Linköping University | Department of Computer and Information Science
Master's thesis, 30 ECTS | Datateknik
202019 | LIU-IDA/LITH-EX-A--2019/001--SE

Trajectory based Arrival Time Prediction using Gaussian Processes

-

Trajektoriebaserad ankomsttidsprediktion med Gaussiska Processer

Sebastian Callh

Supervisor : Mattias Tiger
Examiner : Fredrik Heintz

Linköpings universitet
SE–581 83 Linköping
+46 13 28 10 00 , www.liu.se

Abstract

Abstract.tex

Acknowledgments

Acknowledgments.tex

Contents

List of Figures

1 Introduction

As cities grow, efficient public transport systems are becoming increasingly important. To offer a more efficient service, public transport companies use systems that predict arrival times of buses, trains and similar vehicles, and present this information to the general public. The accuracy and reliability of these predictions are paramount, since many people depend on them, and erroneous predictions reflect badly on the public transport companies.

Machine learning algorithms have been applied with great promise to predict arrival times [3, 8, 7], and research is still ongoing. Modern approaches have seen heavy use of Recursive Neural Networks (RNN) to directly model arrival times from the current state of public transport vehicles, an approach which has shown very good prediction accuracy but have major drawbacks:

1. It does not quantify prediction uncertainties.
2. It is severely lacking in explainability.

Knowing the uncertainty of predictions is important to avoid putting too much faith in an ambiguous model, and being able to explain how a model makes its predictions is equally (if not more) important, since it allows reasoning about why a model made a certain prediction, and about how a model would act in previously unseen scenarios. To address the first problem, a generative model can be used instead of RNNs, but solving the second one and achieving a satisfying level of explainability requires a much richer model of the data altogether. One promising approach to this is *trajectory learning*, where the goal is to learn a *motion pattern* (also known as *activity pattern* or *activity model*) which represents a cluster of individually observed trajectories. Trajectory learning is a problem which spans several fields, such as computer vision [6, 15, 4, 1], pattern recognition [10], autonomous vehicles [2], and health informatics [9], and a lot of different approaches have been explored.

This thesis project aims to...

–small pitch on what this thesis project will actually do–

1.1 Aim

–This is subject to change. Currently written to cover most things talked about in the meeting from first day –

The aim of this thesis project is to model motion patterns of public transport vehicles using Gaussian Processes (GPs), and use these models to make arrival time predictions with an accuracy that is competitive to current state of the art models. Furthermore, the project aims to detect specific characteristics from motion patterns, such as “the driver had to emergency break”, or “the vehicles speed was very slow”, to see if certain characteristics are more common in motion patterns where the vehicle is too early or too late, compared to publicly available time schedules. Finally, this project aims to investigate how irregular motion patterns can be detected.

1.2 Research questions

–this is subject to change. Currently written to cover most things talked about in the meeting from first day–

1. How can Gaussian Processes be used to capture the motion patterns of trajectories?
2. How can similar trajectories be clustered into one representative motion pattern?
3. How can specific characteristics be automatically detected in a motion pattern?
4. How can irregular motion patterns be detected?

1.3 Delimitations

This is where the main delimitations are described. For example, this could be that one has focused the study on a specific application domain or target user group. In the normal case, the delimitations need not be justified.

1.4 Report Outline

2 Related Work

This section covers related work done on arrival time prediction and motion pattern modeling using spatio-temporal data. It also covers related work relevant to clustering of trajectory data and to detection of characteristics in motion patterns.

2.1 Arrival Time Prediction

Arrival time prediction is, in a nutshell, the problem of answering the question "When does my bus arrive?". In recent years, machine learning techniques have been very successful at this task, and Long Short-Term Memory Networks (LSTMs) in particular have proven extremely effective. J.pang et al. predicted bus arrival times to the next station in a continuous setting using LSTMs given the current position of a bus and static domain knowledge about its last stop [8]. D. Nguyen et al. also used LSTMs, but with entire trajectories [7]. They predicted arrival times of ships by training a model to generate continuations of trajectories, and when the model was presented with a new unfinished trajectory, it generated a probable continuation until it arrived at a port. This was then used to make its predictions.

While these approaches do perform admirably on test data, they lack explainability and a way to measure the models certainty.

2.2 Motion Pattern Learning

Motion pattern learning is the problem of learning motion patterns from a set of trajectories, such that each pattern captures a different characteristic of the trajectories. An example with synthetic data can be seen in Figure 2.1. Note that the term *Trajectory learning* is often used in the literature for the same problem, however, the term "trajectory" is ambiguous so in the scope of this thesis the name motion pattern learning will be used.

