DeepETA: A Spatial-temporal Sequential Neural Network Model for Estimating Time of Arrival in Package Delivery System

Motivation

- Over 100 million packages are delivered each day in China
- Estimating time of arrival (ETA) of packages at any time during delivery
  - reducing the anxiety of customers
  - improving the customer experience
  - measuring the service ability and quality of couriers

Challenge

Delivery route

- Multiple destinations
  - Delivery sequence and locations
- Time-variant delivery status
  - The latest route, the delivery pattern
- Time-invariant delivery features
  - Geographical and package properties

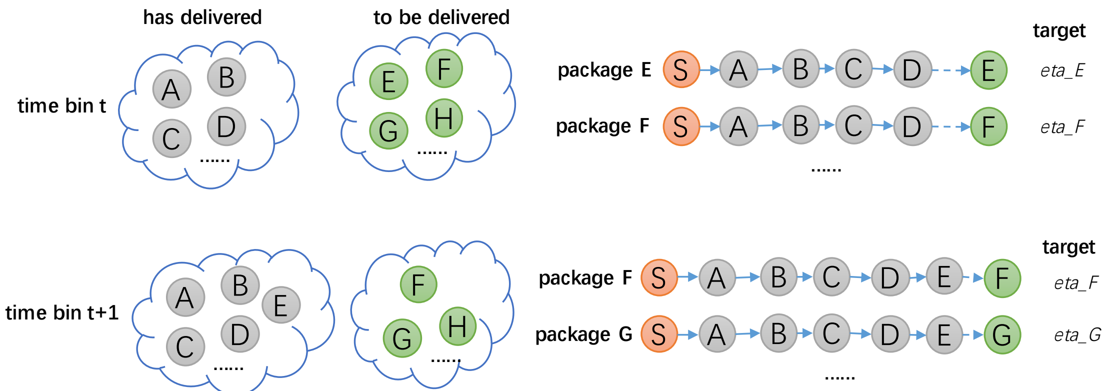
Problem Definition

- The latest route $Route^{d}_{a_s,a_c}$ and the predicted location a_i
- Relative routes from history that are similar with the current route symbolized as $\mathcal{H}_{route}$
- Undelivered location set $Set^{d}_{t_i}$ and all similar sets from history as $\mathcal{F}_{set}$
- Find the most possible delivery time dt_{a_i} from historical relative routes:

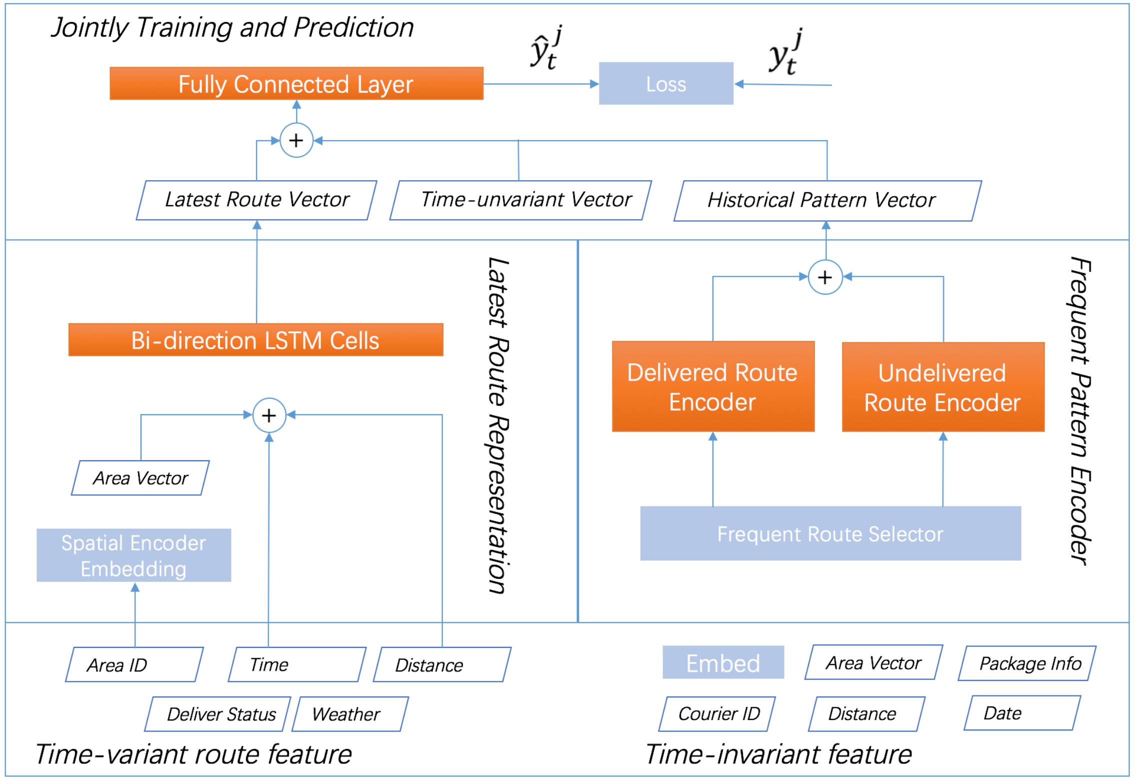
$$dt_{a_i} = f(\mathcal{H}_{route}, \mathcal{F}_{set})$$

