Supplementary Material

APPENDIX 1. K-MEANS CLUSTERING ALGORITHM

The pseudocode of the K-means clustering algorithm is as follows:

Algorithm A1 K-means Clustering Algorithm

Input: Data set $\mathscr{Z} = \{\boldsymbol{z}_i : 1 \leq i \leq N\}$, number of clusters K, maximum iterations T.
Output: Clustering result $\sigma^t(\mathscr{Z}) = \langle \mathscr{Z}_1^t, \mathscr{Z}_2^t, \cdots, \mathscr{Z}_k^t, \cdots, \mathscr{Z}_K^t \rangle$.

procedure KMEANS($\mathscr{Z}, K, T$)
  Randomly initialize K centroids $\{\boldsymbol{c}_1^1, \boldsymbol{c}_2^1, \cdots, \boldsymbol{c}_K^1\}$;
  $t \leftarrow 1$;
  $C(\sigma^0) \leftarrow \infty$;
  do
    $\sigma^t(\mathscr{Z}) \leftarrow \langle \emptyset, \emptyset, \cdots, \emptyset \rangle$;
    $C(\sigma^t) \leftarrow 0$;
    for $i \in \langle 1, 2, \cdots, N \rangle$ **do**
      for $k \in \langle 1, 2, \cdots, K \rangle$ **do**
        $\rho_{ik}^t \leftarrow \left\| \boldsymbol{z}_i - \boldsymbol{c}_k^t \right\|_2$;
      end for
      $k \leftarrow \arg \min_{1 \leq j \leq K} d_{ij}^t$;
      $\mathscr{Z}_k^t \leftarrow \mathscr{Z}_k^t \cup \{\boldsymbol{z}_i\}$;
      $C(\sigma^t) \leftarrow C(\sigma^t) + \min_{1 \leq j \leq K} d_{ij}^t$;
    end for
    for $k \in \langle 1, 2, \cdots, K \rangle$ **do**
      $n_k \leftarrow \left| \mathscr{Z}_k^t \right|$;
      $\boldsymbol{c}_k^{t+1} \leftarrow \sum_{\boldsymbol{z} \in \mathscr{Z}_k} \boldsymbol{z} / n_k$;
    end for
    $t \leftarrow t + 1$;
  while ($C(\sigma^t) \neq C(\sigma^{t-1})$ and $t \leq T$)
  return $\sigma^t(\mathscr{Z})$;
end procedure

APPENDIX 2. FLIGHT SCHEDULE DATASET

The dataset is accessible at www.variflight.com. VariFlight is Chinese largest aviation data service provider. These data have 11 valid attributes, which are illustrated in **TABLE** B1.

To ensure the accuracy and validity of the data, the method of processing missing values, abnormal values, or duplicate data is to delete the samples containing missing data. After the data preconditioning stage, there are $4,056,736$ valid data samples left for further analysis, as shown in the data sample in **TABLE** B2.

TABLE B1 VariFlight flight schedule data description

Variable type	Variable name
Time information	Departure time information: scheduled departure time, actual departure time
	Arrival time information: scheduled arrival time, actual arrival time
Delay information	Departure/Arrival delay duration (minute)
Airport information	Departure airport information: IATA airport code
	Arrival airport information: IATA airport code
	Airport location information: airport longitude, airport latitude, city

TABLE B2 VariFlight flight schedule representation example data

No.	forg	fdst	sche_dep	sche_arr	act_dep	act_arr
0	SJW	KMG	2018-01-01 05:55:00	2018-01-01 08:55:00	2018-01-01 05:56:00	2018-01-01 08:55:00
1	HGH	FUG	2018-01-01 06:00:00	2018-01-01 07:00:00	2018-01-01 06:01:00	2018-01-01 06:58:00
2	PEK	WUH	2018-01-01 06:00:00	2018-01-01 08:25:00	2018-01-01 06:01:00	2018-01-01 07:48:00
3	PEK	XIY	2018-01-01 05:50:00	2018-01-01 08:05:00	2018-01-01 06:03:00	2018-01-01 07:50:00
4	SYX	CAN	2018-01-01 06:15:00	2018-01-01 07:45:00	2018-01-01 06:04:00	2018-01-01 07:07:00
5	CKG	LHW	2018-01-01 06:10:00	2018-01-01 08:15:00	2018-01-01 06:05:00	2018-01-01 07:34:00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
4056731	LJG	CKG	2018-12-31 23:55:00	2019-01-01 01:25:00	2019-01-01 01:30:00	2019-01-01 02:39:00
4056732	KHN	TNA	2018-12-31 21:00:00	2018-12-31 22:45:00	2019-01-01 01:46:00	2019-01-01 03:11:00
4056733	HAK	HGH	2018-12-31 22:45:00	2019-01-01 01:20:00	2019-01-01 01:54:00	2019-01-01 03:49:00
4056734	CGO	KWE	2018-12-31 22:55:00	2019-01-01 01:20:00	2019-01-01 02:21:00	2019-01-01 04:16:00
4056735	KHN	TAO	2018-12-31 23:50:00	2019-01-01 01:45:00	2019-01-01 02:26:00	2019-01-01 03:42:00