Maintaining Vehicle Driver's State Using Personalized Interventions

Kirill Uvarov[*], Andrew Ponomarev[†‡]
[*]Saint Petersburg Electrotechnical University "LETI", St.Petersburg, Russia
[†]ITMO University, St.Petersburg, Russia
[‡]St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St.Petersburg, Russia
uvarovkirill73@gmail.com, ponomarev@iias.spb.su

Abstract— **Distracted driving or driving in a tired state causes many road accidents. One of the approaches to address this problem is to use an automated solution that detects driver state and executes some actions, aimed on maintaining driver's attention focus. Personalized intervention systems are actively developing in the field of medicine, however, in the field of transport, no similar studies have been found. This paper examines the idea of using reinforcement learning to generate personalized driver interventions. The contribution of the paper is twofold. First, it proposes a conceptual model and general reinforcement learning formulation of the problem. Second, it describes a driver simulation model that can be used to train the personalized intervention policy. Experimental study shows, that the proposed formulation allows one to train a personalized policy that can be used to effectively maintain the desired state of the vehicle driver. The implementation of such method in a driver monitoring system can be a very promising direction and help to reduce the number of road accidents for the above reasons.**

I. Introduction

Driver inattention is one of the most common and serious road safety problems. According to the World Health Organization (WHO), more than 1.3 million people die annually in road accidents [1]. Inattentive and distracted driving are common causes of traffic accidents. According to Regan, inattention is defined as “insufficient or no attention to activities critical for safe driving” [2]. The distribution of mobile devices and smartphones is a growing risk factor that can have serious consequences. Using a mobile phone increases the probability of getting into an accident by four times. Sending a text message while driving increases the risk of an accident by 23 times. In addition, the driver's reaction time when using the phone is 50% slower than without it. The problem with driving in a tired state is also quite significant. According to the National Sleep Foundation, approximately 32% of drivers in the United States get behind the wheel in a state of fatigue at least once a month [3]. To combat problems such as driver distraction or driving in a tired state, there are legislative measures that regulate the use of smartphones while driving.

The problem of driver inattention or driving in a tired state is not new. Therefore, developments in this direction have been underway for several years. In order to prevent road accidents due to the driver being in "not optimal" condition, a possible solution is to monitor the condition of drivers, and identify anomalies at an early stage. These functions are typically implemented in intelligent driver monitoring systems. The driver monitoring system consists of two main parts: a detection system and a counteraction system. The detection system analyzes data from various sources (for example, video cameras, wearable devices, car sensors, etc.) to determine the driver's condition. For example, Jaguar is working on the development of an active security system that reads human brain waves. This system allows you to determine whether the driver is careless or very sleepy. The counteraction system uses various countermeasures to increase a person's alertness while driving a car. Existing available counteraction systems can provide feedback in the form of notifications or warnings on the dashboard or infotainment system. An example of such warning may be the display of a "cup of coffee", which may be accompanied by an audible alert. In addition, some systems can simply inform the driver about the levels of alertness, displaying this information in the form of a vigilance scale. In this study, we are developing a solution within the framework of a driver monitoring system. The main purpose of this study is to develop a personalized driver intervention system. Interventions are short-term actions on the driver that can cheer up the driver or draw his/her attention to the road. As mentioned above, examples of such effects can be light or sound indication, tactile actions, for example, vibration of the steering wheel, etc.

Each person is unique and can react differently to interventions. A person's reaction depends on many parameters, which relate to both physiological and cognitive state. In this case, we will consider various tasks in which any intervention on a person is required: for example, attracting the driver to attention while driving or reminding about physical controls when a person is in a sitting position for a long time. It is impossible to make a universal system for everyone that will work equally effectively. Hence, it is necessary to move in the direction of personalized systems that will individually choose the intervention for each user. Personalized strategies for influencing the user (adaptive interventions) are one of the promising areas that allows one to interact with the user carefully and without unnecessary persistence, directing him to various useful habits. Instead of developing such complex strategies manually, reinforcement learning can be used to adaptively optimize intervention strategies based on the user's state and situation context.

This paper discusses the concept of building a system of personalized recommendations to influence the driver. Section II discusses current developments on the topic of personalized interventions in various fields. Section III presents the conceptual model describing the main components of the system and their interaction. Section IV describes the formulation of the reinforcement learning problem. Section V describes the experiment conducted with the developed model.

II. Related work

Currently, personalized intervention systems are actively developing in the field of medicine. Adaptive interventions have become a new perspective of prevention and treatment in healthcare. Just-In-Time Adaptive Interventions (JITAI) is a concept of adaptive intervention aimed at providing the right type/amount of support at the right time based on changes in the internal and external state of a person [4]. The basic principles of JITAI and the design structure of JITAI applications were presented by the authors Nahum-Shani et al. [5]. Mobile healthcare systems with JITAI have proven to be effective in preventing certain health threats (for example, overeating [6], smoking [7] and prolonged sedentary lifestyle [8]) and obtaining positive health outcomes (for example, increased physical activity [9] and condition support associated with chronic diseases [10]). There are various approaches to the implementation of JITAI, for example, rule-based systems, supervised learning, etc. This section briefly describes the most relevant approaches in the field of JTAI, that are based on reinforcement learning.

