Project: Indycar Rank Prediction
Progress Report
10142019-10282019

I. Abstract

The major progress in the past two weeks is in searching for related work. The most relevant papers are [1][2], which apply machine learning to the NASCAR race. They forecast the decision-to-decision loss in rank position for each racer. It provides a promising direction to follow. In this report, I present the main ideas and results in this paper and some preliminary exploratory data analysis results on the IndyCar dataset.

II. Summary of the related work on NASCAR prediction

Paper[1][2] are a series of work supervised by Prof. Rudin when she was at MIT. However, no further progress had been reported after 2015.

[1] describes how they leveraged expert knowledge of the domain to produce a real-time decision system for tire changes within a NASCAR race. Major results and findings include:

- ‘‘this work started with the hypothesis that a data-driven prediction engine operating in real time may be able to assist team captains in making these critical tire change decisions.’’
- ‘‘unfortunately, we cannot perform randomized controlled trials in order to measure the effect of a decision; we are limited by what we can do with the historical data.’’
- ‘‘our evidence suggests that it is possible to generalize across races; that is, we can borrow strength from the data of similar races to make improved predictions.’’
- ‘‘we can see that the machine learning methods are significantly better than the baseline methods.’’
- ‘‘expert commentaries that are typically stated either before or after the race can also be used to qualitatively validate the inferences of our modeling approach.’’

A. NASCAR dataset

NASCAR race usually contains 44 racers, with average number of pit stops per racer varies from 4 to 8.9, and the number of caution laps(yellow flags) varies from 3 to 14. The dataset contains 17 races with the number of laps between 160 to 500.

Tire change decision in pit stop is a major strategic decision for each team. A two-tire change saves about 6 seconds on average than a Four-tire change. But a car with two-tire change is slower than a fresh four-tire change car. In NASCAR, the phenomenon of tire wear down over the course of racing is an obvious pattern which impact the lap time for all the racers.

B. Main Ideas

The first important idea in this work is to use **“stage” as the unit of input data**. As in Fig below, a race is segmented into serval stages which are split by pitstop or caution laps(yellow flags). The timing of each sub-segments are determined by the performance of either pit crew or driver and car.
By using a long stage as input, the influences of those artificial rank position changes caused by pitstop or caution lap temporarily can be mitigated.

