Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Machine Learning Approach to Traffic Flow Prediction Using CP Tensor Decomposition

Thomas Steffen *, Gerwald Lichtenberg **

**Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3TU, UK (Tel: +44 1509 227224, e-mail: t.steffen@lboro.ac.uk, orcid:0000-0002-4667-3601)*
***Faculty Life Sciences, University of Applied Sciences Hamburg, Ulmenliet 20, 21033 Hamburg, Germany (Tel +49.40.428 75-6433, gerwald.lichtenberg@haw-hamburg.de)*

Abstract: This paper deals with the prediction of highway traffic flow based on historic data. The methodology is based on canonical polyadic (CP) tensor decompositions of traffic flow data. This step captures the regular elements of the traffic signal based on daily and weekly rhythms and typical geographical distributions of the traffic, while significantly reducing the amount of data required to describe these. The key factors are then extrapolated into the future, and the traffic data is reconstructed from the decomposition. Applied to traffic flow data from the M62 in the North of England in October 2019, this approach provides a surprisingly accurate prediction based on a very compact model, which is a distinct advantage compared to conventional machine learning approaches. Using 4 factors, the prediction captures 90% of the signal energy, which beats existing rolling average prediction techniques.

Keywords: modeling and simulation of transportation systems, information processing and decision support; intelligent transportation systems, tensor analysis, canonical polyadic decomposition (CPD), traffic flow, prediction, machine learning, disturbance prediction

1. INTRODUCTION

On modern motorways and highways, point measurement and automatic number place recognition (ANPR) are used to constantly monitor traffic density and speed as part of a highway management strategy. This data is used for setting speed limits, predicting journey times, informing route planners, schedule maintenance, analyse highway safety, and plan future investments into the highway infrastructure. Many of these uses require predictions of future traffic flows based on historical data. This paper proposes a novel approach to the prediction based on a tensor analysis of the traffic data.

Tensor analysis is a generalisation of matrix analysis, and it is based on a multi-dimensional view of the data. Let T be a four-dimensional tensor in the dimensions i, j, k, and l of size

$$T \in \mathbb{R}^{n_i \times n_j \times n_k \times n_l}$$

with $n_i n_j n_k n_l$ real-valued elements denoted by $t_{i,j,k,l}$.

The canonical polyadic decomposition (CPD) of this tensor is a generalisation of the singular value decomposition (SVD), and it approximates the tensor using n factors defined by 4 vectors a, b, c, d and a scaling factor λ each (one vector for each dimension of the tensor):

$$t_{i,j,k,l} = \sum_{m=1}^{n} \lambda_m \, a_{m,i} b_{m,j} c_{m,k} d_{m,l} \quad (1)$$

The n vectors are typically stored in matrix form, leading to four matrices A, B, C, and D with respective dimensions (Kolda and Bader 2009).

This is called a rank n decomposition. Like the SVD, it is usually able to capture most of variability of a structured tensor in a low rank, requiring significantly less storage space. Unlike the SVD, finding the CPD is not a convex optimisation problem: several different algorithms exist to find numerical approximations of the optimal decomposition and toolboxes are available (www.tensorlab.net, www.tensortoolbox.org). *Remark:* In contrast to matrix SVD, which always can be truncated to find a lower rank SVD, this is in general not possibly for a tensor CPD, where truncating factors of higher ranks give much poorer results than a CPD with a lower rank.

The paper is organised as follows: Section 2 presents a brief overview of existing approaches, Section 3 defines the research question, Section 4 details the methodology, and Section 5 shows the numerical results, leading to the conclusions in Section 6.

2. BACKGROUND

2.1 Tensor Analysis

Tensor analysis has been successfully used in a number of areas where the data can be easily represented in multiple dimensions. A typical example is calendar based data, which is subject to daily, weekly, and annual cycles.

An example of a similar analysis is presented in (Sewe 2017), where the heating energy demand of a building is analysed for the purpose of fault detection.