In work [9], a JITAI system was developed to increase physical activity for patients with type 2 diabetes mellitus. A mobile application was developed that collected data on patient activity in the background and was sent to a central server where it was analyzed. Based on these data, the reinforcement learning algorithm chose which impact on a particular user to apply in these conditions to increase physical activity the next day. An SMS message was as the impact. As an algorithm, the authors used an algorithm similar to the contextual bandit algorithm. The reward for the algorithm was the number of actions performed by the patient since the last message was sent to him or her. As a result, the level of physical activity in the group that received interventions from the reinforcement learning algorithm was higher than in the control group. The authors in [11] investigated the use of JITAI to assess patients' weight loss. 52 patients were randomly divided into 2 groups: the group that receives maladaptive interventions, the group that receives adaptive interventions. In turn, the group that receives adaptive interventions was divided into 2 more groups: patients who receive individually personalized interventions and patients who receive group personalized interventions. The UCB1 algorithm was chosen by the researchers as a reinforcement learning algorithm. The reward for the agent was calculated every 3-4 days and calculated as a value that depends on: self-control of weight, self-control in food, daily calorie goal, daily goal of physical activity and weight loss divided by the number of days in this intervention period. The result of the experiment showed. The result of the experiment showed that the group that received individual personalized interventions showed not only the best result in weight loss, but also such a result was achieved for the minimum cost of training hours per person. The article [12] presents MyBehavior, a smartphone application that uses a new approach to create deeply personalized health feedback. MyBehavior automatically studies the physical activity and dietary behavior of the user and strategically suggests changes in this behavior for a healthier lifestyle. The authors use the multi-armed bandit algorithm (EXP3) as a reinforcement learning algorithm. For validation, a 14-week study was conducted, which indicates a significant improvement compared to the control group, which continued after the initial phase of novelty.

In addition to the field of personalized medicine, similar systems of personalized effects on the user are beginning to be developed in other areas. For example, the authors in [13] presented an innovative personalized training system for social companion robots that uses verbal and nonverbal affective signals of children to modulate their involvement and maximize their long-term learning outcomes. The study suggests a reinforcement learning approach to develop an individual policy for each student during the educational process, where the child and the robot tell stories to each other. Using a personalized policy, the robot selects stories optimized for the involvement of each child and the development of linguistic skills. The authors use Q-Learning as an algorithm. In the experiment, the children were divided into 3 groups: a group for which the robot selected personalized strategies for each child, a group without a personalized strategy and in which the children followed a fixed curriculum, and a group that did not interact with the robot. Personalized groups of children showed significantly better learning outcomes and higher engagement than the base group and better results than the group without a personalized strategy.

Based on the review, personalized strategies for influencing the user show a high-quality result compared to the effects calculated for the “average” user. The development of such systems is gaining popularity. Currently, such systems are not common in the field of transport, and we offer a conceptual solution for a personalized system of interaction with the driver.

III. Conceptual model of the solution

Fig. 1 shows a diagram of the conceptual model of the proposed solution. It consists of several main components:

- Environment.
- Monitoring system.
- Agent.

A. Environment and monitoring system

The environment is the entire space from which the agent receives data and with which it can interact. The environment consists of 2 participants – a car and a driver. High-quality data acquisition, as well as the correct interpretation of the environment is a very important point in driver monitoring systems. In order to reliably determine the need for a particular intervention, it requires the ability to determine its state. The state of the environment consists of the states of its constituents:

- The state of the car's dynamics.
- Driver's state.

1) The state of the car's dynamics: Modern cars have a large number of sensors and microcomputers on board, which are connected to one common internal network. These sensors exchange large data streams in real time. Each microcomputer is responsible for one of the subsystems of the car, for example, a brake system, an engine or a multimedia system. The interaction of these systems ensures safety and comfort for passengers and the driver during the trip, as well as the efficient operation of the engine. If you connect to the car's internal network, you can receive data from sensors and use this data for further analysis. Driving habits are unique for each driver and these features can be used for various purposes. Examples of applied results of the analysis of these data can be the solution of the problem of driver identification. With the help of driver identification, it is possible to save the results of agent training for each driver within the driver monitoring system. This feature will work great when more than one driver can drive the car, for example, if it is a taxi company or public transport. In addition, an example of using car data may be the task of profiling a driver to determine his driving style. Determining the driving style will allow the agent to act more personalized on the driver. There are also systems that can determine the level of driver fatigue based only on car data. Bosch Group of Companies offer a solution, based on data from the steering angle sensor to determine fatigue and drowsiness to prevent accidents [14]. Thus, the use of dynamics data from car sensors can bring a significant part of the information into understanding about the driver's condition, as well as about any personal habits that may become important when building a personalized strategy.

