Fighting Game Balance Through Feature Engineering & Machine Learning

Nick Chen, 2019; Princeton University COS 424

ABSTRACT

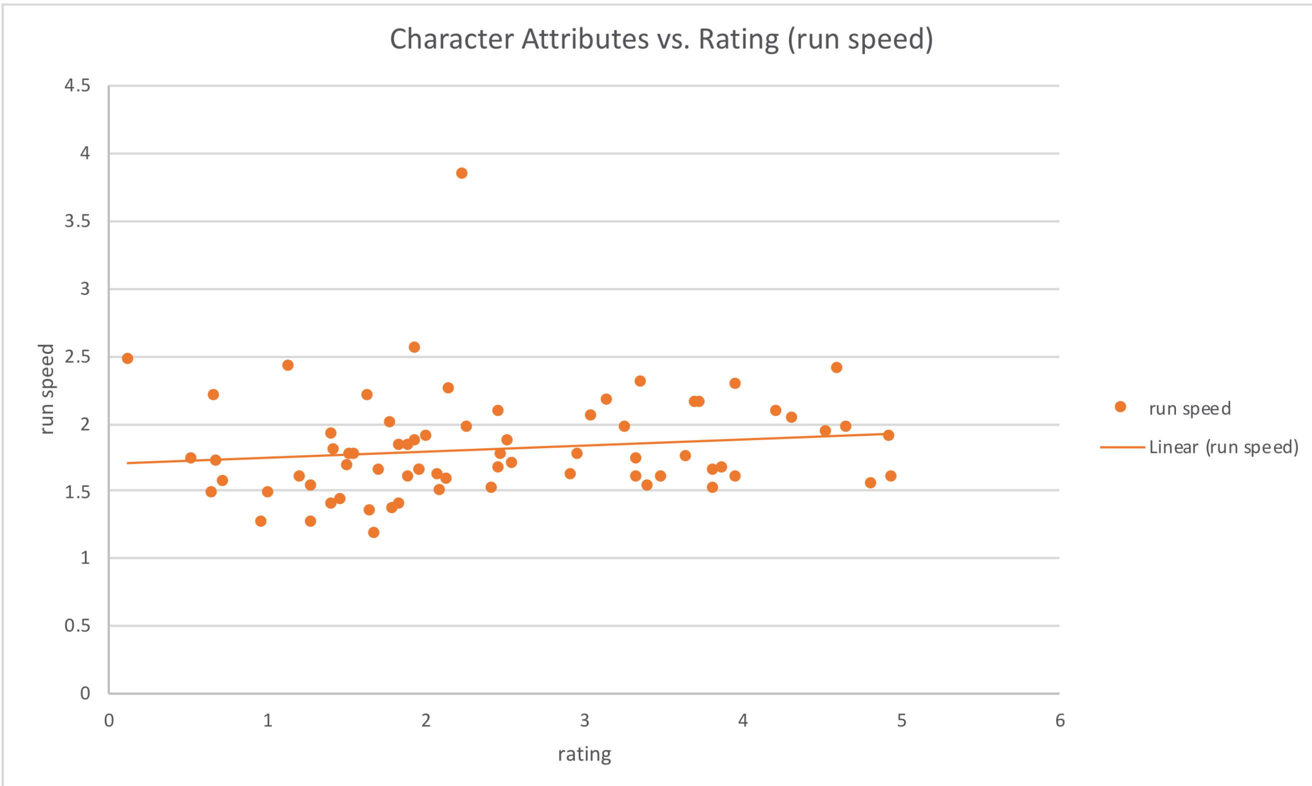
- a baseline study for analyzing the structure and balance of fighting video games
- feature engineering and machine learning methods are applied to a data set describing characters from the Super Smash Bros. Ultimate video game, where the characters' properties and frame data are the features, and their ratings on a scale of 0.0 to 5.0, estimated by a group of professional players, is the target
- because the data set is very small, regression methods produced heavily overfitted results, not only due to a minimal number of samples, but also because the scale of target ratings was entirely subjective
  - despite this, the initial models were partially successful at determining which characters were better than others, i.e. order of "goodness" between characters was largely preserved
- the task of predicting a character's overall performance was redefined as a classification problem, where the target was a character's "tier," A through E, as opposed to an exact numeric value; this change resulted in significantly reduced variance in the model

