Doctoral Dissertation extended summary

Eleni Mantouka, Civil Engineer NTUA

“Deep Reinforcement Leaning Traffic Models for Personalized Driving Recommendations”

The urban transportation landscape is facing many challenges due to the introduction of a variety of mobility solutions for travelers which together with innovations of Information and communication Technologies (ICT) subvert traffic management policies. Policy makers have to reconsider the applied traffic management measures in a way that automation and cooperation requirements of today’s services are taken into account together with the always increasing needs for green and sustainable mobility solutions. Nevertheless, even in this ever-changing transportation system, drivers remain the protagonists. Therefore, the understanding of decision-making while driving, as well as the investigation of driving habits adopted by drivers remains an active field of research for more than a decade.

Abnormal driving has been linked with increased crash risk and, thus, the improvement of driving behavior is considered critical for improving road safety. In addition, previous research has implied that the improvement of individual driving behavior may also result in an improvement of traffic conditions. Nevertheless, no evidence has been provided to support this statement, and the consequences of adjusting individual driving behavior on a network-level still remain unclear.

Within this context, the work contained in this dissertation is motivated by two main driving forces: i) the need to develop a driving recommendation system that treats each driver as an individual and proposes actions that meet his/her own driving preferences and, ii) the need to explore the actual impact of applying a personalized recommendation system on the road network.

Main objectives and research questions

The main objective of this dissertation is to design a personalized driving recommendation system which is based on deep reinforcement learning algorithms and aims at enhancing driving safety through the mitigation of aggressiveness and other unsafe driving habits. Subsequently, the impact of controlling individual driving behavior is assessed with regards to network performance and road safety, as well as the levels of harmful emissions by properly adjusting parameters of traffic models in a city-wide scenario setting using microsimulation. The above-described overarching goal of this dissertation can be divided in three major objectives as described below:

1. Exploit smartphone sensed data to understand driving behavior

2. Develop a traffic theory compatible personalized recommendation framework for improving driving behavior
3. Assess the impact of the recommendation system in traffic, safety and emissions

The concept of driving behavior analysis is not new, and, thus, a thorough review of the literature was conducted, at first, with the aim to identify research gaps and highlight the challenges and caveats that arise when smartphone crowd-sensed data are exploited for this purpose. The review of the literature resulted in the formation of the following 7 research questions.

Question 1 (Q1): Which are the main driving profiles that cover the wide range of driving behavior and how can they be identified by exploiting smartphone data?

Question 2 (Q2): Is it possible to classify the overall driving behavior of drivers into groups that share common driving characteristics, and, if so, to what extent could it be classified?

Question 3 (Q3): Could Artificial Intelligence techniques be exploited within the framework of a driving recommendation system and ensure the requires degree of personalization of the produced recommended actions?

Question 4 (Q4): Which is the most appropriate Reinforcement Learning algorithm for supporting human decision making?

Question 5 (Q5): Is there a link between raising self-awareness and improving conditions of the entire network? To what extent could the improvement of individual behavior affect traffic conditions?

Question 6 (Q6): What kind of impact does the controlling of individual driving behavior have on driving and road safety?

Question 7 (Q7): How are emissions affected by the controlling of individual driving behavior? Is there a significant change on environmental conditions when drivers improve their behavior?

Methodological approach

The recommendation system proposed within this dissertation is basically a decision support system for drivers that aims at mitigating aggressiveness and riskiness. Driving is a complex task since it requires from the driver to take both strategic and dynamic decisions as well as adapt their behavior to emerging conditions of the network. Contrary to the already developed ADAS, the system here has the following three state-of-the-art characteristics:

1. It is personalized, which means that it recommends the best driving actions to each individual taking into account their specific requirements and driving preferences.

2. It is self-aware, which means that the system takes into account previous behavior of each individual driver in order to propose the most suitable driving recommendations.

3. It is autonomous, meaning that it does not require any external input from the network or the traffic. Driving recommendations aim to improve individual driving behavior on its core, namely acceleration and deceleration decisions.

The development of the recommendation system is based on a Reinforcement Learning algorithm which is capable of producing the optimal behavior alteration for each driver given the way they have drove over their last trip.

In order to answer the research questions and achieve the overarching goal of the dissertation, an inclusive methodological framework is proposed which is based on a mixture of unsupervised learning and Deep Reinforcement Learning algorithms as depicted in Figure I.

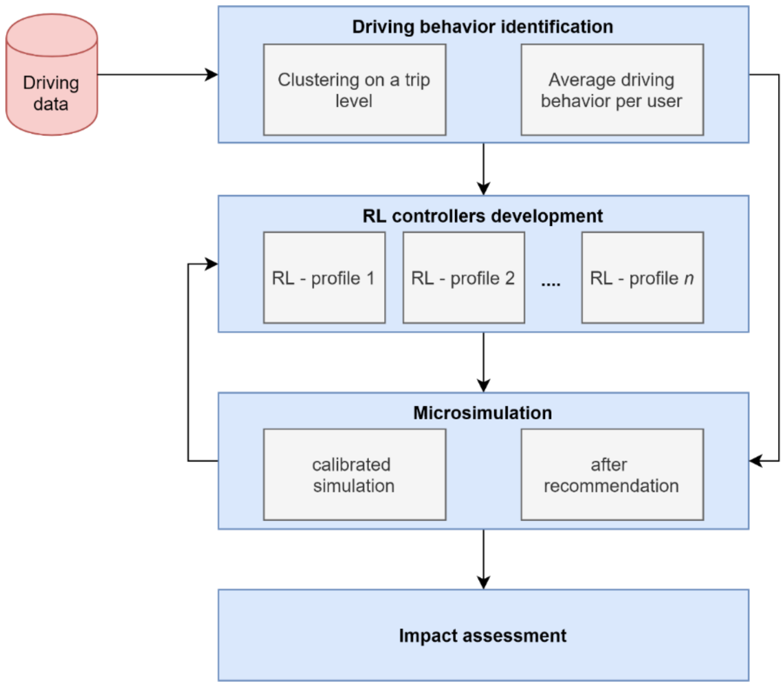

Figure I. Overview of the methodological framework

Starting from raw measurements of GPS location, acceleration and speed, as provided by a telematics application established on smartphone devices, driving features are defined that describe short-term and long-term driving behavior. Following, these features are utilized in an unsupervised learning framework to identify driving profiles that can be used to describe each driver's overall driving behavior (Q1, Q2). Driving behavior is defined at:

- a trip level, which corresponds to the way the driver performed a specific trip, and
- a user level, which corresponds to the overall driving behavior of a specific driver in all of his trips (driving footprint).

A two-level k-means clustering algorithm is implemented, in a selection of driving features, in order to distinguish aggressive from non-aggressive trips within the first level, and then further distinguish between risky and distracted driving at the second level of clustering. After this procedure, each trip was assigned to a specific driving profile (Q1), and then, using statistical measurements, the overall driving behavior of each driver is identified (Q2).

