CS229 – Machine Learning Project
Identifying Gas Savings from Driver Behavior

Arthur Lacoste, Marie Pellat, Thomas Piani

Abstract – This project aimed to identify potential gas savings in a driver's behavior and suggest changes. The steps we took were threefold. We first built a model predicting instantaneous gas consumption from driving data, then extracted driving constraints (trajectory, min & max speed, turns, stops), and finally we devised a working technique to minimize fuel consumption under these constraints. Using machine learning techniques and physics insights, we managed to predict instantaneous gas consumption with very good accuracy (R^2 around 90%) using only 3 distinct features in a linear model based on noisy and limited driving data. Our optimization technique provided applicable driving patterns that drastically reduced gas consumption (by 15 to 20%) while following driving constraints.

I. Introduction

Americans spend around $500 Billion on gas every year, an astounding number that keeps growing with gas prices. Gas consumption, however, is not equally shared between drivers. Indeed, it is common knowledge that driving habits can impact gas consumption, and that two driving patterns on the same route can lead to very different fuel consumptions, directly impacting the cost for drivers. But devising a good driving behavior over the course of an entire trip is a tough problem that requires linking driving data to fuel consumption and providing solutions that are feasible in real life. To tackle this issue, we used datasets of driving data that were provided by Metromile, a company that seeks to make use of driving data to save drivers time and money. Their first offering is a per-mile car insurance that charges drivers based on actual miles driven. For this project, the company gave us access to the driving data (speed, accelerations, heading and instantaneous gas consumption) from 18 employee cars collected using an embarked device.

II. Data Preprocessing

1. Change of referential

The device measuring accelerations is fixed in the car but its orientation in unknown and a first step is to determine the corresponding referential change matrix. The vertical direction is easily identified as the average of all accelerations, gravity being a very strong signal. We used two different methods to determine the x and y axes:

First method: We isolated sections of the trajectory where the heading remains steady and the vehicle is speeding up or slowing down significantly. In these situations, the acceleration should be on the x axis.

Second Method: We computed theoretical axial and lateral accelerations based on the vehicle's speed, speed variation and heading variations. We then ran a modified version of ordinary least squares to determine the x and y axes for which the measured accelerations would best fit the theoretical ones. (see milestone for the closed form solution)

2. Dealing with gaps in the data

Our datasets have time gaps. Long gaps separate different trips and short ones are just missing measures. We decided to fill in gaps under 10 seconds (linearly between the two extreme data points) and regard bigger gaps as a separation between two independent trips.

3. Objective function

The consumption measurement in the datasets is in miles per gallon, but we though the instantaneous consumption (in gallons per unit of time) would be much easier to predict. In addition, as that quantity showed high frequency variations that did not exist in the rest of the data, we smoothed it over a 7 second window (which has no impact on total fuel consumption).

III. Instantaneous Consumption Predictor

1. Initial features and algorithm

We used an ordinary least squares regression as our learning algorithm (70 % training data, 30% testing data), and the R^2 coefficient as our performance measurement to avoid acute sensitivity to the smoothing of the gas consumption and to easily measure the relevance of a feature. We first tried to guess relevant features through physic intuition:

- Speed v (laminar friction losses)
- v^2 (turbulent friction losses)
- Acceleration a_x
- $v * a_x$ (energy variation from speed or elevation variation)

In addition, we added features corresponding to the past and future values of these quantities over a -10s to +10s window, for a resulting R^2 = 75% (train & test). Trying out different combinations of these features, we realized that speed (whether v or v^2) and energy variation were the most important factors, and indeed we reached the same performance of R^2 = 75% using only v and $v * a_x$.

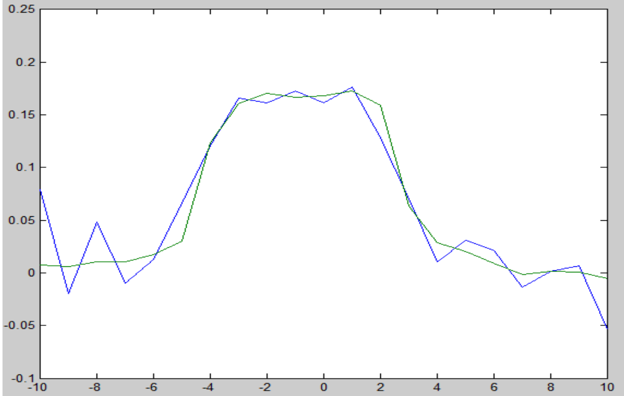


Figure 1: Learned weights of $v(now + \Delta t)$ (blue) and $(v * a_x)(now + \Delta t)$ (green) as a function of Δt

2. Categorization

With no sign of overfitting and in order to further improve our model's performance, we introduced a categorization of the data in different speed (10mph & less, 10 to 30 mph, 30 to 50 mph, 50 mph and more) and acceleration (speeding up, cruising, slowing down) regimes. We fitted a different model for each tuple of speed and acceleration regime. This improved R^2 up to 92% (on some datasets), but introduces discontinuities in the instantaneous consumption prediction which result in aberrant optimal driving behaviour. Therefore, we went back to an uncategorized model.

