Car-following models, Part IV: Nonparametric
Traffic flow theory (TFT)

Zhengbing He, Ph.D.,
http://zhengbing.weebly.com

School of traffic and transportation, Beijing Jiaotong University

October 11, 2015
Outline

1. Introduction
2. Background
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor
3. The nonparametric car-following model
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database
4. Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
Existing car-following model, such as Gipps, IDM, Full velocity model

- Ordinary differential equation (ODE) with speed and acceleration variables.
- Fundamental diagrams and driver’s behavior parameters.
- Calibration is needed before practical usage.
- Should be able to reproduce major traffic characteristics.
Contributions of this paper:

- A nonparametric car-following model based on k-nearest neighbor
- Neither mathematical equation nor calibration is needed
- Neither fundamental diagrams nor driver’s behavior parameters is assumed
- The model is simple and parsimonious, because there is only one parameter
- The model is able to well reproduce important traffic characteristics
Introduction

Background

NGSIM dataset

A nonparametric approach: \( k \)-nearest neighbor

The nonparametric car-following model

The model

Determination of \( k \) and similarity

Analysis on avoiding collisions

Transferability of the model and the database

Simulation scenarios

Scenario 1: Following empirical leaders

Scenario 2: Rubbernecking

Scenario 3: Driver errors
NGSIM dataset

- Open-source: about 1 hour trajectory dataset, 2005
NGSIM dataset

wave speed = -16 km/h
A nonparametric approach: \( k \)-nearest neighbor

- Nonparametric approach/model
  - Parametric approach: based on \textit{mathematic formulas}
  - Nonparametric approach: \textit{no model}, driven by historical data
- \( k \)-nearest neighbor: very simple but works well
  - \textit{History is repeating}: Most similar conditions (input) highly likely result in similar outcome (output).
  - \textit{For example}:
    - Wind, temperature, traffic \( \rightarrow \) magnitude of haze.
    - Looking up similar days in history based on wind, temperature, traffic today
    - Taking \textit{the average magnitudes of haze in these days} as \textit{the estimator of the haze today}. 
The approach selects the most similar historical cases, and takes the average of their outputs as the estimate of this time.

Specifically, the approach estimates $y_0$ in focal $(x_0, y_0)$ as follows.

$$\hat{y}_0 = \frac{\sum_{i=1}^{k} y_i}{k}$$

(1)

where $x_i$ with respect to $y_i$ is one of the $k$ most similar samples to $x_0$. 
Outline

1. Introduction

2. Background
   - NGSIM dataset
   - A nonparametric approach: \( k \)-nearest neighbor

3. The nonparametric car-following model
   - The model
   - Determination of \( k \) and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database

4. Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
A key to understand the nonparametric model:

- **Input**: leader’s two-step speed + follower’s two-step spacing
- **Output**: follower’s speed
The model: mathematical expression

- **Input:**

\[ x_n(t + \tau) = (d_{n-1}(t + \tau), \ d_{n-1}(t), \ s_n(t), \ s_n(t - \tau)) \]. \hspace{0.5cm} (2)

where \( \tau \) is the simulation time step; \( (n - 1) \) is the leader of vehicle \( n \); \( d_{n-1}(t) \) is the moving distance of the leader between time \( (t - \tau) \) and \( t \)

- **Output:** moving distance of vehicle \( n \)

\[ y_n(t + \tau) = d_n(t + \tau) \] \hspace{0.5cm} (3)
Because of strong autocorrelation of time-series trajectory data, it is easy for all $k$ samples coming from a leader-follower pair, in particular when the leader and follower move in a constant speed. Such dominance could reduce the reliability of the model. To overcome this issue, we make all $k$ samples selected from different leader-follower pairs.
Distance between two data samples

- “Ordinary” and simple scaled Euclidean distance, i.e., adjusting the input $x_{ji} \in x_i$ by its mean $\bar{x}_j$ and standard deviation $S_j$ before calculating Euclidean distance
- The model reads

$$D(x_i, x_0) = \sqrt{\sum_{j=1}^{J} (z_{ji} - z_{j0})^2}$$  \hspace{1cm} (4)

where

$$z_{ji} = \frac{x_{ji} - \bar{x}_j}{S_j},$$  \hspace{1cm} (5)

and $J$ is the total number of all elements in an input vector.
Outline

1. Introduction

2. Background
   - NGSIM dataset
   - A nonparametric approach: k-nearest neighbor

3. The nonparametric car-following model
   - The model
   - Determination of k and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database

4. Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
**Determination of $k$: Introduction**

- $k$ in $k$NN: **the number of the historical cases** that are considered to be **similar** to the estimated case.
- Usually, $k$ is estimated by **experience**.
- Here, we determine $k$ by **comparing estimation errors** under different $k$-values.
- Employ the scaled Euclidean distance from $k$th nearest sample to the estimated case, which is the **longest distance** in all $k$ samples, Denote by $D_k$. 