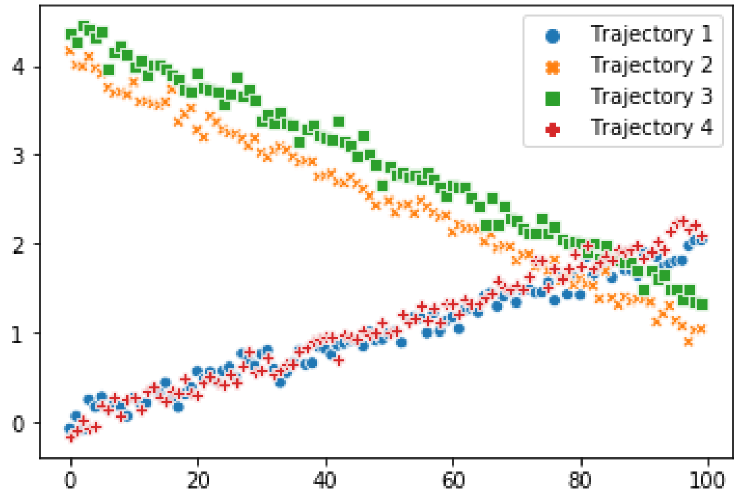

Figure 2.1: Synthetic data showing two motion patterns with two trajectories in each. Trajectory 2 and 3 belong to one motion pattern and Trajectory 1 and 4 belong to a second motion pattern.

Motion pattern learning has a natural interpretation as a clustering problem, but clustering trajectories is difficult, since different trajectories can have different lengths. This means that they do not exist in the same vector space and are not naturally comparable using similarity metrics based on Euclidan distance. Because of this, more advanced similarity metrics need to be used. Several different approaches have been explored, mainly divided by being either deterministic or probabilistic. Common deterministic techniques include using spectral clustering with Dynamic Time Warping (DTW) [10] or Longest Common Subsequence (LCSS) [13] as similarity metrics. These similarity metrics are however computationally expensive, making them unsuitable for real time processing [15].

Probabilistic Approaches

The probabilistic approach instead fits a model to each motion pattern and infers its parameters from data. Two popular approaches for this is fitting Gaussian processes (GPs) and hierarchical Bayesian modeling

Gaussian Processes

GPs have been used in several domains for modeling trajectories. M. Pimentel et al. used GPs to model the vital signs of patients [9], and were successfully able to cluster the data using hierarchical clustering and the GP model likelihoods. They then modeled the motion pattern for a cluster as a GP with the average mean and variance for all GPs in it. K. Kim. et al. used GPs to model the motion patterns of cars from surveillance camera [4]. They introduce the concept of a *frame*, in which the trajectories are discretisised before fitting GPs to them. Having discrete trajectories meant that the local likelihood for observed data x_t, y_t in time step t could be computed as $P(y_t|x_t, M_k)$ for model M_k, which were aggregated to compute a global similarity metric L_k, which in turn was used to cluster the trajectories. To compute the motion patterns of clusters, a sampling scheme was used. In each time point, three GPs were drawn uniformly without replacement from the cluster. The mean value of all drawn GPs were then used as data points to fit a GP for the clusters motion pattern.

Q. Tran and J. Firl used data with both spatial position and velocity, and used GPs to model a vector field for each observed trajectory [12]. Before GPs were trained, the trajectories were normalised using spline interpolation and a sampling scheme. They did not perform any clustering. Instead, they constructed their motion patterns by driving their own test vehicle. When presented with a new trajectory, the model could then compute the most likely motion pattern, and then use a particle filter for the vector field to predict continuations of motion patterns.

The work of M. Tiger and F. Heintz aimed to improve upon the approach of K. Kim et al. who had implicitly assumed that all trajectories were generated from one underlying processes, while a more accurate model is that they are generated from several different, but dependent processes. Assuming a single underlying process causes the model to underestimate its variance, which in turn causes the models likelihood to be too small for data that is still fairly close. They aggregated clustered GPs by considering the mixture of Gaussian distributions they form in a "slice", orthogonal to progression. These slices were then combined to form a single Gaussian distribution. They then generated synthetic data and learned hyperparameters for a single GP to approximate the combined Gaussian distribution, a process they call "inverse Gaussian process regression".

C. Leysen et al. also aim to improve on the work of K. Kim et al [5]. Instead of fitting a GP to re-sampled trajectories, they simply select the trajectory in a cluster which maximises the overall likelihood of the data points in the cluster. While this is less ad hoc, their approach still assumes a single underlying function for all trajectory data, and consequently still underestimates model variance.