Once driving behavior per trip is identified, and all drivers were separated into groups based on their overall driving behavior, the driving recommendation framework is designed and the appropriate algorithms are developed, using state-of-the-art Reinforcement Learning models (Q3). The aim of the Reinforcement Learning algorithm is to learn the optimal policy and suggest the appropriate action that leads to the best possible behavior. Specifically, when dealing with driving behavior recommendations, every action refers to an adjustment of the vehicle's kinematic characteristics including the adjustment of the vehicle's speed and acceleration, which span within a continuous range of values. To this end, the RL algorithm developed in this work should retain one extra property, the ability to handle continuous state and action spaces (Q4). The RL agents follow an actor-critic approach based on the Deep Deterministic Policy Gradient algorithm and are both implemented as deep artificial neural networks, the hyperparameters and the structure of which emerge after an exhaustive grid search. The algorithms are trained using sequences of driving trips of the same driver as input, while the output of each RL controller is the optimal alteration in the acceleration of each driver, given the way they drove in their previous trip.

The structure of the system is such that there is a full mapping to microscopic standardization and traffic flow control logic. In correspondence with widely used car following models, the proposed algorithm acts as an estimation and prediction function of the acceleration at which the vehicle should move.

Finally, the impact of improving individual driving behavior is assessed through a comparative before-after microsimulation analysis, with respect to road safety, traffic and the environment (Q5, Q6, Q7). Using the road network of Athens, Greece, a microsimulation scenario for the morning rush hour demand, was set. For the initial conditions of the network the vehicles move according to the characteristics governing each of the driving behaviors detected in the first step of the methodological framework, while the traffic composition is based on the actual distribution of trips over the driving profiles. In this way, driving diversity is ensured between the vehicles and the traffic conditions in the network are simulated as realistically as possible.

The data

For the purpose of the specific research, data were collected through an innovative smartphone application developed by Oseven Telematics. The naturalistic driving database included 153,953 trips made from 696 unique drivers from December 2017 to August 2019. The trips were performed all around Greece, nevertheless the majority of them were conducted within the Region of Attica. For each trip, a variety of variables are available which include statistical measurements of acceleration and deceleration during a trip, speeding measurements that describe smoothly and with speed excess driving, as well as mobile usage indicators that describe how cautious the driver is. Table I presents the driving parameters used in the specific research.

Table I. Driving parameters per trip

Variable	Description	Unit
harsh_acc_per_min	Average number of harsh accelerations performed per minute	events/min
acc_avg	Average acceleration	m/s^2
acc_std	Standard deviation of acceleration	m/s^2
acc_q90	90% percentile of acceleration	m/s^2
acc_max	Maximum acceleration	m/s^2
harsh_brk_per_min	Average number of harsh decelerations performed per minute	events/min
dec_avg	Average deceleration	m/s^2
dec_std	Standard deviation of deceleration	m/s^2
dec_q90	90% percentile of deceleration	m/s^2
dec_max	Maximum deceleration	m/s^2
speed_max	Maximum speed	km/h
mbu	Percentage of driving with mobile usage	%
speeding_percentage	Percentage of driving with speed over the speed limit	%

All data provided by Oseven are in a fully anonymized format. The main characteristics of the sample used in the specific research are presented in Table II.

Table II. Main characteristics of the sample used in this research

	Total	Safe	Unsafe
Number of trips	153,953	66,566	87,387
Number of drivers	696	197	499
Average number of trips per driver		221	
Minimum number of trips per driver		16	
Average km travelled per driver		2,510 km	

Driving behavior analysis

In order to achieve the first objective of this dissertation, which is to exploit smartphone sensed data to understand driving behavior, a k-means clustering algorithm is implemented in to distinct levels.

For the first level of clustering, the number of clusters is set to k=2 and clustering is implemented on Euclidean distance matrix. Two of the variables that are used for the above procedure describe the number of harsh alterations of the longitudinal position of the vehicle (acceleration and deceleration), while the rest of them are essentially indices of the average acceleration and deceleration of the trip. The results of this first implementation of the k-means clustering are presented in Table III.

Table III. 1st level clustering results

	Harsh acceleration per min	Harsh brake per min	Average acceleration	Standard deviation of acceleration	Maximum acceleration	Average deceleration	Standard deviation of deceleration	Maximum deceleration	Number of trips
Aggressive trips	0.150	0.2081	1.748	1.525	3.847	-1.968	1.843	-4.547	71263
Non-aggressive trips	0.028	0.051	1.137	1.052	2.503	-1.282	1.286	-2.926	82690

Based on the clusters' centers, the trips can be distinguished between aggressive and non–aggressive driving, since trips belonging to the first cluster are featured by aggressive driving characteristics, such as great acceleration and deceleration metrics and significantly higher rates of harsh events per minute of driving.

The second level of k-means clustering was applied separately to the two groups that emerged from the first level of clustering using two driving parameters: the percentage of driving with mobile usage and the percentage of driving with speed over the speed limit. Results of this second level of clustering are presented in the table below (Table IV).

Table IV. 2nd level clustering results

	Percentage of mobile usage	Percentage of driving with speed over the speed limit	Number of trips
		Aggressive trips	
Distracted	0.511	0.062	4505 (2.9%)
Aggressive	0.019	0.032	54394 (35.3%)
Risky	0.023	0.269	12364 (8%)
		Non-aggressive trips	
Risky	0.021	0.306	12494 (8.1%)
Moderate	0.014	0.029	66566 (43.2%)
Distracted	0.514	0.057	3630(2.4%)

The resulting clusters seem to reveal richer driving profiles: distracted driving is recognized by higher values of the percentage of mobile usage while driving, while risky driving is identified through higher values of percentage of driving with speed over the speed limit. The two remaining clusters which have the lower values in both measures are annotated as "aggressive" and "moderate" for the aggressive and non-aggressive trips subsets respectively.