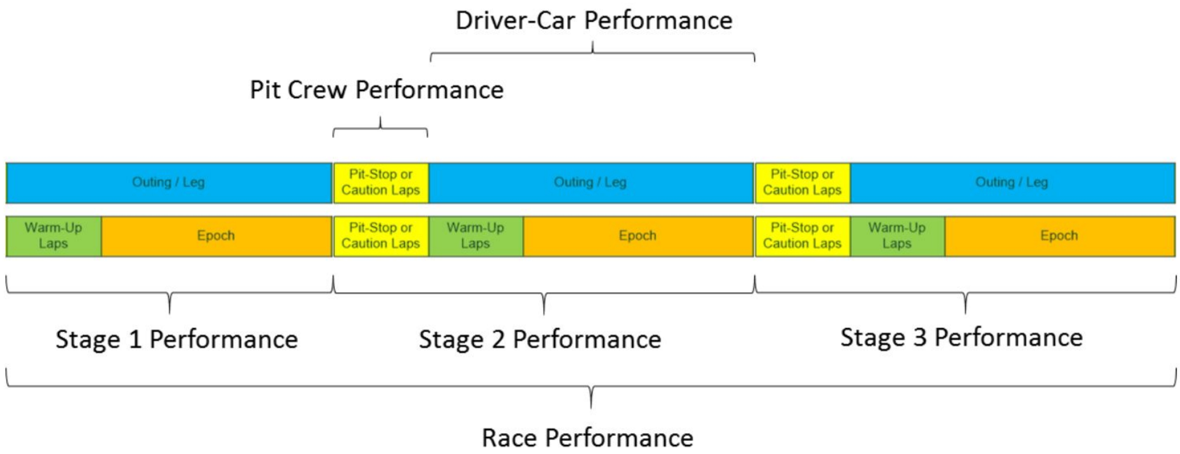


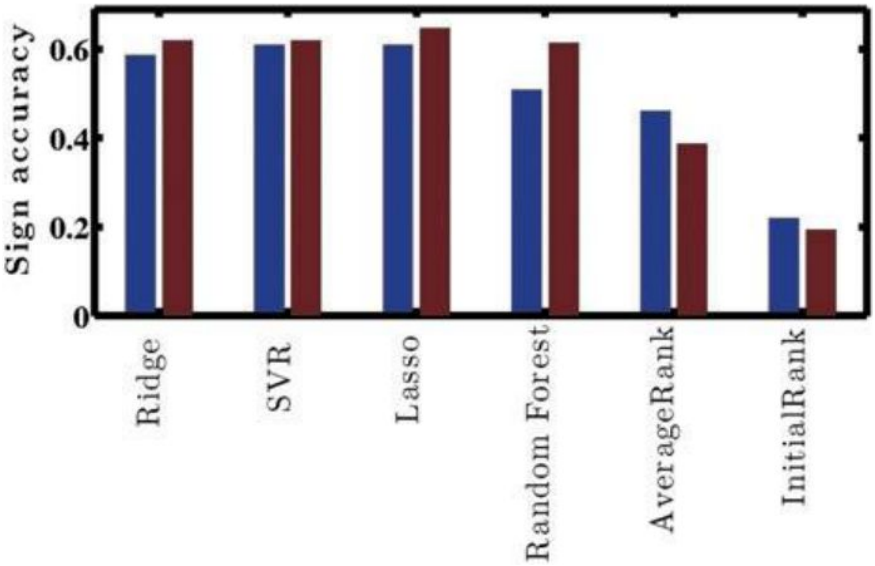
Figure 5. Composition of Race Performance based on Leg, Driver-Car, and Pit Crew based on a hypothetical Race comprising 3 Stages

Secondly, the way of problem setup in this work is inspiring. They chose to model **the change in rank position** and not the other functions. They also avoid predicting the rank position directly since it is complicated due to its dependency on the timing of other racers’ pit stops.

Thirdly, they developed over 100 features with the help of domain experts. Some of them are explained in the paper, therefore can be used if they are available in the IndyCar dataset.

C. Results

The model predicts the rank position change for the coming stage before a car enters into it. Baselines include the predictor using current rank or average rank. As shown in the following figure, the accuracy of the sign of prediction(rank increase or decrease) over different models. Machine learning models are able to achieve accuracy between 50%-60%, significantly better than baselines(20%-50%).

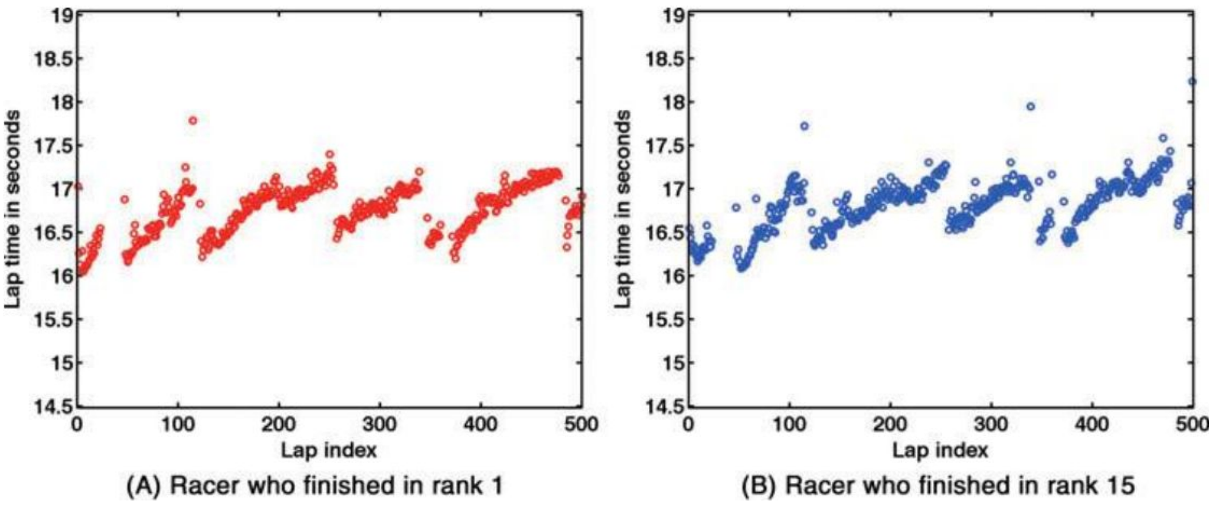


Since the tire-change strategy is one of the features of the model. Once the model is trained, the prediction result can then be used to evaluate which strategy should be chosen for a racer during the race.