Hierarchical Bayesian Models

The idea behind Hierarchical Bayesian Models used for trajectory learning is borrowed from the natural language processing field, where they are known as topic modeling. Topic modeling are unsupervised techniques for clustering documents into different topics, and by considering spatio-temporal trajectories as documents, their observations as words, and the motion patterns they belong to as topics, the same techniques can be used to model motion pattern. These models are usually based on Latent Dirichlet Allocation (LDA) or a generalisation thereof known as Hierarchical Dirichlet Process (HDP) introduced by Teh et al. [11], which are both so called *bag- of-words*-models. A bag-of-words-model assumes independence between words in a document, which in the domain of trajectories translates to the assumption that observed data points are independently drawn.

Both LDA and HDP require a set amount of clusters, which is a great weakness. Wang et al. proposed a model called *Dual Hierarchical Dirichlet Process* (Dual-HDP) [14], which improves upon the HDP model by allowing the model to learn the number of topics and documents from data. Zhou et al. improved upon HDP and LDA by using Markov random fields to encode prior beliefs in a model called *Random Field Topic* (RFT) [16].

– this is more recent but I have not been able to figure this paper out yet– LC-LDA [17]

3 Method

In this chapter, the method is described in a way which shows how the work was actually carried out. The description must be precise and well thought through. Consider the scientific term replicability. Replicability means that someone reading a scientific report should be able to follow the method description and then carry out the same study and check whether the results obtained are similar. Achieving replicability is not always relevant, but precision and clarity is.

Sometimes the work is separated into different parts, e.g. pre-study, implementation and evaluation. In such cases it is recommended that the method chapter is structured accordingly with suitable named sub-headings.

4 Results

This chapter presents the results. Note that the results are presented factually, striving for objectivity as far as possible. The results shall not be analyzed, discussed or evaluated. This is left for the discussion chapter.

In case the method chapter has been divided into subheadings such as pre-study, implementation and evaluation, the result chapter should have the same sub-headings. This gives a clear structure and makes the chapter easier to write.

In case results are presented from a process (e.g. an implementation process), the main decisions made during the process must be clearly presented and justified. Normally, alternative attempts, etc, have already been described in the theory chapter, making it possible to refer to it as part of the justification.

5 Discussion

This chapter contains the following sub-headings.

5.1 Results

Are there anything in the results that stand out and need be analyzed and commented on? How do the results relate to the material covered in the theory chapter? What does the theory imply about the meaning of the results? For example, what does it mean that a certain system got a certain numeric value in a usability evaluation; how good or bad is it? Is there something in the results that is unexpected based on the literature review, or is everything as one would theoretically expect?

5.2 Method

This is where the applied method is discussed and criticized. Taking a self-critical stance to the method used is an important part of the scientific approach.

A study is rarely perfect. There are almost always things one could have done differently if the study could be repeated or with extra resources. Go through the most important limitations with your method and discuss potential consequences for the results. Connect back to the method theory presented in the theory chapter. Refer explicitly to relevant sources.

The discussion shall also demonstrate an awareness of methodological concepts such as replicability, reliability, and validity. The concept of replicability has already been discussed in the Method chapter (3). Reliability is a term for whether one can expect to get the same results if a study is repeated with the same method. A study with a high degree of reliability has a large probability of leading to similar results if repeated. The concept of validity is, somewhat simplified, concerned with whether a performed measurement actually measures what one thinks is being measured. A study with a high degree of validity thus has a high level of credibility. A discussion of these concepts must be transferred to the actual context of the study.

The method discussion shall also contain a paragraph of source criticism. This is where the authors' point of view on the use and selection of sources is described.

In certain contexts it may be the case that the most relevant information for the study is not to be found in scientific literature but rather with individual software developers and open source projects. It must then be clearly stated that efforts have been made to gain access to this information, e.g. by direct communication with developers and/or through discussion forums, etc. Efforts must also be made to indicate the lack of relevant research literature. The precise manner of such investigations must be clearly specified in a method section. The paragraph on source criticism must critically discuss these approaches.

Usually however, there are always relevant related research. If not about the actual research questions, there is certainly important information about the domain under study.

5.3 The work in a wider context

There must be a section discussing ethical and societal aspects related to the work. This is important for the authors to demonstrate a professional maturity and also for achieving the education goals. If the work, for some reason, completely lacks a connection to ethical or societal aspects this must be explicitly stated and justified in the section Delimitations in the introduction chapter.

In the discussion chapter, one must explicitly refer to sources relevant to the discussion.

6 Conclusion

This chapter contains a summarization of the purpose and the research questions. To what extent has the aim been achieved, and what are the answers to the research questions?

The consequences for the target audience (and possibly for researchers and practitioners) must also be described. There should be a section on future work where ideas for continued work are described. If the conclusion chapter contains such a section, the ideas described therein must be concrete and well thought through.

