LSTM encoder-predictor for short-term train load forecasting

Kevin Pasini [1,2] Mostepha Khouadjia [1] Allou Samé[2] Fabrice Ganansia [3] Latifa Oukhellou [2]

[1]Université Paris-Est, IFSTTAR, Cosys-Grettia, Champs-sur-Marne, France

[2]IRT SystemX, Saclay, France

[3]SNCF- Innovation & Recherche, St Denis, France

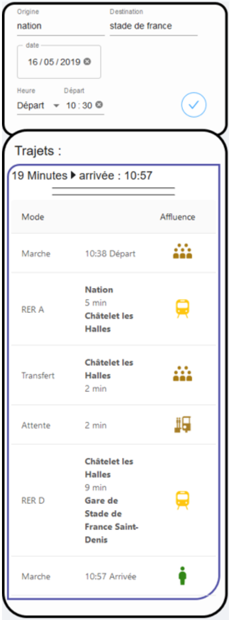
September 25, 2020

Objectives:

- Apply machine learning methods on public transport mobility data
- Train load forecast

Interests:

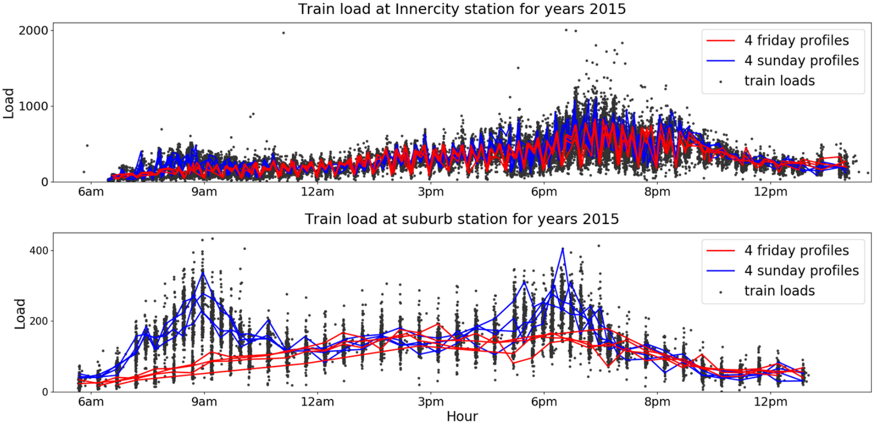
- Enrich public transport information.
- Improve route planner.
- Allow passengers to better plan their daily trips
- Help to transport regulation.



Summary

Issues

Two examples of passenger loads:



- Contextual variance
- Station specificity
- Transportation plan

Work specificities:

- Scale of train passage
  $\Rightarrow$ *No aggregation (Noise and variance)*
- Numerous influence factors
  $\Rightarrow$ *Calendar, contextual and hidden factors*
- Transportation schedule
  $\Rightarrow$ *Variability on Temporal structure*
- Recent data source for rail-infrastructure
  $\Rightarrow$ *Few ambitious study, Some data quality issues*

Data sources

Raw data :

- **Transport Supply**: Timetable information and Automatic Vehicle Location (AVL)
- **Supply vs demand**: Count of boarding and alighting passengers.
- **calendar information**: Day, Month, holidays

Extract and refined features:

Long-term features LT (Planned) :

- Calendar information
- Theoretical hour of train passage
- Train services

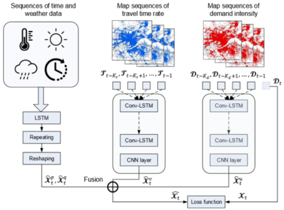
Short-term features ST (measured):

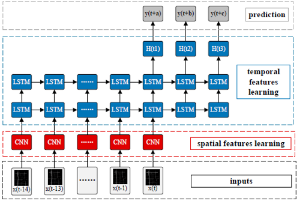
- Delays at the station
- loads passenger of previous trains

Focus: Forecasting affluence on aggregated data

Ke, Jintao, et al (2017)

Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach.