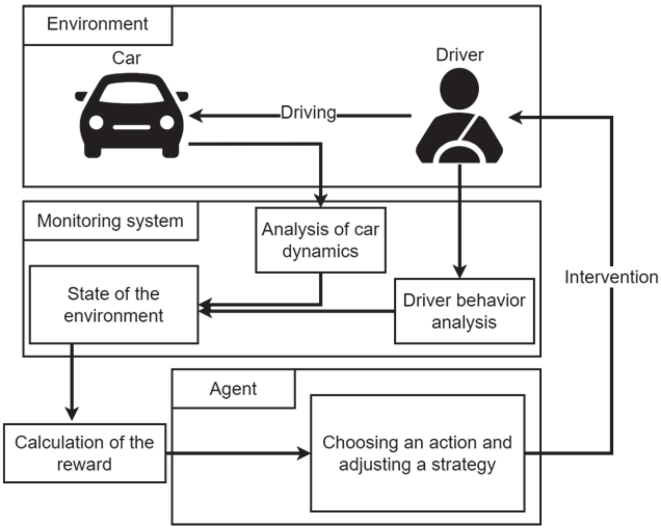

Fig. 1. Conceptual model of the proposed solution

2) Driver's condition: Driver monitoring is a very important task. As mentioned earlier, a large percentage of road accidents occur for reasons that are associated with the driver being in a suboptimal condition. Therefore, it is very important to be able to detect this condition. In the case of driving, it is required to track not only any physiological characteristics of a person, but also to analyze his actions, gaze direction, and hand movements. This is important because a person can be physiologically in a normal state, i.e. be cheerful, rested, but he can be distracted by interaction with a smartphone, an information and multimedia system or communicate with a passenger. All these distractions greatly increase the chance of getting into an accident. Therefore, various technologies and solutions are used to solve such a complex task as driver monitoring. There are several basic ways to analyze driver behavior. The first of them is the use of computer vision methods. At the moment, there are enough studies that use computer vision methods to determine the driver's distraction or level of attentiveness [15]. Such studies are actively developing and achieve high accuracy of guessing distractions on ready-made datasets. Also, such systems can operate under conditions of different lighting levels and regardless of the time of day. Another method of determining fatigue is to use EEG results to determine fatigue. Similar studies are also conducted to determine a person's fatigue [16], [17], [18]. The heart rate can serve as one of the sources of information about the driver's condition. Omron is developing a solution that can receive heart rate data directly in the car, wirelessly. Similar data can also be used to detect a person's drowsiness [19], [20].

B. Agent

Under the agent in the environment, a reinforcement learning algorithm is presented. This algorithm, based on the state of the environment, experience and response from the intervention, decides which intervention should be applied in a particular situation and whether it should be taken at all. The agent is the basis of the whole solution. *The* main feature of the agent is that it is constantly learning, throughout the entire time of interaction with the user. The use of reinforcement learning allows one to create a very flexible algorithm that will interact with the user point-by-point, applying to him exactly the effects to which the user will respond better. This mechanism is implemented based on feedback in the form of a reward that the agent receives after interacting with the driver. The reward calculation strategy is a very important component and requires a separate study.

There are many different types of effects or countermeasures to combat drowsiness or fatigue when driving. Notification via messages on the dashboard, the use of sound or light signals, vibro-tactile exposure through the seat or steering wheel. In most cases, drivers have to confirm the warning messages by pressing a button to clear them. The application of each of these types of interventions on the driver can be effective. In the article [21] the authors investigated the effect of sound warnings on drowsiness and concluded that simple sound warnings lead to improved lane retention. Fairclough and van Winvum showed that visual warnings improved lane retention compared to the lack of feedback [22]. Arimitsu and colleagues showed that exposure to seat belt vibrations led to improved lane retention and reduced subjective drowsiness [23]. Modern commercial warning systems mainly consist of individual warnings (for example, a coffee cup icon that appears when sleepiness is detected) or step-by-step feedback, for example, a scale showing the driver's level of attentiveness. The use of a combination of auditory, visual and tactile influences has great potential to improve driving performance and reduce drowsiness [24].

IV. Problem statement

The task of monitoring the driver and developing warnings/recommendations in order to keep the driver in a certain state (awake, focused on the traffic situation) can be reduced to the task of forming an optimal strategy for influencing the driver of warnings/recommendations. The development of ways to form an optimal strategy for influencing some object based on the experience of the consequences of these impacts is the main subject of reinforcement learning.

Reinforcement learning is one of three categories of machine learning (the other two are supervised learning and unsupervised learning). The main participants in reinforcement learning are the environment and the agent. The agent interacts with the environment, receiving a reward after the interaction, that is, reinforcement, as a signal about how good or bad the action taken is. After exposure to the environment, the environment assumes a new state. The main task of reinforcement training is to train the agent to get the maximum reward. During the training, the agent optimizes its behavior based on the experience of interacting with the system. The agent knows the current state of the environment, and it chooses the actions that should bring the greatest reward based on the experience he has. This interaction between the agent and the environment is shown in Fig. 2.

