Short-term 4D Trajectory Prediction Using Machine Learning Methods

Zhengyi WANG
Man LIANG
Daniel DELAHAYE
ENAC, 31055, Toulouse, France
4D trajectory prediction

❖ Influence factors:
  • Aircraft weight
  • Pilot actions
  • Wind and temperature
  • etc.

❖ Two categories of prediction:
  • short-term: in minutes
  • long-term: in hours

uncertainty  accuracy  efficiency
Previous works

4D trajectory prediction model

Aircraft performance model
- Physics-based approach
- Deterministic model
  - aircraft motion: BADA
  - planned flight routes
  - predicted atmosphere condition
  - Aircraft Intent

Machine learning model
- Data-driven approach
- Learning model
  - input data: surveillance data and meteorological data
  - learning algorithm
  - data-mining
Objective

❖ We propose a novel short-term trajectory prediction model in Terminal Maneuvering Area (TMA) to predict Estimated Time of Arrival (ETA).

❖ This model can be divided into two main parts:

1. Preprocessing part
   ❖ to identify the 4D trajectories into different clusters and remove the noises in an efficient way.

2. Machine learning part
   ❖ to apply Multi-Cells Neural Networks (MCNN) technique to process different traffic data prediction.
Methodology

Preprocessing
- data clearing
- data clustering

Machine learning:
- parallel processing
- Neural Network

Cell 1
Cell 2
Cell 3
Cell n

4D trajectory
- Resampling and data augmentation
- Formatted 4D trajectory
- Dimensionality reduction by PCA
- Vectors containing principle components values
- Clustering by DBSCAN
- 1-st cluster of trajectories
- n-th cluster of trajectories
- Noise
- Preprocessing

Machine learning:
- 1-st learning model
- 2-st learning model
- 3-rd learning model
- n-th learning model
- Input trajectory data
- Classification by entry point

Output
The available dataset includes ADS-B records in July, 2017 over the TMA of Beijing Capital International Airport (BCIA).

Each record of data contains:
1. Type of operation (departure/arrival)
2. Record beginning time
3. Aircraft number
4. Position(X,Y,Z)
5. Heading
6. Horizontal velocity
7. Vertical velocity

Each record with the same aircraft number belongs to an aircraft i, and the collection of all records for that aircraft forms the trajectory.

We select the dataset with 36288 flights and 3242384 trajectory points.
Runways 18R/36L and 18L/36R traffic samples in Beijing Capital International Airport
Clustering-based preprocessing

❖ Four steps:

1) **Data cleaning and formatting**: trajectories less than 50 points are eliminated. All trajectories are resampled to same length for clustering.

2) **Dimensionality augmentation**: to improve the clustering performance. The final features are: position \((X, Y, Z)\), heading \(\Psi\), distance from the reference point \(R\), distance from the corner point \(D\), angular position from the reference point \(\Theta\).

3) **Principal component analysis**: to reduce redundant elements which may decrease computational efficiency, even lead to larger errors.

4) **Clustering via DBSCAN***: to divide the dataset into several clusters and noises.

*DBSCAN: Density-Based Spatial Clustering of Applications with Noise
Learning model

Objective: Estimated Time of Arrival (ETA) prediction

Parallel learning

Preprocessing:
- clearing
- clustering

Machine learning:
- parallel processing
- Neural Network

In each cell: Neural network

Supervised learning

Learning algorithm:
- Training data belongs to 1-st cluster
- Input trajectory data
- Output
- Dimensionality reduction
- Noise
- Clustering by DBSCAN
- Resampling and data augmentation
- Preprocessing

Objective: Estimated Time of Arrival (ETA) prediction

Supervised learning:
- Learning model
- Learning algorithm
- Neural network

Machine learning:
- parallel processing
- Neural Network

Input layer
- X-coordinate
- Y-coordinate
- Z-coordinate
- Distance from the reference point
- Distance from the corner point
- Sine value of angular position from the reference point
- Sine value of heading
- Cosine value of heading
- Airspeed
- Vertical speed

Hidden layer
- Predicted ETA

Output layer
- Input node
- Hidden node
- Output node
- Bias
Validation of learning algorithm

- Nested cross validation
  - The **inner** loop, for parameters selection, such as learning rate, number of hidden nodes;
  - The **outer** loop, to validate the robustness of our prediction model.

