PREDICTING TRAFFIC CONGESTION
WITH DRIVING BEHAVIOR

Author: Yi Zhao,
Big Data Developer & Analyst
Introduction

How many hours a week do your drivers spend stuck in traffic?

Traffic congestion imposes a significant economic impact on industrialized society. The prediction of traffic congestion has been actively researched as one of the most effective solutions, to support proactive traffic management and planning. This article provides an overview of traffic congestion and the challenges of tackling congestion through different models of prediction.

Background

The Negative Impact of Traffic Congestion

Traffic congestion is a problem that began more than a century ago when the introduction of the assembly line in the automotive industry made it possible to mass-produce automobiles. The problem of traffic congestion has not yet been resolved and still causes countless losses in most industrial countries.

The negative effects of traffic congestion:

1. Congestion decreases the traffic volume that can use the roadway.
2. Delays due to congestion cost Americans $121 billion in the form of 5.5 billion lost hours and 2.9 billion gallons of wasted fuel in the USA alone.[1]
3. Congestion makes it hard to estimate traffic time. As shown in Figure 1, in the USA, people have to spend a surprising amount of time on trip planning due to congestion.

![Figure 1. Extra time to make important trips [1].](image-url)
Ways to Reduce Congestion

Three methods of reducing congestion are commonly considered \cite{2}:

1. **Increase transportation capacity by constructing more freeways.** This solution considers road saturation as a natural way to explain traffic congestions. The demand for space on the roads is greater than what is available. However, research uncovered a phenomenon called generated traffic\cite{18} - i.e. increasing road capacity can even stimulate more demands than newly increased capacity can take, thus may lead to even worse traffic. Moreover, further construction of freeways to alleviate congestion is unlikely in many cities today.

2. **Promoting public transport in large cities.** While this may be a good solution, it is not always convenient for people to use.

3. **Employ proactive traffic management and planning.** Under such circumstances, proactive traffic management and planning to improve traffic efficiency is considered to be a major way of alleviating traffic congestion. This requires determining the future state of road segments, i.e. predict potential congestions before real congestion is reached.

Proactive Congestion Detection and Prediction

Traffic dynamic models are the foundation of predictions. Both macro and micro levels of traffic dynamic has been intensively studied. Macroscopic dynamic models are good at explaining recurring traffic congestion patterns on a large-scale network perspective while microscopic dynamic models explore individual vehicle behaviors and are very effective in stimulating the instabilities of traffic waves and non-recurring congestions caused by unpredictable incidents.

Despite the popularity of traffic dynamic models, not so many practical methods of congestion prediction have been developed and deployed, especially for non-recurring congestions which are responsible for more than half of congestion cost. Based on literature reviews, this article discusses a novel method of real-time, short-term non-recurring congestion prediction, which integrates individual vehicle behavior data collected by telematics technologies and research results of modern traffic dynamic models.
Proactive congestion detection and prediction was virtually an untouched area before Kaysi et al.\cite{3} set up a framework for a real-time Advanced Traveller Information System (ATIS). Real-time diversion of traffic is expected to be implemented through this system. In brief, wireless communication infrastructures and navigation technologies are exploited to:

1. Collect real-time data from surveillance system,
2. Infer relevant information concerning congestion, and
3. Provide congestion information through a control and routing (CAR) module to each single driver.

Congestion prediction is critical for the effectiveness of ATIS.

The challenges of congestion prediction are rooted in the complex traffic flow dynamic. Real-world traffic flow is a multi-loop, multi-state, non-linear feedback system that reacts to the individual decision maker’s or driver’s actions in both anticipated and unanticipated ways. Traffic dynamic has been intensively studied since 1950’s. Two perspectives are widely used to simulate traffic dynamic: the microscopic view and macroscopic view.

1. **Macroscopic traffic flow** simulates the behaviors of the traffic stream overall (road density, vehicle velocity and volume).
2. **Microscopic traffic flow** simulates the behaviors of individual vehicles.