The canonical polyadic decomposition (CPD) is the main analysis tool, and it is similar to the singular value decomposition (SVD) or principal component analysis (PCP) for 2 dimensional data. Other decompositions exist, most notably the Tucker decomposition and the tensor train (TT) decomposition (Li 2019).

Applications for tensor analysis are like the PCP, and they include the visualisation and understanding of large datasets, data reduction, clustering, prediction, categorisation and fault detection, as well as the restauration of incomplete data.

2.2 Traffic Flow Prediction

The prediction of traffic flow data is an important question for the management of traffic and the calculation of optimal routes. Unlike the prediction of energy demand (Mody 2017), traffic is a strongly localised phenomenon, and therefore the amount of available data is significantly higher.

There is a good amount of literature, but it focuses on a few common approaches, such as regression (Wu 2015), autocorrelation (Kumar 2015) and deep neural networks (Huang 2014, Lv 2015). The research for traffic analysis in computer networks is much wider (e.g. Joshi 2015).

3. RESEARCH QUESTION

The aim of this paper is to produce a prediction of the traffic on a highway based on historical data. In other words, a model of the traffic data is to be created, which can then be used to extrapolate from known data into the future.

The model is typically an abstraction and achieving the right level of abstraction can be critical for a good prediction. Too detailed a model may not generalise well (overfit). If the model is too coarse, it cannot adequately describe the behaviour in the first place (underfit). Identifying the best level of abstraction is part of the research question.

4. METHODOLOGY

The data used here is from the M62 motorway in the North of England, from 1 October 2019 to 28 October 2019 in 15 minute long intervals. It can be found on the webtris system (highways england 2019). This data is given as the number of passing cars per hour. The same methodology should be applicable to any set of traffic data.

The traffic flow data is partitioned into a training data set and a verification data set (both two weeks). To maintain causality, the learning data set is before the verification data set.

The original data is presented as a matrix, where the first dimension is the location of the measurement, and the second dimension the time in regular intervals. This is reformatted as a tensor, with the following four dimensions:

1. $n_i = 150$ locations
2. $n_j = 96$ periods during a day (24 hours, 4 per hour)
3. $n_k = 7$ days in a week starting Wednesday (1.10.19)
4. $n_l = 2$ weeks (of the training data)

The CPD is performed in MATLAB, using the cpd function from the tensorlab toolbox version 3 (Vervliet 2016). The follow code explains this key step:

```
T=reshape(D,[150 96 7 2]); % rearrange as tensor
Tcpd=cpd(T,7);             % perform a rank 3 decomposition
Tgen=cpdget(Tcpd);         % restore from the decomposition
```

The tensorlab toolbox does not use the factor λ, which is assumed to be one, and instead includes this factor in the vectors. To the make results easier to interpret, the vectors are normalised, and λ is calculated according to equation (1). This makes no difference to the model, but it simplifies the analysis.

The tensor is reconstructed from the decomposition and compared to the original. The mean square error (MSE) is calculated as a measure of quality of the approximation, to measure whether the data has been captured in the decomposition.

To generate the prediction, the vector d needs to be extended from the 2 weeks of the training dataset to cover the additional 2 weeks of the verification data set. The average of the 2 training weeks is filled in.

A better prediction at this point would be desirable, e.g. using context information or data from the previous year. Obviously, the best data would be the verification dataset itself, but this cannot be used without violating the causality of the prediction. The reconstruction of the extended tensor decomposition is then compared against the verification dataset. The full code is available on github (Steffen 2020).

5. RESULTS

Figure 1 shows a carpet plot of the original traffic data, with location on the vertical and time on the horizontal axis. The daily and weekly rhythm is clearly visible in the plot. The striking orthogonal features of this plot indicate that it is highly structured and suitable for a decomposition along the dimensions. The left half of this plot is the training data, and the right half is the verification dataset. The full dataset consists of 403200 values, with 2093 missing values, an average of 458.6, and a standard deviation of 374.6.