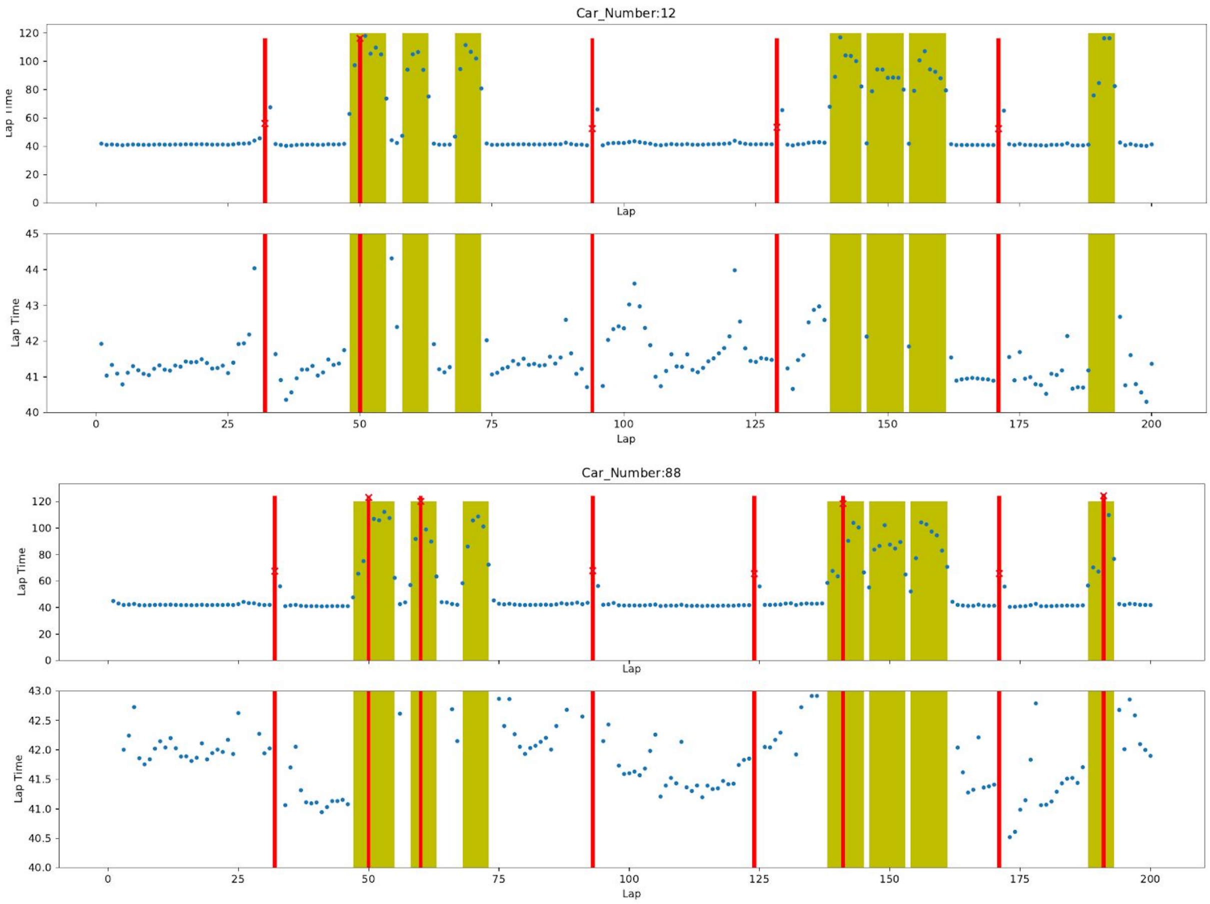
III. Exploratory Data Analysis of IndyCar

In order to apply the ideas learned from the work of NASCAR prediction on IndyCar, we first run exploratory data analysis comparing the two datasets.

