Trajectory Bayesian Indexing: The Airport Ground Traffic Case

Cynthia Delauney, Nicolas Baskiotis and Vincent Guigue

IEEE 19th International Conference on Intelligent Transportation Systems
Rio de Janeiro, Brazil

November 2nd, 2016
**Trace** = set of measures (id, time, location, *contextual info*)

**Issues:**
- Clustering/categorization [Jiang et al. 08]
- Anomaly detection [Bu et al. 09]
- **Indexing** [Guttman et al. 84, Chakka et al. 03, Zheng et al. 11]

**Challenges:**
- Variable size
- Noise(s)
- Data amount
**Main goal: light & rich indexing**

**Use cases:**

- **Query** What is close to a **given situation**?
- **Analysis** What are the **common features** shared by close trajectories?
- **Predict** Does the current trajectory **become closer to a risk situation**?

- Which trajectory **representation**?
- Which **metrics** between trajectories?
Whole dataset:

1 year ~ 130 000 trajectories
~ 350 Gb (with a rich context)
$|T_k| \sim 1000$ in average

Trajectory samples:

$T_k = \{c, (t_1, \ell_1, \ldots, t_{|T_k|}, \ell_{|T_k|})\}$
\[ t \in \mathbb{R}, \ell \in \mathbb{R}^2 \]
$c$ : context, $t_i$ : time, $\ell_i$ : location
**DISCRETIZATION & BAG OF WORDS**

Word definition:

\[ w_i = (\ell, v, d) \in \mathbb{N}^3 \]

location, velocity, direction

\[ T_k = \{c, w\}, \quad w \in \mathbb{N}^Z \]

Frequency normalization:

\[ w_i \Rightarrow w_i^f = \frac{w_i}{\sum_j w_j} \in \mathbb{R}_+ \]

\[ T_k = \{c, w^f\} \]

\( S \times 6 \text{ velocites} \times 8 \text{ directions} \Rightarrow \text{Fixed dimensions } Z \)

\( S = 30 \times 30 \Rightarrow Z = 43200 \)
**DISCRETIZATION & BAG OF WORDS**

**Word definition:**

\[ w_i = (\ell, v, d) \in \mathbb{N}^3 \]

\( \ell \) - location, \( v \) - velocity, \( d \) - direction

**Frequency normalization:**

\[ w_i \Rightarrow w_{fi} = \frac{w_i}{\sum_{j} w_j} \in \mathbb{R}^+ \]

**Trajectory:**

\[ T_k = \{ c, w^J \} \]

**Grid discretization:**

\[ S = 30 \times 30 \Rightarrow Z = 43200 \]
**Naive Bayes Modeling**

**Multinomial model:**

\[
\Theta = \begin{bmatrix}
\vdots \\
\theta_i = p(w_i|\ell) \\
\vdots 
\end{bmatrix} \in \mathbb{R}^Z
\]

\[
p(w_i|\ell) = \frac{\sum_k w_i^{(k)}}{\sum_k \sum_{\{j|\ell \in w_j\}} w_j^{(k)}}
\]

\[Z = 43200\]
Introduction

Representation

Query

Traj. analysis

Conclusion

Naive Bayes Modeling

Multinomial model:

\[
Z = 43200
\]

\[
\Theta = \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\vdots \\
\theta_n
\end{bmatrix}
\]

\[
\theta_i = p(w_i|\ell)
\]

\[
p(w_i|\ell) = \sum_k w(k)_i \sum_j \{ \ell \in w_j \} w(j)_k
\]

Vincent Guigue
**Entropy Issue**: A normalization is required

Parking (green)  
High entropy

Runway (yellow)  
Low entropy

Local normalization procedure:

\[ \theta_i = \frac{p(w_i|\ell)}{\max\{p(w_i|\ell)\}_{i|\ell \in w_i^k}} \]

\[ \theta_i = p(w_i|\ell) \]

⇒

locally norm. likelihood
**LOCAL BEHAVIOR DESCRIPTIONS**

- **Yellow**
- **Blue**
- **Magenta**
- **Cyan**
- **Green**
- **Red**
LOCAL BEHAVIOR DESCRIPTIONS

Spatial characterization:
Simple framework:

- **Query**: 1 trajectory
- **Answers**: \( k (=3) \) Nearest Neighbors (Euclidian distance)
Simple framework:
- Query: 1 trajectory
- Answers: $k(=3)$ Nearest Neighbors (Euclidean distance)

Smart query:
- Query = region $\ell$ (all veloc./dir.)
- Sorted answers: 4 Lowest likelihood

Query in the original representation space

Query in representation space + likelihood
CONSISTENCY OF THE REPRESENTATIONS

1 dot = 1 (take-off) trajectory

- Unsupervised learning... difficult to evaluate
- Colors = airport configurations
  - 4 runways
  - East or west direction

⇒ Clear latent space division

T-SNE projection (2D)
Introduction

Representation

Query

Traj. analysis

Conclusion

CONSISTENCY OF THE REPRESENTATIONS

1 dot = 1 (take-off) trajectory

T-SNE projection (2D)

Fine analysis of the magenta cluster:

○ left sub-cluster
○ right sub-cluster
Protocol:

1. Clustering of the parkings

⇒ 10 clusters
Protocol:

1. Clustering of the parkings
2. Taxiing duration pdf estimate

Raw estimate

Smoothed estimate (Parzen)
Protocol:

1. Clustering of the parkings
2. Taxiing duration pdf estimate
3. Late = last percentile

Smoothed estimate (Parzen) + last percentiles of each cluster
Circled dot = late trajectory

We detect some regularities in late trajectories

Outliers (often) correspond to late trajectories

T-SNE projection (2D)
(Re-)introducing **time** in the analysis:

Trajectory = series of **words** \( \Rightarrow \) series of **likelihoods**

\[
T = \{w_{t_1}, \ldots, w_{t_{|T|}}\} \Rightarrow \{\mathcal{L}(w_{t_1}), \ldots, \mathcal{L}(w_{t_{|T|}})\}
\]

Likelihood course of a late trajectory:
The plane had an abnormal low velocity in 3 spatial tiles of the grid.
Finding trajectories with:

anomaly in the region $\ell$
& velocity $> \text{ML velocity}$
Conclusion

- **Very light** way to index trajectories
- Consistent
- (Local) **likelihood**
- Many possible coding (presence, frequency, tf-idf...)

*inspired from text indexing*

Perspectives

- Indexing \(\Rightarrow\) categorization with **continuous modeling** (neural network)
- Identifying **precursory events** of abnormal situations
- Trajectory \(\Rightarrow\) **Situation** (multiple vehicles)

*bigram?*

Many thanks to AWACS FUI Grant partners: **SafetyLine, IFSTTAR, ADP**

Vincent Guigue