Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

EFFICIENT ROUTING WITH INVERSE REINFORCEMENT LEARNING

Srikrishnan Subramanian
Student, Dept of Computer Science & Engg
SRM University, Chennai, India
Email: srikrishnan_subramanian@srmuniv.edu.in

Adithya Raam Sankar
Graduate Student Institute for Artificial Intelligence
University of Georgia Athens, GA, USA
Email: adithya.raam@uga.edu

ABSTRACT:

Transportation has undergone a lot of evolution since the invention of the wheel. With more and more sophisti-cated manufacturing methods being implemented, the production time of new vehicles has reduced drastically. This has led to a substantial increase in vehicular traffic in the last 3 decades. The beginning of personalized transportation has ushered in a new dimension in the understanding of traffic. The initial approach in managing spatial areas was to merge the roads and visualize it as a graphed network, where every stretch of road is visualized as an edge and the digressions/splitting of traffic being nodes leading to other edges. Increased vehicular traffic with almost constant mapped spatial area causes an unstable equilibrium, leading to congestion and gridlocks. This equilibrium requires an effective balancing/routing strategy to maintain stability among the network. The correlation between road networks and com-puter networks has been exploited to solve this problem, expecting minimal deviation from ideal behavior. Networking protocols are unable to handle deviations that occur due to natural human behavior. Machine Learning Techniques can be implemented to understand these deviations and obtain patterns in real time. The proposed system approaches the routing problem with the aim of learning optimal reward functions by observing regular human behavior for a set of actions. These functions are pivotal in maximizing utility for every agent involved in the procedure by adopting a cooperative and interactive approach.

Keywords—Inverse Reinforcement Learning, Traffic Regula-tion, Congestion Reduction, Re-Routing.

I. INTRODUCTION

Since the time man started moving around, he was con-stantly searching for a better method of transportation. The invention of the wheel was a major turning point in this jour-ney. It was followed by the development of various vehicles for personal as well as commercial commutation. Today, we have a variety of vehicles ranging from the eco-friendly bicycles for short distances to spaceships to reach other planets. Though both of these do not create much of a problem, the increase in the use of privately owned vehicles, like motorbikes and cars, has led to the increase in the congestion of roadways. In turn, this leads to the faster depletion of fossil fuels which provide the primary source of energy for the vehicles.

More than travelling a shorter distance and saving time, the need for saving fuel and eventually saving money has become a rising concern. Deciding the trade-off between time and fuel is a very crucial task. Any traveller will not want to compromise on time or distance as that is what they know as a factor for fuel consumption. Whereas in reality, traffic delays and poor routing[1] also influences the fuel efficiency. One of the possible solutions for this situation would be to create an ideal network of roads that reduce the number of traffic signals and in turn lessen the amount of vehicular crowding. But, given that the road network has already been laid out, any alterations would involve colossal destruction and a great deal of money. This raises the need for routing algorithms that act in real time to increase mileage in the existing system of roads. The proposed system intends to achieve the same by finding alternate paths effectively to eliminate on road congestion.

II. RELATED WORK

Researchers started approaching this problem from an essentially mathematical background. Two main approaches that included computer networking and economic models.The key aspect in which the current system is being proposed is the demarcation that exists between road traffic and network traffic. Network Traffic involves ideal participants while road traffic involves agents exhibiting erratic rational behavior.

The initial approach to route included baseline computer networking algorithms to chart possible paths. Dijkstra's Al-gorithm to obtain the shortest path is the most widely imple-mented algorithm. Location extraction and management can be achieved through Global Positioning Systems, which assure a good accuracy rate.Even if the network has a few sites that are down, this algorithm can be used to easily manage among the existing paths.

The second approach to solving traffic was to utilize eco-nomic principles. This was a feasible approach as economics

considers competition and social equilibrium. A Wardrop equi-librium [2] denotes a strategy profile in which all used paths by drivers between a given origin-destination pair have equal and minimal latency. The first rule of Wardrop Equilibrium states that no agent can decrease its experienced latency by unilaterally deviating to another path.[3]) Selfish routing would just bring about a string of latency issues for all agents involved. Also, conversion of economic principles into stable real-time self-balancing models requires additional computa-tional principles.