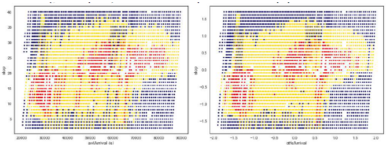




Junbo, Zhang, et al (2017)

Deep spatio-temporal residual networks for citywide crowd flows prediction

Focus: Forecasting train load on non-aggregated data



Heydenrijk-Ottens, Leonie, et al (2018)

Supervised learning: Predicting passenger load in public transport

Ding, Chuan, et al (2016)

Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees

Subway Station	Performance for Different Models (Measured by Root Mean Squared Error (RMSE) and R^2)							
	NN		SVM		RF		GBDT	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
DWL	134.2033	0.9599	171.4534	0.9346	107.6754	0.9742	65.9933	0.9806
FXM	60.9258	0.9825	88.1399	0.9633	68.2797	0.9780	37.4414	0.9893
HLG	99.4166	0.9837	149.4753	0.9631	125.6164	0.9739	64.0564	0.9916

Note: NN = BP-neural network, SVM = support vector machine, and RF = random forest.

Our Proposition:

- Based on RNN encoder decoder structure for translation

  Cho, Kyunghyun, et al. (2017) "Learning phrase representations using RNN encoder-decoder for statistical machine translation.

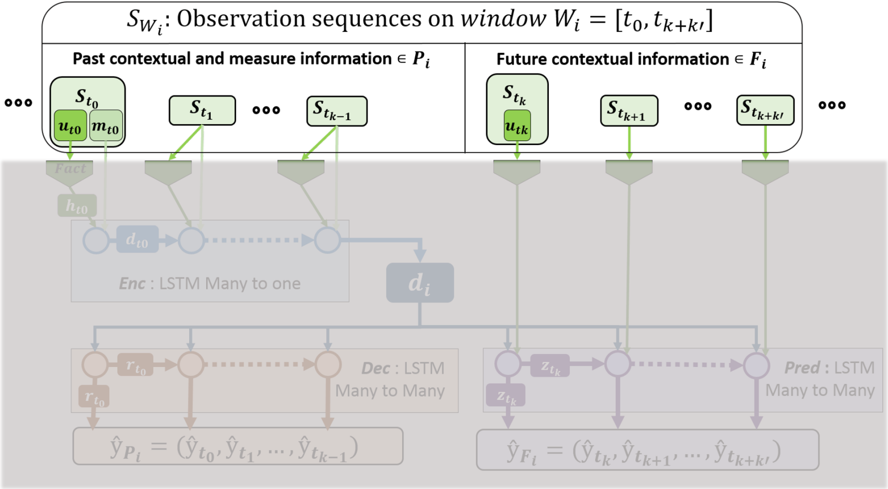
- Overcome the temporal variability by learn "contextual representation"

  Bengio, Yoshia. (2013) "Representation learning: A review and new perspectives".

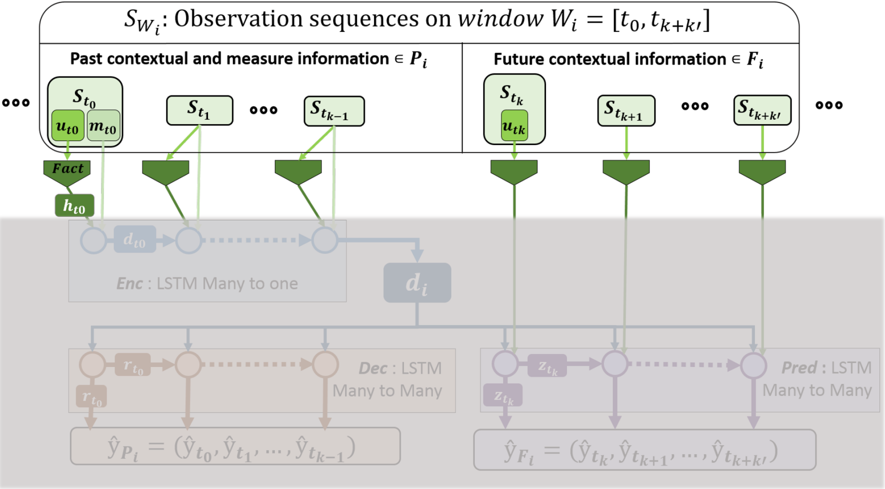
Motivations:

- Better exploit sequential structure (capture past dynamic).
- Better exploit contextual features (Non-linear combination).
- Multi step-prediction with inference of futur dynamic.

