Agent-Based Integrated Decision Making for Autonomous Vehicles in Urban Traffic

Maksims Fiosins, Jelena Fiosina, Jörg P. Müller and Jana Görmer

Abstract We present an approach for integrated decision making of vehicle agents in urban traffic systems. Planning process for a vehicle agent is separated into two stages: strategic planning for selection of the optimal route and tactical planning for passing the current street in the most optimal manner. Vehicle routing is considered as a stochastic shortest path problem with imperfect knowledge about network conditions. Tactical planning is considered as a problem of collaborative learning with neighbor vehicles. We present planning algorithms for both stages and demonstrate interconnections between them; as well, an example illustrates how the proposed approach may reduce travel time of vehicle agents in urban traffic.

1 Introduction

The application of multi-agent modeling and simulation to traffic management and control problems becomes more relevant as intelligent assistant functions and car-to-X communication pave the way to a new generation of intelligent networked traffic infrastructure. Typically traffic environments are regulated in a centralized manner using traffic lights, traffic signs and other control elements. Multi-agent traffic systems are modeled with autonomous participants (vehicles), which intend to reach their goals (destinations) and act individually according to their own interests.

Previous research in this area has mostly concentrated on traffic lights regulation methods, traffic lights agent architecture, coordination and decision making mechanisms ([5]). Multi-agent reinforcement learning (MARL) for coordination of traffic lights was applied by Bazzan, Lauer and others ([3], [4]).

Maksims Fiosins, Jelena Fiosina, Jörg Müller and Jana Görmer
Clausthal University of Technology, Julius-Albert Str. 4, D-38678 Clausthal-Zellerfeld, Germany
e-mail: {maksims.fiosins,joerg.mueller,jana.goermer}@tu-clausthal.de
Maksims Fiosins and Jana Görmer are supported by the Lower Saxony Technical University (NTH) project ”Planning and Decision Making for Autonomous Actors in Traffic” (PLANETS)

In contrast, there is less research on individual driver behavior and architectures of ”intelligent vehicle” agents: existing research is mostly focused on mesoscopic models for travel demand planning [2] or adaptive cruise control [6].

We consider a structure of decision making of a vehicle agent in an urban traffic environment. A vehicle environment is presented as a directed graph $G = (V,E)$, where nodes and edges represent intersections and streets correspondingly. Denote $N(e_i) \subset E$ a set of edges, which start from the node, where the edge e_i ends. We consider the discrete linear model time $t \in 0,1,2,\ldots$.

We suppose that each vehicle j at any time t is located on some edge $e^j(t) \in E$. A relative position of the vehicle j on the edge $e^j(t)$ at time t is defined as a distance to the end of the edge $x^j(t) \in 0,\ldots,d(e^j(t))-1$. Let $l^j(t) \in 1,\ldots,l(e^j(t))$ be a lane, $v^j(t)$ be a speed of the vehicle j at time t.

The complete state of a vehicle is given by a tuple, consisting of edge, relative position on the edge, lane, speed and traffic light time:

$$s^j(t) =< e^j(t), x^j(t), l^j(t), v^j(t), tl^j(t) > . \tag{1}$$

The goal of a vehicle is to reach its destination as quickly as possible.

Planning process for a vehicle agent is separated into two stages: strategic planning (SP) for selection of the optimal route and tactical planning (TP) for passing the current street in the most optimal manner.

During SP, vehicle agents plan the optimal strategic policy $\pi^{*j}_{str}(e^j(t), I^j(t)) \in N(e_j(t))$, which gives the next edge in the fastest path after the edge $e^j(t)$. Vehicles plan their routes individually, based on historical and actual information about edge travel times, applying a modification of Stochastic Shortest Path Problem.

During TP, vehicle agents plan their operative decisions together with other agents. Vehicles on one edge plan their actions $a =< \Delta v^j, \Delta l^j >$, where Δv^j is a speed change, $\Delta l^j \in \{-1,0,1\}$ is a lane change, in order to minimize travel time of the whole group by applying DEC-MARL to learn the optimal tactical policy π^{*j}_{tact}.

The integrated policy of a vehicle j consists of strategic and tactical policy $\pi^{*j} =< \pi^{*j}_{str}, \pi^{*j}_{tact} >$.

The paper is organized as follows. In Section 2 we consider underlying planning algorithms: Section 2.1 describes SP, Section 2.2 TP. In Section 3 we provide first experimental results. Section 4 concludes the paper and suggests future work.

