Future location and Trajectory prediction methods

Harris Georgiou, Petros Petrou, Panagiotis Tampakis, Stylianos Sideridis, Eva Chondrodima, Nikos Pelekis, Yannis Theodoridis

Extended Summary

Mobility analytics is one of the most commonly addressed task in the general context of geolocation data mining, usually involving mobility patterns, mobility graphs, points of interest, hotspot detection, etc. The increasing use of portable devices such as navigation systems and the wide range of location-aware applications have led to a huge amount of mobility data being produced on a daily basis. At the same time, huge fleets of maritime vessels and aircraft can be tracked in real-time throughout the world via dedicated geolocation services and automated transponder devices , so that their movements can be stored and analyzed later on offline. In real-time scenarios where new positions are traced in streaming mode, prediction algorithms evaluate whether the moving object remains on route or deviates, e.g. from the flight plan or the common maritime routes. The algorithms take advantage of offline data analytics results to produce accurate and reliable predictions regarding future movements and events.

This chapter evolves around three main aspects: (a) a thorough but compact formalization of the main components and problems regarding Future Location Prediction (FLP) and Trajectory Prediction (TP); (b) recent developments and representative novel approaches; and (c) related works, comparable approaches and discussion on the results of the proposed methods.

There are two major directions when dealing with the FLP problem: (a) vector-based prediction and (b) pattern-based prediction or the data mining & machine learning approach. The vector-based approaches, inspired by the spatial database management domain, aim to model current locations and perhaps a short history of objects as motion functions, in order to be able to predict future locations by some kind of extrapolation. The pattern-based approaches, inspired by the spatial data mining domain, identify and exploit motion patterns by analyzing historic data of moving objects, i.e., classification models, repetitive patterns, clusters of `similar' movements, etc, based upon a history of movements.

For TP, predictive analytics over mobility data involves applications where moving objects are tracked in order to compute predictions for an entire trajectory. This is preformed either in the short-term, exploiting the immediate state and motion parameters of the moving object (e.g. location, speed, acceleration) to predict the evolution of its movement, or in the long-term, exploiting previous historic data and analytics that produce predictive models for

the entire trajectory (e.g. most probable route and destination). In this context, external factors may be taken into account as data enrichments, for example localized weather conditions or seasonal trends, in order to further refine these models and make them more adaptive.

There are various aspects and tradeoffs in the design of FLP methods, primarily with regard to `memory' versus look-ahead time and, subsequently, simpler/faster versus more complex/demanding algorithms, lower or higher volume/rate of available input data, etc. In this study, two major viewpoints of FLP are presented; namely, one for short-term routes-agnostic (limited `memory') FLP and one for long-term network-aware (increased `memory') FLP.

In the case of short-term FLP, a Recurrent Neural Network (RNN) model is presented as a very robust and efficient algorithmic choice. Over the last decade, Neural Networks (NN) has attracted a renewed research interest, in order to reveal their true power on forecasting aircrafts' locations. Recent literature review has highlighted that a large number of papers employ the power of the special RNN architecture, which have become the state-of-the-art for sequence modeling, due to the fact that they naturally handle temporal and sequential data. Motivated by very recent developments in the current state-of-the-art in mobility analytics, we propose a LSTM-based (Long Short-term Memory) framework for the FLP task. Although various FLP methods have been proposed in the aviation field, limited research has been conducted on employing the power of RNN-based architectures. Contrary to static NN, which are not tailored to work with temporal and sequential data, an RNN is inherently equipped to handle data with time dependency. Aiming at performing effective prediction in a time horizon up to several minutes, the proposed LSTM network is trained using massive historical data from past aircraft movements. This framework provides predictions of aircrafts' future locations by using only running location data, i.e., without requiring any other information such as flight plans. The advantage of using such deep learning approaches is that the motion pattern identification is performed automatically by the LSTM architecture, specifically designed to process the entire trajectory of the moving object as it is generated, in order to predict its future evolution in an arbitrary look-ahead time frame.

On the other hand, if a large volume of historic data is available, analytics over this entire database of trajectories can provide enough insight for long-term FLP models. In this study, a two-stage approach is proposed: (a) pattern extraction, where the mobility trends are identified and modeled over the historic database, and (b) predictive evaluation, where these models are used for long-term FLP. The first stage is a `network discovery' process, where frequent patterns and spatio-temporal trends are identified from the historic data. Thus, it can be performed offline and periodically, whenever the statistics of the database are changed significantly by new data or when seasonal trends are present. In the core of this stage, a robust clustering algorithm is presented in detail, employing spatio-temporal (instead of only spatial) grouping of trajectories and capacity to handle variable-rate input data. The second stage is a `network-based' FLP, exploiting the discovered underlying `network' of frequent routes and mobility patterns, thus enabling the prediction in the long-term context. In practice, this FLP approach described here for the long-term context is inherently intuitive and self-explanatory. It relies on past routes of the same or similar objects in order to forecast how a specific object will move while it is already residing on a specific frequently-traversed route. Moreover, the proposed approach is implemented and evaluated as an online algorithm, i.e., with a distributed architecture and the capacity to handle streaming input data.