---

Zhengbing He, Ph.D., [http://zhengbing.weebly.com](http://zhengbing.weebly.com)  
Car-following models, Part IV: Nonparametric
Determination of $k$: Four databases

We first build three databases with different sizes by using the datasets collected on Lane 2 and 3. Such databases are employed to estimate the movement of the followers on Lane 1.

- **Database 1**: the small-size database containing 22,202 input-output samples, which is built by using the dataset collected on Lane 2 during the first 15 minutes;

- **Database 2**: the medium-size database containing 78,683 samples, which is built by using the dataset collected on Lane 2 during the all study 45 minutes;

- **Database 3**: the large-size database containing 152,637 samples, which is built by using the dataset collected on Lane 2 and 3 during the all study 45 minutes.
**Introduction**

**Background**

The nonparametric car-following model

**Simulation scenarios**

The model

Determination of $k$ and similarity

Analysis on avoiding collisions

Transferability of the model and the database

---

**Determination of $k$: First scenario**

- We estimate the followers of two typical vehicles who traverse stop-and-go oscillations on Lane 1
- Database 3 is used, and the absolute errors with different $k$-values are compared

---

(a) Following Vehicle 422 on Lane 1

(b) Following Vehicle 1989 on Lane 1
Determination of $k$: First scenario

- **Upper plot (absolute errors):** smaller $k$ basically makes smaller errors, but the errors may fluctuate more
  - the closer historical samples better reflect estimated case, but averaging a small number of samples results in instability
  - General tendency but not always true.

(a) Following Vehicle 422 on Lane 1
(b) Following Vehicle 1989 on Lane 1
Determination of \( k \): First scenario

- **Lower plot (distance \( D_k \)):** smaller \( k \) basically makes smaller distance (more similar)
  - For Database 3, when \( k = 10 \), the estimations are usually good with the distance smaller than 0.2, i.e. \( D_k < 0.2 \)

![Graphs](image-url)
Determination of $k$: Second scenario

To make a more general conclusion,

- Using different $k$-values and Database 3, estimate **randomly-selected** 50 followers.
- Calculate the mean and standard deviation of absolute errors
- It can be seen:
  - Upper plot: for the dataset, the optimal $k$-value is about 10
  - Lower plot: usually, $D_k < 0.2$
Determination of $k$: Second scenario

- Fixing $k = 10$, and using different databases,
- It can be seen
  - The larger Database 3 results in better estimations (i.e., lower mean and standard deviation of the absolute errors)
  - Further shortens the distance $D_k$. 

![Graph showing mean and standard deviation for databases 1, 2, and 3.](image)

Zhengbing He, Ph.D., http://zhengbing.weebly.com
Car-following models, Part IV: Nonparametric
Determination of $k$: Conclusion

- Specify $k = 10$ for the database built on the US-101 dataset.
- An estimation is considered to be satisfied if $D_k < 0.2$.
- This determination is a premise to apply $k$NN, because the optimal $k$-value is highly related to the underlying database.
## Outline

1. **Introduction**
2. **Background**
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor
3. **The nonparametric car-following model**
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database
4. **Simulation scenarios**
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
Generally speaking, averaging similar collision-free historical cases would not lead to a collision. However, as a data-driven method, it may be difficult to prove it mathematically. To show collision-free statistically, we plot the relative errors between ground-true and estimated space headway.
Analysis on avoiding collisions
Outline

1. Introduction
2. Background
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor
3. The nonparametric car-following model
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database
4. Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
Transferability of the model and the database

- Basic assumption of $k$NN: drivers repeat their behavior in similar circumstances
- An database could be well transferred to any site with similar circumstances including driving habits, roadway geometry, etc.
- To show this, we estimate vehicles on I-80 using US-101 database.
Transferability of the model and the database

(a) Lane 2, I-80

(b) Lane 3, I-80
Outline

1 Introduction

2 Background
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor

3 The nonparametric car-following model
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database

4 Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors

Zhengbing He, Ph.D., http://zhengbing.weebly.com

Car-following models, Part IV: Nonparametric
Outline

1 Introduction

2 Background
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor

3 The nonparametric car-following model
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database

4 Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
A platoon with real boundary conditions

Scenario 1: Following empirical leaders
Scenario 2: Rubbernecking
Scenario 3: Driver errors

**Empirical data**
- (a) Real platoon following Vehicle 422
- (c) Real platoon following Vehicle 1989

**Simulation**
- (b) Simulated platoon following Vehicle 422
- (d) Simulated platoon following Vehicle 1989

Zhengbing He, Ph.D., http://zhengbing.weebly.com
A platoon with different boundary conditions

