CE810 GD2 50-8-02-22 2018-05-25

CE810 - Game Design 2 Evaluating Performance

> Joseph Walton-Rivers & Piers Williams Tuesday, 15 May 2018

CE810 - Game Design 2

Evaluating Performance

Joseph Walton-Rivers & Piers Williams

Tuesday, 15 May 2018

University of Essex

Player experience

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

CE810 GD2 57-50 8002 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

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 CE810 GD2

What is Player Experience?

Jeffrey is playing an online RTS g friend online against two other p	
Question Which of these are a part of the are not?	player experience and whi
Losing a Unit Laundry Finishing Collecting resource New message in chat window Unit Moving	

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?

Yes

CE810) GD2
5-25	
18-0	
201	└─What is Player Experience?

What is Player Experience?

Scenario Jeffrey is playing an online RTS ga friend online against two other pe	
Question Which of these are a part of the p are not?	layer experience and which
Losing a Unit Laundry Finishing Collecting resource New message in chat window Unit Moving	Yes

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	

	CE810 GD2
25	
)5-	
018	└──What is Player Experience?
2	

What is Player Experi	ience?	
-----------------------	--------	--

Jeffrey is playing an online RTS gas friend online against two other pe	
Question Which of these are a part of the pl are not?	ayer experience and whic
Losing a Unit Laundry Finishing Collecting resource New message in chat window Unit Moving	Yes No

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	Yes No
Collecting resource	Yes
New message in chat window	
Unit Moving	

	CE810 GD2
-25	
05-	
-0	
201	└─What is Player Experience?

What is	Player	Experience?
---------	--------	-------------

Jeffrey is playing an online RTS gar friend online against two other per	
Question	
Which of these are a part of the pl	ayer experience and whi
are not?	
Losing a Unit	Yes
Laundry Finishing	No
Collecting resource	Yes
New message in chat window	
Unit Moving	

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	

CE8	10 GD2	
.25		
05-		
18-0		
201	└─What is Player Experience?	

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 Al?

What is Player Experience?

are not?

friend online against two other people. Question

New message in chat window

Which of these are a part of the player experience and which

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
Laundry FinishingNoCollecting resourceYes
New message in chat window Yes
Unit Moving Yes

(CE810 GD2
5-25	
18-0	
20	└─What is Player Experience?

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?

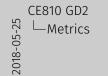
What is Player Experience?

are not?

friend online against two other people. Question

New message in chat window

Which of these are a part of the player experience and which



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?

High-level human experience

• High-level human experience

5

CE810 GD2

└_ Metrics

└─ Data from humans

2018-05-25

High-level human experience
 Final game scores?

- High-level human experience
 - Final game scores?

Data from humans

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

CE810 GD2 Hetrics CE810 GD2 Metrics CE810 GD2 CE810 GD2

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

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

CE810 GD2 Hetrics CE810 GD2 Metrics CE810 GD2 CE810 GD2 Metrics

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

CE810 GD2 ^{K7} ⁴⁰ ⁸⁰

Data from humans

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

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

CE810 GD2 Hetrics CE810 GD2 Metrics CE810 GD2 CE810 GD2 Metrics

- High-level human experience

Data from humans

Final game scores?
 How long did they play for?
Biosignals
 Where did they look?
 Galvanic skin response

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

CE810 GD2 Herrics CE810 GD2 Metrics CE810 GD2 CE810 GD2

High-level human experience
 Final game scores?

Data from humans

How long did they play for?
 Biosignals

Where did they look?
 Galvanic skin response
 BCI

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

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

CE810 GD2

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

CE810 GD2 CE810 GD2 -Metrics -B -B -Data from humans

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

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?

• Internal State



• What does it mean if it doesn't make full use of the AI?

Data from bots

Internal State

- Difficult Choice: MCTS near identical branches, GA No Convergence
- What does it signify about the game?

• Internal State

• Will depend on bot architecture

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?

Internal State
 Will depend on bot architecture
 Measure state visits in PSM

• 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?

CE810 GD2

-Metrics

└─ Data from bots

2018-05-25

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?

- 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?

CE810 GD2 ²⁷-5 ²⁶ ²⁶ ²⁷-5 ²⁶ ²⁷-5 ²⁶ ²⁷-5 ²⁶ ²⁷-7 ²⁶ ²⁷-7 ²⁶ ²⁷-7 ²⁷-7 ²⁶ ²⁷-7 ²⁷



Data from bots

Will depend on bot architecture
 Moacure 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?

- 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?

• Final Score distribution



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

Data from either

- 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?

- Final Score distribution
- Game Duration

CE810 GD2 CE810 GD2 Metrics BC CE810 GD2 CE810 GD2

Data from either

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?

7

- Final Score distribution
- Game Duration
- Score "Drama"

CE810 GD2 ²⁷⁻ ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ^{40-</sub> ^{40-</sub> ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ^{40-</sub> ⁴⁰⁻ ^{40-</sub>}}}}}}}}}</sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup></sup>

Final Score distributio
 Game Duration
 Score "Drama"

Data from either

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?