Methology

- End to end structure
- Latest route encoder
- The frequent pattern encoder
- The prediction module

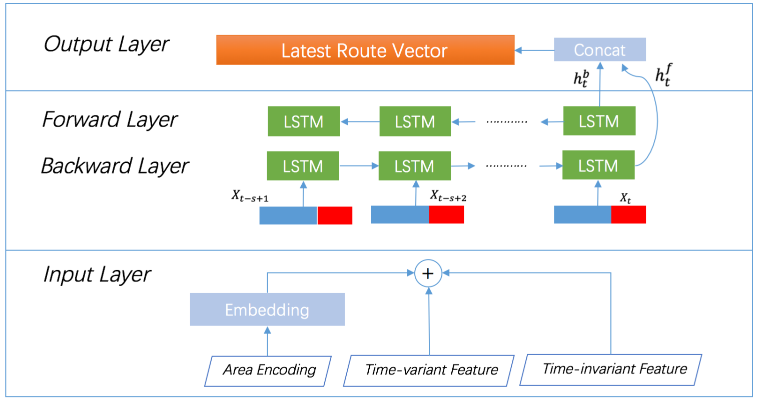


Delivery route



Model structure

Latest Route Encoder



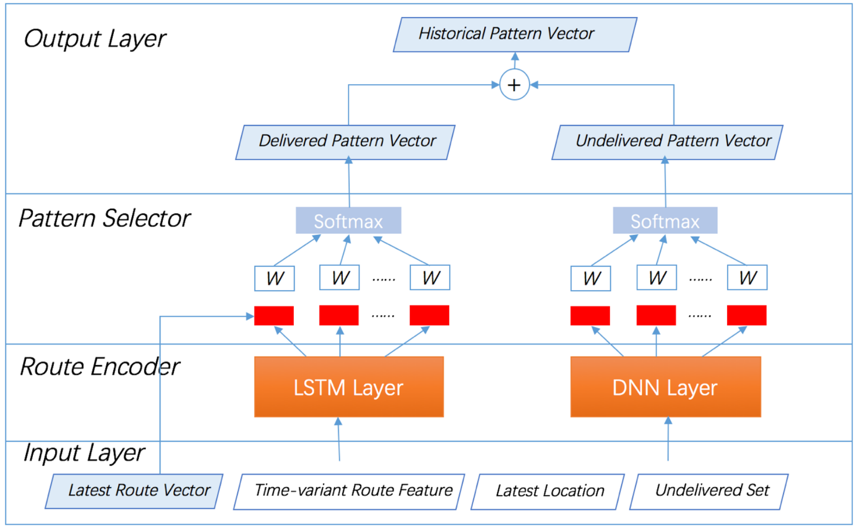
- Spatial encoder
  - Location a_i is transformed into Geohash encoding, namely G_{a_i}

$$V_{a_i} = f\left(W_a G_{a_i} + b_a\right)$$

- BiLSTM
  - Spatial encoder V_{a_i}, time-variant features V_{tv} and time-invariant features V_{ti} as $X_t = [V_{a_i}, V_{tv}, V_{ti}]$

$$h_t = LSTM(h_{t-1}, X_t)$$

Frequent Pattern Encoder



- Route encoder
  - $\tilde{\mathcal{H}}_t^k = \mathrm{LSTM}\left(\tilde{\mathcal{H}}_{t-1}^k, \mathcal{H}_{route}\right)$ and $\tilde{\mathcal{F}}_t^k = DNN\left(\tilde{\mathcal{F}}_{t-1}^k, \mathcal{F}_{set}\right)$
- Pattern selector
  - An attention-based layer
  - $f\left(\tilde{\mathcal{H}}_t^k, h_t\right) = tanh\left(\tilde{\mathcal{H}}_t^k W_{score} h_t\right)$, $\sigma(z)_k = \frac{e^{z_k}}{\sum_{i=1}^{K} e^{z_i}}$ and $V_{fp} = \sum\left(\sigma(z)_j \tilde{\mathcal{H}}_t^j\right)$

Jointly Training and Prediction

- Loss function
  - The fully connected layer: $\tilde{y}_t = \sigma\left(W_{fc}\left[h_t, V_{fp}, V_{ti}\right] + b_{fc}\right)$

$$\mathcal{L}(\theta) = \sum_{i=1}^{N}\left((\tilde{y}_t - y_t)^2 + \lambda\left(\frac{\tilde{y}_t - y_t}{y_t}\right)^2\right)$$

Experiment

Methods	RMSE (min)	MAPE (%)
LR	144.18	43.5
DNN	127.58	37.4
XGBoost	123.66	38.2
LSTM	110.37	34.9
DeepTTE	97.85	29.7
DeepMove	72.43	24.3
DeepETA	**63.58**	**20.6**

- Real world delivery dataset
- Evaluation metric

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_t^j - y_t^j\right|}{y_t^j}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_t^j - y_t^j\right)^2}$$

- DeepETA achieves the lowest RMSE (63.58 minutes) and the lowest MAPE which improves the best performance of the baseline methods by 13.8% (RMSE) and 16.5% (MAPE)