CE810 - Game Design 2

Evaluating Performance

Joseph Walton-Rivers & Piers Williams

Tuesday, 15 May 2018

University of Essex

CE810 - Game Design 2
Evaluating Performance
Joseph Walton-Rivers & Piers Williams
Tuesday, 15 May 2018
University of Essex

What is Player Experience?

Player experience

Collection of events that **occur** to the player **during** the game

What is Player Experience?

Player experience
Collection of events that occur to the player during the game

Should be clear - it is only the events that occur because of the game that are important

What is Player Experience?

Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

- Losing a Unit
- Laundry Finishing
- Collecting resource
- New message in chat window
- Unit Moving

What is Player Experience?
Scenario
Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.
Question
Which of these are a part of the player experience and which are not?
Losing a Unit
Laundry Finishing
Collecting resource
New message in chat window
Unit Moving

All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

What is Player Experience?

Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

Losing a Unit	Yes
Laundry Finishing	
Collecting resource	
New message in chat window	
Unit Moving	

What is Player Experience?
Scenario
Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.
Question
Which of these are a part of the player experience and which are not?
Losing a Unit | Yes
Laundry Finishing
Collecting resource
New message in chat window
Unit Moving

All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

What is Player Experience?

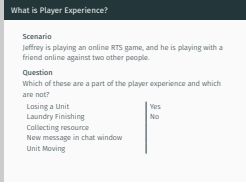
Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

Losing a Unit	Yes
Laundry Finishing	No
Collecting resource	
New message in chat window	
Unit Moving	



All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

What is Player Experience?

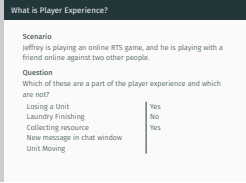
Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

Losing a Unit	Yes
Laundry Finishing	No
Collecting resource	Yes
New message in chat window	
Unit Moving	



All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

What is Player Experience?

Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

Losing a Unit	Yes
Laundry Finishing	No
Collecting resource	Yes
New message in chat window	Yes
Unit Moving	

What is Player Experience?
Scenario
Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.
Question
Which of these are a part of the player experience and which are not?
Losing a Unit — Yes
Laundry Finishing — No
Collecting resource — Yes
New message in chat window — Yes
Unit Moving

All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

What is Player Experience?

Scenario

Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.

Question

Which of these are a part of the player experience and which are not?

Losing a Unit	Yes
Laundry Finishing	No
Collecting resource	Yes
New message in chat window	Yes
Unit Moving	Yes

What is Player Experience?
Scenario
Jeffrey is playing an online RTS game, and he is playing with a friend online against two other people.
Question
Which of these are a part of the player experience and which are not?
Losing a Unit Yes
Laundry Finishing No
Collecting resource Yes
New message in chat window Yes
Unit Moving Yes

All happen while the person is playing a game

Anything that occurs during the game and as part of the game is part of the player experience. Which of these can be detected by an AI?

Metrics

Metrics

Collect data on how players/bots work

Activity

What kinds of features can we collect?

Collect data on how players/bots work
Activity
What kinds of features can we collect?

Data from humans

- High-level human experience

Data from humans
- High-level human experience

Data from humans

- High-level human experience
  - Final game scores?

Data from humans
- High-level human experience
  - Final game scores?

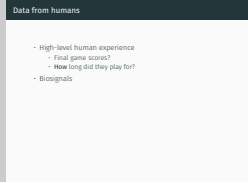
Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?

Data from humans
- High-level human experience
  - Final game scores?
  - **How** long did they play for?

Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals



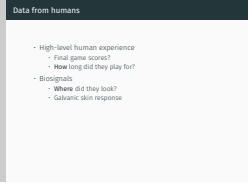
Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?

Data from humans
- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?