A. Laptime Analysis



(a) NASCAR, Sawtooth profile of typical racers in a race.



(b) IndyCar. Car#12 finished in rank 1 and #88 finished in rank 14. Yellow bar indicates caution laps and red bar indicates pit stop.

In the above figure(a), typical racers in a NASCAR race demonstrate a sawtooth shape of lap time distribution over time. A strong influence of tire wear happens in these races. While in IndyCar, tire wear seems to be not an issue. As in figure(b), no consistent pattern can be observed.

B. Pitstop Analysis

Comparing the pit stop strategy in NASCAR and IndyCar, as shown in the following figure, the variance in NASCAR is smaller than that in IndyCar. Obvious waved pitstops pattern can be observed in (a)NASCAR. And most of the cars will adopt pitstop early in the caution laps. This kind of regularity helps to make the idea of ‘stage’ work because of fewer impacts of the different timing of pit stops on the rank position. IndyCar is more challenging to deal with the dynamics, as shown in (b).

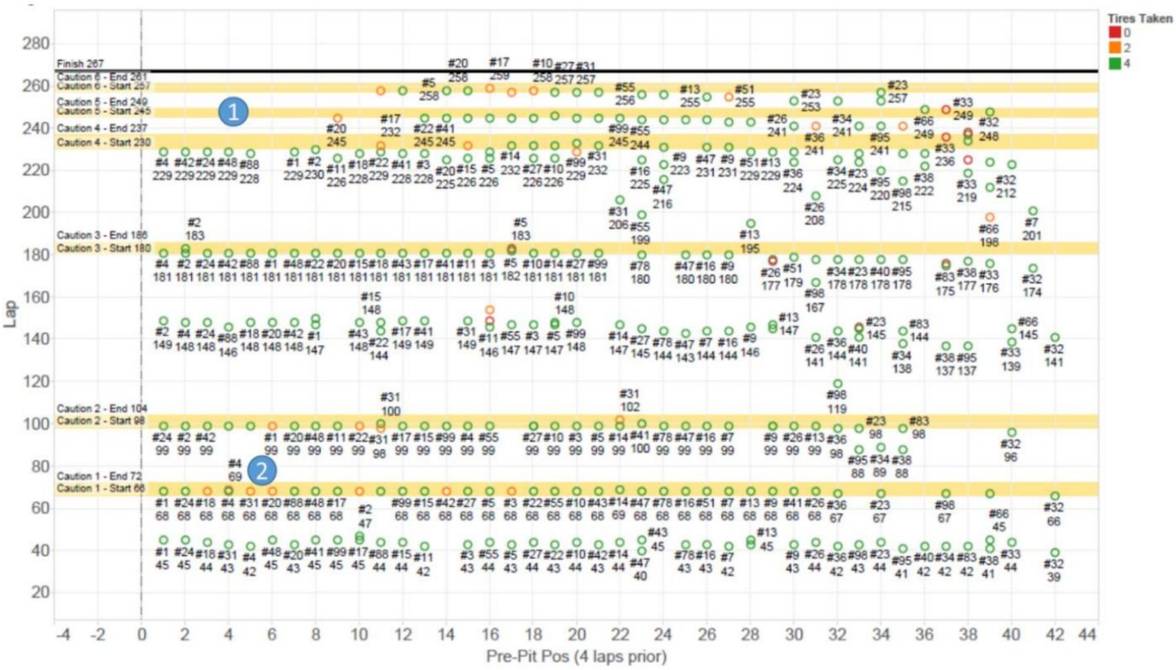
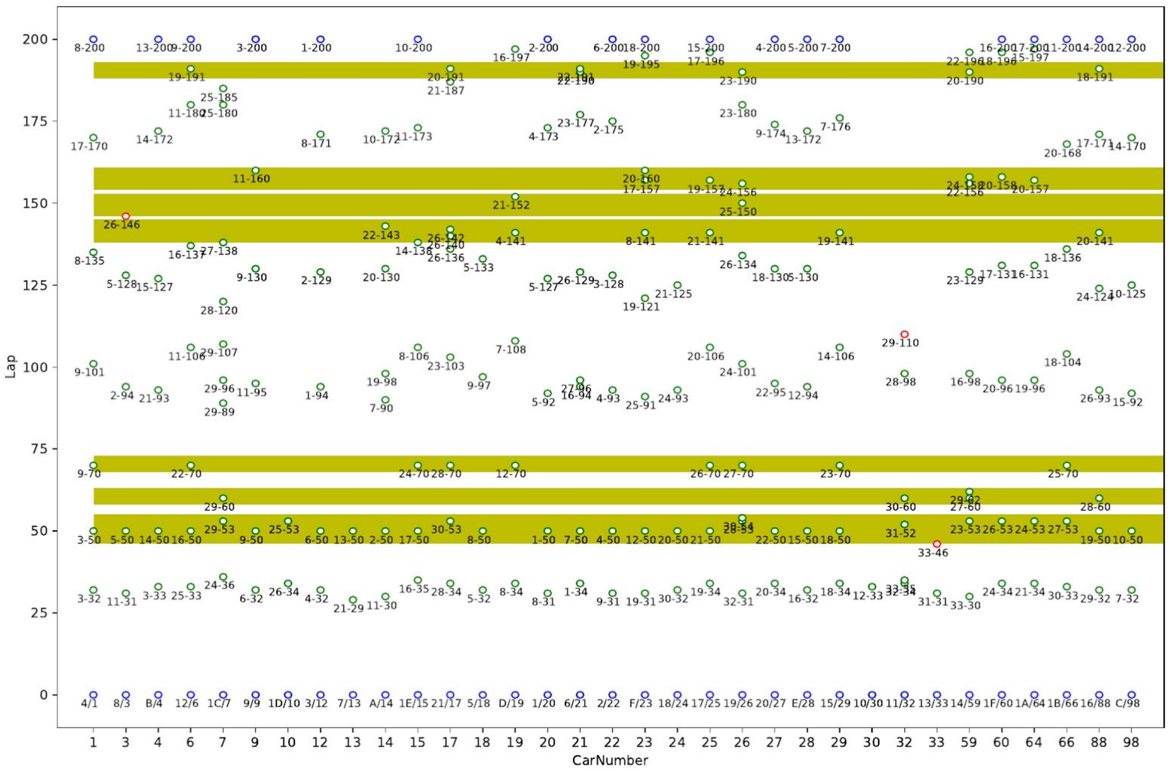