Application of decentralization and considering every ve-hicle as a potential data node was analyzed. Wireless Ad-hoc networks and Vehicular Ad-hoc networks(VANETs) have been implemented with such an overview [4]. Such networking approaches help in facilitating inter-vehicular transmission and aggregation of data that characterize the network under consideration.

Significant development began with creating mathematical definitions of traffic flow, both from an economic and a statistical standpoint. The measures of statistical aggregation were the initial metrics of defining the system. This was the essential characteristic of a Markov chain and understanding the transitions of the system. Hence, Markov chains were applied to further enhance the understanding of the traffic network.[5]

In recent studies, the properties of stochastic processes have been widely applied to understand complex human interac-tions. The same was extended to the understanding of traffic flows. The stochastic extension of Markov chains was Markov Decision Processes. were articulate in representing traffic flow. Given the essential characteristics of a stochastic process, such models were analyzed with greater detail for accuracy.

Markov Decision Processes explains state wise transitions and can help in an efficient maximization of the utility involved during the path, with solutions to the Markov decision process described in Bellman Equations [6]. These equations provide an effective solution for long term maximized utility. The primary requisite for solving a Markov Decision Process is to have a well-defined reward function in place, that can be maximized in the solution given by the Bellman equations. As-suming arbitrary reward functions, the results are not accurate and do not take into consideration, competition among agents that would pose changes to the rewards/payoff obtained.The branch of Inverse Reinforcement Learning began, as a result, to enable the learning of reward functions, over a long term period of observation of expert system, whose behaviour we intend to simulate.

III. PROPOSED SYSTEM

As we have seen the two pronged approaches of networking algorithms and economic models, it is increasingly clear about the shortcomings. These shortcomings can be overcome by merging both the systems. The proposed system is a hybrid approach to the routing of vehicular traffic. It employs a congestion aware mechanism that utilizes the knowledge of real time traffic. This availability of knowledge maps the traffic flowing in the streets and effectively merges the data into the decision making process of the system, thus making effective routing choices. The initial functioning of the system involves extraction of the source and destination under consideration. Following this, all possible paths are retrieved. These paths undergo an extensive analysis to identify possible levels of congestion at each “edge”. If found, systems will identify the alternative path to ensure the driver reaches the destination with no loss in rewards. The entire sequence of the system is governed by the second principle of Wardrop’s Equilibrium that drivers cooperate in a multi-agent environment for maxi-mized reward. The approach of inverse reinforcement learning to solve cooperative games enables in learning a more optimal reward function. Inverse Reinforcement learning algorithms as specified in [7],specifies that, reward functions are essentially quantifiers of tasks performed by the agents in the system.

Such systems would then alleviate road congestion, pre-vent future congestion and ensure that the agents that are participating in the routing mechanism have maximized utility. The methods of machine learning are utilized to enhance the routing efficiency and provide a real time best path for the agent under consideration. This system , as described in Figure 1, builds on the idea of applying reinforcement learning to learn the metrics of the flow. The system describes the real time updating and maintenance of the network of roads as main-tained.The set of Reinforcement learning algorithms balance exploration and exploitation. Exploration is trying different things to see if they are in face better than what has been tried before. Exploitation involves making greedy decisions based on local parameters. The advantages of Reinforcement Learning over standard supervised learning algorithms is that the latter doesn’t perform this balance. They generally are purely exploitative. (Bayesian algorithms implicitly balance exploration and exploitation by integrating over the posterior.) The approach of inverse reinforcement learning to solve co-operative games enables in learning a more optimal reward function. The crucial terminology involved in any reinforce-ment learning system include environment, agents, rewards and utility function. The environment is the network of roads under consideration upon which drivers function. Drivers are the agents that operate on the environment. Rewards would include any benefits that adds to the user’s usability levels. The system involves the context of real time information through regular inclusion into decision making , thus making the whole process of route prediction concurrent and progressive.