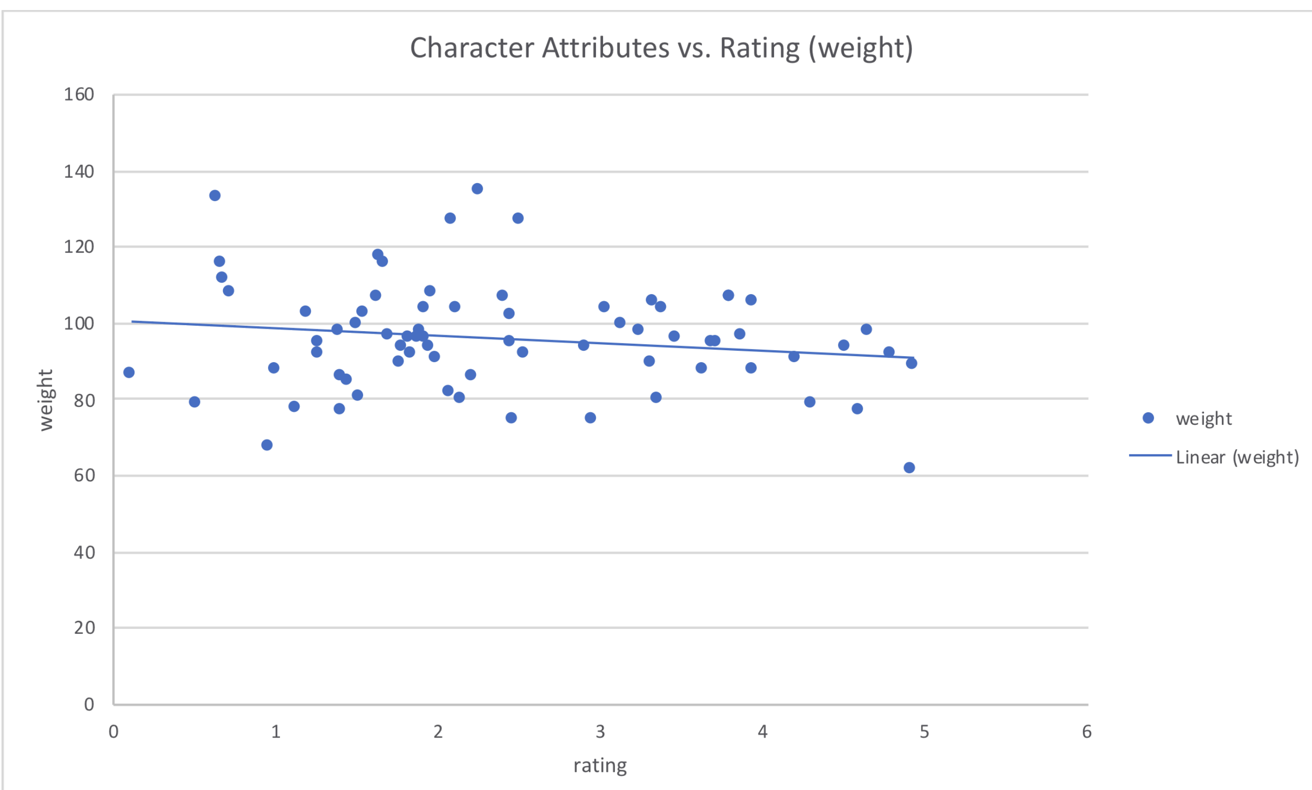
BACKGROUND

- game balance is one of the most difficult tasks facing video game developers today
  - increased popularity and profitability of professional "Esports" requires clearly structured and fair gameplay
- traditional game balance methods are based on human play testing, where developers gain an intuitive "feel" for the game by playing it and then making changes
- games should not be structured so that they are “solvable,” but there should still be room for quantitative investigation
- it benefits developers to use more concrete methods to balance their games, as their popularity, marketability, and longevity rely on their appeal to a competitive fanbase [7]

