New data mining technique for multidimensional aircraft trajectories analysis

Maria Solntseva-Chaley[1,*]

[1]MIPT, Moscow, Russia

Abstract. Under conditions of growing airport workload, airspace sectorization is necessary for accidents prevention. Sectorization should be performed taking into account regular traffic of aircrafts. A new data mining technique, solving the problem, is described. It allows to fulfill stable partition of aircraft intent trajectory sample into the groups (asymptotically converged beams), corresponding to the same runway approaches. Method is taking into account special geometric characteristics (curvature, torsion and multiple intersections) of multidimensional space trajectories of aircrafts.

1 Introduction

Currently, due to increasing data volume, to design methods and tools for big data rapid and automatic processing becomes a problem of great importance. New approaches are required for large data sets analysis (like unsupervised data mining or other machine learning techniques), that can recognize hidden patterns of motion and identify moving objects with similar characteristics and/or the same final targets.

In many areas, particularly, in aviation, it is necessary to process huge data sets of trajectories for monitoring, control or other purposes. Aircraft trajectory is represented by four-dimensional description of the aircraft's states with time, where state may include a position of aircraft center of mass and other characteristics of motion, like velocity, attitude and weight. So, in general, space trajectory is a multidimensional description of moving object path.

According to statistics, large number of aviation accidents takes place in extended airport area due to increased work load of air traffic management (ATM) systems. So, while air traffic is permanently increased, modernization and optimization of ATM-systems are needed to maintain high safety level.

Data mining techniques are widely used in the investigations on work load optimization in extended airport area, in particular, for air space sectorization. Sectorization is space partitioning into the sectors or zones of different controllers responsibility.

Air space sectorization should be done taking into account regular flows that are described by the samples of aircraft trajectories.

---

[*] Corresponding author: solnceva.chalei@gmail.com

Up to date, sectorization problem is solved by clustering [1] 2D-projection of aircraft 4D-trajectories, or with the help of space partitioning into the regions like Voronoi-diagrams [2].

Development of 3D-sectorization algorithm is at initial stage and associated with the problem of trajectory beams revealing in 3D-space. Trajectory beams are corresponding to the groups of 3D-trajectories with similar characteristics. For partitioning of trajectory sample into the beams such methods as PCA [3], nonparametric approach based on Dynamic Bayesian Networks [4] and spectral clustering [5] are used. The methods are based on reduction of analyzed data space dimensions.

Complexity of analyzed data space geometry which has specific curvature, torsion and multiple space intersections is demonstrated in Fig.1a. Example of two intent trajectory beams on different runways is shown in Fig.1b.

In the present work the details of a new approach to airport space sectorization which has been proposed earlier [6] are described. Current technique is based on determination of geometric asymptote tangent to multidimensional trajectory beam in the lower dimension space. Then trajectory beam is separated in the result of reverse transition to the original data space. As it will be shown further, such approach, while being applied to the analyses of radar data, allows getting stable results.
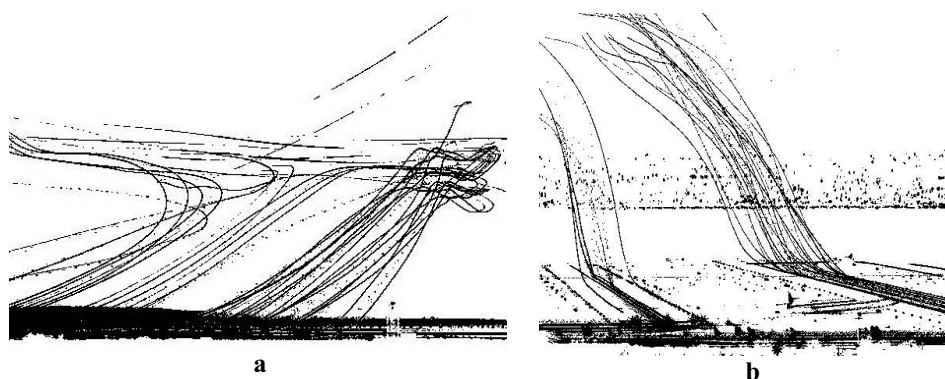

a   b

Fig.1 a) Intent aircraft trajectories are shown. b) Example of intent trajectory beams is presented