(a) Vehicle 422, entry gap 30 m  
(b) Vehicle 422, entry gap 40 m  
(c) Vehicle 422, entry gap 50 m
Necessity of each input

(a) Inputs without $d_{n-1}(t - \tau)$

(b) Inputs without $d_{n-1}(t)$

(c) Inputs without $s_n(t - \tau)$

(d) Inputs without $s_n(t)$
Outline

1 Introduction

2 Background
   - NGSIM dataset
   - A nonparametric approach: $k$-nearest neighbor

3 The nonparametric car-following model
   - The model
   - Determination of $k$ and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database

4 Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors
Constructing a rubberneckering scenario

- A 1.25 km one-lane roadway is simulated for 1 hour;
- New vehicle enters with an initial speed of 54 km h$^{-1}$, when its leader has left the entrance 30 m away;
- The rubberneckering zone is located at section [1, 1.05] km.
- A probability $r$ to rubberneck and then slow down by a percentage of $(1 - p)$.
- If rubbernecking occurs, it will occur at most once.
- The database contains all data collected from Lane 1, 2, and 3 during all 45 minutes.
- No assumption or calibration
Simulation results: Time-space plots

(a) Time-space diagram coloured by speed

(b) Time-space diagram coloured by distance
Simulation results: Time-space plots

(c) Region 1

(d) Region 2

(e) Region 3

(f) Region 4
Simulation results: Fundamental diagrams

Suppose that virtual detectors are installed in the roadside, and the traffic flow $q$, density $\rho$, and speed $v$ within a time period $T$ are measured as standard models.

$$q = \frac{N}{T}, \quad \rho = \frac{\sum_{n=1}^{N} \frac{1}{v_n}}{T}, \quad \text{and} \quad v = \frac{q}{\rho} = \frac{N}{\sum_{n=1}^{N} \frac{1}{v_n}} \quad (6)$$

where $N$ is the count of the vehicles passing the detection location within the time period $T$, and $v_n$ is the passing speed of a detected vehicle.
Simulation results: Fundamental diagrams

(a) flow-density diagram for empirical traffic

(b) flow-density diagram for simulated traffic

(c) speed-flow diagram for empirical traffic

(d) speed-flow diagram for simulated traffic

(e) speed-density diagram for empirical traffic

(f) speed-density diagram for simulated traffic
Simulation results: Periods and amplitudes
Introduction

Background

The nonparametric car-following model

Simulation scenarios

Outline

1. Introduction
2. Background
   - NGSIM dataset
   - A nonparametric approach: k-nearest neighbor
3. The nonparametric car-following model
   - The model
   - Determination of k and similarity
   - Analysis on avoiding collisions
   - Transferability of the model and the database
4. Simulation scenarios
   - Scenario 1: Following empirical leaders
   - Scenario 2: Rubbernecking
   - Scenario 3: Driver errors

Zhengbing He, Ph.D., http://zhengbing.weebly.com

Car-following models, Part IV: Nonparametric
We model the driver errors in a form of a white Gaussian noise with diffusion coefficient $\sigma^2$.

\[
\ddot{d}_n(t + \tau) = d_n(t + \tau) + W(\varepsilon)
\]

(7)

where

\[
W(\varepsilon) = \begin{cases} 
\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\sigma^2}}, & d_{jam} < d_n(t + \tau) < d_{free} \\
0, & \text{otherwise}
\end{cases}
\]

(8)

where $d_{jam}$ and $d_{free}$ are the moving distance/speed around jam density and free-flow conditions, respectively.
Constructing a driver-error scenario

- A 900 m one-lane roadway is simulated for 1 h.
- The white noise is only added when a vehicle is moving in the section between 300 m and 600 m. This is analogous to an uphill section.
- It is set that $d_{\text{free}} = 54 \text{ km h}^{-1}$ and $d_{\text{jam}} = 15 \text{ km h}^{-1}$
- The entry speed and gap are 54 km h$^{-1}$ and 20 m, respectively.
Simulation results: Time-space plots

(a) Time-space diagram of trajectories ($\sigma = 0.2$)

(b) Time-space diagram of trajectories ($\sigma = 0.5$)
Simulation results: Time-space plots

(c) Region 1

(d) Region 2
Conclusion

- Neither mathematical equation nor calibration is needed to be concerned in the model;
- Neither the fundamental diagrams nor driver’s behaviour parameters is assumed;
- The model is simple and parsimonious particularly in the conceptual point of view, and the only parameter is $k$;
- All inputs and outputs are based on vehicle positions, which are straightforward to reproduce traffic dynamics in computer simulations;
- The model is able to well reproduce traffic characteristics contained by the underlying database, such as all stages of stop-and-go oscillations, fundamental diagrams, periods and amplitudes of oscillations.
Thank you!