- The proportion of training sets, validation sets and test sets is 64% / 16% / 20%.
Experiment and result

We use dataset containing 8677 arrival flights of QFU 36.

Result 1: Principal component
Experiment and result

Result 2: Clustering with $\varepsilon = 1.8$ and MinPts = 200

($\varepsilon$ affects the size of cluster, MinPts affects the noise identification)
Result 3: Noise results in clustering with $\varepsilon = 1.8$ and MinPts = 200

($\varepsilon$ affects the size of cluster, MinPts affects the noise identification)
Result 4: Clustered partitions for each iteration in outer loop
We use **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** to evaluate our trajectory prediction model performance. We compare our NN prediction model with **Multiple Linear Regression (MLR)** model which is the most common form of regression analysis, frequently applied to prediction.

**Result 5: Performance on ETA prediction of NN and MLR with preprocessing step**

<table>
<thead>
<tr>
<th>Partition number</th>
<th>percentage</th>
<th>MAE for NN+P. (s)</th>
<th>RMSE for NN+P. (s)</th>
<th>MAE for MLR+P. (s)</th>
<th>RMSE for MLR+P. (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>13.85%</td>
<td>106.08</td>
<td>141.51</td>
<td>113.67</td>
<td>150.20</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>5.62%</td>
<td>82.91</td>
<td>108.08</td>
<td>92.99</td>
<td>118.59</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>58.39%</td>
<td>61.68</td>
<td>97.81</td>
<td>82.48</td>
<td>117.14</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>13.64%</td>
<td>46.00</td>
<td>69.39</td>
<td>51.09</td>
<td>75.12</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>8.51%</td>
<td>88.76</td>
<td>124.31</td>
<td>97.42</td>
<td>132.62</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>69.19</td>
<td>104.82</td>
<td>84.37</td>
<td>119.13</td>
</tr>
</tbody>
</table>

**Result 6: Performance on ETA prediction of NN and MLR without preprocessing step**

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (s)</th>
<th>RMSE (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR without P.</td>
<td>108.03</td>
<td>160.40</td>
</tr>
<tr>
<td>NN without P.</td>
<td>76.28</td>
<td>127.76</td>
</tr>
</tbody>
</table>
Result 7: The distribution of ETA prediction errors with different methods.

- **(a) NN without Preprocessing**
- **(b) MLR without Preprocessing**
- **(c) NN with Preprocessing**
- **(d) MLR with Preprocessing**

**Y**: Frequency, which presents the percentage of trajectories on the associated error.

**X**: Prediction error
Experiment and result

**Result 8:** Mean absolute error of ETA prediction with the time to destination

<table>
<thead>
<tr>
<th>Time to destination (s)</th>
<th>Absolute error (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NN with preprocessing</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td>400</td>
<td>100</td>
</tr>
<tr>
<td>600</td>
<td>150</td>
</tr>
<tr>
<td>800</td>
<td>200</td>
</tr>
<tr>
<td>1000</td>
<td>250</td>
</tr>
<tr>
<td>1200</td>
<td>300</td>
</tr>
<tr>
<td>1400</td>
<td>350</td>
</tr>
<tr>
<td>1600</td>
<td>400</td>
</tr>
<tr>
<td>1800</td>
<td>450</td>
</tr>
<tr>
<td>2000</td>
<td>500</td>
</tr>
</tbody>
</table>

Graph showing the comparison of mean absolute error for different models with and without preprocessing.
Conclusion

1. We propose a novel hybrid 4D trajectory prediction model based on clustering and MCNN.

2. Our proposed model is robust. The preprocessing part can provide high-quality inputs to the prediction part.

3. Our proposed model can improve the accuracy of prediction, compared with Multiple Linear Regression (MLR) model.
Future work could be conducted in different look-ahead times, on a comparison with results from model-based methods, as well as on studying prediction accuracy for other trajectory variables besides ETA. Moreover, more complex prediction model, such as deep learning approaches, would be very valuable.
End of presentation

Thanks for your attentions!

Questions?
cauc_wzy@163.com
man.liang@enac.fr
daniel@recherche.enac.fr