![Macroscopic versus microscopic](image-url)
Different models have been developed to explain congestion based on these two perspectives. Macroscopic models are quite effective in explaining recurring congestion. This type of congestion reflects the day-to-day build-up of traffic on urban expressways and arterials, notably during the morning and afternoon commuter peak periods. The regularity of this congestion allows its measurement relatively straightforward. Common measurements include macroscopic features such as traverse time, average speed and density.

Macroscopic models have been intensively researched and a few of them have been implemented in real traffic management systems. They helped to discover recurrent congestion patterns of large scale traffic networks in question. However, macroscopic models are less sensitive to another type of congestion, non-recurring congestion, which is responsible for two-thirds of traffic delays\textsuperscript{[1,7]}. 

Non-recurring congestion is caused by roadway incidents or other unexpected events that temporarily disrupt the flow of traffic on a segment of a roadway.

When an incident occurs, there is a temporary decrease in the capacity of the roadway and if current demand exceeds this reduced capacity it results in congestion, thus forming queues.
When vehicle density is large enough (~30/km), a random event such as a fluctuation in a leading car's velocity can induce a transition from free flow to jam. This explains the phenomenon where for no apparent reason, traffic will slow to a crawl or even a halt and after several minutes return to normal speeds. Such stop-and-go roads only carry ½ to ⅔ of the vehicles as a smoothly flowing road. Macroscopic aggregate data is less straightforward to fit in the explanations.\(^4\)

Due to complex human factors, it is only in recent years that microscopic models have been well established to simulate non-recurring congestion phenomena. Mathematicians from the University of Exeter developed a math model\(^6\) revealing that by slowing down below a critical speed when reacting to such an event, a driver would force the car behind to slow down further and the next car back to reduce its speed further still. The result of this is that several miles back, cars would finally grind to a halt. Flynn et al.\(^5\) coined the term phantom traffic jams to describe jams that arise in the absence of any obstacles.

They found that the fundamental features of traffic waves (jamitons) in phantom traffic jams are similar to detonation waves produced by explosions and argued that the exact shape and the speed of propagation of jamitons could be predicted.

However, no practical prediction methods have yet been developed to apply to non-recurring congestions. A fundamental challenge is the general lack of data at microscopic levels for incident detection.\(^7\)
How Vehicle Telematics Helps Drivers Manage Traffic Congestion

Vehicle Telematics has been defined as the convergence of telecommunications, information processing and vehicle technologies and road transportation\(^8\). GPS fleet management and vehicle tracking can play an important role in handling traffic congestion. For example, vehicles equipped with Geotab technology have successfully realized safe trips and significant cost savings benefits. **Figure 4** illustrates a high-level architecture of the Geotab fleet management system.

![Figure 4](image)

**Figure 4.** A high-level picture of the architecture of Geotab fleet management system.

In this architecture, each vehicle (referred as “Geotab probe vehicle” in the following) has one Geotab GO device plugged into the on-board diagnostics (OBD) interface. After several minutes of warm-up at the first time of being installed, the GO device starts to retrieve real-time OBD information, such as:

- Engine status
- Fuel level
- Mileage
- Speed
- Seat belt usage status
- Time when brake is peddled
- and more
The GO device is also embedded with a GPS chip and an accelerometer to detect host vehicle’s location and acceleration. In every second, over 10,000 of such real-time track records are transferred upward to Geotab servers for further processing. Derived info around driving behaviors, such as unexpected harsh speeding or decelerating, turning radius and idling without ignition off, as well as vehicle model specific features, such as inertia and fuel consumption, are calculated. As of now, Geotab allows its end users to view this information in real-time within Geotab’s web-hosted reporting environment and in turn achieve smarter fleet management. GO devices are also capable of transferring downward route plan instruction from fleet managers to matched Garmin Navigation System\(^9\) on host vehicles.

Geotab fleet management architecture unintentionally conforms with ATIS framework.