3. Physics-based model

a. Insight

We had identified that energy variation was a key feature to predict gas consumption, but realized that using $v * a_x$ meant we treated positive and negative energy variations in the same way, which did not make much physical sense. Indeed, any energy gain directly translates into a cost in gas, but not all energy losses do (for instance, when braking hard). We decided to incorporate this by using only the positive part of energy variation in our linear regression: $(dE/dt)_+ = (v * a_x)_+$ In addition, physical intuition and our previous results both suggest to also use a term depending only on v, $F(v)$, that would represent the baseline consumption when the car maintains speed v.

b. Choosing $F(v)$

Since we were not sure of what type of function to choose for $F(v)$, we decided to discretize our data with respect to speed to get the value of $F(v)$ at different speeds. In practice, we used LMS for our whole dataset with the features $(dE/dt)_+$ and $I_{\{(k-1)v_0 \le v < kv_0\}}$ with $v_0 = 2$ mph and $k = 1 \ldots 50$ buckets. The plotted curve [Figure 2] of the weights of the indicator features deliver $F(v)$, for which a polynomial would be a good approximation. We rand LMS with different polynomial models (replacing the indicator features) and all of them fit $F(v)$ decently. However, the $a + bv^2 + cv^4$ and $a + bv + cv^2$ models both sometimes (depending on the dataset) show negative weights (respectively for v^4 and for v), which does not make physical sense (a negative weight for v means that you consume less fuel per unit of time driving slowly than not moving at all). Therefore, we decided to only retain the very significant v^2 term and settled on the model $F(v) = \alpha + \beta v^2$.

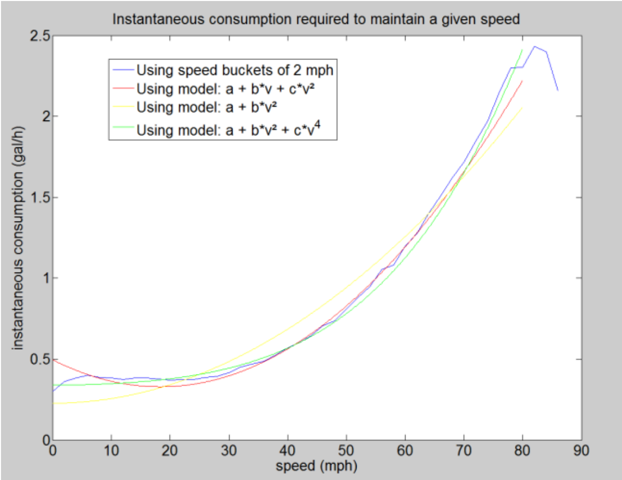


Figure 2: $F(v)$ as a function of speed for different models

c. Model and results

Our model is: $C_{gal/s} = \alpha + \beta v^2 + \gamma (dE/dt)_+$ where α is the gas consumption just from the engine running, βv^2 is the fuel cost to compensate losses at speed v and $\gamma (dE/dt)_+$ is the fuel cost of increasing the car's kinetic or potential energy. We originally added all [-10s +10s] values of $(dE/dt)_+$ but the weights fitted proved that not all of them actually matter. Indeed, we were able to reach almost the same performance of R^2 = 89% using only the average of $(dE/dt)_+$ over a [-3s +4s] time window where the weights were the highest. We believe there are two reasons for this spread:

- Measurements of a_x are very noisy, and smoothing a_x would probably have been a valid approach too.
- The objective function itself is smoothed over [-3s +3s], which ties the consumption at time t to that of neighboring points. We confirmed that increasing the smoothing window for the instantaneous gas consumption increases that spread.