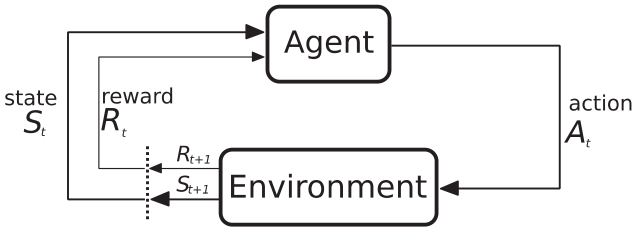

Fig 2. Reinforcement learning scheme

Consider the task of issuing personal interventions to the driver to support the "normal" state, as a reinforcement learning task:

- The *environment* is a driver whose condition is various physiological and cognitive parameters that he possesses at a certain point in time.
- *Agent* is a system for issuing personal recommendations.
- *Observations* are the parameters that the agent observes from all the parameters of the state of the environment.
- *Actions* are possible interventions that can influence the driver's vigilance. Such interventions can be sounds of different loudness and height or a light indication. Since each person is individual, each such influence affects him/her with an individual intensity and the value of vigilance that is obtained after exposure is unique to a person.
- *Reward* is the difference between the state after exposure and before exposure. Thus, the agent receives information about which effects are most effective for this particular person.

V. Description of the experiment

The reinforcement learning process is an iterative process. In order for an agent to interact with the environment well enough, a lot of interactions are required, because initially the agent does not know how to interact with the environment, his actions will be random and may not bring any reward. As part of the first step of research and development of such a system, we will train the agent on a computational model of the driver.

A. Modeling the environment

The ability to predict the performance and fatigue in a given situation is in demand by people of very different backgrounds. This includes individuals trying to make the most of their time and efficiency, employers doing the same for their employees, and security personnel trying to determine if an operator is fit to do their job. For example, a model is required to predict fatigue for the aviation industry, since the deterioration of flight abilities can lead to a large loss of money and even life. For several years, research has been conducted on ways to model and predict fatigue and performance.

Driving is a complex multitasking activity that requires the driver to distribute his attention to various tasks, for example, keeping the car in the lane, controlling the speed of the car. Also, drivers while driving can be distracted by passengers, interaction with a multimedia system or smartphone. To simulate human behavior, in addition to the characteristics that can be calculated, it is required to know the cognitive and psychological laws by which the human brain works and by which it makes various decisions. In general, all real models that simulate a person's attention or fatigue can be divided into 2 types.

- Models that are created using cognitive architectures.
- Biomathematical models of alertness.

1) Models created using cognitive architectures: To model complex tasks that are related to human behavior, various cognitive architectures are actively used that allow us to describe the decision-making process. Cognitive architecture is a general framework for defining computational behavioral models of human cognitive activity. Architecture embodies both the capabilities and limitations of the human system – for example, abilities such as forgetting and remembering, learning, perception and motor actions; and limitations, such as memory impairment, foveal and peripheral vision, and limited motor activity. Thus, the cognitive architecture helps to ensure that the cognitive models developed within this system are rigorous and psychologically sound, thereby respecting all the limitations of the human system. Cognitive architectures have demonstrated the ability to model tasks ranging from basic laboratory tasks to a higher level of cognition and decision-making in complex dynamic tasks (for example, piloting fighter jets). Examples of similar architectures: ACT-R [25], QN-MHP [26], CASCaS [27], etc.

Based on the cognitive architecture of ACT-R, Ganzelmann et al. developed a model that simulates the effects of sleep-induced fatigue and circadian rhythms by changing the parameters of the ACT-R module [28], [29]. The model could predict a person's performance in various cognitive tasks (for example, the task of sustained attention, the task of addition/subtraction) as a consequence of fatigue.

Many different models have been built based on the cognitive architecture of QN-MHP. One of the models is the driver model, which evaluates the efficiency and load of the driver [16]. This model stimulates mental workload, measured using six scales of the workload index of the National Aeronautics and Space Administration (NASA-TLX). The model also simulates driving performance, reflecting the mental load based on subjective and performance-based indicators. In addition, it simulates age differences in workload and productivity and allows one to visualize the mental load of the driver in real time.

COSMODRIVE (COgnitive Simulation MOdel of the DRIVEr) is a cognitive simulation model of a driver designed to simulate the mental activity of drivers [17]. This model simulates, from perception functions to behavioral characteristics carried out while driving. The overall goal of this research program is to design, develop and implement a computational model of a car driver capable of driving a virtual car in a virtual road environment.

2) Biomathematical models of alertness: The second type of models that *allow* you to model cognitive parameters such as fatigue and alertness are biomathematical models. In this type of models, the formulas of biological processes that occur in the human body are mathematically set. Examples of such models are:

- Three-process model of alertness [30].
- FAST [31].
- SAFE [32].
- BAM [33].

Almost all models of this type take the values of a person's sleep schedule as input parameters and form his alertness levels for several days. As you can see, since the model generates the values of a person's alertness for several days, then such a model does not provide for any possibility of increasing the level of alertness.

