Car-traffic forecasting: A representation learning approach

Ali Ziat, Gabriella Contardo, Nicolas Baskiotis, Ludovic Denoyer FIRSTNAME.LASTNAME@LIP6.FR
Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France

Abstract

We address the problem of learning over multiple inter-dependent temporal sequences where dependencies are modeled by a graph. We propose a model that is able to simultaneously fill in missing values and predict future ones. This approach is based on representation learning techniques, where temporal data are represented in a latent vector space. Information completion (missing values) and prediction are then performed on this latent representation. In particular, the model allows us to perform both tasks using a unique formalism, whereas most often they are addressed separately using different methods. The model has been tested for a concrete application: car-traffic forecasting where each time series characterizes a particular road and where the graph structure corresponds to the road map of the city.

1. Introduction

Traffic data has particular characteristics that can not be fully handled by classical sequential and temporal models: they contain multiple missing values, and one has to consider simultaneously multiple sources that can be somehow related, by spatial proximity for example.

We propose a novel method that aims at integrating these aspects in one single model. The proposed approach is based on representation learning techniques aiming at projecting the observations in a continuous latent space, each sequence being modeled at each time-step by a point in this space. It has many advantages w.r.t existing techniques: it is able to simultaneously learn how to fill missing values and to predict the future of the observed temporal data, avoiding to use two different models, and it naturally allows one to deal with information sources that are organized among a graph structure. Moreover, the model is based on continuous optimization schemes, allowing a fast optimization over large scale datasets.

Proceedings of the 2^{nd} International Workshop on Mining Urban Data, Lille, France, 2015.

2. Context and Model

2.1. Notations and Tasks

Let us consider a set of n temporal sequences $\mathbf{x_1}, ..\mathbf{x_n}$ such that $x_i^{(t)} \in \mathcal{X}$ is the value of the i-th sequence at time t defined by $\mathbf{x_i} = (x_i^{(1)}, .., x_i^{(T)})$. In the case where $\mathcal{X}$ is $\mathbb{R}^m$, the context corresponds to multiple multivariate time series. The sequences contain missing values so we also define a mask $m_i^{(t)}$ such that $m_i^{(t)} = 1$ if value $x_i^{(t)}$ is observed - and thus available for training the system - and $m_i^{(t)} = 0$ if $x_i^{(t)}$ is missing - and thus has to be predicted by the model. In addition, we consider that there exists a set of relations between the sequences which correspond to an external information, like spatial proximity for example when $\mathcal{X}$ is discrete. The sequences are thus organized in a graph $\mathcal{G} = \{e_{i,j}\}$ such that $e_{i,j} = 1$ means that $\mathbf{x_i}$ and $\mathbf{x_j}$ are related, and $e_{i,j} = 0$ elsewhere.

2.2. Model

The *RepresentAtIoN-baSed TempOral Relational Model* (RAINSTORM) is a loss-based model which is described through a continuous derivable loss function that will be optimized using classical optimization techniques.

Let us define $\mathcal{L}(\theta, \gamma, \mathbf{z})$ the loss function to minimize where $\mathbf{z}$ is the set of all the vectors $z_i^{(t)}$ for $i \in [1..n]$ and $t \in [1..T]$, T being the size of the observed time windows i.e. the history of the time series. We define $\mathcal{L}$ as:

$$\begin{aligned}\mathcal{L}(\theta, \gamma, \mathbf{z}) = & \frac{1}{O} \sum_{i=1}^{n} \sum_{t=1}^{T} m_i^{(t)} \Delta(f_\theta(z_i^{(t)}), x_i^{(t)}) \text{ (term 1)} \\ & + \lambda_{dyn} \sum_{i=1}^{n} \sum_{t=1}^{T-1} ||z_i^{(t+1)} - h_\gamma(z_i^{(t)})||^2 \text{ (term 2)} \\ & + \lambda_{struct} \sum_{i,j \in [1..N]^2} \sum_{t=1}^{T} e_{i,j} ||z_i^{(t)} - z_j^{(t)}||^2 \text{ (term 3)}\end{aligned} \tag{1}$$

where O is the number of observed values i.e. values such that $m_i^{(t)} = 1$.