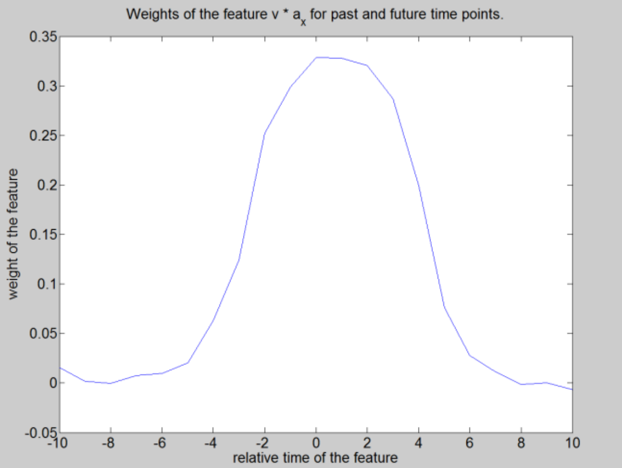


Figure 3: weight of $(v * a_x)(t + \Delta t)$ as a function of Δt

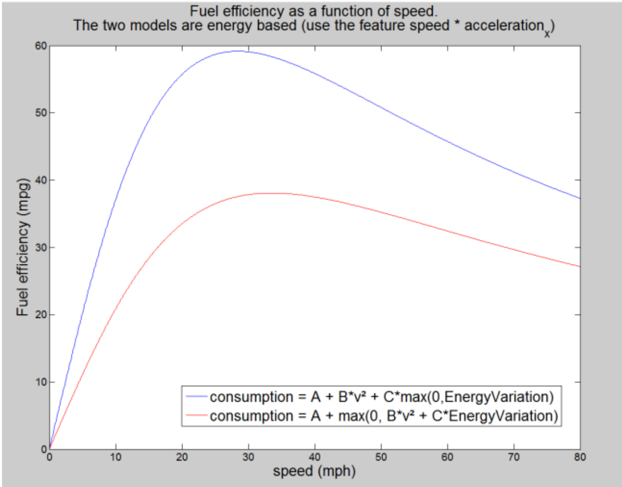


Figure 4: mpg(speed) for the two trained models

4. Model improvement

a. Insight

Our previous model makes the assumption that there is no way to transform negative energy variations into gas savings, which is not entirely true. Indeed, braking wastefully dissipates energy, but inevitable losses (the βv^2 term) can be compensated either by fuel input or by an energy decrease. Thus it is possible to save up to βv^2 instantaneous fuel usage from negative energy variations, but no more (the idling consumption α cannot be negated). This yields the following model:

$$C_{gal/s} = \alpha + (\beta v^2 + \gamma\, dE/dt)_+$$
$$= \alpha + \gamma(dE/dt)_+ + (\beta v^2 - \gamma(dE/dt)_-)_+$$

b. Training strategy

This type of model cannot be directly fitted using LMS because some parameters are inside the positive part. Instead, we used the following training strategy:

fit $y = C_{gal/s}$ using $\alpha + \beta v^2 + \gamma(dE/dt)_+$ to get initial (α, β, γ) values

repeat until convergence:

set $y' = y - \alpha + (\beta v^2 + \gamma\, dE/dt)_+$

fit y' using $\alpha' + \beta' v^2 + \gamma'(dE/dt)_+$

update $(\alpha, \beta, \gamma) \leftarrow (\alpha, \beta, \gamma) + (\alpha', \beta', \gamma')$

c. Results

This model yielded slightly better results than the first model R^2 = 90%, but has the added advantage of more realistically handling energy decrease phases. Using both models, we plotted the consumption in miles per gallon to maintain the car at speed v: $C_{mile/gal} = \frac{v}{C_{gal/s}} = \frac{v}{\alpha + \beta v^2}$.

Figure 4 shows there is an optimal speed v^* at which fuel efficiency is maximal, so that the optimal unconstrained driving behavior over a long enough trip will be to cruise at speed v^*. The fuel efficiencies predicted by both models are significantly different for the following reason: under the no-savings model, consumption in energy decrease phases are overestimated, which LMS partially compensates by fitting lower values for α and β, resulting in abnormally high fuel efficiencies.

IV. Driving Behavior Optimization

1. General methodology

Given driving data, our end goal is to provide gas-saving driving behavior suggestions. Using a simple physics model, we can reconstruct the driver's trajectory from that data as well as generate simulated driving data for any speed profile over that trajectory. We can then feed that simulated driving data to our trained instantaneous consumption predictor to estimate the total consumption.

2. Trajectory reconstruction

The most important constraint for an optimal driving pattern is that it follows the same trajectory as the original data. We used the following equations to reconstruct (x, y, z) for every data point:

$$\sin(\theta_t) = (a_x)_t - \frac{dv_t}{dt}$$
$$x_t = x_{t-1} + v_{t-1} * \cos\theta_{t-1} * \cos\phi_{t-1}$$
$$y_t = y_{t-1} + v_{t-1} * \cos\theta_{t-1} * \sin\phi_{t-1}$$
$$z_t = z_{t-1} + v_{t-1} * \sin\theta_{t-1}$$

where θ is the slope of the trajectory and ϕ is the heading of the car (included in the data).