In addition to modeling the level of alertness, it also requires modeling a person's reaction to any impact, as well as a mechanism for getting used to any impact. In psychology, there are such concepts as the theory of arousal and the theory of habituation. These theories conceptually describe the processes that occur when a person is exposed to some stimulus, how his level of alertness changes and how he gets used to the same influences. However, no mathematically adequate models have been found during the literature review.

Therefore, in order to train an intervention algorithm an original driver alertness model has been developed. The model consists of 2 parts. Part 1 is responsible for modeling alertness during the day and night. This part is implemented using the three-process model (e.g., [30]). Part 2 is responsible for modeling interventions, as well as the mechanism of getting used to these interventions. The three-process model of alertness uses the waking and sleeping schedules as input. The model determines the level of vigilance based on 3 components:

- S component;
- C component;
- W component.

C component is a process that describes the effect of circadian rhythm on drowsiness. It has a sinusoidal shape with a peak in the evening hours and a nadir in the early morning hours process C is calculated using the formula:

$$C = M * \cos\left(\frac{\pi}{12} * (t - p)\right),$$

where M – amplitude (2.5), t - time of day (in decimal hours), p - acrophase (in decimal hours).

S component reflects the homeostatic components (the amount of time since waking up and the amount of previous sleep) and is an exponential function (1). The process S is continuously changing — it has a peak immediately after waking up from sleep, and decreases with continued wakefulness, while the curve is smoothed as it approaches the lower asymptote:

$$S = la + (sw - la) * e^{(d * dt)}, \quad (1)$$

where la = low asymptote (2.4), sw - S at waking up, d - decay (–0.0353) and dt - delta time since waking up, in decimal hours.

At the beginning of sleep, this component is denoted S' to represent the reverse process (recovery during sleep):

$$S' = ha - (ha - ss) * e^{-0.381*dt},$$

where ss - S at falling asleep and dt - delta time since falling asleep, in decimal hours.

W component is the third component in the model, which denotes the inertia of sleep. Inertia is represented by drowsiness after sleep and drowsiness, which usually manifests itself during the first 20-30 minutes of wakefulness. Formula for calculating the W value:

$$W = 5.7 * e^{-0.65692*t},$$

where t – time since awaking.

To obtain the vigilance value, all components are summed up:

$$Alertness = S + C + W$$

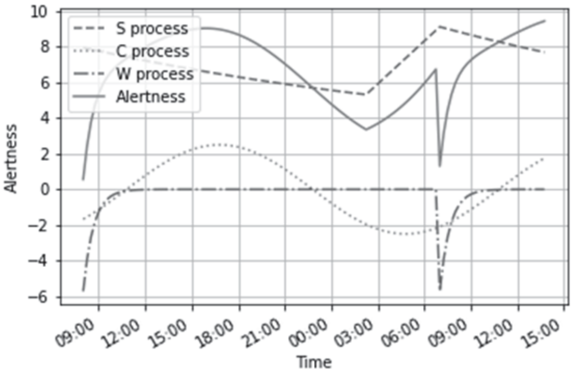

Fig 3. Three-process model of alertness

The resulting value of alertness ranges from 1 to 16, in practice "3" corresponds to extreme drowsiness, "14" - high alertness, and "7" - the threshold of drowsiness. Figure 3 shows an example of alertness modeling using a three-process alertness model. The parameters for this model were a schedule in which a person was awake from 7.30 to 23.30 and slept from 23.30 to 7.30.

A set of actions (interventions) has been defined to interact with the environment. Each action can be described by an intensity value – the strength of the impact. In real life, such actions can be sounds of different loudness, light, vibro-tactile effects or a combination of them. The set of actions consists of several elements with different intensity values. As mentioned earlier, a mechanism of habituation was developed for the experiment. Explanations of how a person is affected by various influences are the subject of arousal theory. The theory of habituation is closely related to the theory of arousal and explains the habituation to non-critical signals during repetitive stimulation in vigilance tasks. Repeated stimulation leads to a decrease in the arousal response. The decrease is usually a negative exponential function of the number of stimuli presented [34]. Changing the stimulation can lead to an immediate improvement in performance. The value by which the intervention increases is calculated by the formula:

$$IncAlert\ = CurrAlert * StimulationCoef,$$

where $CurrAlert$ – the value of current alertness, $StimulationCoef$- this value that is calculated based on the mechanism of stimulation and habituation:

$$StimulationCoef = \frac{(1\ -\ HabitValue) *\ IntensityAct}{100},$$

where $IntensityAct$ – is the value of the intensity of the action.

$$HabitValue = \mathrm{e}^{\ (-(CurrTime\ -\ PrevTime)\ /\ (7.5\ *\ 1.2^{N}))},$$

where $CurrTime$ – current time when the excitation occurs in minutes, $PrevTime$ – time of the previous excitation in minutes, N - number of excitations of this type that have already been made. If this is the first excitation, then $HabitValue$ is 0.

