

Modèle LSTM encodeur-prédicteur pour la prévision court-terme de l'affluence dans les transports collectifs

Kévin Pasini^{1,2}, Mostepha Khouadja², Allou Samé¹, Latifa Oukhellou¹

¹Grettia - IFSTTAR - Université Paris-Est, ²IRT SystemX

1. Contexte des travaux :

Partenaires : SNCF et Ile-de-France mobilités.

Thèse : Analyse de traces numériques de mobilités.

Projet : IRT- System-X IVA : Information voyageur augmentée

- Mieux comprendre l'impact de l'information voyageur

- Fournir des indicateurs prédictifs sur l'état du réseau

Objectif des travaux :

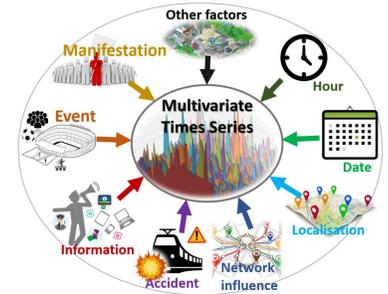
- Valoriser les données brutes pour en extraire de l'information
- Mieux comprendre les comportements de mobilité
- Analyser l'impact des perturbations

2. Problématique : Prévision d'affluence à l'échelle des trains par gare

Mieux comprendre les comportements de mobilités en confrontant l'offre (plan de transport) et la demande (charge) par apprentissage statistique.

Challenges :

- Données fortement structurées
- Hétérogénéité des pas temporels (passages de trains)
- Variabilité liée au plan de transport (mission/horaires)
- Capture de l'influence des facteurs calendaires
- Facteurs d'influence inconnus (événement, incident,...)
- Comportement spécifique par gare



3. Travaux :

Périmètre d'étude : train Paris-banlieue

Ligne H : multi-branches ~40 gares

Année : janvier 2015 – juin 2016

Données brutes :

- Information calendaire
- Plan de transport
- Charges de train

Features extraites et raffinées :

- Information calendaire (mois, jours, vacances)
 - Information horaire (heure théorique à la gare)
 - Information de plan transport (mission, desserte)
 - Retard à la gare
 - Charges des précédents trains
- LT Long-terme
CT Court-terme



Tâche : Prédiction long/court terme par gare

Plage temporelles : exhaustive (7j/7 24h/24)

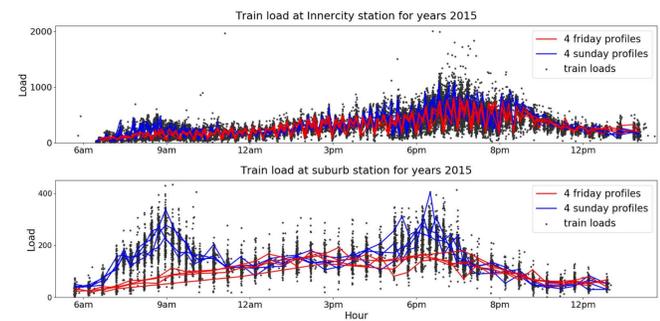
Nombre observations : plusieurs 10000 de train

Data set : deux gares (banlieue et centre-ville)

Train : année 2015 / **Test :** début d'année 2016

Modèles :

- DV : Last Values
- MC : Moyenne Contextuelle
- XGB LT : Modèle ensembliste avec features LT.
- XGB ST : Modèle ensembliste avec features CT + LT
- LSTM EP : Réseau récurrent encodeur-prédicteur



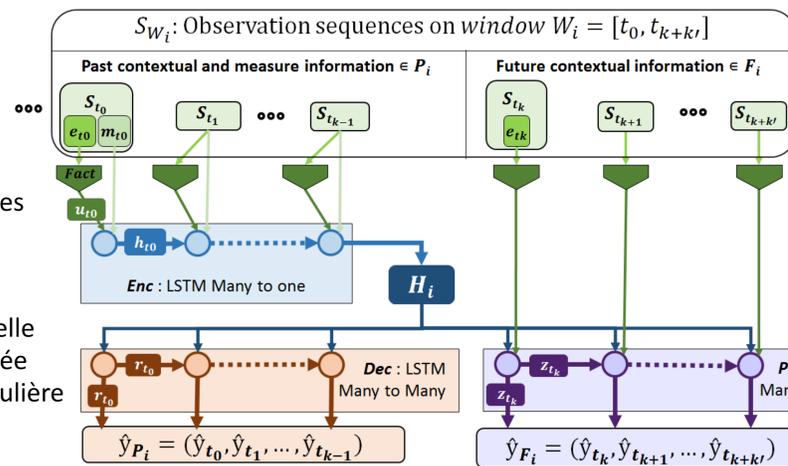
4. LSTM Encodeur-Prédicteur :

Motivation :

- Prise en compte de la structure séquentielle
- Meilleure abstraction dans la manipulation des features
- Prédiction sur un horizon temporel

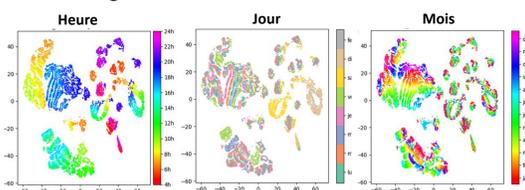
Choix :

- Réseau récurrent pour exploiter la structure séquentielle
- Encodage-décodage pour capturer la dynamique passée
- Embedding contextuel pour pallier à la structure irrégulière

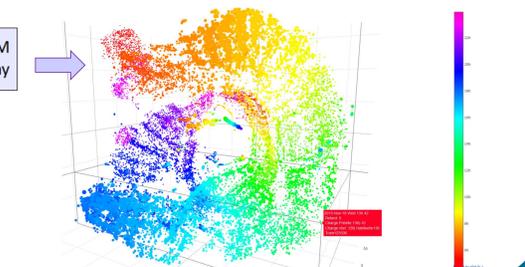


Visualisation des espaces latents par TSNE :

Embedding contextuel U réduit confronté à 3 contextes



Embedding prédictif Z réduit en 3D confronté aux horaires



1. Usine contextuelle
Embedding de contexte

Encoder
Capture de la dynamique passée

Objectif de Reconstruction
Structure encodeur-décodage

Objectif de Prédiction
Inférence de la dynamique future

4. Résultats : Performance prédiction t+1

Modèle	Banlieue		Centre-ville	
	WAPE	RMSE	WAPE	RMSE
Score d'apprentissage				
DV	17.9	35.8	41.9	186.7
MC	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 EP	10.7	22.1	10.9	57.7
Score de Test				
DV	24.1	47.2	46.9	205.0
MC	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 EP	16.0	33.8	12.9	72.4

Performance prédiction Mutli-step (Set de test)

Modèle	Gare de banlieue					
	t+1	t+2	t+3	t+4	t+5	t+6
Time interval	14-32	29-62	44-92	59-122	75-152	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 EP	33.8	34.0	34.1	34.4	34.7	34.9

Modèle	Gare de ville					
	t+1	t+2	t+3	t+4	t+5	t+6
Time interval	2-13	5-23	9-31	12-43	15-53	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 EP	72.4	72.1	72.1	72.2	72.6	72.8

5. Perspectives :

Détection d'anomalie de charge

Analyse de l'impact d'une perturbation

Prédiction à l'échelle du réseau