A. Initial Data

The starting module in the entire system involves initial-ization of the system structure and baseline training. Train-ing process involves the basic understanding of the system

processes and executing forthcoming processes with the same accuracy. The data for the initial training process is obtained by leveraging existing data collection methods. This data is used to create the initialization for the network and construction of a probability table with respect to the data observed. The most widely used methods of aggregating location based information with respect to vehicles are the availability of Geographical Positioning Systems(GPS). GPS data can be remotely tracked from mobile devices associated with the vehicle. A real-time aggregation of such data can happen with a regular poll of the GPS coordinates. Thus, obtaining periodic location information helps in analyzing the exact route taken by the vehicle and also time is taken to travel the distance. The environment handler accepts routing requests and performs the iterative process of routing . Environment Handler , as mentioned in Figure 1 , performs the process of apprenticeship learning where the requests are associated with an agent and a separate set of initializations performed.

A futuristic technique would be to take advantage of the developments in Internet of Things. This approach would involve the consideration that all vehicles are interconnected and there is consistent transfer of data among the network. Data collection using Internet of Things can be achieved using a ”checkpoint” strategy where vehicles log their traversal on a given road at data collecting devices positioned strategically. Increased availability of vehicles with built-in GPS system al-lows the decentralization of the path finding mechanism. Once the user enters the destination, the vehicles can find path to be taken by themselves and send it to the server. This minimizes the load on the algorithm and makes implementation much easier. This technique is advantageous in maintaining a stable persistent system when it comes to real time updates of traffic flows. The data gets mapped into a database to maintain the information with respect to the clusters.

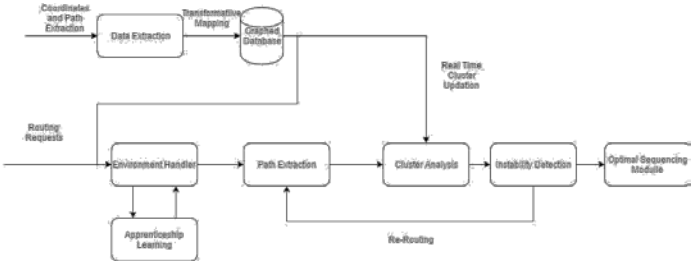

Fig. 1. Architecture of the System

B. Clustering Mechanism

This module is responsible for mining patterns and analyz-ing road clusters that exhibit similar traffic patterns. The train-ing data can be used to understand the historical patterns and deviations in traffic flow. Interesting patterns with relevance to this system would involve traffic flow among clusters. Once the system has been trained, it attains appropriate confidence levels that helps it get a solid understanding with respect to the reward functions used and the allocation of said rewards. The initial requirement in routing is to obtain all the possible paths from the source to the destination, achievable through single source shortest path algorithms. These paths would have common linkages with several clusters. The essentiality of the module is to ensure that the path optimally chosen provides a balanced traffic flow. These patterns may be based on a number of parameters that include the number of vehicles passing through, delay time, bottleneck issues etc. When the data with respect to the roads are obtained, the algorithm will cluster such roads to monitor for similar or deviant behaviour of traffic. Essential Analysis of such competing clusters is required to analyze possible sources of congestion .

Algorithm 1: Training and Reward Extraction

input: Coordinates of agents,C
1 M= Markov Decision Process;
2 C_{a1} = Monitored Coordinates of vehicle a_1 ;
3 T = Transition Probability Table of M;
4 foreach c_a 2 C do
5Extract the paths from coordinates;
6Update T
7 end
8 Constructfor the observed M such that
$E[^{P}_{t=0}\ ^{1}_{t}\ R(s_t)j\ )] > E[^{P}_{t=0}\ ^{1}_{t}\ R(s_t)j\ )]$8

Algorithm 2: Clustering Algorithm

input: graph G
1 P = All possible paths in the network;
2 initialize clusters;
3 foreach p 2 P do
4n_p = number of vehicles on p;
5t_p = average time spent on traversal
6 end
7 f lowrate$_p$=n_p/t_p; 8
while p do
9path$_p$ = paths with the same flowrate as current path ;
10 l_p=linking path between current path and paths in path$_p$;
11 if l_p then
12 append path p to cluster
13 end
14 create new cluster and assign path p to it.
15 end

C. Stability and Driver Interaction

The system functions essentially on the premise of the second rule of Wardrop Equilibrium [2].