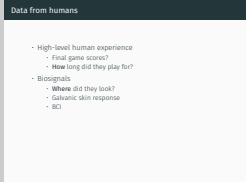
Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response



Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI



Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI
- Surveys and interviews

Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI
- Surveys and interviews

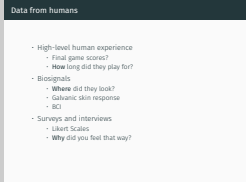
Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI
- Surveys and interviews
  - Likert Scales

Data from humans
- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI
- Surveys and interviews
  - Likert Scales

Data from humans

- High-level human experience
  - Final game scores?
  - **How** long did they play for?
- Biosignals
  - **Where** did they look?
  - Galvanic skin response
  - BCI
- Surveys and interviews
  - Likert Scales
  - **Why** did you feel that way?



Data from bots

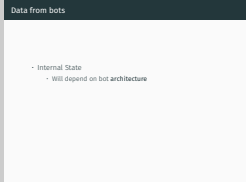
- Internal State

Data from bots
- Internal State

- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from bots

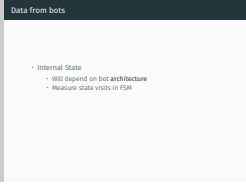
- Internal State
  - Will depend on bot **architecture**



- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from bots

- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM



- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from bots

- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?

Data from bots
- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?

- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from bots

- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?
- How many times does a bot face a **difficult** choice?

Data from bots

- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?
- How many times does a bot face a **difficult** choice?

- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from bots

- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?
- How many times does a bot face a **difficult** choice?
  - What is a difficult choice?

Data from bots
- Internal State
  - Will depend on bot **architecture**
  - Measure state visits in FSM
  - Did the game make **full** use of the AI?
- How many times does a bot face a **difficult** choice?
  - What is a difficult choice?

- What does it mean if it doesn't make full use of the AI?
- Difficult Choice: MCTS - near identical branches, GA - No Convergence
- What does it signify about the game?

Data from either

- Final Score distribution

Data from either
- Final Score distribution

Some things can be measured regardless of if a human or AI is playing

- How high, variation?
- Length, range of lengths
- Runaway victory?, keep changing hands? loop?
- Some states not used at all? Some overused?
- How to measure this?

Data from either

- Final Score distribution
- Game Duration

Data from either
- Final Score distribution
- Game Duration

Some things can be measured regardless of if a human or AI is playing

- How high, variation?
- Length, range of lengths
- Runaway victory?, keep changing hands? loop?
- Some states not used at all? Some overused?
- How to measure this?

Data from either

- Final Score distribution
- Game Duration
- Score “Drama”

Data from either
- Final Score distribution
- Game Duration
- Score "Drama"

Some things can be measured regardless of if a human or AI is playing

- How high, variation?
- Length, range of lengths
- Runaway victory?, keep changing hands? loop?
- Some states not used at all? Some overused?
- How to measure this?

Data from either

- Final Score distribution
- Game Duration
- Score “Drama”
- Statistical distribution of states

Data from either
- Final Score distribution
- Game Duration
- Score “Drama”
- Statistical distribution of states

Some things can be measured regardless of if a human or AI is playing

- How high, variation?
- Length, range of lengths
- Runaway victory?, keep changing hands? loop?
- Some states not used at all? Some overused?
- How to measure this?

Data from either

- Final Score distribution
- Game Duration
- Score “Drama”
- Statistical distribution of states
- Degree of challenge

Data from either
- Final Score distribution
- Game Duration
- Score “Drama”
- Statistical distribution of states
- Degree of challenge

Some things can be measured regardless of if a human or AI is playing

- How high, variation?
- Length, range of lengths
- Runaway victory?, keep changing hands? loop?
- Some states not used at all? Some overused?
- How to measure this?

Data from populations

- Variability of scores
- Skill-depth

Data from populations
- Variability of scores
- Skill-depth

Action Sequences

Data from either

- Actions taken
- Record the sequence of button-pushes

Data from either
- Actions taken
- Record the sequence of button-pushes

Entropy

- Sometimes used to **interpret** aspects of player experience

Entropy
- Sometimes used to **interpret** aspects of player experience

- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i) \log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0) \log_2 P(x_0)) + (P(x_1) \log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9 \log_2 0.9) + (0.1 \log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

Entropy

- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$

Entropy
- Sometimes used to interpret aspects of player experience
- $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$

- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9\log_2 0.9) + (0.1\log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

Entropy

- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?