2 Description of analysed data

Fig. 2a illustrates analyzed sample of intent trajectories. The sample contains 116 trajectories of the aircrafts landed at the airports of San Francisco bay area 1 January of 2005 year (data is freely available at https://c3.nasa.gov/dashlink/resources/132/). Analyzed data were recorded by TRACON radar system. Reference point coincides with the radar location. Time interval between trajectory points is about 5 sec. Each trajectory consists of 160 points, as it was described in [6, 7].
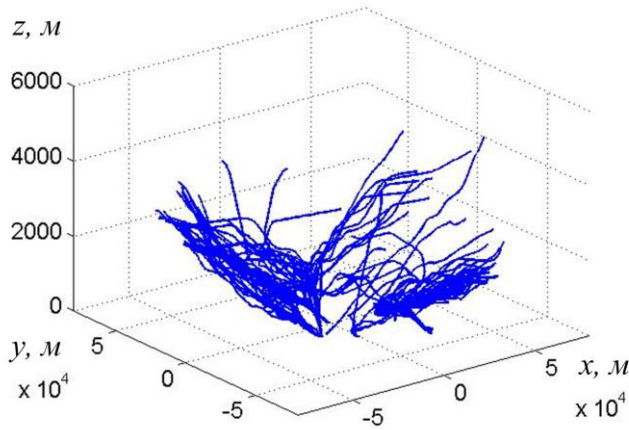
Fig.2 Analyzed sample of 4D aircraft intent trajectories is shown in 3D-space

3 Methods

3.1 Assessment of beam's asymptote

Trajectory beam $N_k, k = \overline{1, K_0}$ (K_0 is empirical parameter) is considered as asymptotically converged with a threshold parameter ε, if for all vectors $\left\{ \mathbf{x}[i] \in R^{3 \times L}, i \in N_k \right\}$ in the beam N_k a condition of asymptotic convergence of the beam is fulfilled

$$\forall (i, j) \in N_k, \quad \left\| \mathbf{x}[L_i; i] - \mathbf{x}[L_j; j] \right\|_2 < \varepsilon, \tag{1}$$

where $\forall i \in N_k$ $\mathbf{x}[L_i; i]$ are coordinates of final trajectory points on a runway. Parameters $L_i, i \in N_k$ are subject to determination, $\|.\|_2$ is Euclidean distance metric in three-dimensional coordinate space R^3, ε is a cutoff parameter with value of no more than runway width.

In considering the proposed approach to determination of number of converged beams, it should be taken in to account that the trajectories in asymptotically converged beams have some typical form (profile) and specific geometric asymptote in the region of convergence (1). Geometric asymptote in converged beam of multidimensional aircraft intent trajectories is a line in R^3 that meets the requirement (1). Trajectories in asymptotically converged beam have tangent line in area around their final points $\forall i \in N_k$, $\mathbf{x}[L_i; i]$. So, asymptotically converged trajectory beams can be identified by determination of their tangential geometric asymptotes in the points of their focuses.

Because discrete points of beam trajectories are tightly located around its' asymptote, to determine number of asymptotically converged beams sample of trajectory vectors $\left\{ \mathbf{x}[i] \in R^{3 \times L}, i = \overline{1, N} \right\}$ is scattered into set of the trajectories' points

$$\left\{ \mathbf{x}[i] \in R^{3 \times L}, i = \overline{1, N} \right\} \Rightarrow \left\{ (x[j, i], y[j, i], z[j, i]) \in R^3, j = \overline{1, L}, i = \overline{1, N} \right\}. \tag{2}$$

Set of the points (2) has to be sorted according to the values of one of the coordinates (in ascending or descending order). At that, other coordinates of points, representing

converged trajectory beam of certain profile, are also ordered. Then, for the scattered three-dimensional data $\left\{ \mathbf{z}_i = (x_i, y_i, z_i), i = \overline{1, L \cdot N} \right\}$ (2) orthogonal linear regression models

$$M\ (\boldsymbol{\theta}) = \left\{ (x, y, z) \in \Box^3 \middle| (a_1 x + b_1 y + c_1 z = d_1) \wedge (a_2 x + b_2 y + c_2 z = d_2) \right\} \qquad (3)$$