Since our axial acceleration data a_x is very noisy, the estimates for $\sin(\theta)$ vary significantly and fast. In order for the reconstructed trajectory to be realistic, we smoothed them over a large time

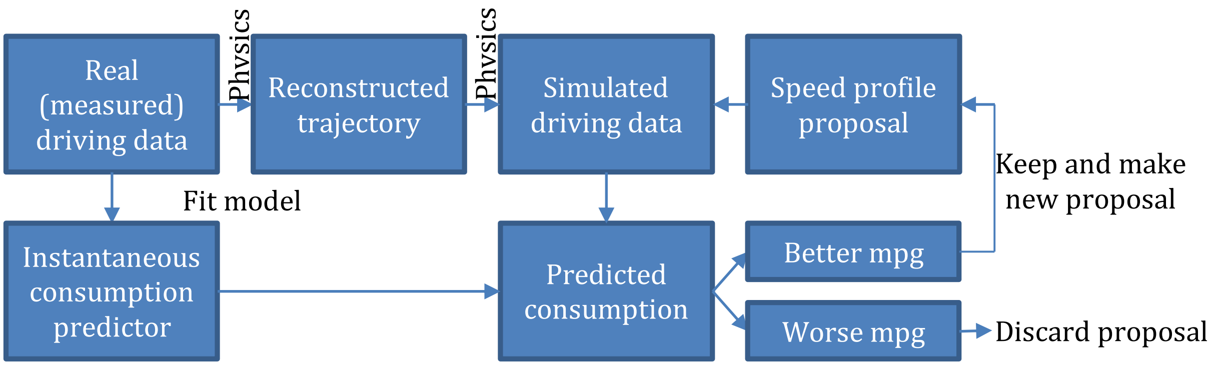


Figure 5: General methodology

window [-15s +15s]. As previously, the integral over time remains the same, so that we filter out high frequency noise but still capture and do not deform actual elevation changes.

3. Constraints

In order to generate useful suggestions, we need to incorporate additional constraints:

- Bounded lateral accelerations (impossible to take sharp turns at high speed).
- Bounded axial accelerations (safety & car limitations)
- Wherever the driver stopped, a valid speed profile must also stop and for the same amount of time. This accurately models stops and lights (under the assumption that the stopping time is unpredictable), but is plain wrong for traffic jam situations.

In reality there are additional constraints that we are unable to infer from the driving behavior, for example another car slowing down forcing you to slow down as well.

4. Optimization techniques

Ideally, we would like to find the optimal (with regard to consumption) speed profile satisfying the constraints. However this is a complex and non-convex optimization problem (a linear combination of valid speed profiles is not guaranteed to be valid), so we tackled it with approximation techniques.

a. Speed profile guesses

In a first approach, we attempted to guess consumption-reducing modifications to the speed profile. We implemented two types of guesses: linearizing the speed between two points of the trajectory, and increasing/decreasing the speed over a section of the trajectory by a flat amount (by a very small amount first to test for the sign of the gradient, followed by a line search if successful). We would then try applying these modifications to random portions (or all portions) of the trajectory. This technique has the advantage of being applicable regardless of the model used for instantaneous consumption, but has a prohibitive computational cost.

b. Point by point optimization

Although our final model for instantaneous consumption uses the averaged energy variation, it is reasonable (when only trying to predict the overall consumption) to scale the γ coefficient and use the instantaneous energy variation instead.

$$\int_t \alpha + \left(\beta v^2 + \gamma \sum_{\Delta t=-3}^{4} \frac{dE}{dt}\bigg|_{t+\Delta t}\right)_+ dt \approx \int_t \alpha + \left(\beta v^2 + \gamma' \frac{dE}{dt}\bigg|_t\right)_+ dt$$

with $\gamma' = \sum_{\Delta t=-3}^{4} \gamma$

Then, modelling the speed as linear between two discretization points, we can express the total consumption between two points of respective speeds v_1 and v_2 separated by a distance d with slope θ:

$$C(v_1, v_2, d, \theta) = \int_{t=0}^{\frac{d}{v_1+v_2}} \alpha + \left(\beta v^2 + \gamma' v(g \sin\theta + a)\right)_+ dt = \int_{v=v_1}^{v_2} \frac{\alpha}{a} + \left(\beta v^2 + \gamma' v(g \sin\theta + a)\right)_+ dv$$

$$\text{with } a = \frac{dv}{dt} = \frac{v_2 - v_1}{\Delta t} = \frac{v_2^2 - v_1^2}{d}$$