C. Agent development

There is a large number of different reinforcement learning algorithms. For our experiment, we used the DQN (Deep Q-Network) algorithm developed by DeepMind in 2013 [35]. The algorithm is based on the classical Q-Learning reinforcement learning algorithm. The DQN algorithm is complemented by deep neural networks and a technique called experience reproduction. This algorithm has obtained good results in solving various problems.

Based on the reward received from the environment, the agent forms a utility function Q, which subsequently gives him the opportunity to choose a behavior strategy not by chance, but to take into account the experience of previous interaction with the environment. The Q-function of a policy π, $Q^{\pi}(s,a)$, measures the expected return or discounted sum of rewards obtained from state s by taking a action first and following policy thereafter. Optimal Q-function $Q^{\pi}(s,a)$ can be determined as the maximum return that can be obtained starting from observation, taking action and following the optimal policy thereafter. The optimal Q-function obeys the following Bellman optimality equation:

$$Q^{*}(s,a) = \ \mathbb{E}[r +\ \gamma max_{a'}Q^{*}(s',a')]$$

For our problem, it is impractical to represent the Q-function as a table containing values for each combination of s and a. Instead, we train a function approximator, such as a neural network with parameters θ, to estimate the Q-values, i.e. $Q(s,a,\theta)\ \approx Q^{*}(s,a)$.

Our developed agent had some set of actions to interact with the environment (e.g., sounds of different volume and pitch, warnings). Each action had a different intensity value - the strength of the impact. In the experiment we consider 11 actions with the following intensities: {1, 1, 2, 2, 4, 4, 5, 5, 7, 7, 13}. For each action, the habituation mechanism is used, which was described in section 5, subsection B. Thus, using the same action several times in a row will significantly less increase the vigilance of the driver model. The agent developed by us accepts additional data for the condition in order to take into account the mechanism of addiction. For each type of exposure, the time difference in minutes between the last exposure of this type and the current time of exposure to the environment is calculated. This data will allow the agent to better take into account the peculiarities of addiction and should allow interacting with the environment more effectively.

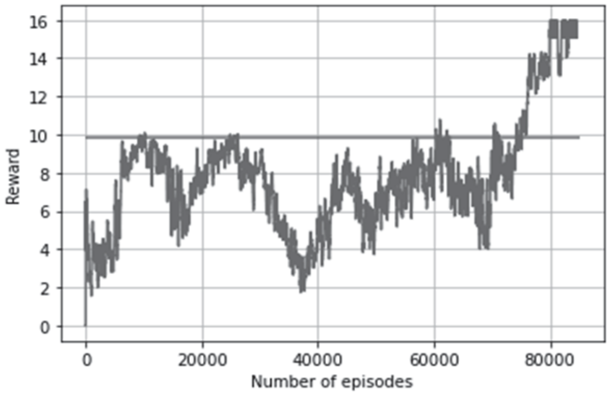

Fig 4. Reward dependence on the number of episodes

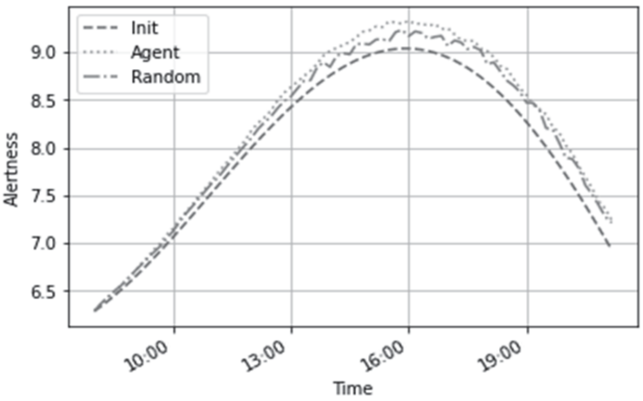

Fig 5. Comparison of the alertness of the trained agent, random policy and initial alertness

The agent's reward strategy is an important component of machine learning models with reinforcement. This strategy determines how the agent learns the patterns in the system and how he will effectively interact with it. After each interaction with the system, the agent receives a reward, the value of which is equal to $IncAlert$. In addition, it is assumed that the agent should not apply the impact in the case when the driver's level of vigilance is not below a certain limit. This boundary was set to a value of 11. If an agent exerts an impact with non-zero intensity, and the level of vigilance is higher than 11, then his reward is equal to = -0.5. That is, the agent receives a penalty. If the agent exerts an impact with zero intensity, i.e. does not affect the driver, then his reward is equal to 0.5.

D. Conducting an experiment

Figure 4 shows a graph of the reward dependence on the number of episodes of the developed model. As Figure 4 shows, the agent tends to increase the reward level. Figure 5 shows a comparison of the alertness levels of a trained agent, a random policy and the initial value of alertness. Figure 5 also shows a straight horizontal line, the value of Reward = 9.84. This is the value of the average reward per episode for 100 episodes with a random agent policy. The policy, that the developed agent has learned, has an advantage over the random policy. Thus, interaction with the user based on the policy of a trained agent can be more effective in the situation of maintaining the driver in a normal state.