This loss function contains three terms, each one associated with one of the constraints that have been presented previously:

- Term 1 aims at simultaneously learn $\mathbf{z}$ and a function f_θ - called **decoding function** - such that, from $z_i^{(t)}$, f_θ can be used to predict the value $x_i^{(t)}$. The function $f_\theta(z_i^{(t)})$ is defined as $f_\theta : \mathbb{R}^N \rightarrow \mathcal{X}$. Δ is used to measure the error between predicting $f_\theta(z_i^{(t)})$ instead of $x_i^{(t)}$, $m_i^{(t)}$ playing the role of a mask restricting to compute this function only on the observed values.
- Term 2 aims at finding values $z_i^{(.)}$ and a dynamic model h_γ such that, when applied to $z_i^{(t)}$, h_γ allows us to predict the representation of the next state of time series i i.e. $z_i^{(t+1)}$. h_γ is the **dynamic function** which models the dynamicity of each series directly in the latent space: $h_\gamma : \mathbb{R}^N \rightarrow \mathbb{R}^N$. The parameters γ will be learned to minimize the mean square error between the prediction $h_\gamma(z_i^{(t)})$ and $z_i^{(t+1)}$.
- At last, term 3 corresponds to a **structural regularity** over the graph structure that encourages the model to learn closer representations for time series that are related. This will force the model to learn representations that reflect the structure of the considered graph.

λ_{dyn} and λ_{struct} are manually defined coefficients that weight the importance of the different elements in the loss function.

The learning problem aims at minimizing the loss function $\mathcal{L}(\theta, \gamma, \mathbf{z})$ simultaneously on θ, γ and $\mathbf{z}$. By restricting the f_θ and h_γ to be continuous derivable functions, we can use gradient-descent based optimization approaches.

3. Traffic Forecasting and Experiments

Experiences have been made on two datasets from the cities of Beijing and Warsaw. The dataset are provided by (Zheng, 2011) and (ICDM) and are not described here for sake of space.

3.1. Models

We propose to compare the RAINSTORM approach to the following baseline models, some baselines being used for data completion, and some others for prediction. For the completion problem we consider the following models:
MF: This correspond to the classical matrix factorization framework for data completion.
MF-with geographic context: This method is the one named TSE (traffic speed estimation) in (Shang et al., 2014).
For the prediction task, we consider:
NeuralNetwork: This is the classical baseline method used in traffic forecasting based on a neural network architecture, described for instance in (Dougherty & Cobbett, 1997).
SAE: This is the method described in (Lv et al., 2014).

We also compare RAINSTORM with a model based on a heuristic able to perform both completion and prediction that we call **RoadMean** and can be described as follow: this model predicts and fills missing value with the mean of observed values on the sequence.

3.2. Experiments and Results

We compare our model with baselines approach for the two tasks of completion and prediction. Results are reported in Table 1. and Table 2.

Table 1. Prediction at $T+1$, comparison between described baselines models and the RAINSTORM model for different size of latent space N with a root mean square error (RMSE)

N	Model/Dataset	Beijing	Warsaw	
		Volume	Volume	Speed
	RoadMean	5.51	5.09	11.02
	NeuralNetwork	4.77	4.27	8.05
	SAE	4.75	4.27	7.85
5	RAINSTORM	4.82	4.28	7.74
10	RAINSTORM	4.78	**4.20**	7.21
20	RAINSTORM	**4.54**	4.21	**7.19**
50	RAINSTORM	4.66	**4.20**	7.60

Table 2. Completion for 50% missing data, comparison between described baselines models and the RAINSTORM model for different sizes N of the latent space with a root mean square error (RMSE)

N	Model/Dataset	Beijing	Warsaw	
		Volume	Volume	Speed
	RoadMean	5.55	5.00	11.10
	MF	3.58	3.16	6.80
	MF-Geo	3.24	2.99	6.49
5	RAINSTORM	2.99	3.12	6.49
10	RAINSTORM	3.03	3.00	6.24
20	RAINSTORM	3.22	2.94	**6.23**
50	RAINSTORM	**2.97**	**2.93**	6.70

4. Conclusion

We have presented a new way to learn over incomplete multiple sources of temporal relational data sources. The RAINSTORM approach is based on representation learning techniques and aims at integrating in a latent space the observed information, the dynamicity of the sequences of data, and their relations. In comparison to baselines models that have been developed for prediction only or completion only, our approach shows interesting performance and is able to simultaneously complete missing values and predict the future evolution of the data.

References

Dougherty, Mark S and Cobbett, Mark R. Short-term inter-urban traffic forecasts using neural networks. *International journal of forecasting*, 13(1):21–31, 1997.

ICDM. I.E.E.E icdm contest: Tomtom traffic prediction for intelligent gps navigation.

Lv, Yisheng, Duan, Yanjie, Kang, Wenwen, Li, Zhengxi, and Wang, F-Y. Traffic flow prediction with big data: A deep learning approach. 2014.

Shang, Jingbo, Zheng, Yu, Tong, Wenzhu, Chang, Eric, and Yu, Yong. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

Zheng, Yu. T-drive trajectory data sample, August 2011. URL http://research.microsoft.com/apps/pubs/default.aspx?id=152883.