In order to separate drivers into groups with the same driving preferences, an average driving profile of each individual was identified by applying a simple rule. All four driving profiles indicating an unsafe driving behavior (Risky, Distracted, Aggressive-risky, Aggressive-distracted) were grouped as the worst class (3), aggressive trip profiles constitute the second class (2), while trips with typical characteristics belong to the first class (1), as shown in Figure II. For each individual driver, an average from all their trips is estimated and drivers are separated into two main groups based on their average behavior, as follows:

- Moderate/typical drivers: *trip average ≤ 1.5*
- Reckless drivers: *trip average > 1.5*

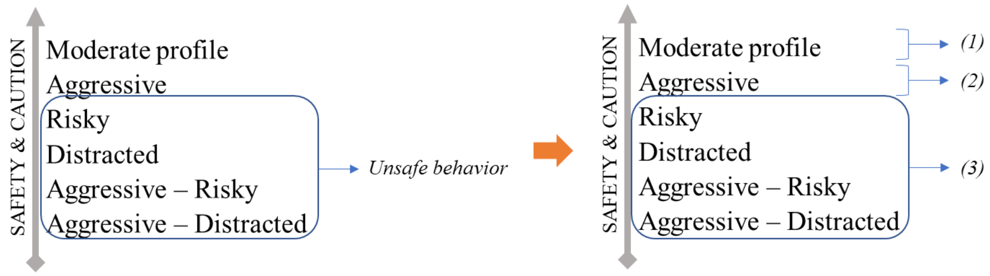

Figure II. Trip profile grouping for drivers' average driving profile estimation

For each individual driver, an average of the annotations from all their trips is estimated, where trip average less than 1.5 implies a moderate/typical driver and trip average greater than 1.5 refers to reckless drivers. Based on some statistical analysis, trip average less than 1.5 indicates that at least 60% of the trips performed by a driver are characterized by "moderate" driving behavior. In order for the developed controller to be as adaptive as possible to each individual's behavior, the proposed framework should be very strict when characterizing a driver as "typical/moderate" in order to avoid suggesting changes in behavior that the driver himself is impossible to follow as they will be far from his own average behavior.

RL: concept, principles and model development

In order to develop the Self-Aware Driving Recommendation Assistant (SADRA), a structured procedure is followed. First, the total trip database is divided into two, based on the average driving profile of each driver. In particular, the first database includes the trips of all drivers belonging to the "typical-safe" drivers, while the second includes all the trips of drivers with unsafe average driving behavior. For the sake of brevity, from this point on, the RL controller that corresponds to the "typical" drivers is referred to as SADRA – I, while the corresponding controller for the reckless drivers is referred to as SADRA – II respectively.

Every RL agent consists of three main components: *states (s)*, *actions (a)* and *rewards (r)*. In each timestep the agent observes the current state of the environment and takes the appropriate action from the set of the possible actions. Then, the agent receives a reward which measures the success or failure of the agent's actions for the given state.

In this study, the environment states are defined through a five-dimensional vector that describes how a driver drove during their trip and includes trip's average acceleration (a_{avg}), 90% percentile of acceleration (a_{90}), average deceleration (d_{avg}), 90% percentile of deceleration (d_{90}) and percentage of driving with speed over the speed limit (*speeding*):

$$\boldsymbol{s} = \{a_{avg}, a_{90}, d_{avg}, d_{90}, speeding\}$$

Our recommendation system is not context-aware which means that its ultimate goal is to improve individual's personal driving style independently from the road setting they are driving in (type of road, traffic conditions, etc.). The selection of the appropriate speed is not independent from the road geometry and road traffic, as well as deceleration decisions are not always independent from the leading vehicle's behavior and traffic signals. Therefore, the only parameter that purely describes one's driving style is the acceleration, as it is only dependent on the driver's perception and preference between smoothly or harshly accelerating. Indeed, in recent literature, a driver's driving style is usually defined by their acceleration profile. To this end, actions that the system produces and are proposed to the driver belong to a continuous action space which is defined by a two-dimensional vector including a change in average acceleration and in the 90% percentile of acceleration, which define the usual/preferred acceleration for the entire trip in regular situations and the value that should not be exceeded, e.g., when performing overtaking maneuvers, except from cases of emergency:

$$\boldsymbol{a} = \{da_{avg}, da_{90}\}$$

For the sake of simplicity from hereon, the 90% quartile of the acceleration may be equally referred to as "maximum acceleration".

A key component of the RL agent is the reward function. The aim of the reward function is twofold; to evaluate the current state and the transition between states. In other words, the driving behavior at each trip, as well as the change in driving behavior between successive trips of the same user are evaluated. For this purpose, a custom driving evaluation function had to be constructed first. The score of each trip was estimated by the distance of this specific trip from the center of the moderate profile (the center of the cluster), in order to quantify how far each individual's behavior is from the typical (moderate) behavior. For the purpose of this analysis, the Mahalanobis distance is used to estimate the distance between each trip and the moderate profile.

Trip evaluation is performed on the basis of the following formula:

$$trip\ score_i = e^{-driving\ profile_i * \frac{\text{Ɱ}_{(i,moderate\ profile)}}{Q_{75}(\text{Ɱ})}}$$

where i is an individual trip and Ɱ is the Mahalanobis distance. Here, the 3rd quartile of the Mahalanobis distance is used instead of the maximum value in order for the score function to be stricter with drivers whose behavior excludes more than 75% of the typical (moderate) behavior.

The reward function for a driver moving from one trip to the next one was established based on the following formula:

$$\boldsymbol{r} = \ trip\ score_{i+1}\left(1 + \frac{trip\ score_{\mathrm{i+1}} - trip\ score_{i}}{100}\right)$$

Once the main components for the development of the RL controllers were estimated, the data were organized in the following format:

$$(state, action, reward, next\ state)$$

For every unique driver in the dataset, their trips were sorted in an ascending order according to each trips starting date. The training samples were tuples of sequential trips of a specific driver along with the corresponding action and reward of the transition from the first trip to the succeeding one. It should be noted that for every distinct driver in the dataset, their first trip was used only as "*state*" while their last trip of was used only as "*next state*". Following this data preparation procedure, 33,440 unique data samples were constructed for training SADRA I and 119,817 unique data samples were used for the training process of SADRA II.

The RL controllers are developed based on the Deep Deterministic Policy Gradient (DDPG) algorithm which implements an actor-critic approach to learn a policy and produce the optimal actions. Thus, for each controller two neural networks are developed; representing the actor and the critic respectively. The actor (μ) and critic (Q) networks for both the safe and unsafe drivers' subsets were trained following the procedure of Algorithm below.

DDPG Algorithm implementation

Initialize critic $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ networks using rewards as Q-values
Set the above as initial target networks (Q' and μ')
Split the sample into M minibatches
for minibatch=1, M **do**
  Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
  Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i\left(y_i - Q(s_i, a_i|\theta^Q)\right)^2$
  Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta_\mu} J \approx \frac{1}{N}\sum_t [\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t)}\ \nabla_{\theta_\mu}\mu(s|\theta^\mu)|_{s=s_t}]$$

  Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

end for

An exhaustive grid search was performed in order to conclude to the final architecture of the two networks. Specifically, all possible combinations of the networks' structures and parameterization, within a range of reasonable values, have been examined and compared in order to detect the optimal one. The parameters that were taken into consideration are: number of hidden layers, number of neurons and activation of each layer, optimization algorithm and learning rate, batch size and number of training epochs, as shown in Table V.