VI. Conclusion and future work

The main purpose of this article is to develop the concept of a personalized strategy of influencing the driver of the car. The development of personalized intervention strategies is a new direction in various fields and is only gaining popularity. At the moment, there are no works on such topics in the field of transport. We propose a concept of a driver assistance system which leverages reinforcement learning to form an optimal strategy for influencing the vehicle driver based on the observed history of prior interventions. As part of the article, an experiment was conducted to test if such policy can be learned. As part of the experiment, an environment was modeled that represents a person and the value of his vigilance. An agent-based reinforcement learning algorithm was also developed. The DQN algorithm was used as a reinforcement learning algorithm. The experiment has shown that an effective intervention policy can successfully be learned by a reinforcement learning algorithm (DQN), which supports the initial hypothesis.

An important limitation of the study is that it relies on a rather simple user alertness model. The future direction of this work is:

- improving the user model, based on psychological theories of arousal and habituation;
- the study of other learning algorithms with reinforcement from a combination of algorithms to solve the problem of personalization of impacts on the driver, as well as various strategies to reward an agent for an action to determine which strategy works best, and the selection of other hyperparameters that can affect the result of the algorithm.

It should be noted that this paper considers the solution to the problem of insufficient alertness of vehicle's driver. The development of autonomous transport technologies will solve this problem, because the human factor in automobile traffic will be mostly eliminated (or severely reduced). However, there is still a lot of time before the full transition to autonomous transport and the problem of human distraction will be relevant for a long time. In addition, the presented solution is an example of the development of personalized intervention strategies in a broader context. Currently, there are very few industries where such solutions are used. Accordingly, the solution can be used as a basis for the development of personalized intervention systems in other application areas, where it is required to maintain a person's attention in a vigilant state.

Acknowledgement

The reported study was funded by Russian Science Foundation, project number 18-71-10065.

References

[1] World health organization, *Global status report on road safety 2018,* Switzerland, 2018.

[2] M. A. Regan, C. Hallett, and C. P. Gordon, «Driver distraction and driver inattention: Definition, relationship and taxonomy», *Accident; Analysis and Prevention*, vol.43 (5), Sept. 2011, pp. 1771–1781.

[3] National Sleep Foundation website, 2008 Sleep in America Poll, Web: https://www.sleepfoundation.org/wp-content/uploads/2009/06/2008_POLL_SOF.pdf.

[4] S. Wang, C. Zhang, B. Kröse, and H. van Hoof, «Optimizing Adaptive Notifications in Mobile Health Interventions Systems: Reinforcement Learning from a Data-driven Behavioral Simulator», *Journal of Medical Systems,* vol. 45(12), Dec. 2021.

[5] I. Nahum-Shani, E. B. Hekler, and D. Spruijt-Metz, «Building Health Behavior Models to Guide Adaptive Interventions: A Pragmatic Framework», *Health Psychology*, vol.34, Dec. 2015, pp. 1209–1219.

[6] S. P. Goldstein, B.C. Evans, D. Flack, A. Juarascio, S. Manasse, F. Zhang and E.M. Forman, «Return of the JITAI: Applying a Just-in-Time Adaptive Intervention Framework to the Development of m-Health Solutions for Addictive Behaviors», *International Journal of Behavioral Medicine*, vol.24(5), Jan. 2017, pp. 673–682, 2017.

[7] H. Sarker , M. Sharmin ,A. Ali ,M. Rahman M, R. Bari , S. Hossain , S. Kumar, «Assessing the availability of users to engage in just-in-time intervention in the natural environment», *in Proc UbiComp,* Sept. 2014, pp. 909–920.

[8] J. Graham Thomas and D. S. Bond, «Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in

obese adults: Implications for JITAI optimization», *Health Psychology*, vol.34, Dec. 2015, pp. 1261–1267.

[9] E. Yom-Tov, G. Feraru, M. Kozdoba, S. Mannor, M. Tennenholtz, and I. Hochberg, «Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system», *Journal of Medical Internet Research*, vol.19 (10), Oct. 2017.

[10] C. Sun, S. Hong, M. Song, J. Shang, and H. Li, «Personalized vital signs control based on continuous action-space reinforcement learning with supervised experience», *Biomedical Signal Processing and Control*, vol.69 (5), Aug. 2021.

[11] E. M. Forman, S. G. Kerrigan, M. L. Butryn, A. S. Juarascio, S. M. Manasse, S. Ontañón, D. H. Dallal, R. J. Crochiere and D. Moskow, «Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss?», *Journal of Behavioral Medicine*, vol.42(2), Apr. 2019, pp. 276–290.

[12] M. Rabbi, M. H. Aung, M. Zhang, and T. Choudhury, «MyBehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones», *in Proc. UbiComp* Sept. 2015, pp. 707–718.

[13] H. W. Park, I. Grover, S. Spaulding, L. Gomez, and C. Breazeal, «A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education», *in Proc. of the AAAI Conference on Artificial Intelligence*, July 2019, pp. 687–694.