2 Planning for the vehicle agent

2.1 Strategic Planning

In this section we present the method for SP of a vehicle agent. We modify the algorithm R-SSPPR [7] for calculation of the Stochastic Shortest Path with imperfect information. Agents make their SP individually, without cooperation.

Let T_i^j be dependent random travel times of the agent j through the edges $e_i \in E$, $T^j = \{T_1^j, T_2^j, \dots, T_{n_e}^j\}$. We assume that the distribution of T^j is unknown, only a sample of travel time realizations $X = \{X_1, X_2, \dots, X_k\}$ is available, where $X_i = \{X_{i,j}\}$, $j = 1, \dots, n_e$ is a set of travel time values for all edges for the i-th historical realization, p_i^j is a probability that i-th realization takes place for the agent j.

Let $I^j(t)$ be the information, available to the vehicle j at time t, which consists of known travel times; it is a set of events $I^j(t) = \bigcup_{e_i \in E_{kn}^j(t)} \{T_i^j = t_i^j\}$, where $E_{kn}^j(t)$ is a set of edges, which travel times are known.

The information $I^j(t)$ is regularly supplemented. Suppose that the travel time at the edge $e_u \in E$ becomes known to the vehicle j at time τ. It can calculate the posterior conditional probabilities $P\{T^j = X_v\}$, $v = 1, \dots, k$ by using Bayes' formula:

$$P\{T^j = X_v | I^j(\tau), I^j(\tau')\} = \frac{1}{Z} P\{T^j = X_v | I^j(\tau')\} P\{I^j(\tau) | T^j = X_v, I^j(\tau')\}. \quad (2)$$

where Z is a normalizing constant, ensuring that the sum of all posterior probabilities is equal to 1, $I^j(\tau) = I^j(\tau') \cup \{T_u^j = t_u^j\}$.

Let us denote $\pi_{str}^j(e^j(t), I^j(t)) \in N(e^j(t))$ a decision rule about the edge after $e^j(t)$ for the agent j in its path. For its calculation, we use dynamic programming in this stochastic case with imperfect information. Denote $V_\pi^j(e_i, I^j(t))$ an expected travel time of the vehicle j from the beginning of the edge e_i to the destination edge e_i^d under the decision rule π. The following recurrent equation is true:

$$V_\pi^j(e_i, I^j(t)) = \begin{cases} t_i^j, & \text{if } e_i = e_j^d, \\ t_i^j + E_{\tilde{I}^j}[V(\pi^j(e_i, I^j(t)), \tilde{I}^j)] & \text{otherwise.} \end{cases} \quad (3)$$

where the expectation is taken over all possible future information sets $\tilde{I}^j$.

Then we need to minimize (3) over all possible next edges for calculation of the optimal $V^{*j}(e_i, I^j(t))$. and the optimal policy $\pi_{str}^{*j}(e_i, I^j(t))$.

However, there is some difficulty in calculation of the expectation $E_{\tilde{I}^j}$ over all possible future information sets $\tilde{I}^j$. For this purpose, we need to consider all possible travel times of the edges $e_i \notin E_{kn}^j(t)$. In order to avoid this difficulty, we use the resampling of future values of travel times ([1]). We go with the probability $2^{-\zeta}$ to ζ steps forward and extract according to the probabilities $P\{T^j = X_i | I^j(t)\}$ one value $X_{w,u}$ for the travel time T_u^j, which is added to the set $\tilde{I}^j$. Then the probabilities $P\{T^j = X_i | I^j(t), \tilde{I}^j\}$ are updated according to (2). This procedure is repeated r times, and an average is accepted as the expectation $E_{\tilde{I}^j} V^{*j}(e_k, \tilde{I}^j)$.

SP consists of two stages: pre-planning and routing. During pre-planning, values $V^{*j}(e_i, I^j(t))$ and $\pi_{str}^{*j}(e_i, I^j(t))$ are calculated for all edges and all possible information sets. During routing, the policy $\pi_{str}^{*j}(e_i, I^j(t))$ is used for the optimal routing. We summarize all above mentioned in the Algorithm 1.