Table V. Hyperparameters of the Critic and Actor networks for both SADRA I and II

Hyperparameters	Critic network	Actor network
	SADRA I – Safe drivers	
Number of hidden layers	6	3
Number of neurons per layer	(64,32,16,16,32,64,1)	(128,64,32,2)
Epochs	200(initial network:110)	200(initial network:110)
Batch size	150(initial network:150)	150(initial network:150)
Activation	ReLU	ReLU
Optimizer	Adam	Adam
Learning rate	0.001	0.0001
	SADRA II – Unsafe drivers	
Number of hidden layers	6	3
Number of neurons per layer	(32,16,8,8,16,32,1)	(128,64,32,2)
Epochs	170(initial network:170)	210(initial network:210)
Batch size	250(initial network:250)	250(initial network:100)
Activation	ReLU	ReLU
Optimizer	Adam	Adam
Learning rate	0.0001	0.0001

Simulation setting

The quantification of the impact of adopting driving recommendations by all drivers on traffic, road safety and emissions was performed under a network-level microscopic simulation scenario. The SUMO simulation software is used and its default car-following model, Krauss model, which is a microscopic, space-continuous model based on the safe speed; the driver of the following car adopts a safe speed which allows them to adapt to the deceleration of the leading vehicle.

The case study for the simulation experiments is the inner-ring network of Athens, Greece. The network consists of 1,293 nodes/intersections and 2,572 edges. The total length of the network is 348 kilometers. The calibration of the network led to the definition of **86,054 vehicles**, achieving a total of 1,393,634 counts (97.47% of the total counts extracted from the Aimsun simulator) and a GEH value below 5 (GEH < 5) for 95.26%.

Two distinct scenarios were designed both corresponding to the demand of the Athens' Road network during the morning peak hour (8:00 – 9:00 AM). First, the initial conditions of the network are simulated in order to estimate the performance of traffic when vehicles move around, based on the characteristics that govern the six identified driving profiles. In order to ensure the robustness of the results, simulation was performed in 10 replications with ten different seed numbers. Stochasticity is an important aspect of reproducing reality in a simulation scenario, since it adds randomness over the distributions of difference aspects of the simulation (e.g., route distributions, vehicle type distributions). Subsequently, driving recommendations were produced offline for every served vehicle based on the way each vehicle performed their trip. The recommendations were produced from the corresponding RL controllers using as input the state of the trip (average acceleration, 90% percentile of acceleration, average deceleration, 90% percentile of deceleration, speeding percentage) and as output the optimal alteration of the maximum acceleration. It should be highlighted here that although the developed RL controllers produce a two-dimensional vector that includes alterations on both the average and the maximum acceleration, only the maximum acceleration was exploited during the simulation runs, since the Krauss model takes into account only the maximum values of acceleration and deceleration.

Finally, a second simulation run was performed, where previously served vehicles follow the proposed recommendations, namely an alternation of their maximum acceleration, while the rest of the traffic follows the distribution among the six driving profiles.

The behavior that implies each driving profile was simulated through the adjustment of the car-following model. The car-following model can be parametrized by a number of parameters: the maximum acceleration of the vehicle (*accel*), the maximum deceleration of the vehicle (*decel*), the maximum velocity of the vehicle (*maxSpeed*), the maximal physically possible deceleration for the vehicle (*emergencyDecel*) and the vehicles' expected multiplicator for lane speed limits (*speedFactor*). At first, the current (initial) state of the road traffic is simulated in SUMO using the six defined driving profiles, whose parameters were introduced to the Krauss model of different vehicle types, as shown in Table VI.

Table VI. Car-following model parameters for each vehicle type

Vehicle types (trip profiles)	**Car-Following Model Parameters**				
	accel (m/s^2)	**decel** (m/s^2)	**emergencyDecel** (m/s^2)	**maxSpeed** (km/h)	**speedFactor** (mean, min, max)
Moderate	2.519	-2.942	-5.909	64.51	(0.029, 0, 0.168)
Aggressive	3.817	-4.483	-18.083	66.93	(0.033, 0, 0.151)
Risky	2.392	-2.824	-5.328	100.28	(0.306, 0.1627, 0.96)
Distracted	2.601	-2.990	-5.112	67.38	(0.057, 0, 0.631)
Aggressive-risky	3.944	-4.825	-25.884	100.8	(0.269, 0.147, 0.907)
Aggressive-distracted	3.939	-4.553	-10.845	71.99	(0.062, 0, 0.744)

For the initial state of the network, the six distinct vehicle types were created in a route file, with the corresponding car-following model's parametrization. The route of each vehicle was also identified in the route file, as it was estimated from the path assignment of Aimsun. In one hour of simulation for the morning peak, about 58% of the total demand was inserted in the network and 28% of the vehicles completed their journey within this time.

Subsequently, for each vehicle that reached their destination the following parameters were estimated for each trip:

- average acceleration
- 90% percentile of acceleration
- average deceleration
- 90% percentile of deceleration
- speeding percentage

These driving characteristics were used as input to the RL controllers which recommend the optimal action for each trip. For the second run of simulation, the exact same vehicles were used, which follow the exact same routes on the same road network, in order to estimate the impact of the recommendation. The proposed actions of each vehicle were introduced as a modification of the car-following model's parameter in the route file. The adoption of this approach enabled hands-on implementation of the recommendation process with direct control over the outcomes.

In this case as well, 10 replications with the same seed values as before, were performed to ensure the robustness of the results. Findings revealed that in one hour of simulation 57% of the demand was served on average, while the corresponding percentage of served vehicles was reduced by 1% compared to the initial conditions.

Impact assessment

Impact assessment of the proposed system is performed using microsimulation and by following a before-after approach. Specifically, for both simulation cycles the Key Performance Indicators of traffic, safety and environmental conditions were estimated, and comparatively assessed so that to quantify the overall impact of adopting personalized driving recommendations which improve each individual's driving behavior. The KPIs used in the analysis for each network's aspect are presented in Table VII.

Table VII. Key Performance Indicators for each network's aspect

Traffic	Safety	Environment
Served demand	Total conflicts	Cumulative amount of emissions (CO_2, CO, PMx, NOx)
MFDs	Total rear-end conflicts	Emissions per vehicle
Travel times	Conflicts per vehicle	

The estimation of traffic-related KPIs was dependent on the outputs of the simulation, which included the number of inserted and served vehicles, as well as edge-based information regarding the three fundamental elements of traffic flow theory (flow, speed and density). Instead of using aggregated measures of the fundamental variables, the Macroscopic Fundamental Diagrams (MFDs) were constructed and significant outcomes were drawn regarding the differences in the performance of the network before and after the application of the recommendation system. The estimation of the harmful air pollutants is based on the emissions' model already integrated into SUMO, the PHEMlight model. PHEMlight is a simplified version of PHEM (Passenger car and Heavy-duty Emission Model), a complete vehicle emissions model developed in Europe since 1999. PHEM is based on extensive emission measurements on vehicles such as passenger cars, light duty vehicles and urban buses. The approximation of the conflicts that constitute an indicator for road safety is based on the SSAM tool, which computes a number of surrogate measures of safety for each conflict (crossings, rear-ends, lane changes) that is identified in the trajectory data and then computes summaries (mean, max, etc.) of each surrogate measure.