are analyzed by the RANSAC (*Random Sample and Consensus*) algorithm [8]. Here symbol $\wedge$ is conjunction, $\boldsymbol{\theta} = \left\{ a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2 \right\}$ is vector of the parameters of the models (3), that is determined under given cutoff of Euclidian distance $\rho_\perp \big(\mathbf{z}, M\ (\boldsymbol{\theta}) \big)$ calculated by orthogonal projection of point $\mathbf{z} = (x, y, z)$ from the set (2) onto a line $M\ (\boldsymbol{\theta})$. By such a way the model (3) is symmetrical relatively to the coordinates x, y, z. Any pair of the points from (2) is sufficiently to put forward a hypothesis about the model of orthogonal linear regression (3). Final model (3) is proved by the greatest relative quantity (percent) of scattered data $\left\{ \mathbf{z}_i = (x_i, y_i, z_i), i = \overline{1, L \cdot N} \right\}$ (2). Algorithm MLESAC (*Maximum Likelihood Estimation Sample Consensus*) [9], that is probabilistic modification of RANSAC [8] algorithm, may be used for these purposes. The algorithm estimates likelihood of the model (3), in representing distance distribution of scattered data $\left\{ \mathbf{z}_i = (x_i, y_i, z_i), i = \overline{1, L \cdot N} \right\}$ from the model $M\ (\boldsymbol{\theta})$ (3) as a mixture of data distributions some of which support the model (3) (inliers), while the rest ones reject it (outlier). Considering the scattered data Z (2) as independent, we obtain a relation for logarithm of likelihood as following

$$
\begin{aligned}
L\big(\rho_\perp \big(Z, M\ (\boldsymbol{\theta}) \big) \big| \boldsymbol{\theta} \big) = \\
= \sum_{i=1}^{L \cdot N} \log \Big(\gamma p \big(\rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \big| \mathbf{z}_i \text{ is inlier} \big) + (1 - \gamma)\, p \big(\rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \big| \mathbf{z}_i \text{ is outlier} \big) \Big),
\end{aligned}
\qquad (4)
$$

where γ is mixing parameter. Distribution of the distances to the data, supporting the model (3), is represented by Gaussian distribution

$$p \big(\rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \big| \mathbf{z}_i \text{ is inlier} \big) \propto \exp \left(- \frac{\big(\rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \big)^2}{2 \sigma^2} \right), \qquad (5)$$

where σ is standard deviation. Distribution of distances to the data, rejecting the model (3), is described by uniform distribution

$$p \big(\rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \big| \mathbf{z}_i \text{ is outlier} \big) = \begin{cases} (2\rho_{max})^{-1}, \rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) < \rho_{max} \\ 0, \ \rho_\perp \big(\mathbf{z}_i, M\ (\boldsymbol{\theta}) \big) \geq \rho_{max} \end{cases}, \quad (6)$$

where ρ_{max} is maximal distance to data (it is defined by the context). Minimization of likelihood logarithm (4) allows to estimate vector of the parameters $\boldsymbol{\theta}$ and mixing parameter γ (see Eq. (21) in [9]). Estimation of the parameters is traditionally done using EM-algorithm [10]. Example of the algorithm practical realization in Matlab can be found in [11].

The most likely linear regression of the scattered data of trajectory sample defines geometric asymptote $M\ (\boldsymbol{\theta})[k], k = \overline{1, K}$ (3) of one of the sample beams under condition (1). Geometric asymptote obtained in such a manner meets the requirement

$$\forall k = \overline{1, K}, \ \forall i \in \mathrm{N}_k, \ \big\| M\ (\boldsymbol{\theta})[L_k, k] - \mathbf{x}[L_i; i] \big\|_2 < \varepsilon .$$