With Geotab technology, comprehensive and detailed info around individual vehicle behaviors becomes available and measurable.

This triggers a great applied research interest in exploring the viability of utilizing data from Geotab devices in real-time, short-term (i.e. 5-30 minutes) congestion predictions, i.e. when and where congestion is likely to occur on the freeway network. Comparing to predictions purely based on macroscopic traffic features, deeply exploring and properly utilizing individual behavior data has advantages in identifying anomalies which leads to stop-and-go traffic and in turn predict congestions at the very early stage. Such work is also expected to put valuable insights into the current Geotab vehicle trip planner and future automatic fleet control system.

Goals of this Article

This article is a pilot study of the said applied research project. The goal is to have a better understanding of:

(i) the state-of-the-art of traffic congestion;

(ii) how Geotab’s data can help in traffic congestion prediction.

The remaining content is organized into three sections. Section 2 provides a literature review around traffic congestion prediction approaches. This article does not attempt to give a complete list of models generated since 1990s, but rather focuses on approaches which can be easily integrated into ATIS. Section 3 gives a qualitative analysis of micro-level traffic flow dynamic as feedback loops. Based on the analysis, Section 4 proposes a novel approach of predicting traffic congestion by integrating traffic flow Geotab data. Section 5 draws a conclusion based on the review of information.
Related Work on Congestion Prediction

Huisken [10] surveyed several different schemes which had been proposed as research design under the ATIS framework. They commonly use motorway data from dual induction loops detectors (magnetic induction devices that sense and report the passage of vehicles periodically) which are installed alongside freeways and on-ramps and off-ramps. Those data are aggregated with time bins from one minute to several minutes to derive further traffic information (speed, volume and occupancy) and then are used as input to prediction models. Those research works all claimed to be promising, but subject to further comparison, ATIS was left only in a conceptual level. It was only until recent years that several approaches that can easily be integrated with ATIS have been implemented.

Nihan et al. [11] reported congestion predicting algorithms based on pattern recognition applied on an on-ramp freeway metering system for freeway network of the Seattle Metropolitan region. They started with a detailed discussion on the definition of “bottleneck” or “forced flow” condition for freeway. A condition with lane occupancy value greater than a pre-set threshold and a positive storage rate which was used before has been proved to oscillate significantly during both congested and uncongested conditions, and thus not efficient. Instead, the ratio of flow over occupancy ($F/O$) was considered to be a better indicator of the transition from uncongested to congested traffic, as it corresponds to speed in a linear way and has consistent average values for both flows.

$$S = F/D$$

where $S =$ space mean speed ( miles/hour), $F =$ flow rate (vehicle/hour) and $D =$ density (vehicle/kilometer). Density can be expressed as a function of lane occupancy by $D = \left(\frac{O}{100}\right)\left(\frac{1000}{L_e}\right) = g(O)$ where $O =$ lane occupancy(%), $L_e =$ average effective vehicle length (meter).

So we have $S = \text{len}/10\ (F/O)$

Experiments with real freeway data revealed that $F/O$ values below certain thresholds (varies to different roadways) were very accurate indicators of existing forced-flow and impending forced-flow. Identifying a “bottleneck” as soon as it occurs using $F/O$ criterion usually gives upstream several minutes of warning before it also becomes congested.