INITIAL DATA TESTING

- there are clear indicators of whether a character is “good” or not





THE DATA

- community-sourced data pulled from various public internet sources
  - extensive data cleaning and feature engineering required
  - small data set; 72 samples, 68 features
  - includes frame data, basic attributes and some constructed variables for every character
- some key character data are not available or difficult to quantify, e.g. hit/hurt box data and special moves
- example data sample:

Full Feature Data for Mario (rating = 2.96)

attributes		startup		advantage		damage				
air accel	0.08	Jab	2	Jab	-29	Jab	8.9	recovery	top hitbox?	1
air speed	1.21	F-Tilt	5	F-Tilt	-13	F-Tilt	7		invulnerabilit	1
max fall spee	1.50	U-Tilt	5	U-Tilt	-18	U-Tilt	5.5		max vertical	12
fast fall spee	2.40	D-Tilt	5	D-Tilt	-15	D-Tilt	7		horizontal m	0
initial dash	1.94	Dash Attack	6	Dash Attack	-17	Dash Attack	8	other	tether?	0
run speed	1.76	F-Smash	15	F-Smash	-22	F-Smash	17.7		projectile?	1
walk speed	1.16	U-Smash	9	U-Smash	-20	U-Smash	14		reflector?	1
weight	75.00	D-Smash	5	D-Smash	-31	D-Smash	10		counter?	0
full hop	36.33	N-Air	3	N-Air	-2	N-Air	8		max jumps	2
short hop	17.54	F-Air	16	F-Air	-12	F-Air	14		weapon?	0
air jump	36.33	B-Air	6	B-Air	-2	B-Air	10.5			
		U-Air	4	U-Air	-3	U-Air	7			
		D-Air	5	D-Air	-11	D-Air	12			
		Grab	6			Pummel	1.3			
		Dash Grab	9			Forward Thr	8			
		Pivot Grab	10			Back Throw	11			
						Up Throw	7			
						Down Throw	5			

REGRESSION

- exact values predicted using the Random Forest model
- overfitting, due to small data and subjective target values
    - other methods yielded worse results

Rating Predictions for Test Set 'B2'

	actual		predicted	difference	
ID	character	rating	rating	point diff	% diff
7	Pikachu	4.30	3.14	-1.16	-0.27
10	Greninja	3.94	2.93	-1.01	-0.26
12	Wario	3.80	2.79	-1.01	-0.27
13	Ike	3.80	2.70	-1.10	-0.29
17	Ivysaur	3.47	2.53	-0.94	-0.27
20	R.O.B.	3.33	2.26	-1.07	-0.32
22	Marth	3.25	3.49	0.24	0.07
26	Ness	2.92	2.09	-0.83	-0.28
32	Simon/Richt	2.41	2.02	-0.39	-0.16
33	Bowser	2.25	1.95	-0.30	-0.13
40	Samus/Dark	1.96	1.84	-0.12	-0.06
54	Ken	1.55	1.97	0.43	0.27
55	Bayonetta	1.52	2.42	0.90	0.60
60	Dr. Mario	1.39	2.28	0.89	0.64
61	Robin	1.27	2.04	0.76	0.60
66	Jigglypuff	0.95	2.59	1.63	1.71
71	Kirby	0.52	2.42	1.91	3.70