The proposed FLP methods are tested with real-world aviation data, evaluating the LSTM-based approach for the short-term and the two-stage clustering/network-based approach for the long-term context. Both experiments illustrate efficient and robust performance, with regard to prediction error against various look-ahead windows, while at the same time provide hints about distributed implementations for streaming data online processing.

The TP problem is slightly different than FLP in the sense that, instead of predicting the immediate evolution of the movement as a sequence of locations, here the entire trajectory of the moving object is to be analyzed or predicted. Additionally, data enrichments, such as localized weather conditions, seasonal trends and the moving object's intrinsic properties (e.g. type, weight, size) are integrated in the models as semantic information.

In this study, the generalized Future Semantic Trajectory Prediction is explored in the context of aviation. More specifically, the pre-flight information that is used as input is the enriched flight plan, which includes a very rough description of the intended route via a limited set of reference waypoints, the predicted localized weather over each one of these, as well as additional semantic data with regard to the specific aircraft type, weight, etc.

Despite the fact that a flight plan is generally considered as a strict guideline for the actual path of the aircraft in civil aviation, surveillance data show that deviations from the reference waypoints are in the range of up to 12-15 km or more, compared to each individual reference waypoint of the submitted flight plan. As a result, the flight plan data are a useful guideline but cannot be used as-is for TP, as the actual route of the aircraft is severely altered by stochastic factors.

In this study, a multi-stage approach is presented for addressing the FSTP problem in the context of aviation. First, the historic data of actual flight routes are analyzed via clustering, similarly to the long-term FLP as described above, but here incorporating all the data enrichments instead of only the spatio-temporal part. Using the clusters' `representatives' (medoids) as guidelines for the most common flight routes between pairs of airports, deviations between these and the corresponding flight plans can be extracted, analyzed and modeled. Since each flight plan is defined by a limited set of reference waypoints and these can be considered `constraints' with regard to the intended flight route (mandatory by law), a per-waypoint approach is designed instead of modeling the entire trajectory.

More specifically, one or more (adjacent) waypoints from the enriched flight plan are used as input in various predictive models, ranging from Hidden Markov Models (HMM) and Linear Regressors (LR) to regression trees (CART) and Neural Network (NN) regressors, trained to predict the deviation from a specific waypoint of the flight plan. In other words, the general flight route is known and described by the flight plan, while the exact deviations from it are predicted by appropriate models trained upon historic data. This method is tested on real-world aviation data and demonstrates almost an order of magnitude better accuracy than the current state-of-the-art in aviation TP, due to the fact that flight plans and their enrichments are exploited to the maximum. Additionally, this multi-stage approach and per-waypoint modeling provides two levels of inherently distributed way to address the FSTP problem: (a) separating the clustering stage that can be performed offline and periodically upon the entire historic database from the actual predicting stage that is lightweight enough to be performed online, and (b) addressing the predictions of the flight plan deviations per-waypoint, i.e., independently for each leg of the flight, thus capable of implementing them naturally for parallel processing.

In summary, this study presents a set of novel multi-stage hybrid approaches for the FLP and TP problems. The FLP task is considered under the short-term / routes-agnostic and the long-term / network-aware variants, while TP is incorporating extensive exploitation of data enrichments and flight plans as constraints in the aviation domain, thus addressing the more generic FSTP task. In FLP, the LSTM-based short-term variant demonstrates how streaming trajectory data can be exploited to train predictive models based on variable-rate `common' mobility patterns that are discovered from the data itself. In contrast, the two-stage clustering/prediction long-term variant demonstrates how a rich set of historic mobility data can be used to model extensive `common' routes in the form of connected nodes, i.e., a routes network, which is subsequently exploited as the base for long-term trajectory predictions for any given moving object within the same region of interest. These two approaches are complementary and can be adapted to various domain-specific tasks in maritime, aviation, etc. Similarly in FSTP, clustering is introduced for grouping together enriched trajectories, using a properly designed semantic-aware similarity function. This enrichment includes localized weather data (e.g. wind speed & direction), aircraft properties (e.g. type, wake category) and other external factors (e.g. weekday). Subsequently, a set of independent predictive models are trained for each cluster, addressing the core task of TP in the context of each reference waypoint of the flight plans. HMM, linear and non-linear regressors were employed as base for the predictive models, exploring the trade-off between having very simple predictors and moderate accuracies versus more complex predictors and higher accuracies. The experimental results proved the feasibility of the proposed FSTP framework in real-world applications, in terms of prediction accuracy versus varying model complexity.

Providing accurate FLP and TP for flights, in the pre-flight phase as well as in-flight, is crucial in maintaining: (a) safe and timely Air Traffic Control (ATC) procedures, (b) prompt scheduling of Air Traffic Management (ATM) procedures with significant improvements of flight delays, and as a result (c) better fuel management and lower emissions. Similarly in the maritime domain, accurate predictions over the routes and mobility patterns in sea vessels is extremely important in maintaining safety, tracking illegal activities (e.g. smuggling), monitoring and preemptive protection of restricted areas (e.g. for fishing), as well as optimizing logistics, time and fuel management in the world-wide scale. The proposed approaches presented in this study provide guidelines and methodologies to these problems, not only in the research but also in the application level, towards big data solutions with distributed streaming online processing.