Marfia et al. [12] defined congestion as a road’s state, which lasts for at least $S$ units of time, during which travel times exceed the time $T^*$ normally incurred under light or free-flow travel conditions. A training phase is needed to collect sufficient sample data and get an optimized $S$ and threshold $T^*$ for the observed road section. After that when a vehicle spends a longer time to traverse that given road section, ATIS will alert all other vehicles that are approaching this area. The objective function for $S$ and $T^*$ as well as algorithm to solve this function were elaborated in [12]. I noticed an unstated presumption in their work that the value of $(S,T^*)$ can be considered as a feature of given road section and does not change frequently, so that using a calculated $(S,T^*)$ to detect congestion state 1 month after the training phase was mentioned in this article. The strength of their algorithm is that it is able to detect and alert recurrent congestions without requiring any prior knowledge regarding a roadway. And it also makes it possible to detect congestion purely with data from probe vehicles without other roadside facilities. The research work was verified to be effective in congestion detecting with experiments on 9 non-consecutive road sections in Los Angeles, California and Pisa, Italy and comparison with congestion. But several important issues relevant to implementation were not touched, for example, if and how the value of threshold varies among different times of the day and a month.
IBM Traffic Prediction Tool (TPT)\textsuperscript{[13]} is a learning and analytics engine based on statistical models for the near-term prediction of traffic conditions. It has been tested in Singapore to provide one-hour traffic predictions. Average volume and speed are considered as the key indices to assess traffic conditions. These data are gathered from road sensors, such as inductive loop detectors. TPT employ historical data to find anomalies, determine the most likely resulted traffic patterns. Their prediction model typically rely on what mathematicians call Lagrangian sensing, a method by which algorithms provide optimized solutions to complex arrays of partial differential equations that represent traffic flows. Pilot application of this systems showed a quite accurate traffic pattern recognition and congestion prediction. However, the detailed model and implementation is not published.

\textbf{Overview}

In existing congestion prediction models (including the above ones) which can be found, initial congestion conditions (bottlenecks) are defined and detected according to different statistical models built on only macroscopic traffic features (density, velocity and volume). Road features and recurring rush hours speak a lot for the most likely traffic flow patterns resulted by bottlenecks. Thus all presented models showed quite satisfying predicting results for recurrent congestions on specific road sections. However, we are often confronted with non-recurrent congestions without any apparent bottlenecks, which slows traffic to a crawl or even brings it to a halt. Intensive reports and studies have proved what we experience is not a random phenomenon but is in fact responsible for more than 60\% of the total congestion cost \textsuperscript{[1,7]}. This type of congestion reflects the delays caused by random incidents, such as stalled vehicles, overreacting hard decelerations, accidents and truck spills. Its randomness makes it difficult to predict. A fundamental challenge is the general lack of data at microscopic levels (second-by-second measurements) \textsuperscript{[7]}. Difficulties also arise from complex microscopic traffic dynamics models that heavily relying on individual driver behaviors \textsuperscript{[14]} and in turn difficult to simulate, and lack of efficient holistic method of associating micro-level incidents with macro-level traffic volumes and other information on the occurrence of incidents \textsuperscript{[7]}. The idea of this article is based on an observation that congestions could be caused by both macro and micro factors. Thus a satisfying congestion prediction could not be achieved without either side. The majority of existing traffic prediction applications are based on macroscopic models. Lacking individual vehicle info, these methods could hardly grasp the causes and in turn the initialization of a congestion. This limits the capability of those models to look ahead of a congestion. Recognizing the events which might cause a congestion can be very helpful to predict a forming congestion even before it really happens. So, this article proposes a holistic congestion predicting method which integrates the two levels, hoping to extend the model capability of looking ahead of time. The presentation here is based on the simplest possible setting, i.e. vehicles following each other on a single lane. In this setting, as shown in Figure 5, vehicles are plotted as discrete entities moving in continuous time and continuous space. At a certain time point, the position of the front bumper of the nth car is denoted by $x_n$, its velocity is $v_n$ and the bumper-to-bumper distance to the vehicle in front (called the headway) is $h_n$.\textsuperscript{[14]}
In the following sections, I describe the traffic dynamics in a holistic perspective involving both macro and micro features, then I discuss how to use Geotab data with the described traffic dynamics to conduct congestion predictions.