Rating Order Predictions for Test Set 'B2'

actual		predicted	
character	rating	character	rating
Pikachu	4.30	Marth	3.48916237
Greninja	3.94	Pikachu	3.14418426
Wario	3.80	Greninja	2.93492858
Ike	3.80	Wario	2.79387247
Ivysaur	3.47	Ike	2.69799972
R.O.B.	3.33	Jigglypuff	2.589465
Marth	3.25	Ivysaur	2.53355449
Ness	2.92	Kirby	2.42381977
Simon/Richt	2.41	Bayonetta	2.42017073
Bowser	2.25	Dr. Mario	2.28042513
Samus/Dark	1.96	R.O.B.	2.26176653
Ken	1.55	Ness	2.0907552
Bayonetta	1.52	Robin	2.03584451
Dr. Mario	1.39	Simon/Richt	2.02250993
Robin	1.27	Ken	1.97293259
Jigglypuff	0.95	Bowser	1.95131773
Kirby	0.52	Samus/Dark	1.84367914

- character order (tier list) is preserved:

CLASSIFICATION

- tier rankings (order) predicted using the Random Forest model
- significant decrease in variance; **coefficient of determination increased from 0.27 to 0.41**
- misclassified samples explained by lack of data e.g. Jigglypuff and Wario

Class/Tier Predictions for Test Set 'B2'

	actual		y2		y3	
character	y2	y3	rand forest	k-neighbors	rand forest	k-neighbors
Pikachu	A	A	D	B	A	A
Greninja	B	A	A	A	B	A
Wario	B	A	C	C	A	C
Ike	B	A	B	B	A	A
Ivysaur	B	B	D	B	D	B
R.O.B.	B	B	D	D	C	D
Marth	B	B	A	A	A	A
Ness	C	B	D	C	C	C
Simon/Richt	C	C	D	B	C	B
Bowser	C	C	D	E	D	E
Samus/Dark	D	C	D	B	D	B
Ken	D	D	D	D	E	E
Bayonetta	D	D	D	D	C	D
Dr. Mario	D	D	D	D	D	E
Robin	D	E	D	D	D	D
Jigglypuff	E	E	D	A	D	A
Kirby	E	E	D	C	D	B

ACKNOWLEDGEMENTS

A huge thank-you to Professor Engelhardt and the TAs for COS 424, for their teaching and support throughout the semester.

REFERENCES

[1] Beefy Smash Doods. “Who Can Climb THE WALL In Smash Ultimate?” *YouTube*, YouTube, 3 May 2019, www.youtube.com/watch?v=Tg8oXDob960.

[2] Branniglenn, Zapp. “Smash Ultimate Frame Data Patch 3.0 (Complete).” *Smashboards*, 29 Apr. 2019, smashboards.com/threads/smash-ultimate-frame-data-patch-3-0-complete.465028/.

[3] Brownlee, Jason. “Discover Feature Engineering, How to Engineer Features and How to Get Good at It.” *Machine Learning Mastery*, 26 Sept. 2018, machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/.

[4] inktivate. “April Ultimate Tier List - a Combination of 18 Different Pros' Lists from the Past 50 Days.” *Reddit*, Apr. 2019, www.reddit.com/r/smashbros/comments/bagz79/april_ultimate_tier_list_a_combination_of_18/.

[5] Kurogane. “Smash Ultimate Frame Data Repository.” *Kurogane Hammer*, 2019, kuroganehammer.com/.

[6] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Sciki-learn: Machine Learnning in Python. Journal of Machine Learning Research. 2011; 12; p. 2825– 2830. https://scikit-learn.org/stable/.

[7] Sirlin, David. “Balancing Multiplayer Competitive Games.” *GDC Vault*, UBM Technology Group, 2009, www.gdcvault.com/play/1570/Balancing-Multiplayer-Competitive.

[8] “SmashWiki.” *SmashWiki*, 2019, www.ssbwiki.com/.

[9] Wonkey. “Harada on Revealing Frame Data in Tekken Games.” *Avoiding The Puddle*, 2014, www.avoidingthepuddle.com/news/2014/2/4/harada-on-revealing-frame-data-in-tekken-games.html.

[10] ガチョック. “Untitled Animation - All Characters High Jump Championship.” *Facebook*, Facebook, Apr. 2019, www.facebook.com/UntitledAnimation/videos/348093746052624/.