[14] Driver drowsiness detection, Web: https://www.bosch-mobility-solutions.com/en/solutions/assistance-systems/driver-drowsiness-detection.

[15] N. Moslemi, M. Soryani and R. Azmi, «Computer vision-based recognition of driver distraction: A review», *Concurrency and Computation Practice and Experience*, vol.33(4), July 2021, pp. 1–2.

[16] N. Zhao, L. Dawei, H. Kechen, C. Meifei, W. Xiangyu, Z. Xiaowei and H. Bin, «Fatigue detection with spatial-temporal fusion method on covariance manifolds of electroencephalography», *Entropy*, vol.23(10), Sept. 2021.

[17] J. Cui, Z. Lan, T. Zheng, Y. Liu, O. Sourina, L. Wang and M. Wolfgang, «Subject-Independent Drowsiness Recognition from Single-Channel EEG with an Interpretable CNN-LSTM model», *unpublished*.

[18] M. Zhu, J. Chen, H. Li, F. Liang, L. Han and Z. Zhang, «Vehicle driver drowsiness detection method using wearable EEG based on convolution neural network», *Neural Computing and Applications*, vol.33, May. 2021.

[19] J. Vicente, P. Laguna, A. Bartra and R. Bailón, «Drowsiness detection using heart rate variability», *Medical & Biological Engineering & Computing*, vol. 54(6), Jun. 2016, pp.927–937.

[20] S.H. Jo, J.M. Kim and D.K. Kim, «Heart Rate Change While Drowsy Driving», *Journal of Korean medical science,* vol. 34(8), Feb. 2019.

[21] C. Berka , D. Lewindowski, , «Implementation of a closed-loop real-time EEG-based drowsiness detection system: effects of feedback alarms on performance in a driving simulator», *in Proc. International Conference on Augmented Cognition*, July 2005.

[22] S. Fairclough, W. V. Winsum, «The Influence of Impairment Feedback on Driver Behavior: A Simulator Study», *Transportation Human Factors,* vol. 2(3), Sep. 2000, pp. 229-246.

[23] S. Arimitsu, K. Sasaki, H. Hosaka, M. Itoh, K. Ishida, A. Ito, «Seat Belt Vibration as a Stimulating Device for Awakening Drivers», *IEEE/ASME Transactions on Mechatronics,* vol. 12(5), Nov. 2007, pp. 511-518.

[24] J. G. Gaspar,T. L. Brown,C. W. Schwarz,J. D. Lee,J. Kang and J. S. Higgins, «Evaluating driver drowsiness countermeasures», *Traffic Injury Prevention,* vol. 18(12), Mar. 2017, pp. 58-63.

[25] J. R. Anderson and C. J. Lebiere, *The Atomic Components of* Thought, Psychology Press, 1998.

[26] Y. Liu, R. Feyen, and O. Tsimhoni, «Queueing Network-Model Human Processor (QN-MHP) A computational architecture for multitask performance in human-machine systems», *ACM Transactions on Computer-Human Interaction,* vol. 13(1), Mar. 2006, pp. 37-70.

[27] A. Lüdtke, J. Osterloh, T. Mioch, F. Rister and R. Looije, «Cognitive Modelling of Pilot Errors and Error Recovery in Flight Management Tasks», *Lecture Notes in Computer Science,* vol. 5962, 2010, pp. 54-67.

[28] G. Gunzelmann, J. B. Gross, K. A. Gluck, and D. F. Dinges, «Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling», *Cognitive Science A Multidisciplinary Journal*, vol.33 (5), July 2009, pp. 880–910.

[29] G. Gunzelmann, L. Richard Moore, D. D. Salvucci, and K. A. Gluck, «Sleep loss and driver performance: Quantitative predictions with zero free parameters», *Cognitive Systems. Research*, vol.12 (2), June 2011, pp. 154–163.

[30] T. Åkerstedt and S. Folkard, «The Three-Process Model of Alertness and Its Extension to Performance, Sleep Latency, and Sleep Length», *Chronobiology International*, vol.14 (2), Apr. 1997, pp. 115–123.

[31] S. R. Hursh, T. J. Balkin, J. C. Miller and D. R. Eddy, «The fatigue avoidance scheduling tool: Modeling to minimize the effects of fatigue on cognitive performance», *SAE Technical Paper*, vol.113 (1), Jun. 2004, pp. 111–119.

[32] M. B. Spencer and K. A. Robertson, «The application of an alertness model to ultra-long-range civil air operations model», *Somnologie*, vol.11, Nov. 2007, pp. 159–166.

[33] Fatigue Risk Management, Web: http://ww1.jeppesen.com/documents/aviation/pdfs/Fatigue_2009-10_Final_II.pdf.

[34] R. F. Thompson, «Habituation: A history», *Neurobiology of Learning and Memory*, vol.92 (2), Sep. 2008, pp. 127–134.

[35] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, «Playing Atari with Deep Reinforcement Learning», NIPS Deep Learning Workshop 2013.