3.2 Separation of trajectory beam

Beams of the trajectories tangential to the corresponding geometric asymptotes are defined in the result of minimization of a cost function

$$J = \arg\min_{\{M\ (\boldsymbol{\theta})[k], k=\overline{1,K}\}} \sum_{i=1}^{N} \sum_{k=1}^{K} r[i;k]\rho_{\cosin}^2\left(\mathbf{x}[i], M\ (\boldsymbol{\theta})[k]\right),$$

where $\left\{r[i;k]\in\{0,1\}, k=\overline{1,K}\right\}, i=\overline{1,N}$ is a set of binary indicator variables (e.g., if vector $\mathbf{x}[i]$ was attributed to the beam k, then $r[i;k]=1$ and $r[i;k]=0$, otherwise). Distance between each geometric asymptote and sample's trajectories is calculated according to cosine measure

$$\rho_{\cosin}\left(\mathbf{x}, M\ (\boldsymbol{\theta})[k]\right) = \left(\mathbf{x}\cdot M\ (\boldsymbol{\theta})[k]\right)\Big/\left(\sqrt{(\mathbf{x}\cdot\mathbf{x})}\sqrt{(M\ (\boldsymbol{\theta})[k]\cdot M\ (\boldsymbol{\theta})[k])}\right).$$

After elimination of the points, representing the trajectories of separated beam, from the scattered data (2), the procedure of geometrical asymptote detection is repeated and the next trajectory beam is separated. In this case, remained scattered data (2) are sorted with respect to another space coordinate (different from previous one), as the model determination should be symmetric relatively to the coordinates x, y, z. The possible dependence of result (3) from coordinate directions is obviated by changing direction of data ordering in (2) from ascending to descending order or vice versa. Analysis of trajectory sample is completed, when all beams in the sample are separated.

In general, the approach described in the present work consists of two stages. Sufficient reduction of data dimensions at the first stage simplifies the revealing of data specific features. In considering 2D-projection of scattered points of 3D trajectories, the most likelihood orthogonal linear regression of the scattered data which corresponds to geometrical asymptote of one of the beams in analyzed sample of the trajectories is defined. At the second stage of the approach, after reverse transferring into the original dimensional space, a certain beam is separated from the analyzed trajectory sample in accordance with proximity (by cosine measure) to defined asymptote. Thus, due to such approach no information about the original data is lost.

4 Demonstration of the results

Earlier [7], sample of multidimensional aircraft intent trajectories was divided into subsamples in 3D-space with the help of method of polynomial regressions [12]. However, the subsamples obtained in the result are not homogeneous, as each subsample consists of several trajectory beams. Besides, the number of beams in the method of polynomial regressions is needed to be specified preliminary and the result obtained is not stable. Method described in the present work is unsupervised, so no knowledge about number of trajectory beams in original sample is necessary. In addition, this method gives stable results.

Fig.3 illustrates the stages of described technique allowing to divide one of the subsample, earlier obtained in [7], into asymptotically converged beams (see Fig.3d). Fig.3a shows two-dimensional projection of scattered data (aligned with time) of analyzed trajectory subsample. Fig.3b illustrates the result of data's linear regression using RANSAC [8] algorithm. The defined asymptote (long blue line) of the first beam is represented in Fig.3c. The beam highlighted in blue in Fig.3d corresponds to the asymptote of the same colour in Fig.3c. It should be noted that in this case the results of