Bibliography

[1] Damian Campo, Mohamad Baydoun, Lucio Marcenaro, Andrea Cavallaro, and Carlo S. Regazzoni. "Modeling and classification of trajectories based on a Gaussian process decomposition into discrete components". In: *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (Aug. 2017), pp. 1–6. DOI: 10.1109/AVSS.2017.8078495.

[2] Sepideh Afkhami Goli, Behrouz H. Far, and Abraham O. Fapojuwo. "Vehicle Trajectory Prediction with Gaussian Process Regression in Connected Vehicle Environment⋆". In: *2018 IEEE Intelligent Vehicles Symposium (IV)* (June 2018), pp. 550–555. ISSN: 1931-0587. DOI: 10.1109/IVS.2018.8500614.

[3] ByeoungDo Kim, Chang Mook Kang, Seung Hi Lee, Hyunmin Chae, Jaekyum Kim, Chung Choo Chung, and Jun Won Choi. "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network". In: *arXiv preprint arXiv:1704.07049* (2017).

[4] Kihwan Kim, Dongryeol Lee, and Irfan Essa. "Gaussian process regression flow for analysis of motion trajectories". In: *2011 International Conference on Computer Vision* (Nov. 2011), pp. 1164–1171. ISSN: 2380-7504. DOI: 10.1109/ICCV.2011.6126365.

[5] Christiaan Leysen, Mathias Verbeke, Pierre Dagnely, and Wannes Meert. "Energy consumption profiling using Gaussian processes". In: *2016 IEEE 8th International Conference on Intelligent Systems (IS)* (Sept. 2016), pp. 470–477. DOI: 10.1109/IS.2016.7737463.

[6] Brendan T. Morris and Mohan M. Trivedi. "Learning and Classification of Trajectories in Dynamic Scenes: A General Framework for Live Video Analysis". In: *2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance* (Sept. 2008), pp. 154–161. DOI: 10.1109/AVSS.2008.65.

[7] Duc-Duy Nguyen, Chan Le Van, and Muhammad Intizar Ali. *Vessel Destination and Arrival Time Prediction with Sequence-to-Sequence Models over Spatial Grid*. ACM, June 2018. ISBN: 978-1-4503-5782-1. DOI: 10.1145/3210284.3220507.

[8] Junbiao Pang, Jing Huang, Yong Du, Haitao Yu, Qingming Huang, and Baocai Yin. "Learning to Predict Bus Arrival Time From Heterogeneous Measurements via Recurrent Neural Network". In: *IEEE Transactions on Intelligent Transportation Systems* (2018).

[9] Marco A. F. Pimentel, David A. Clifton, and Lionel Tarassenko. *Gaussian process clustering for the functional characterisation of vital-sign trajectories*. IEEE, Sept. 2013. DOI: `10.1109/MLSP.2013.6661947`.

[10] Jingren Tang, Hong Cheng, Yang Zhao, and Hongliang Guo. "Structured dynamic time warping for continuous hand trajectory gesture recognition". In: *Pattern Recognit.* 80 (Aug. 2018), pp. 21–31. ISSN: 0031-3203. DOI: `10.1016/j.patcog.2018.02.011`.

[11] Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. "Sharing clusters among related groups: Hierarchical Dirichlet processes". In: *Advances in neural information processing systems*. 2005, pp. 1385–1392.

[12] Quan Tran and Jonas Firl. *Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression*. IEEE, June 2014. DOI: `10.1109/IVS.2014.6856480`.

[13] M. Vlachos, G. Kollios, and D. Gunopulos. *Discovering similar multidimensional trajectories*. IEEE, Feb. 2002. DOI: `10.1109/ICDE.2002.994784`.

[14] Xiaogang Wang, Keng Teck Ma, Gee-Wah Ng, and W. Eric L. Grimson. "Trajectory analysis and semantic region modeling using a nonparametric Bayesian model". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition* (June 2008), pp. 1–8. ISSN: 1063-6919. DOI: `10.1109/CVPR.2008.4587718`.

[15] Zhang Zhang, Kaiqi Huang, and Tieniu Tan. *Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes*. Vol. 3. Aug. 2006. DOI: `10.1109/ICPR.2006.392`.

[16] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. "Random field topic model for semantic region analysis in crowded scenes from tracklets". In: *CVPR 2011* (June 2011), pp. 3441–3448. ISSN: 1063-6919. DOI: `10.1109/CVPR.2011.5995459`.

[17] Jialing Zou, Qixiang Ye, Yanting Cui, Fang Wan, Kun Fu, and Jianbin Jiao. "Collective motion pattern inference via Locally Consistent Latent Dirichlet Allocation". In: *Neurocomputing* 184 (Apr. 2016), pp. 221–231. ISSN: 0925-2312. DOI: `10.1016/j.neucom.2015.08.108`.