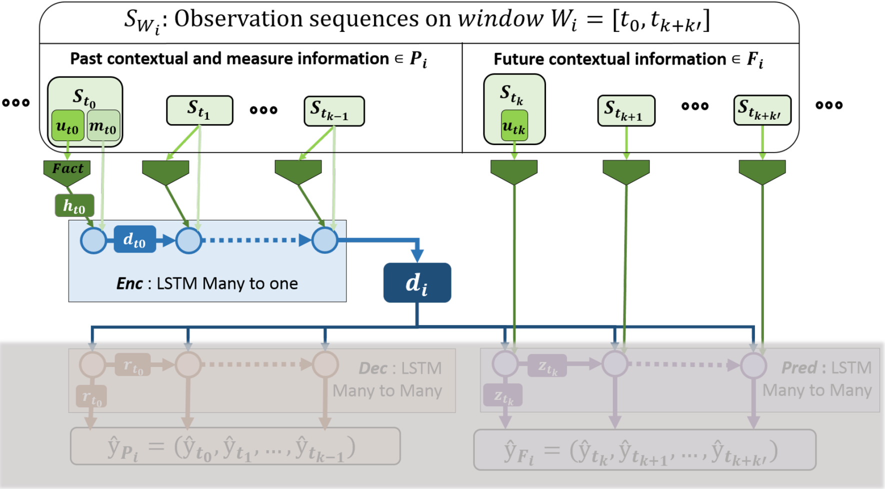
Forecasting Model : LSTM Encoder-Predictor



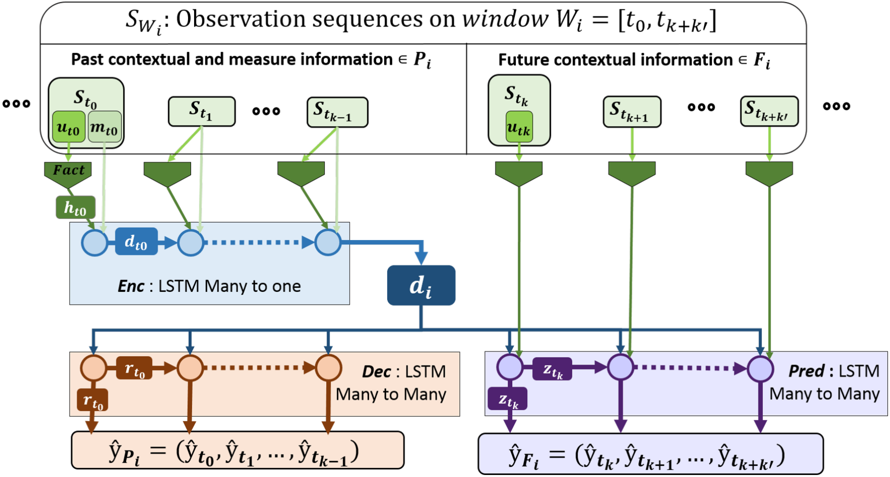
Forecasting Model : LSTM Encoder-Predictor



Forecasting Model : LSTM Encoder-Predictor



Forecasting Model : LSTM Encoder-Predictor



Work perimeter

- Study perimeter: Suburb North of Paris
- Line Characteristics: :
  - 50 stations & 500 daily train
  - Multiple branches.
  - 4 suburb terminus.
  - Train services heterogeneity.
- Historical depth: January 2015 - June 2016
- Days studied: All week
- Daytime: 5am to 2am of next day

Train load Forecasting

- **Task**: Forecast the next train load at a station (non-aggregated data)
- **Train**: Year 2015 / **Test** : first half of year 2016
- **Station**: Suburb station (30,000) and a innercity station (60,000)
- **Metrics**: RMSE + WAPE on train and test sets.
- **Models**
  - DV: last Values (LT)
  - MC: Contextual average (LT)
  - XGB LT: Ensemble model using LT features (LT)
  - XGB ST: Ensemble model using both ST + LT features (ST)
  - LSTM: Standards recurrent neural networks (ST)
  - LSTM EP: LSTM encoder-predictor network (ST)