orthogonal linear regression of scattered and full trajectory data coincide. Trajectories of the first beam (see Fig.3d) are separated from the subsample mentioned above according to proximity of the trajectories to the asymptote (see Fig3c) in according with cosine measure.
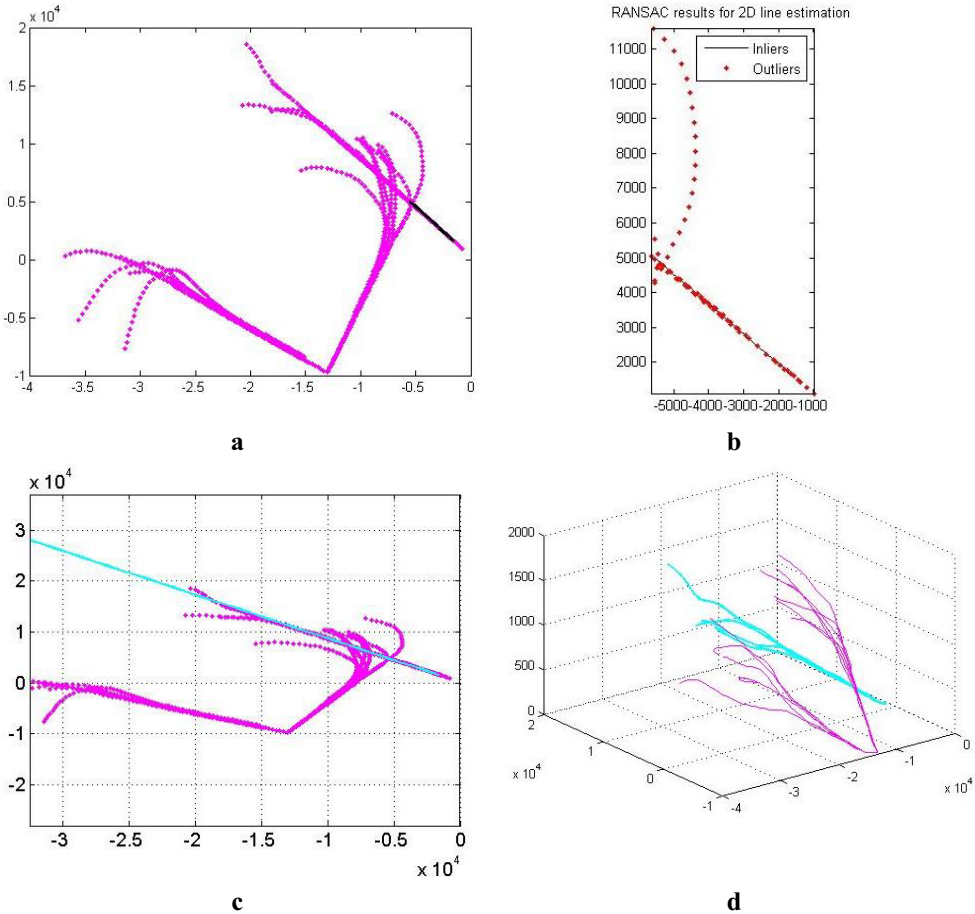

a

b

c

d

Fig.3. Results obtained at different stages of the algorithm are presented. a) Two-dimensional projection of scattered data (aligned with time); b) Linear regression of analysed scattered data obtained with RANSAC [8] algorithm; c) Asymptote (long blue line) of the first separated trajectory beam; d) Three finally separated trajectory beams in original 3D-space. The beam shown in blue corresponds to the asymptote represented in Fig.3c.

Conclusions

New data mining technique for analysis of multidimensional trajectories of landing aircrafts in 3D, described in the work, allows to obtain stable results in partitioning of the analyzed trajectory sample into asymptotically converged beams of the trajectories. Thus, 3D airspace sectorization of extended airport area may be performed.

References

1. A. McFadyen, T. Martin, M. O'Flynn, D. Campbell, *2016 IEEE Aerospace Conference* (IEEE, Big Sky, MT, 2016)
2. M. Sergeeva, D. Delahaye, C. Mancel, L. Zerrouki, N. Schede, *International Conference on Models and Technologies for Intelligent Transportation Systems* (IEEE, Budapest, 2015)
3. M. Gariel, A. Srivastava, E. Feron, *IEEE Transactions on Intelligent Transportation Systems* (IEEE, 2011) **12** (4), 1511-1524.
4. V. Bastani, L. Marcenaro, C. Regazzoni, *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE, 2014), p. 1-6
5. M. Enriquez, C. Kurcz, 5th *International Conferences on Research in Air Transportation (ICRAT)* (FAA, Berkley, Ca., 2012)
6. B. Kukharenko, M. Solntseva-Chalei, Informacionnye tehnologii **22** (2), p.83-89 (Novtech, Moscow, 2016)
7. B. Kukharenko, M. Solntseva, Informacionnye tehnologii (5), p.3-7 (Novtech, Moscow, 2014)
8. M. Fischler, R. Bolles, Communications of the ACM, **24** (6), p. 381-395 (1981)
9. P. Torr, A. Zisserman, Journal of Computer Vision and Image Understanding **78** (1), 138-156 (2000)
10. A. Dempster, N. Laird, and D. Rubin, Journal of the royal statistical society, p. 1-38 (1977)
11. I. Nabney, NETLAB: algorithms for pattern recognition. Springer Science & Business Media (2002).
12. S. Gaffney, P. Smyth, In Advances in neural information processing systems, p. 473-480 (2004)