Figure 45. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Midwest_B Race 2014

(a) NASCAR(2014-Midwest_B, green circle denotes pitstop labeled with carno-lap)



(b)IndyCar (2018-Indy500, green circle denotes pitstop labeled with rank-lap, blue circles are the start/end position)

IV. Work TODO

Although IndyCar data show quite different characteristics compared with NASCAR, the idea of building a model on rank change over a stage is still the most promising direction. In the next two weeks, I will continue working on it, including:

- Extract stage data from IndyCar dataset.
- Correlation analysis on the rank position change on the stage dataset.

References

[1]T. Tulabandhula and C. Rudin, “Tire changes, fresh air, and yellow flags: challenges in predictive analytics for professional racing,” Big data, vol. 2, no. 2, pp. 97–112, 2014.

[2]C. L. W. Choo, “Real-time decision making in motorsports: analytics for improving professional car race strategy,” PhD Thesis, Massachusetts Institute of Technology, 2015.

Appendix:

1. 2018 Indianapolis 500

https://www.racing-reference.info/race/2018_Indianapolis_500/O

Time of race: 2:59:43
Average speed: 166.935 mph
Pole speed: 229.618 mph

Cautions: 7 for 41 laps
Margin of victory: 3.159 sec
Attendance: n/a
Lead changes: 30