Uni-step Forecasting results

Model	Suburb			Inner city	
	WAPE	RMSE		WAPE	RMSE
LV	17.9	35.8	Train score	41.9	186.7
CA	13.7	28.7		14.2	73.1
XGB LT	8.4	17.2		8.3	44.75
XGB ST	7.5	15.1		8.2	43.5
LSTM	11.5	24.3		8.9	51.5
LSTM EP	10.7	22.1		10.9	57.7
LV	24.1	47.2	Test score	46.9	205.0
CA	19.0	40.0		18.5	96.5
XGB LT	18.8	38.9		13.4	76.0
XGB ST	16.8	35.7		12.7	73.0
LSTM	16.2	34.0		13.7	75.3
LSTM EP	**16.0**	**33.8**		**12.9**	**72.4**

- ML models outperform baseline models.
- Short-term features gives small gain.
- LSTM weaks to face service heterogeneity.
- LSTM EP slightly outperforms other approaches.

Multi-step Forecasting Results

Task Predict the load on future horizon of 6 next train passage.

Suburb station

Model Time interval*	t+1 14-32	t+2 29-62	t+3 44-92	t+4 59-122	t+5 75-152	t+6 90-182
XGB LT	38.9	38.9	38.9	38.9	38.9	38.9
XGB ST	35.7	36.6	36.7	36.7	37.6	38.1
LSTM	34.0	34.4	34.8	35.5	36.3	36.9
LSTM EP	**33.8**	**34.0**	**34.1**	**34.4**	**34.7**	**34.9**

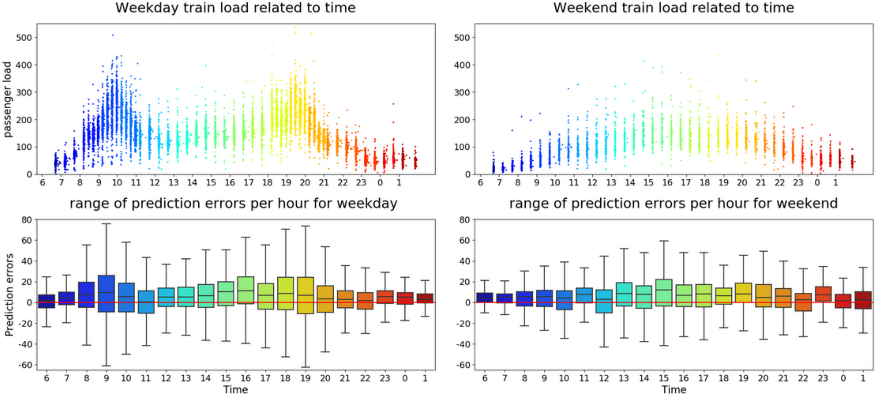
Inner-city station

Model Time interval*	t+1 2-13	t+2 5-23	t+3 9-31	t+4 12-43	t+5 15-53	t+6 18-61
XGB LT	76.0	76.0	76.0	76.0	76.0	76.0
XGB ST	73.0	72.8	73.3	73.8	73.4	73.5
LSTM	75.3	75.4	80.2	83.9	90.5	92.9
LSTM EP	**72.4**	**72.1**	**72.1**	**72.2**	**72.6**	**72.8**

- Expected Deterioration & Small deterioration for LSTM EP

Hourly residual variance (LSTM EP)

Figure: Prediction errors depend on hour class for suburb station



- Same order of errors for weekday and weekend.
- Peak hours are more difficult to predict due to variance.

Conclusion

Contribution :

- Forecast load at train passage scale on real data.
- Compare baselines and several machine learning models.
- Propose a model based on LSTM E-D architecture for mutlistep prediction with contextual features.

Perspectives :

- Confront prediction residual to abnormal situation.
- Better exploit spatial structure network.

Contextual anomaly detection for multivariate time series

Manuscript submitted to Data mining and knowledge discovery

Motivation : context-invariant anomaly score.