Results: Driving recommendations

The two versions of the trained DDPG algorithm were used to produce driving recommendations with respect to two categories of drivers; typical drivers who exhibit a moderate average behavior (SADRA I) and unsafe drivers who interchange their behavior among various unsafe driving habits (SADRA II). The recommendations are in the form of driving alterations that refer to the optimal driving actions that the specific driver can adopt in order to improve their driving based on their current behavior.

A comparison between the outputs of the two controllers revealed that both of them are trained to generate recommendations that move drivers closer to the average safe behavior of a typical driver, which has an average acceleration equal to 1.137 m/s^2 and a maximum acceleration equal to 2.503 m/s^2. Based on the indicative samples of the table below (Table VIII), the mean recommended average acceleration was estimated 1.145 m/s^2, while the mean value of the proposed maximum accelerations was 2.507 m/s^2 respectively. It can therefore be concluded that a universal application of the proposed recommendation system would lead to the harmonization of the acceleration profiles for the entire fleet of vehicles.

Table VIII. Example of RL input and output and the produced recommendations

Profiling and Scoring		RL input					RL output		Recommendation	
Driving profile	**Score**	**Acceleration average**	**Deceleration average**	**Acceleration Q90**	**Deceleration Q90**	**Speeding percentage**	**action_1 (max)**	**action_2 (avg)**	**new Acceleration Q90**	**new Acceleration average**
Aggressive-distracted	9	1.512	-1.58	3.600	-3.72	0.0090	↓ -0.99	↓ -0.31	2.610	1.197
Aggressive-distracted	2	1.888	-1.86	3.336	-3.95	0.0660	↓ -0.29	↓ -0.51	3.046	1.377
Aggressive-risky	6	1.779	-1.63	3.432	-3.82	0.3235	↓ -0.65	↓ -0.49	2.778	1.288
Aggressive-risky	2	1.574	-1.73	3.480	-4.20	0.5693	↓ -0.96	↓ -0.40	2.519	1.171
Aggressive	2	1.905	-1.68	3.600	-3.18	0.0000	↓ -0.77	↓ -0.63	2.834	1.275
Aggressive	17	1.948	-2.57	3.768	-5.69	0.0822	↓ -1.10	↓ -0.66	2.669	1.289
Aggressive	10	1.774	-2.17	4.368	-4.08	0.0000	↓ -1.69	↓ -0.54	2.676	1.230
Distracted	2	1.405	-1.33	4.104	-2.88	0.0133	↓ -1.38	↓ -0.21	2.727	1.193
Distracted	24	1.504	-1.52	3.324	-3.36	0.0000	↓ -0.64	↓ -0.27	2.679	1.236
Distracted	35	1.054	-1.08	2.489	-2.67	0.0000	↑ 0.06	↑ 0.11	2.548	1.166
Moderate	73	1.386	-1.59	2.760	-3.28	0.0226	↓ -0.46	↓ -0.33	2.299	1.055
Moderate	58	0.592	-0.92	1.560	-2.81	0.0294	↑ 0.56	↑ 0.39	2.123	0.980
Moderate	86	1.366	-1.43	2.880	-3.36	0.0000	↓ -0.45	↓ -0.26	2.426	1.106
Moderate	90	0.826	-1.03	1.984	-2.43	0.0000	↑ 0.26	↑ 0.18	2.240	1.006
Moderate	67	0.967	-1.43	2.040	-3.67	0.0674	↑ 0.05	↓ -0.01	2.091	0.954
Moderate	78	1.047	-1.23	2.070	-2.33	0.0000	↑ 0.21	↓ -0.02	2.284	1.030
Moderate	77	0.666	-0.69	1.440	-1.68	0.0367	↑ 0.76	↑ 0.32	2.195	0.989
Moderate	94	1.124	-1.24	2.568	-2.64	0.0000	↓ -0.06	↓ -0.01	2.511	1.118
Moderate	58	0.871	-1.24	1.920	-3.56	0.1667	↑ 0.27	↑ 0.13	2.187	1.003
Risky	52	1.367	-1.36	3.093	-3.06	0.2392	↓ -0.49	↓ -0.17	2.602	1.196
Risky	33	1.098	-1.43	2.280	-3.24	0.2857	↑ 0.32	↑ 0.09	2.600	1.192

Figure III provides some indicative examples of the recommendations produced by the two controllers given the same input (first state). Findings revealed that although the recommendations of the controller concerning unsafe drivers (SADRA II) lead to significantly lower average accelerations for the next trip (next state) compared to the previous trip (initial state), they maintain a significant distance upwards for the respective recommendations produced from the typical drivers' RL controller (SADRA I). Nevertheless, it should be noticed that both the controllers lead to a smoother acceleration profile for the entire traffic.

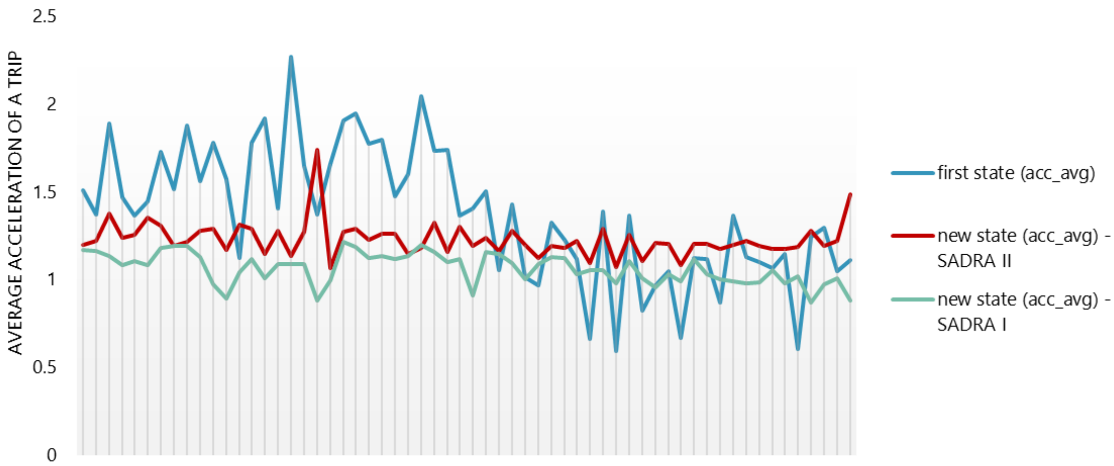

Figure III. Comparison of the new state's average acceleration as it emerged from the Typical and the Unsafe RL controllers.

Results: Simulation and Impact

The quantification of the impact of applying the proposed recommendation system and in consequence, of the adoption of an improved driving behavior by all drivers is of great importance both for researchers as well as practitioners and can lead to significant findings regarding the usefulness of improving individual driving behavior. The assessment of the recommendation system is performed by utilizing specific Key Performance Indicators that correspond to three areas of interest: traffic, safety and emissions. Each of the simulation rounds was done in 10 replications to enhance the validity and robustness of the results. In total, the trained SADRA I controller was used to produce recommendations for 43% of the vehicles, while the rest of the vehicles followed the recommendations produced by SADRA II.