Algorithm 1 Pre-planning stage of the strategic planning

$e_i \leftarrow e_d^j$
while $prev(e_i) \neq \emptyset$ **do**
  $I' \leftarrow \bigcup_{u \in N(e_i)} I(e_u)$
  for all $\zeta \in 1, \ldots, k$ **do**
    $I'' \leftarrow I' \cup \{T_i^j = X_{\zeta,i}\}$
    for all $\nu \in 1, \ldots, k$ **do**
      $P\{T^j = X_\nu | I''\} \leftarrow P\{T^j = X_\nu | I'\} P\{I'' | T^j = X_\nu, I'\}$
    end for
    for all $e_u \in N(e_i)$ **do**
      for all $\eta \in 1, \ldots, r$ **do**
        $V^{(\eta)}(e_u, I'') \leftarrow RESAMPLE(I'')$
      end for
    end for
    $V^*(e_i, I'') \leftarrow X_{\zeta,i} + \min_{e_u \in N(e_i)} \mathrm{avg}_{\eta \in 1,\ldots,r} V^{(\eta)}(e_u, I'')$
    $\pi^*(e_i, I'') \leftarrow \mathrm{argmin}_{e_u \in N(e_i)} \mathrm{avg}_{\eta \in 1,\ldots,r} V^{(\eta)}(e_u, I'')$
  end for
  $e_i \leftarrow \mathrm{prev}(e_i)$
end while

2.2 Tactical Planning

According to SP, a vehicle enters some edge together with other vehicles. Its TP allows sharing an edge with other vehicles by selecting appropriate speed and lane changes to pass through the edge as quickly as possible.

A state of the vehicle $s^j(t)$ is described by its edge, relative position at the edge, lane, speed and traffic light time, S a set of all possible states. Vehicle actions consist of pairs $a = < \Delta v, \Delta l > \in A$, which correspond to speed and lane change. So

$$s^j(t+1) = \begin{cases} < e^j(t), x^j(t) - v^j(t), l^j(t) + \Delta l, v^j(t) + \Delta v, tl^j(t+1) >, \\ \text{if } x^j(t) - v^j(t) > 0, \\ < \pi_{str}^{*j}(e^j(t)), x^j(t) - v^j(t) + d(e^j(t)), l^j(t) + \Delta l, v^j(t) + \Delta v, \\ tl^j(t+1) > \text{otherwise.} \end{cases} \tag{4}$$

Note that the second situation is only possible if the traffic light signal for the desired direction is green, so $tl^j(t) \in G_i(\pi_{str}^{*j}(e^j(t)))$.

We assume that for each state $s^j(t) \in S$ a corresponding reward $r(s^j)$ is available. We further assume that the reward structure is fully additive:

$$r(s^j(t)) = r^x(x^j(t)) + r^l(l^j(t)) + r^v(v^j(t)) + r^{tl}(tl(t)). \tag{5}$$

The position part $r^x(\cdot)$ has bigger values at the end of the edge; the lane part $r^l(\cdot)$ has bigger values for the lane, which has a turn to the next edge in the vehicle route; the speed part $r^v(\cdot)$ has bigger values for bigger speeds; the traffic light part $r^{tl}(\cdot)$ has a big negative value for $tl \notin G(L)$ and small $x^j(t)$.

Let us apply DEC-MARL algorithm [4] for solving the cooperative task of multiple agents. Let $gr_i(t)$ be a set of agents indices, which are located at edge $e_i \in E$ at time t. Let $s_{gr} \in S^{|gr_i(t)|}$ be a joint state, which includes states of all such agents. A local state-action value function $\tilde{Q}^j(s_{gr}, a^j)$ depends on the action of the j-th agent and joint state s_{gr} of the group. $\tilde{Q}^j(s_{gr}, a^j)$ is updated to ensure a maximum of joint-action Q-functions [4]. A learning procedure is given in the Algorithm 2.