Glossary **2018 Verizon IndyCar Series results** / **2018 standings**

Fin				Sponsor / Owner	C/E/T				
1	3	12	Will Power	**Verizon** (Roger Penske)	D/C/F	200	running	59	108
2	1	20	Ed Carpenter	**Fuzzy's Vodka** (Ed Carpenter Racing)	D/C/F	200	running	65	92
3	9	9	Scott Dixon	**PNC Bank / NTT Data** (Chip Ganassi)	D/H/F	200	running	0	71
4	32	27	Alexander Rossi	**NAPA Auto Parts** (Andretti Autosport)	D/H/F	200	running	1	65
5	14	28	Ryan Hunter-Reay	**DHL** (Andretti Autosport)	D/H/F	200	running	1	61
6	2	22	Simon Pagenaud	**Menards / Verizon** (Roger Penske)	D/C/F	200	running	1	65
7	21	29	Carlos Munoz	**Ruoff Home Mortgage** (Andretti Autosport)	D/H/F	200	running	4	53
8	4	1	Josef Newgarden	**Verizon** (Roger Penske)	D/C/F	200	running	3	55
9	18	6	Robert Wickens	**Lucas Oil** (Schmidt Peterson Motorsports)	D/H/F	200	running	2	45
10	30	15	Graham Rahal	**United Rentals / Mi-Jack** (Rahal Letterman Lanigan Racing)	D/H/F	200	running	12	41
11	27	66	J.R. Hildebrand	**Salesforce / Mecum Auctions** (Dreyer & Reinbold Racing)	D/C/F	200	running	0	38
12	12	98	Marco Andretti	**U.S. Concrete / Curb** (AHA with Curb-Agajanian)	D/H/F	200	running	0	36
13	11	4	Matheus Leist	**ABC Supply Co.** (A.J. Foyt)	D/C/F	200	running	0	34
14	22	88	Gabby Chaves	**Harding Group** (Harding Racing)	D/C/F	200	running	0	32
15	23	25	Stefan Wilson	**#Driven2SaveLives** (Andretti Autosport)	D/H/F	200	running	3	31
16	31	60	Jack Harvey	**AutoNation / SiriusXM** (Michael Shank with SPM)	D/H/F	200	running	0	28
17	26	64	Oriol Servia	**Grove by Manitowoc** (Scuderia Corsa with RLL)	D/H/F	200	running	16	27

18	15	23	Charlie Kimball	**Fiasp** (Trevor Carlin)	D/C/F	200	running	0	24
19	13	19	Zachary Claman DeMelo	**Paysafe** (Dale Coyne)	D/H/F	199	running	7	23
20	6	21	Spencer Pigot	**Preferred Freezer Service / Direct Supply** (Ed Carpenter Racing)	D/C/F	199	running	3	25
21	33	17	Conor Daly	**United States Air Force** (Dale Coyne)	D/H/F	199	running	0	18
22	20	59	Max Chilton	**Gallagher** (Trevor Carlin)	D/C/F	198	running	0	16
23	25	26	Zach Veach	**Group One Thousand One / Relay** (Andretti Autosport)	D/H/F	198	running	0	14
24	28	7	Jay Howard	**OneCure** (Schmidt Peterson Motorsports)	D/H/F	193	running	0	12
25	10	14	Tony Kanaan	**ABC Supply Co.** (A.J. Foyt)	D/C/F	187	**crash**	19	11
26	24	24	Sage Karam	**Wix Filters / Comfort Revolution** (Dreyer & Reinbold Racing)	D/C/F	154	**crash**	0	10
27	8	3	Helio Castroneves	**Pennzoil / Verizon** (Roger Penske)	D/C/F	145	**crash**	0	12
28	5	18	Sebastien Bourdais	**SealMaster / Sport Clips** (Dale Coyne w/Vasser-Sullivan)	D/H/F	137	**crash**	4	16
29	17	32	Kyle Kaiser	**NFP** (Juncos Racing)	D/C/F	110	**mechanical**	0	10
30	7	13	Danica Patrick	**GoDaddy** (Ed Carpenter Racing)	D/C/F	67	**crash**	0	13
31	29	10	Ed Jones	**NTT Data** (Chip Ganassi)	D/H/F	57	**crash**	0	10
32	16	30	Takuma Sato	**Mi-Jack / Panasonic** (Rahal Letterman Lanigan Racing)	D/H/F	46	**crash**	0	10
33	19	33	James Davison	**Jonathan Byrd's 502 East** (Foyt-Byrd-Hollinger-Belardi)	D/C/F	45	**crash**	0	10

C/E/T = Chassis / Engine / Tire codes
Chassis: D - Dallara
Engine: C - Chevrolet; H - Honda
Tires: F - Firestone