All replications of the same simulation round present mutual results which are slightly reduced after the application of the recommendation system. On average, 2.9% less vehicles were served based on the results of the second round of the simulation. However, results of the statistical hypothesis test *t-test* indicated that there are no significant differences between the means of the served vehicles before and after the recommendations in 95% confidence interval. A greater investigation of the traffic flow properties together with aggregated metrics of driving behavior was conducted to further quantify the impact on the other dimensions of the road network as well. The application of the personalized recommendation system had a substantial impact on the maximum acceleration of the vehicles, as shown in Figure IV. When all vehicles followed the suggestions generated by the two RL controllers, the mean value of the maximum acceleration was somewhat increased from 2.83 m/s^2 to 2.96 m/s^2, mostly because the majority of the vehicles who adopted a very small maximum acceleration, which was far lower from the corresponding acceleration of the "moderate/typical" behavior, they were suggested to slightly increase their acceleration. However, the condensation of the interquartile range is evident after the recommendations, which indicates the harmonization of the acceleration profiles of all vehicles in the simulation. Finally, the maximum value of the observed maximum accelerations remained at the same level of 3.94 m/s^2 after the application of the proposed system.

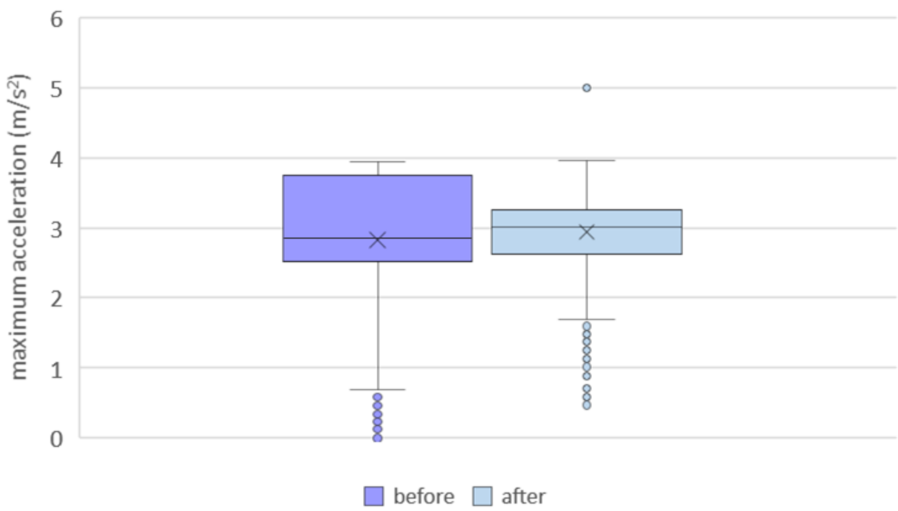

Figure IV. Boxplot of maximum acceleration before and after recommendation

The differences observed in the magnitude of the average speed are minimal, since in both situations the vehicles adopt an average speed of around 25 km/h, while the maximum average speed that is observed is approximately 55 km/h.

Alterations on the speed of vehicles resulted on changes of the rest traffic flows properties, namely flow and density. Microscopic fundamental diagrams were calculated to provide a thorough graphical representation of these variables' relations for the initial conditions as well as the conditions emerged after the recommendations. All three fundamental diagrams (Figures V – VII) demonstrate the relationships between traffic flow properties, namely mean vehicle flow, mean density and mean speed, as they emerged from the simulation based on aggregated measurements of all edges for the 10 replications. Results indicate that the implementation of self-aware driving suggestions although it leads to safer and less aggressive driving behavior for each individual, it does not improve the performance of the road network. More specifically, self-improvement is evident from the lower mean density values which indicates that vehicles keep greater distances from the leading vehicles. Additionally, lower speeds are also observed after the adaptation of the recommended accelerations with the difference from the initial conditions being more significant in the case of saturated network flow (Figure V).

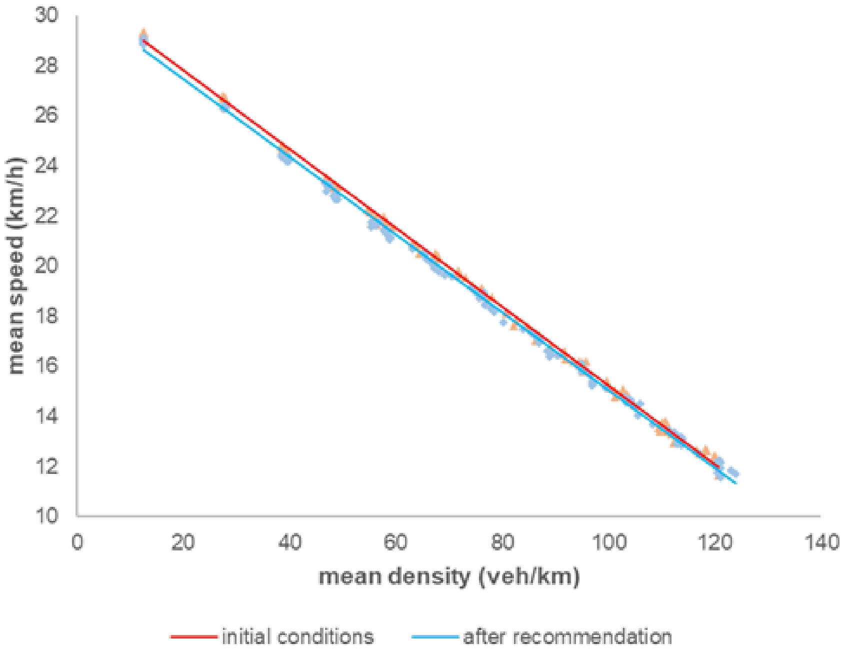

Figure V. Fundamental diagram of speed-density before and after driving recommendations, based on simulation results

Individual driving safety is augmented, yet the impact on traffic conditions is not similarly positive. The vehicles that move at lower speeds and with a lower density worsen traffic flow conditions, since fewer vehicles are served per time unit compared to the initial conditions. Nonetheless, this decrement of mean flow may be considered acceptable if assessed in conjunction with the positive effects on driving safety. However, based on the findings of this research, it can no way be concluded that the improvement of personal driving behavior is associated with a significant improvement in traffic conditions and therefore, the imposition of soft policy measures, such raising self-awareness with respect to individual driving safety and performance, it cannot be considered as a key measure for traffic management.

The fundamental diagram of flow-density seems to depict a uniformity between the initial and the final conditions, although some minor differences are observed with respect to the absolute value of capacity flow (Figure VI). Specifically, for the value of critical density, which was estimated 33.1 veh/km, the corresponding values of traffic flow are 360 veh/h and 358 veh/h for the initial conditions and after the recommendations respectively.