Entropy
- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?

- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9\log_2 0.9) + (0.1\log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

Entropy

- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?
  - $H(fairCoint) = \sum_{i=1}^{2}(\frac{1}{2})\log_2(\frac{1}{2}) = -\sum_{i=1}^{2}(\frac{1}{2}) \times (-1) = 1$

Entropy
- Sometimes used to interpret aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?
  - $H(fairCoint) = \sum_{i=1}^{2}(\frac{1}{2})\log_2(\frac{1}{2}) = -\sum_{i=1}^{2}(\frac{1}{2}) \times (-1) = 1$

- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9\log_2 0.9) + (0.1\log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

Entropy

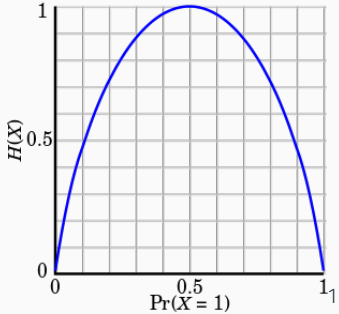
- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?
  - $H(fairCoint) = \sum_{i=1}^{2}(\frac{1}{2})\log_2(\frac{1}{2}) = -\sum_{i=1}^{2}(\frac{1}{2}) \times (-1) = 1$
  - How about an unfair coin? What is the entropy for a coin of probability 0.9?

Entropy
- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$
  - Take a fair coin - how much entropy?
  - $H(fairCoint) = \sum_{i=1}^{2}(\frac{1}{2})\log_2(\frac{1}{2}) = -\sum_{i=1}^{2}(\frac{1}{2}) \times (-1) = 1$
  - How about an unfair coin? What is the entropy for a coin of probability 0.9?

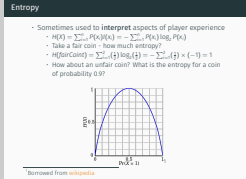
- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9\log_2 0.9) + (0.1\log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

Entropy

- Sometimes used to **interpret** aspects of player experience
  - $H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$
  - Take a fair coin - how much entropy?
  - $H(fairCoint) = \sum_{i=1}^{2} (\frac{1}{2}) \log_2 (\frac{1}{2}) = -\sum_{i=1}^{2} (\frac{1}{2}) \times (-1) = 1$
  - How about an unfair coin? What is the entropy for a coin of probability 0.9?



[1]Borrowed from wikipedia



- We won't worry too much about the middle definition
- Because it is a fair coin - each toss can tell us nothing
- Whiteboard time if students stuck:
  - $P(x_0) = 0.9, P(x_1) = 0.1$
  - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i) \log_2 P(x_i) =$
  - Continued: $-\Big((P(x_0) \log_2 P(x_0)) + (P(x_1) \log_2 P(x_1))\Big) = 0.47$
  - Continued: $-\Big((0.9 \log_2 0.9) + (0.1 \log_2 0.1)\Big) = 0.47$
- Check our answer from earlier matches the diagram

A Game Example

loc	visits	p(loc)	calc
0,0	10	0.067	$0.067 \log_2(0.067)$

loc	0	1	2
0	10	20	15
1	12	35	13
2	15	20	10

150	Total:	

A Game Example

- Some sample 2D location visit counts
- Converted into visit counts as fraction of total and then into probability of having visited that location
- Then we just perform the math as a giant summation. Computers are good at this
- Except computers are not keen on 0's

A Game Example

loc	0	1	2
0	10	20	15
1	12	35	13
2	15	20	10

loc	visits	p(loc)	calc
0,0	10	0.067	$0.067 \log_2(0.067)$
0,1	12	0.08	$0.080 \log_2(0.008)$
0,2	15	0.1	$0.100 \log_2(0.100)$
1,0	20	0.134	$0.134 \log_2(0.134)$
1,1	35	0.234	$0.234 \log_2(0.234)$
1,2	20	0.134	$0.134 \log_2(0.134)$
2,0	15	0.1	$0.100 \log_2(0.100)$
2,1	13	0.0867	$0.0867 \log_2(0.0867)$
2,2	10	0.067	$0.067 \log_2(0.067)$
	150	Total:	