Figure 2 shows the components found by CPD with rank 7. This represents the 150*96*7*2=201600 data points from the training set using only 7*(150+96+7+2)=1785 values. Several components are very similar, and they cancel each other out to some degree. Note that the weekly factors are set to 1 for Week 3 and 4, because those are extrapolated, as only Week 1 and 2 are used for training.

Figure 3 shows the reconstructed traffic data, and Figure 4 shows the absolute error. The errors are generally small, with deviations being quite localised. The longest deviation is around location 90 at time step starting from 1500. The deviation is negative (less traffic than predicted), it last for

Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

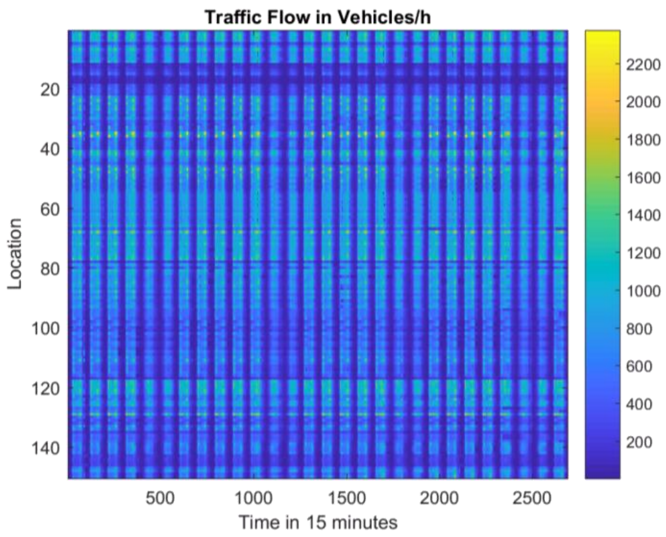

Fig. 1. Carpet Plot of the Original Traffic Data

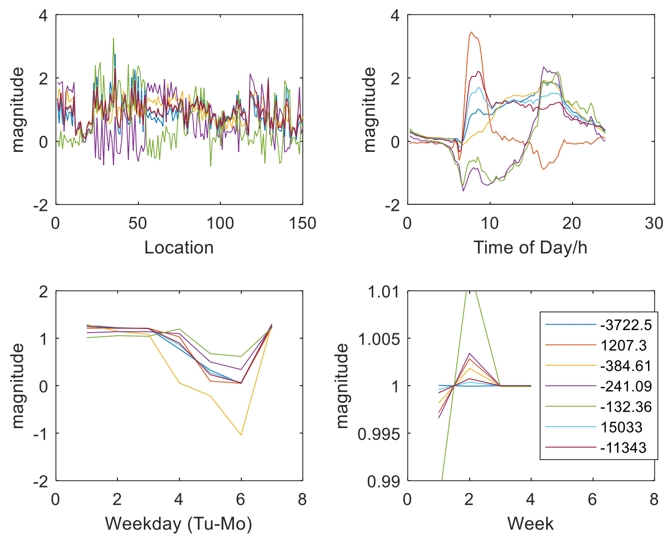

Fig. 2. Components of the Canonical Polyadic Decomposition

most of the day, and it aligns with recorded road works, so it is reasonable to assume that it is caused by a traffic jam forcing road users onto a diversion route.

To assess the quality of the tensor model, the root mean square error (RMSE) for the training and the validation data set are plotted over the CPD rank in Figure 5. As expected, the training error reduces with increasing rank, but the validation error seems to level off from around rank 7, with a good prediction reached from rank 3. This indicates that overfitting may be happening from about rank 10, and for the demonstration, the rank 7 was used as a compromise between prediction quality and complexity.