Algorithm 2 Multi-agent tactical learning algorithm at the edge $e_i \in E$

while not end of the simulation **do**
  for all $j \in gr_i(t)$ **do**
    observe reward $r^j(s'_{gr})$
    $a'^j \leftarrow \tilde{\pi}^{*j}_{tact}(s'_{gr})$, take action a'^j, observe next state s''^j
  end for
  $s''_{gr} \leftarrow \{s''^j\},\ j \in gr_i(t)$
  for all $j \in gr_i(t)$ **do**
    $\tilde{Q}'^j(s_{gr}, a^j) = \max\{\tilde{Q}^j(s_{gr}, a^j), r(s'_{gr}) + \gamma \max_{a'^j} \tilde{Q}^j(s'_{gr}, a'^j)\}$
    $\tilde{\pi}^{*j}_{tact}(s_{gr}) = \begin{cases} a'^j, & \text{if } \max_{a^i} \tilde{Q}'^j(s_{gr}, a^j) > \max_{a^j} Q^j(s_{gr}, a^j) \\ \tilde{\pi}^{*j}_{tact}(s_{gr}) & \text{otherwise} \end{cases}$
    $Q(s, a_i) \leftarrow Q'(s, a_i),\ a^j \leftarrow a'^j$
  end for
  $s_{gr} \leftarrow s'_{gr},\ s'_{gr} \leftarrow s''_{gr}$
end while

3 Experiments and Results

We simulate a traffic network in Hannover (Germany) in AimSun, specialized simulation software for traffic applications. The road network is shown in the Fig. 1.

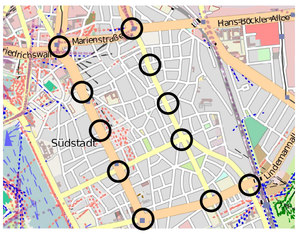

Fig. 1 Street network

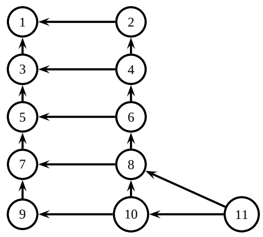

Fig. 2 Considered graph

All intersections are regulated by traffic lights with fixed control plans, known to vehicles. We use the realistic traffic flows, collected in given region of Hannover in morning rush hours. There are traffic flows in all directions; we are interested in the flow $9 \rightarrow 1$; these vehicles use the graph, shown on the Fig. 2, for their decisions.

In our model, we divide each street to cells of 4 m length. The possible speeds of vehicles are: $\{0,5,\ldots,50\}$km/h. One simulation step corresponds to about 1/3 sec.

Experimental results are summarized in Table 1. We calculate travel times depending on the ratio to the flows in Hannover in morning hours.

Table 1 Travel Times of the Route $A \to E$ depending on vehicle generation probability per time

Flows ratio	0.5	0.9	1	1.2	2
Without planning	93.4	102.0	112.7	138.9	168.4
With strategic planning	88.2	95.1	105.5	128.4	156.7
With tactical planning	90.1	96.5	110.6	132.0	162.9
With strategic and tactical planning	87.3	93.3	102.1	126.2	160.4

We conclude that the application of integrated planning allows reducing the travel time of vehicles to about 10%; this is more than SP or TP separately.

4 Conclusions

We proposed an integrated planning process for vehicle agents, which includes both SP and TP. For SP, we showed how to apply existing information for effective solving of a routing problem under imperfect information. For TP we used a modification of DEC-MARL, which allows vehicles to collaborate inside a road segment in order to traverse it in the most quick way. First experiments show that the demonstrated approach allows to reduce the travel time of vehicles.

The proposed approach is used to simulate a larger traffic network in Hannover. In future we will work on an integration of the approach with centralized regulations from TMC, as well as more dynamic agents group formation.

References

1. Andronov, A., Fioshina, H., Fioshin, M.: Statistical estimation for a failure model with damage accumulation in a case of small samples. J. of Stat. Planning and Inf. **139**, 1685–1692 (2009)
2. Balmer, M., Cetin, N., Nagel, K., Raney, B.: Towards truly agent-based traffic and mobility simulations. In: Proceedings of AAMAS 2004, pp. 60–67 (2004)
3. Bazzan, A.L., de Oliveira, D., da Silva, B.C.: Learning in groups of traffic signals. Engineering Applications of Artificial Intelligence **23**(4), 560 – 568 (2010)
4. Lauer, M., Riedmiller, M.: An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In: Proc. of 17-th Int. Conf. on Machine Learning, pp. 535 – 542 (2000)
5. Masterton, T., Topiwala, D.: Multi-agent traffic light optimisation and coordination. In: Thales Research and Technology: White Papers, vol. 2 (2008)
6. Naranjo, J.E., Sotelo, M.A., Gonzalez, C., Garcia, R., de Pedro, T.: Using fuzzy logic in automated vehicle control. IEEE Intelligent Systems **22**, 36–45 (2007)
7. Polychronopoulos, G.H., Tsitsiklis, J.N.: Stochastic shortest path problems with recourse. Networks **27**, 133–143 (1996)