Data : Montreal metro ridership - Multivariate Regular Time Series

Interest : Anomaly characterization , Impact Analysis

Approach : Score based on variance-normalized prediction residual

Concept : Capture Known and latent factors as a dynamic context

$$s_t = \frac{(Y_t - \hat{Y}_t) - \hat{B}_t}{2\hat{\sigma}_t} \begin{cases} \hat{Y}_t : \text{Prediction model} \\ \hat{B}_t : \text{Contextual bias estimator} \\ \hat{\sigma}_t : \text{Contextual variance estimator} \end{cases}$$

Proposed estimators based on :

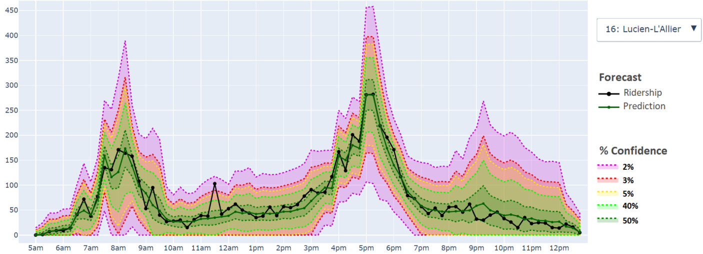
- Prior sampling
- ML Model (RF) on prediction residuals
- Extraction from RF prediction model
- Extraction from DEEP prediction model

Variance analysis

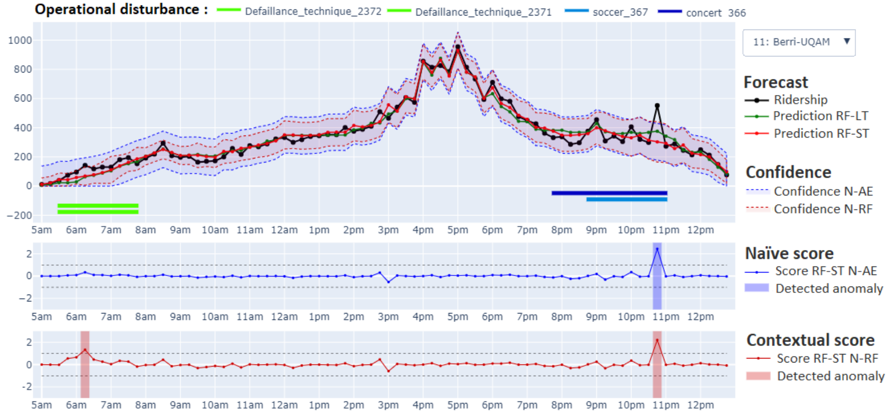
Sampling analogy : Build homogeneous sampling on heteregenous data structured by a context dynamic.

Exploitation :

- Exploration of data variability.
- Prediction model confidence.
- Contextual anomaly detection.

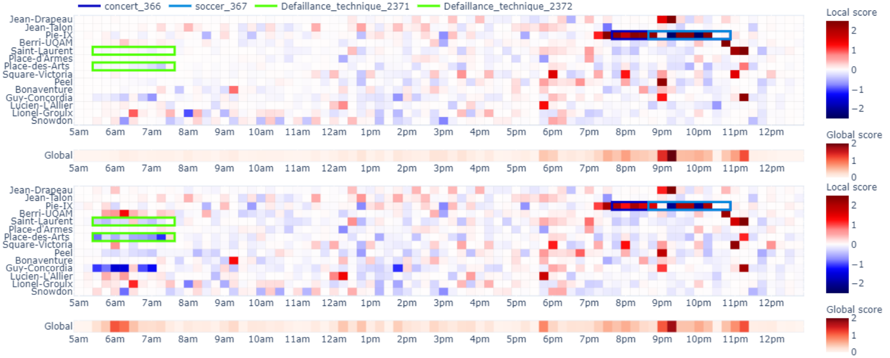


Contextual normality based on contextual confidence interval



Mutlivariate context-normalised anomaly score

Absolute anomaly score



Context-normalized anomaly score

Synthesis

Handle structural data for prediction :

- Learn on structure if regular.
- Translate into contextual attribute else.

Time series in a dynamic environment :

- Analogy : Homogeneous contextual sampling.
- ML algorithm perform well (Specially ensemblist method).

Take into account the contextual variance :

- Variability exploration.
- Confidence of prediction.
- Contextual anomaly detection.