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

- Final Score distribution
- Game Duration
- Score "Drama"
- \cdot Statistical distribution of states
- $\cdot\,$ Degree of challenge

CE810 GD2 ^{K2} Data from either

Final Score distribution Game Duration Score "Drama" Statistical distribution of states Deeree of challense

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

- How high, variation?
- $\cdot\,$ 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

• Variability of scores

• Skill-depth

CE810 GD2 -Action Sequences

Action Sequences

Action Sequences

Actions taken
 Record the sequence of button-oushes

- Actions taken
- Record the sequence of button-pushes

Entropy

• Sometimes used to interpret aspects of player experience

CE810 GD2 -Action Sequences

- \cdot 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: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))) = 0.47$
 - Continued: $-((0.9 \log_2 0.9) + (0.1 \log_2 0.1)) = 0.47$

 \cdot Check our answer from earlier matches the diagram

Entropy

• Sometimes used to **interpret** aspects of player experience • $H(X) = \sum_{i=1}^{n} P(x_i) l(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$ CE810 GD2 CE810 GD2 Action Sequences

- \cdot 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: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))) = 0.47$
 - Continued: $-((0.9 \log_2 0.9) + (0.1 \log_2 0.1)) = 0.47$
- $\cdot\,$ Check our answer from earlier matches the diagram

- 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?

CF810 GD2 2018-05-25 Action Sequences -Entropy

 $H(X) = \sum_{i=1}^{n} P(x_i) l(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$

- \cdot 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: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))) = 0.47$
 - Continued: $-((0.9 \log_2 0.9) + (0.1 \log_2 0.1)) = 0.47$
- $\cdot\,$ Check our answer from earlier matches the diagram

- 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?
 - $H(fairCoint) = \sum_{i=1}^{2} (\frac{1}{2}) \log_2(\frac{1}{2}) = -\sum_{i=1}^{2} (\frac{1}{2}) \times (-1) = 1$

CE810 GD2 CE810 Sequences

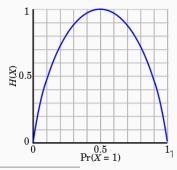
- \cdot We won't worry too much about the middle definition
- Because it is a fair coin each toss can tell us nothing
- Whiteboard time if students stuck:
 - $P(x_0) = 0.9, P(x_1) = 0.1$
 - Answer is: $H(dodgyCoin) = -\sum_{i=1}^{2} P(x_i) \log_2 P(x_i) =$
 - Continued: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2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

- 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?

ametimes used to interpret aspects of player experience \cdot H(X) = $\sum_{i=1}^{n} P(X_i) F(X) = -\sum_{i=1}^{n} P_i(X_i) F(X)$. Take a fair coin - how much entropy? \cdot H(fairCoint) = $\sum_{i=1}^{n} (\frac{1}{2}) \log_{2}(\frac{1}{2}) = -\sum_{i=1}^{n} (\frac{1}{2}) \times (-1) = 1$ \cdot How about an unfair coin? What is the entropy for a coin of probability 0.9?

- \cdot 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: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))) = 0.47$
 - Continued: $-((0.9 \log_2 0.9) + (0.1 \log_2 0.1)) = 0.47$
- $\cdot\,$ Check our answer from earlier matches the diagram

- 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?



CE810 GD2 CE810 GD2 Action Sequences



$f_{\text{rescaled}}^{1} \xrightarrow{p_{1} + 1} p_{2} + 1}$

- \cdot 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: $-((P(x_0)\log_2 P(x_0)) + (P(x_1)\log_2 P(x_1))) = 0.47$
 - Continued: $-((0.9 \log_2 0.9) + (0.1 \log_2 0.1)) = 0.47$
- $\cdot\,$ Check our answer from earlier matches the diagram

¹Borrowed from wikipedia

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:

- Some sample 2D location visit counts
- Converted into visit counts as fraction of total and then into probability of having visited that location

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 log₂(0.067)

150 Total:

- 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

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	0.0867log ₂ (0.0867)
2,2	10	0.067	0.067log ₂ (0.067)
	150	Total:	

CE810 GD2 Action Sequences

0	0 10 12	1 20 35	2 15 13	loc 0,0 0,1 0,2 1,0 1,1 1,2	visits 10 12 15 20 35 20	p(loc) 0.067 0.08 0.1 0.134 0.234 0.134	calc 0.067 log ₂ (0.067) 0.080 log ₂ (0.008) 0.100 log ₂ (0.100) 0.134 log ₃ (0.134) 0.234 log ₃ (0.234) 0.134 log ₃ (0.134)
2	15	20	10	2,0	15	0.1	0.100 log2(0.100)
				2,1	13	0.0867	0.0867 log2(0.0867)
				2,2	10	0.067	0.067 log2(0.067)
					150	Total:	

- 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

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

Exercise

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

• 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?

Skill Ratings

How good is a player?

- 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

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

CE810 GD2 CE810 GD2 Action Sequences So So So CE810 GD2 CE810 GD2 CE810 GD2 CE810 GD2 CE810 GD2 CE810 GD2 CE810 Sequences

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

Skill Ratings

- 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

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

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

Skill Ratings

- 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

 \cdot 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 is based on probability
- Elo(A) Elo(B) = P(A beats B)

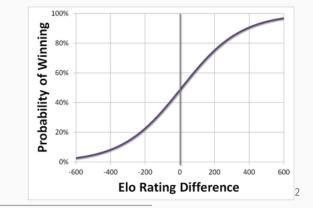
Elo is based on probability
 Elo(A) – Elo(B) = P(A beats B)

Elo Ratings

- 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

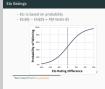
Elo Ratings

- Elo is based on probability
- Elo(A) Elo(B) = P(A beats B)



²Borrowed from liquipedia

CE810 GD2 CE810 GD2 Action Sequences CE810 GD2 CE81



- 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