification
- Motivation for data scientists
- Taxonomy of players
- Other games analyzed: Clash Royale
Fogs of War project summary

1. Part 1: game AI agents dealing with uncertainty and partial information
   a. Deducing missing information from current beliefs of AI agents: logical and probabilistic reasoning
   a. Making decisions based on incomplete or uncertain data: modified implementations of Utility AI, Monte Carlo Tree Search and other techniques
   a. Cooperation and group belief formation
   a. Explaining game AI behavior in a way understandable to humans (developers and players)

1. Part 2: Automated Turing test - an ML model for discerning human players from AI bots, based on game logs and additional data
THE NEW CHALLENGE AT KnowledgePit
Our competitions related to video games (so far)

➢ **Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction**

➢ **AAIA'18 Data Mining Challenge: Predicting Win-rates of Hearthstone Decks**

➢ **AAIA'17 Data Mining Challenge: Helping AI to Play Hearthstone**

➢ All of the above competitions were organized in association with the FedCSIS conference series.

➢ Data sets are available on request.
HOW DOES IT WORK?

A typical competition schema

1. The available data set is divided into the training and test parts.
2. Target values (e.g. labels) for the test set are hidden from participants – they have to be predicted.
3. Participants submit solutions which are assessed on a sample from the test set.
4. Participants select their most reliable models and write short reports.
5. The final solutions are evaluated on the remaining test data.
IEEE BigData 2021 Cup: Predicting Victories in Video Games

➢ April 30, 2021: start of the competition, datasets became available,
➢ August 31, 2021: deadline for submitting the solutions,
➢ September 5, 2021: reports due, end of the competition,
➢ Awards sponsored by QED Software:
  ○ 1750 USD cash prize pool for the top 3 teams
  ○ 3x costs of IEEE BigData 2021 registrations
  ○ publication opportunity for authors of interesting solutions
Multimodal data representations

The three main data representation types available for participants:
➢ Truncated game logs
➢ Flattened log format
➢ Tabular (aggregated) format

Additional data about units, gadgets, and maps is available as metadata in separate files.

The main challenge:
how to utilize multimodal data views to facilitate the prediction of winners at a given time point in a game?
Truncated logs

➢ Full game logs truncated at the decision time - allow to reproduce the course of the game.
➢ 7.5GB + 3.8GB of data.
➢ Separate JSON file for each game.
➢ Logs contain each taken action and unit status change.
Flattened logs

- Additional 1.5GB training + 0.8GB test data.
- JSON files whose names correspond to game ids
- Detailed description of the situation on a map (including positions of each troop, etc.).
- Detailed description of JSONs is available in competition files.
Tabular data

- 38,658 training games + 20,000 test instances.
- Id column allows joining data from other representations.
- A mixture of 197 numeric and symbolic features.
- Simple characteristics of each troop on the map.
- Additional information on units and weapons in meta-data files.
Data acquisition process
We will release one more dataset - game screens.
Each screen will show the situation on a map at the prediction time.
We are considering adding additional screens from preceding game turns - is it worth it?
planned release date: May 31
Participation so far...

➢ 61 registered teams
➢ enrolled participants from 15 countries
➢ 260 submitted solutions
➢ we expect those numbers to grow :-) 
➢ currently, the best Leaderboard result is 0.8854 (AUC)
Uses only the tabular representation.
Standardized and mean-filled numerical features.
One-hots of categorical features with a few distinct values (<10 distinct categories)
Ordinal-encodes the remaining symbolic features.
XGBoost hyper-parameter search
0.8733 preliminary score
Most important feature: HP (obviously)
Summary

● The main goal is to create efficient models for predicting game outcomes in Tactical Troops: Anthracite Shift
  ○ results can be deployed in the Tactical Troops Analytics portal empowered by the SENSEI system
  ○ trained models can be used to improve AI bots
● The big question is whether the multimodal representations of data can help in the construction of better models
● The competition is attracting attention from around the world but we still hope for more :-)
Thank you for your attention!

Andrzej Janusz
andrzej.janusz@qed.pl

Maciej Świechowski
maciej.swiechowski@qed.pl

Rafał Tyl
rafal.tyl@qed.pl