For comparison, the graph also contains the RMSE for a prediction based on the average weekly profile from the training set (the average of Week 1 and 2). These are

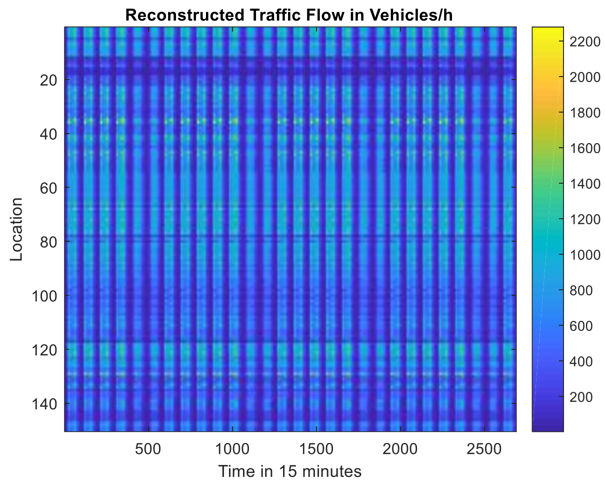

Fig. 3. Carpet Plot of the Reconstructed Traffic Data

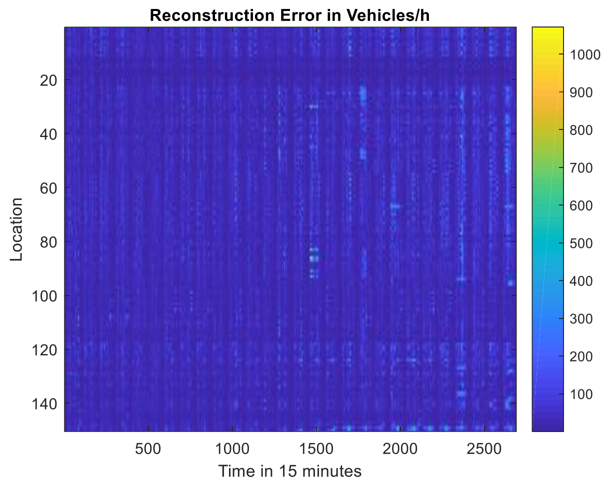

Fig. 4. Absolute Error of the Reconstructed Traffic Data

representative of a conventional regression or filtering approach and are shown as horizontal dashed lines in the graph. It is remarkable that the tensor decomposition achieves a better prediction with a much lower amount of data. This is presumably because the tensor decomposition can find basic recurrent structures of the data – due to reasonably chosen dimensions – and creates an abstract model that covers typical behaviour but ignores minor abnormalities in the data.

Even better predictions should be possible if a prediction of the weekly factors (the second half of the last graph from Figure 2) is added, for example based on external factors. Further tensorization, i.e. finding more than 4 dimensions, e.g. splitting the time to two dimensions hour of the day and quarter of the hour could improve the compression factors but will not lead to improved errors because in traffic, there is nothing special e.g. about 15 min past the hour in contrast to 30 min past the full hour.

Finally, Figure 6 shows the autocorrelation of the original traffic data, and the autocorrelation of the prediction error.

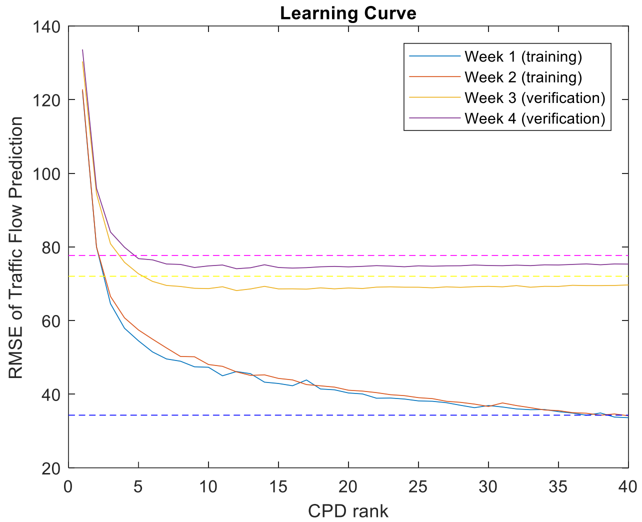

Fig. 5. Prediction Error by Week over CPD Rank (horizontal lines denote averages for comparison)