A Game Example

- Some sample 2D location visit counts
- Converted into visit counts as fraction of total and then into probability of having visited that location
- Then we just perform the math as a giant summation. Computers are good at this
- Except computers are not keen on 0's

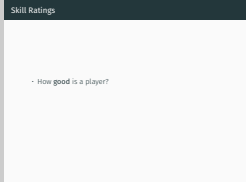
Exercise

Now you try - in Java. Download the here and calculate the entropy

A Game Example
Exercise
Now you try - in Java. Download the here and calculate the entropy

Skill Ratings

- How **good** is a player?



- And how do we represent this?
- Based on observations, was it enough? Watch F1 at one track and use those observations for another?
- Usually this is the case in games
- Does close win rates prove a lack of skill depth? No, current set of players doesn't demonstrate it. Like me and Joe playing Pool

Skill Ratings

- How **good** is a player?
- What is the **issue** with win rates?

Skill Ratings
- How **good** is a player?
- What is the **issue** with win rates?

- And how do we represent this?
- Based on observations, was it enough? Watch F1 at one track and use those observations for another?
- Usually this is the case in games
- Does close win rates prove a lack of skill depth? No, current set of players doesn't demonstrate it. Like me and Joe playing Pool

Skill Ratings

- How **good** is a player?
- What is the **issue** with win rates?
- If A > B and B > C is A >C?

Skill Ratings
- How **good** is a player?
- What is the **issue** with win rates?
- If A > B and B > C is A >C?

- And how do we represent this?
- Based on observations, was it enough? Watch F1 at one track and use those observations for another?
- Usually this is the case in games
- Does close win rates prove a lack of skill depth? No, current set of players doesn't demonstrate it. Like me and Joe playing Pool

Elo Ratings

- Elo is based on probability

Elo Ratings
- Elo is based on probability

- Designed for chess
- Point difference between players denotes the probability of winning
- Advantage of 100 points = 64% chance of winning
- Advantage of 200 points = 76% chance of winning
- Works by taking points from the loser and giving them to the winner. Number transfered proportional to difference between points

Elo Ratings

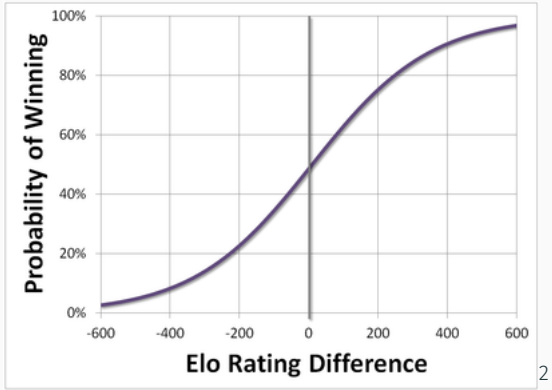
- Elo is based on probability
- $Elo(A) - Elo(B) = P(A \text{ beats } B)$

Elo Ratings
- Elo is based on probability
- $Elo(A) - Elo(B) = P(A \text{ beats } B)$

- Designed for chess
- Point difference between players denotes the probability of winning
- Advantage of 100 points = 64% chance of winning
- Advantage of 200 points = 76% chance of winning
- Works by taking points from the loser and giving them to the winner. Number transfered proportional to difference between points

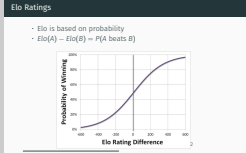
Elo Ratings

- Elo is based on probability
- $Elo(A) - Elo(B) = P(A \text{ beats } B)$



[2]

[2] Borrowed from liquipedia



- Designed for chess
- Point difference between players denotes the probability of winning
- Advantage of 100 points = 64% chance of winning
- Advantage of 200 points = 76% chance of winning
- Works by taking points from the loser and giving them to the winner. Number transfered proportional to difference between points