**Traffic Dynamics Analysis**

It is very helpful to think about the traffic dynamics as a set of interacting feedback loops. My approach begins with exploring a utopian world where traffic equilibrium exists where everyone maintains a reasonable speed, keeps a safe distance between cars, and signals their intent before acting. My contention is that the entire system would flow smoothly with every vehicle moving toward downstream at the same speed, with a minimum of collisions and stop-and-go incidents.
The equilibrium is shown in Figure 6 as a balance feedback loop. From the perspective of vehicle n, whenever vehicle n+1 decelerates (steps on the brake), the relative speed between vehicle n+1 and n goes to negative and in turn the headway hn decreases. According to basic observation as well as simulations with mathematical models\cite{14}, the smaller the headway is, the slower the driver intend to go and vice versa. So vehicle n will brake to slow down. If it slows down too much, the headway begins to increase and in turn drives n tends to to increase its speed a bit back.

For an ideal vehicle, we assume that the whole feedback completes very quickly and accurately so that it will not cause the following car to stop or crash into it. The same feedback loop also happens on vehicle n+1, n+2 and so on. Every driver has a comfortable headway value but as a whole, as the result of this balance loop, all vehicles in this lane will move at a same speed and the whole traffic flow will look to be smooth.

However, this ideal equilibrium rarely happens nor is able to last in the real world. When vehicle n+1 decelerates, it is straightforward that headway hn decreases immediately, but the following dynamic is way more complex than what Figure 6 illustrated. Several other important factors should be considered, as shown in Figure 7 with yellow links:

\hspace{1cm}

\begin{itemize}
  \item Deceleration of vehicle n+1
  \item Velocity of vehicle n+1
  \item Relative speed between vehicle n+1 and n \((vn_+1-vn)\)
  \item Headway
  \item Velocity of vehicle n
  \item Deceleration of vehicle n \([m/s^2]\)
  \item Vehicle inertia
  \item Brake intent
  \item Desired headway
  \item Comfortable deceleration
\end{itemize}

Figure 7. Including driver’s intent into traffic equilibrium.
• When the change of headway is noticed by driver \( n \), a desired headway will be compared to it in driver \( n \)'s mental model according to this driver’s experience. The bigger the difference is, the harder driver \( n \) intends to brake. How big the difference could be positively correlates to driver \( n \)'s reaction time, i.e. the dead time required to process information about stimuli and to initiate actions. A longer reaction time tends to cause harder braking. Each person’s reaction time to the distance and to the velocity difference is different, but almost falls into the range from 0.5 to 1.5 seconds\(^{[14]}\).

• Drivers intend to limit the force of brake within their comfort zone (represented by comfortable deceleration).

• However hard driver \( n \) brakes, the maximum deceleration vehicle \( n \) can achieve is limited by its mechanical capacity. Also, vehicle \( n \)'s weight give it an inertia which slows down the speed of deceleration.

We can consider a further assumption. Drivers choose speed based on safety. As proposed in collision avoidance models\(^{[15]}\), we assume drivers on freeway will keep a safe time distance from the leader. In mental model, people are fundamentally estimate time distance from its leading car. People feel a headway is safe when they believe they have a certain amount of time for their own car to stop without crashing into the next car. Although everyone has their own concept of what is a safe time. Thus Figure 6 can be further extended with desired time distance and velocity felt by driver \( n \), as shown in Figure 8.

**Figure 8.** Extend traffic equilibrium for the sake of computation.
All in all, generally speaking, when vehicle n+1 brakes, vehicle n’s achieved deceleration is a function of its maximum deceleration, the driver’s comfortable deceleration, desired maximum velocity, desired time gap, and the driver’s recognition of current speed and headway. Detailed mathematical functions and explanation can be found in [14].

A Unique Approach to Traffic Congestion Prediction

As mentioned above, an overreaction, or harsh braking, in a non-free traffic flow will cause a traffic jam miles back. Whether we can predict this kind of congestion once we detect harsh braking is a question of interest. We don’t need to wait for congestion indicators such as a high density or long traverse time to become significant before sending congestion alarms to vehicles heading to this area.

Giving drivers a warning even one minute ahead of the next off-ramp and incoming vehicles could allow them to leave the freeway and avoid the traffic jam.