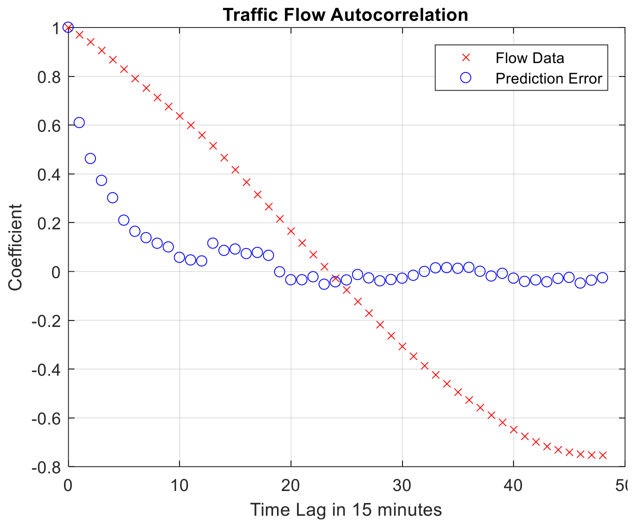

Fig. 6. Autocorrelation of the Original Flow Data and the Prediction Error

This shows that the prediction is successful, and the correlation of the prediction error is remarkable short.

Further research should investigate the stochastic behaviour of this deviation and create an observer that allows real time traffic prediction based on the prediction and a model of the error fed by real time traffic data.

6. CONCLUSIONS

This paper shows that traffic flow data can be analysed using a tensor decomposition. Even a low rank decomposition captures a large share of the data, and it can be used to compress the data, analyse the data, highlight deviations from nominal situations, and to predict future traffic. This prediction is superior to a more data insensitive filtering approach, because it manages to separate a regular component of the data from irregular occurrences. It is therefore a worthwhile alternative or addition to a conventional machine learning regression model or a moving average or recursive filter.

The improved prediction has far ranging applications in traffic management, traffic control, maintenance schedule, and route planning that may affect all traffic participants. Further work is required to study these applications in detail, and the applicability to wider datasets.

REFERENCES

Chen, X., He, Z., and Wang (2018). Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition. pp 59-77 (86). Transportation Research Part C: Emerging Technologies. doi:10.1016/j.trc.2017.10.023

highways england (2019), WebTRIS Traffic Flow API. http://webtris.highwaysengland.co.uk/ .

Huang, W., et al (2014). Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Trans. on Intelligent Transportation Systems*. Vol. 15, no. 5, pp. 2191-2201, Oct. 2014, doi: 10.1109/TITS.2014.2311123.

Joshi, M., Hadi, T.H. (2015). A Review of Network Traffic Analysis and Prediction Techniques. arXiv:1507.05722

Kolda, T, Bader, B. (2009). Tensor Decompositions and Applications, SIAM Review, 51(3), 455–50

Kumar, S.V., Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. Eur. Transp. Res. Rev. 7, 21. doi:10.1007/s12544-015-0170-8

Li, L., Yu, W., Batselier, K. (2019). Faster Tensor Train Decomposition for Sparse Data. arXiv:1908.02721 .

Lv, Y., Duan, Y, Kang, W., Li, Z., and Wang, F (2015). Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*. vol. 16, no. 2, pp. 865-873.

Mody, S., and Steffen, T. (2017). Benefits of stochastic optimisation with grid price prediction for electric vehicle charging. SAE World Congress. doi:10.4271/2018-01-0953

Sewe, E., Pangalos, G., Lichtenberg, G. (2017). Fault Detection for Heating Systems using Tensor Decompositions of Multi-linear Models. SIMULTECH pp 27-35.

Steffen, T. (2020). Tensoranalysis. https://github.com/tsteffenlboro/tensoranalysis .

Vervliet, N., Debals, O., Sorber, L., Van Barel, M. and De Lathauwer, L. (2016). Tensorlab 3.0. https://www.tensorlab.net/ .

Wu, Y., Tan, H., Peter, J., Shen, B., and Ran, B. (2015). Short-Term Traffic Flow Prediction Based on Multilinear Analysis and k-Nearest Neighbor Regression. 15th COTA International Conference of Transportation.