Introduction

IMGD 2905

What is data analysis for game development?

• Using game data to inform the game development process
• Where does this data come from?
  → Users playing game
  – Quantitative (instrumented)
  – Qualitative (subjective evaluation)
  – (But often lots more of the former!)

What can game analysis do for game development?

• Improve level design – e.g., see where players are getting stuck
• Focus development on critical content – e.g., see what game modes or characters are not used
• Balance gameplay – e.g., tune parameters for more competitive and fun combat
• Broaden appeal – e.g., hear if content/story is engaging or repulsing

Why is data analysis for game development needed?
Why is data analysis for game development needed?

- **Challenge**
  - Games gotten larger and more complex
  - Number of reachable states, characters
  - Need for metrics to make sense of player behavior has increased
- **Opportunity**
  - New technologies enable aggregation, access and analysis

IMGD 2905 – Doing Data Analysis for Game Development

- **Data analysis pipeline** – get data from games, through analysis, to stakeholders
- **Summary statistics** – central tendencies of data
- **Visualization of data** – how to display analysis, illustrate messages
- **Statistical tests** – quantitatively determine relationships (e.g., correlation)
  - Probability needed as foundation
- **Regression** – model relationships
- More advanced topics (e.g., ML, Data management...)

For this class:
- Described in lecture
- Read about in book
- Applied in projects

Foundations for Data Analysis for Game Development at WPI

- **Statistics classes**
  - MA 2610 Applied Statistics for Life Sciences
  - MA 2611 Applied Statistics I
  - MA 2612 Applied Statistics II
- **Probability classes**
  - MA 2621 Probability for Applications
- **Data Science minor**
  - CS 4445 Data Mining and Knowledge Discovery in Databases
- **Other**
  - CS 3004 Introduction to Programming for Non-Majors
  - CS 3431 Database Systems I

Outline

- **Overview** *(done)*
- **Game Analytics Pipeline** *(next)*
- **Game Data Analysis Examples**

Sources of Game Data

**Quantitative (Objective)**
- Internal Testing
  - Developers
  - QA
- External Testing
  - Usability testing
  - Beta tests
  - Long-term play data

**Qualitative (Subjective)**
- Surveys
- Reviews
- Online communities
- Post mortems

How to get from data to dissemination?

Game analytics pipeline

Game Analytics Pipeline

- Game
- Extracted Data
- Analysis
- Exploration
- Charts and Tables
- Statistical Tests
- Dissemination
  - Presentation
  - Report
Game Analytics Pipeline - Example

**Game Analytics Tools**

- **Games** – breadth of experience with games, specific experience with game to be analyzed
- **Statistics** – measures of central tendency, measures of spread, statistical tests
- **Probability** – rules, distributions
- **Data Visualization** – bar chart, scatter plot, histogram, error bars
- **Technical Writing and Presentation** – white paper, technical talk; audience is peer group, developers, boss

**Game Analytics Pipeline - Example**

- **Dissemination**
  - PowerPoint
  - Word

**Outline**

- **Overview** (done)
- **Game Analytics Pipeline** (done)
- **Game Data Analysis Examples** (next)

**Example:** Project Gotham Racing 4

- **Publisher** – Microsoft 2007
  - 134 vehicles, 9 locations, 10 game modes
- **Analyzed data**
  - (Authors worked at Microsoft)
  - 3.1 million log entries, 1000s of users

**Game Mode**

- **OFFLINE_CAREER**
  - Races: 1479586, % Total: 47.63%
- **PGR_ARCADE**
  - Races: 566705, % Total: 18.24%
- **NETWORK_PLAY**
  - Races: 584201, % Total: 18.81%
- **SINGLEPLAYER_PLAY**
  - Races: 185415, % Total: 5.97%
- **NET_TOURNY_ELIM**
  - Races: 2713, % Total: 0.09%

**Group**

- **STREET_RACE**
  - Races: 795334, % Total: 25.60%
- **NET_STREET_RACE**
  - Races: 543491, % Total: 17.50%
- **ELIMINATION**
  - Races: 216042, % Total: 6.95%
- **HOTLAP**
  - Races: 195949, % Total: 6.31%

**Thoughts?**

- What are some main messages?

**Project Gotham Racing 4: Results**

- **Mode**
  - Offline career dominates
  - Network tournament hardly used
- **Events**
  - Street race and network street race dominate
  - Cat and mouse never used
- **Vehicles (not shown)**
  - 1/3 used in less than 0.1% of races
Project Gotham Racing 4: Conclusion

• Content underused - 30-40% of content in less than 1% of races
• Use to shift emphases for DLC, next version
  – Asset creation costs significant, so even 25% reduction noticeable
• Other (not shown)
  – Encouraging new players to play career mode
  • Increasing likelihood of continuing play
  – Encouraging new players to stay with F Class longer
  • Rather than move to more difficult to control A Class

Halo 3: Results

• Thoughts?
• What are some main messages?

Example: League of Legends

• Publisher – Riot Games 2009
  – Rank: ~5 Tiers, 5 divisions each → 25
• User study (~2 players)
  – Play LoL in controlled environment
  – Record objective data
  • (e.g., player rank and game stats)
  – Provide survey for subjective data
  • (e.g., match balance and enjoyment)

League of Legends: Results

• Publisher – Microsoft 2007
  – Achievements: single player missions, challenges such as finding skulls, multiplayer accomplishments...
• Analyzed data
  – (Author worked at Microsoft)
  – 18,000 players
League of Legends: Results

Objective

- Most teams are balanced
  - About 10% more than 3 from mean
- Game is balanced
- Win?
- Lose?

Subjective

- Win?
  - Game is fun (70%), never not fun
- Lose?
  - Game is almost never fun (90%)

League of Legends: Conclusion

Objective

- Teams are balanced
  - 50% players within 1 rank of each other
- Games are balanced
  - 80% teams within 1 average rank of each other

Subjective

- Games are not balanced
- When players win, perceive slight imbalance
- When players lose, perceive large imbalance
- Players enjoy winning more than losing (no surprise)
- (Surprise!) Players most enjoy matches imbalanced in their favor!

Summary

- Data analysis for games increasingly important
  - Has potential to improve game development
- Knowledge and skills required
  - Scripting
  - Statistics
  - Data analysis
  - Writing and presentation

Matchmaking systems may want to consider - e.g., balance not so important, as long as player not always on imbalanced side.