Based on dynamic models in the previous section, this article proposes a unique congestion prediction method, consisting of the following:

- **Deceleration info from a leading car.** A leading car could be any vehicle equipped with a Geotab GO device which is able to compute vehicle deceleration with motion sensors and accelerometers.

- **A model conforming with traffic dynamics computing how the deceleration of a leading car affects the following ones.** That is, given a deceleration value of vehicle n+1, possible velocity values of vehicle n and other following cars can be calculated. Probably the most difficult and complicated part is how to model and compute driver’s behavior. Deep statistics and fields study are needed to form a proper computable model. Relative discussion is out of the scope of this article. However, this article would refer to [14] for candidate models as a realistic starting point. Some critical points regarding the computability of this model are identified and discussed below.

- **A training module** tuning model parameters, using historical data.

- **A prediction module** taking real-time vehicle deceleration values as inputs and predicting if, when and where a congestion might happen.
Integration with Macroscopic Data Source

Given the deceleration of vehicle n+1, microscopic vehicle data collected by Geotab GO devices alone is not sufficient to accomplish dynamic calculation. Macroscopic flow data (density) is needed.

- Headways($h_n, h_{n-1}, h_{n-2}, \ldots$) correlate to traffic density, which is collected by inductive loop detectors.
- Desired headway depends on desired time distance\cite{16} and density-dependent optimal velocity\cite{14}. These parameters along with drivers’ comfortable deceleration and vehicle inertia should be carefully estimated by field studies and stats.

Driver Behavior

With all the above values prepared, deceleration of the following cars can be calculated. In existing dynamic models, the deceleration has been prescribed as

$$\ddot{v}_n(t) = f(h_n(t - \tau), h_n(t - \sigma), v_n(t - k))$$

where $\tau, \sigma, k$ represent driver reaction times to different stimuli. A simplification is commonly used in the literature: “Human driver set-up’: $\tau = \sigma > 0, k=0$. This set-up represents that drivers react to the distance and to the velocity difference with the same delay, but they are aware of their own velocity immediately\cite{17}.

A driver’s reaction time, which could be in the order of seconds, plays a crucial role in traffic congestion. When the leading car brakes, time distance will decrease immediately, and there is a reaction delay for the follower to take further actions. This means the following car should decelerate more than the leading car, otherwise the time distance will be shorter than before which violates our assumption about the driver intending to keep a safe time distance. As every follower brakes harder than the one before, at some point the velocity of the followers will drop down to a stop.

Threshold for Deceleration Value

Everyday, approximately 200 million acceleration records are collected by Geotab from hundreds of thousands of vehicles across more than 50 countries.

A threshold value of deceleration (a congestion-prone brake) should be carefully chosen in order to trigger a necessary congestion prediction process and send an alarm accordingly.

In a free-flow or very low density road, a brake won’t necessarily cause congestion, but on a high density road it will. So, field studies are necessary to discover the quantitative correlation between road density and minimum congestion-prone brake.
Threshold for Congestion Alarm

When a forming stop-and-go is predicted, the system should wisely decide whether to trigger a congestion alert and how far away this alarm should reach. As an initial stop-and-go might cause a series of secondary stop-and-go situations, creating a proper alarm is quite tricky.

Conclusion

A number of congestion prediction models have been established. In these models, macroscopic features of traffic, such as density, volume and flow rate, are proven to be necessary and quite effective. At the same time, other work demonstrated that microscopic data of individual vehicles such as driver’s delayed reaction time and overreaction in heavy traffic are significant factors contributing to traffic jam. However, the potential of utilizing microscopic data to predict non-recurring congestions in real-time has not been fully exploited. This article starts from analyzing a traffic dynamic loop and then discusses a novel method of integrating real-time probe vehicle data with this loop to predict non-recurring congestions in a qualitative way. Future work is needed to develop systematic quantitative prediction models and field studies are necessary to verify the effectiveness of prediction.
References


Copyright, Geotab 2015. All Rights Reserved.