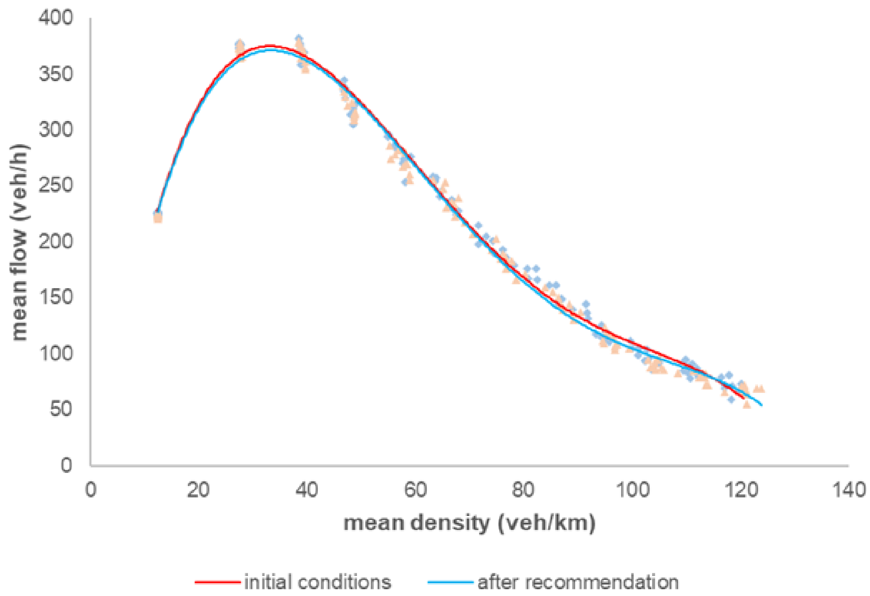

Figure VI. Fundamental diagram of flow-density before and after driving recommendations, based on simulation results

The flow-speed diagram is used to determine the speed at which the optimum flow occurs. For the initial conditions of the road network, the optimum flow occurs when vehicles move with 26.1 km/h, while the corresponding speed after the recommendation is reduced 3.4% with its absolute value estimated 25.2 km/h (Figure VII).

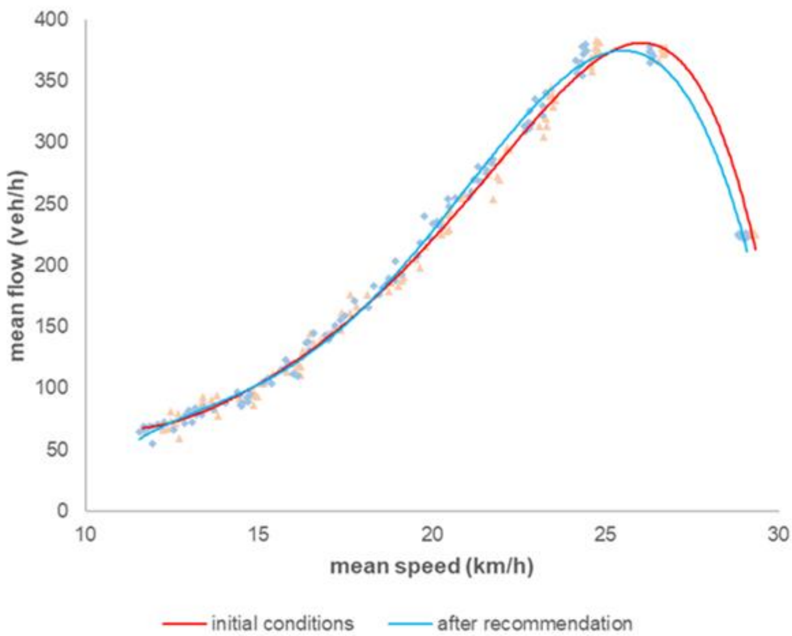

Figure VII. Fundamental diagram of flow-speed before and after driving recommendations, based on simulation results

Except for the performance of the network, another key performance indicator is safety. The assessment of the applied recommendations with respect to safety was performed by calculating the number of conflicts occurred between the vehicles during the simulation. Table IX presents the number of conflicts that were observed for the entire traffic before and after the recommendation. There are three types of conflicts that can be identified from vehicles' trajectories, which are crossings, rear-ends and lane changes. Here, a special focus on rear-ends is given since the proposed recommendations only affect the car-following behavior of each driver.

Table IX. Safety performance indicators in Athens Network before and after applying driving recommendations

	Initial conditions	After recommendation [% difference]
Vehicles served	23,990	23,302
(in one hour of simulation)	*(27.88% of demand)*	*(27.08% of demand)*
Total number of conflicts	2.86 conflicts/vehicle	2.75 conflicts/vehicle [-4.2%]
Rear - ends	2.01 rear-ends/vehicle	1.90 rear-ends/vehicle [-5.5%]

A reduction of 4.2% of the total number of conflicts was observed when vehicles followed the corresponding driving recommendations, while the corresponding percentage of elimination for the rear-end conflicts is 5.5%. Although these percentages may not seem very high, the absolute number of conflicts that was calculated after the recommendation is significantly reduced by approximately 6,000 conflicts for the one hour of simulation. Rear-ends constitute about 33% of the total number of conflicts, which indicates that each driver gets involved in all different kind of conflicts during driving.

Some indicative results on the impact of the proposed recommendation system on emissions is provided. The corresponding Key Performance Indicator is the level of emissions for all different kind of air pollutants, namely Carbon Dioxide (CO_2), Carbon Monoxide (CO), Particulate Matter (PM_x) and Oxides of Nitrogen (NO_x). A significant reduction in all categories of emissions is observed compared to the initial conditions of the network, as shown in Table X. Findings revealed that the homogenization of acceleration profile for the entire traffic has led to a slightly reduced emissions per vehicle. Specifically, the reduction in all categories of emissions is estimated as follows: 2.5% in CO_2, 0.3% in CO, 1.3% in PM_x and 3.3% in NO_x. It should be noted that this improvement in the environmental conditions is very important since the proposed recommendation system had a positive impact on emissions despite the fact that the controller was not trained towards this direction.