Once the system receives the data with respect to the locations in consideration, the system checks all possible paths available for the journey. Once the path ideal for the user has been chosen, the algorithm now will analyze all possible allied paths that would influence the traffic throughout the network of paths that the vehicle would encounter. As by our initial definition, each vehicle is idealized as an agent in a multi-agent

environment. The common resource that all agents want would vary. It is least time and maximized fuel benefit that most drivers target. When the travel is considered, the resource would also include the right of traversing on a road. When agents compete for the same resource, we enter into a stand-off with two or more systems requesting access to the same resource. Given user preferences that change, certain agents may prefer optimizing on fuel costs over time at times. If the parameters considered in allocation are similar for the both the systems, it essentially reduces to a function requiring a feasible solution(maximized utility with respect to those parameters). Feasible solutions are defined as optimal allocation procedures. Hence, optimization procedures such as the particle swarm Optimization can be utilized to perform the analysis to obtain an optimized path. If the parameters that the agents prioritize are different in nature, we enter into a heterogeneous competition. Thus, resolving involves the concept of Nash's Equilibrium. We consider a bimatrix game with agents competing. Nash's Equilibrium will define the optimal utility that the agents can avail so as to not lose the possible utility.

When the introduction of the vehicle into the prescribed path causes an imbalance, the system would now have to route in an alternate manner. Thus, we would have to chart alternate paths from the driver's location towards the destination. This would require the passing of the geo-coordinates back to the path extraction. The entire systems now go through a rerouting phase, which is a second pass of the same algorithm. Once an alternate path that maximizes driver utility is obtained, the system enters the optimal sequencing module and the driver is guided along the same path and the stability of the network is restored.

Algorithm 3: Path Extractor Algorithm

```
input: source s, destination d
P = All possible paths from s to d;
foreach p 2 P do
n_p = number of vehicles on p; 4 end
Pick the path with least n_p;
while currentLocation is not d do
if inclusion into path causes no instability then
    n_p = n_p + 1;
else
    Calculate utility ;
    if utility < threshold then
        route := True;
        pathExtractor(currentLocation, destination)
    end
end
end
```

IV. CONCLUSION

The proposed system attempts to provide an congestion aware efficient routing mechanism using Inverse Reinforcement Learning. This routing system optimally provides us with the best possible path that can be learnt from the vehicular data. This system uses a hybrid model to determine congestion in the system by analyzing cluster based flows. The algorithm is devised in such a way that the complex characteristics of human driving are captured and reward functions are suitably formulated. This makes sure that efficiency of the system can be made better every time there is new areas or vehicles. In the process of traffic engineering, providing congestion-aware routing is always been a daunting task, this system takes a step forward in providing content which satisfies their interest and also helping the transport community to handle and construct traffic flows better.

ACKNOWLEDGMENT

The authors would like to thank the institution for allowing us to carry out this research work. The staff and coordinators were very much supportive in bringing out this system. Their queries, comments and suggestions were helpful in shaping the system to yield more effective results.

REFERENCES

[1] J. R. Pierce, "The Fuel Consumption of Automobiles," Scientific Amer-ican(ISSN: 0036-8733), vol. 232, 1975.

[2] J. G. Wardrop, "Some Theoretical Aspects Of Road Traffic Research," Proceedings of the Institution of Civil Engineers, vol. 1, 2002.

[3] F. S. G. Carlier, C. Jimenez, "Optimal Transportation with Traffic Congestion and Wardrop Equilibria," SIAM Journal on Control and Optimization, vol. 47, 2008.

[4] O. W. Y. B.K.Mohandas, R.Liscano, "Vehicle traffic congestion manage-ment in vehicular ad-hoc networks," Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on, vol. 47, 2009.

[5] E.Indrei, "Markov Chains and Traffic Analysis," The Rose-Hulman Undergraduate Mathematics Journal, vol. 8, 2006.

[6] P. E.Rachelson, F.Garcia, "Extending the Bellman equation for MDPs to continuous actions and continuous time in the discounted case," The International Symposium on Artificial Intelligence and Mathematics, 2008.

[7] S. R. Andrew Ng, "Algorithms for Inverse Reinforcement Learning," http://ai.stanford.edu/ ang/papers, vol. 1, 2000.