This allows to optimize the total consumption point-by-point (coordinate descent) by fixating the speeds of neighbour points and taking:

$$v = \arg\min_v C(v_{before}, v, d_{befo\Box e}, \theta_{before}) + C(v, v_{after}, d_{after}, \theta_{after})$$

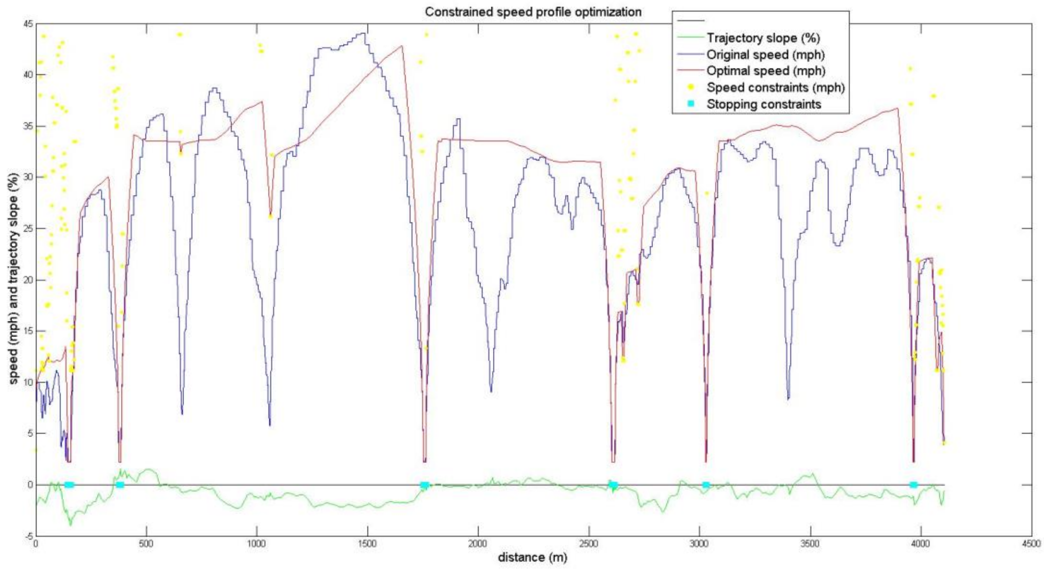
This technique is much more efficient efficient, and could be adapted to work with any instantaneous consumption model (it would require to simulate driving data locally around each point to optimize instead of just evaluating an integral).

5. Results and limitations

Figure 6 shows the results of the point-by-point optimization technique applied to a real trip in one of our datasets. On this example, the optimization improves the overall fuel efficiency for the trip from 23.61 mpg up to 27.26 mpg (an increase of ~15%). This demonstrates there are indeed important savings to be made by improving driving behavior. However, there is no guarantee on the optimality of the solution the algorithm converges to (the problem is not convex).

For example, point-by-point optimization handles stopping points strangely: it always prefers cruising at speed v^* and instantly slowing down to speed zero (or saturating the deceleration constraint) when reaching the stopping point. By doing so, it wastes all of its kinetic energy instead of slowing down over time to convert it into fuel savings. Although the optimal solution of this stopping problem under the consumption model $C_{gal/s} = \alpha + (\beta v^2 + \gamma\, dE/dt)_+$ is not easy to determine, we proved that cruising at speed v^* and instantly stopping is suboptimal. We are still not certain what the reason for that strange behaviour is: the cruise-and-stop profile being a local optimum of that problem or discretization modifying the problem (and its optimum) in a subtle way.

Figure 6: This figure was obtained using the point-by-point optimization technique with the following constraints: $|a_{lateral}| \leq 0.4g \qquad -0.2g \leq \frac{dv}{dt} \leq 0.2g$



V. Conclusion

Using fairly limited and noisy driving data, we managed to build a gas consumption predictor that achieved very high accuracy (R^2 around 90%) using a linear structure and only 3 different features. We were able to extract trajectory and constraints from the driving data and, using our optimization strategy, to devise optimized driving patterns that would be feasible in a real setting and would yield important gas savings. It is interesting to note that throughout this project, we consistently achieved the best results when we managed to incorporate physics insights about the problem in machine learning strategies. This is an example of the importance of human intelligence and guidance in the success of computer-assisted techniques.

VI. Acknowledgements

We would like to thank Danny Goodman, Director of Data Science at Metromile, for his invaluable help and guidance throughout the project. Danny made himself extremely available for us and consistently offered great advice that proved instrumental in the success of our project.