Table X. Difference in vehicle emissions before and after applying driving recommendations

Emissions	Initial conditions	After recommendation [% difference]
CO_2	0.704 kg/vehicle	0.686 kg/vehicle [-2.5%]
CO	0.027 kg/vehicle	0.026 kg/vehicle [-0.3%]
PM_x	0.0133 g/vehicle	0.0131 g/vehicle [-1.3%]
NO_x	0.296 g/vehicle	0.287 g/vehicle [-3.3%]

Conclusions and main contributions

The main findings of the dissertation can be summarized in the following points:

- A two-level clustering approach can provide great insights on the characteristics that govern aggressiveness during driving and can be further exploited to distinguish safe from unsafe driving patterns.
- Six distinct driving profiles are able to describe the overall driving behavior that someone performs during their trip.
- There are two categories of drivers according to the average behavior of each driver resulting from how they drove in all their trips. In the first category drivers usually drive in a typical manner while in the second category drivers perform a number of unsafe driving actions or drive in an aggressive manner in the majority of their trips.
- The Actor-critic approach from the family of reinforcement learning algorithms can be exploited to find the best possible driving action for each dividual driver given the way they drove in their previous trip.
- When a controller provides driving recommendations to a fleet of vehicles, the acceleration profile of the entire fleet is harmonized on a value which is close enough to the acceleration decisions of a typical – safe driver.
- The application of a personalized recommendation system to a city's road network does not have a significant impact on traffic conditions.
- When each driver improves their own behavior, road safety is enhanced on the network. Specifically, critical conflicts between vehicles are significantly reduced after the application of the proposed system.
- The level of emissions for all different kinds of air pollutants is reduced which indicates that harmonization of the accelerations for the entire traffic can have an important positive impact on the environmental conditions.

Concluding, it should be noted that the deterioration of traffic may be considered acceptable if one takes into account the compensation through the benefits of adopting smoother driving behavior in road safety and emissions. To this end, policy makers and researchers should not neglect the real impact on all network's dimensions when planning traffic management strategies and applying soft and hard policy measures.

The present doctoral dissertation offers significant innovative contributions in five areas:

1. It makes use of an innovative naturalistic driving dataset. A great volume of data was available with high temporal resolution from real driving, enriched with a variety of factors that describe driving behavior, environment and other external attributes for each trip.

2. It proposes a methodological framework to extract driving profiles straight from the data, which describe the entire range of driving behavior. A data-driven approach is followed to classify critical driving patterns that appear during a trip by exploiting k-means clustering as being the most appropriate tool for this purpose.
3. It develops novel Reinforcement Learning algorithms to solve a real-world problem, this of assisting driving behavior. A deep Reinforcement Learning algorithm was chosen as the most suitable tool to learn the optimal policy and suggest the appropriate action that leads to the best possible driving behavior for each individual driver.
4. It proposes a methodology which is capable of recognizing individual driving preferences and produce personalized driving actions to each driver. Specifically, an inclusive methodological framework is implemented which incorporates tools and methods that first recognize driving behavior of every user, then assigns every user to the corresponding RL controller version based on their overall behavior and finally produces personalized driving actions that mitigate aggressiveness and riskiness of driving.
5. It evaluates the large-scale network effects of implementing a personalized driving recommendation system on three areas of interest using specific KPIs, precisely on traffic, safety and emissions. Impact assessment of the proposed recommendation system is performed using a real-world scenario that of the Athens' Road network through microsimulation and by applying a before-after methodology to compare the values of the KPIs before and after the application of the system.

Limitations, impact and future research

As any other data-driven approach, this research as well, relied on some limitations with regards to problem setup and adaptation. Firstly, some limitations emerged from the need to match the RL output with the simulation properties. More specifically, one of the two components of the recommended action, the average acceleration of each driver, could not be imported into the microsimulation car-following model, which is parametrized by the acceleration ability of vehicles and therefore only the maximum acceleration is adopted within the simulation. Nevertheless, due to the nature of the phenomenon of driving, all parameters describing how a driver chooses to drive over a trip are inextricably linked with each other and therefore, the neglection of the average acceleration was not expected to have a significant impact over the results of the simulation. Besides car-following behavior, a driver during their trips takes actions regarding lane change, priority concession and other decisions concerning interactions with other road users. However, in this research the focus was explicitly on the car-following behavior as the ultimate goal was to create a user-centric system that looks only at the driver and does not require any external information from the road network in order to be trained and implemented. Thus, the proposed actions refer on the way the driver drives along the road, namely the way they choose to hit the acceleration pedal, which depends only on the personal preferences and perceptions of the driver. The lack of information about the environment can be considered as a limitation of the developed system, since its transformation into a context-aware system would give other perspectives both to the system itself and to the possibilities of its use as a traffic management tool.

An extension of the above limitation is the fact that since the system ignores the state of the environment it cannot operate real-time. In other words, the proposed methodology is not able to produce recommendations real-time, namely during a trip. Instead, an offline system is developed which suggests alterations on driving behavior in a sequence of trips for each driver. The integration of external information into the system would allow, at least conceptually, the real-time provision of driving recommendations.

Lastly, another limitation, which applies to all data driven approaches, is the generalization and transferability of the developed model and the corresponding outcomes. In most cases it is unclear whether the sample used to train the model is representative of the entire population and also whether its characteristics are similar to those of a different population. In this work, a big naturalistic driving dataset is used to develop the RL models which includes trips performed by a great number of drivers, nevertheless, it cannot be said that the results can be generalized and spatially transferred to another road network.

Besides the limitations described above, the outputs produced within this dissertation may have a significant impact on several aspects of both research (R), technology (T) and policy-making (P). Future research can benefit and significantly evolve by further examining the conclusions drawn with regards to the following points:

- (R) Aggressiveness does not necessarily constitute an unsafe driving habit and can be detected either as an individual behavior or in combination with other unsafe behaviors.
- (R) Reinforcement learning algorithms can be implemented in real-world problems and specifically, the DDPG algorithm can learn how to make human-like decisions on complex and high-dimensional environments.
- (R & T) The identified human driving profiles can provide great insights for human-like autonomous driving.

Technological advancements can be achieved in case the proposed recommendation system is incorporated in already developed software, such as insurance telematics apps and ADAS. Such system can be revolutionized, become more human friendly and adopt a more personalized way of supporting human decision making.

Moreover, policy makers could take advantage of the results of this dissertation to redesign soft policy measures and redefine the role of drivers in the current traffic management strategies, since in this work it was shown that the improvement of driving behavior on an individual level can have significant impact on road safety and emissions, but not a noteworthy impact on traffic conditions.

Finally, it can be understood that findings of this work can have far reaching implications for future research. Although this research provides significant contributions on driving behavior analysis, there is still much room in the exploration of driving behavior dynamics and thus, further research should be conducted in that direction involving enriched driving datasets and additional driving behaviors and parameters (e.g., cornering, tasks that cause distraction except from mobile usage). Moreover, the dedicated study of the dynamic evolution of driving behavior is also very important to provide answers to the question of how much and how rapidly driving profiles are altering over time. Another direction of future research concerns the recommendation system, which should investigate the way the produced recommendation should be passed to the driver in order for them to be understood by the user and then to be accepted by him. Furthermore, the identification of the required specifications that will enable the real-time operation of the system could also be a part of future research. Towards this direction, the most significant future research objective would be the modification of the proposed system in a way that it becomes context-aware, meaning that the system can interact with environment in which the agent takes decisions and have a full view of its dynamics and alterations. In this way, the proposed system could be implemented in real-time, and additionally it could also act as a traffic management tool which uses driving behavior as a key force of enhancing traffic efficiency.