



HAL
open science

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

Kevin Pasini, Mostepha Khouadjia, Allou Same, Fabrice Ganansia, Patrice Aknin, Latifa Oukhellou

► **To cite this version:**

Kevin Pasini, Mostepha Khouadjia, Allou Same, Fabrice Ganansia, Patrice Aknin, et al.. Modèle LSTM encodeur-prédicteur pour la prévision court-terme de l'&apost;affluence dans les transports collectifs. CAP 2019, Conférence sur l'&apost;Apprentissage Automatique, Jul 2019, Toulouse, France. hal-02298337

HAL Id: hal-02298337

<https://hal.science/hal-02298337v1>

Submitted on 26 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Kevin Pasini^{1,2}, Mostepha Khouadja², Allou Samé¹, Fabrice Ganansia³, Patrice Aknin², and Latifa Oukhellou¹

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

²IRT SystemX, Paris-Saclay, France.

³SNCF- Innovation & Recherche, St Denis, France

Résumé

Les possibilités offertes en termes de collecte et de stockage de données permettent de renouveler les approches de modélisation dans le domaine du transport. L’exploitation croisée de différentes sources de données a pour vocation la création de services à forte valeur ajoutée pour l’usager. Les travaux détaillés dans cet article portent sur le développement de modèles de prévision à base de méthodes d’apprentissage notamment profond, pour la prévision court-terme de la charge (nombre de passagers) des trains. Cette prévision de l’affluence dans les trains peut servir à enrichir l’information voyageur à destination des usagers des transports collectifs qui peuvent ainsi mieux planifier leur déplacement. Elle peut également servir aux opérateurs de transport pour une régulation ”à la demande” de l’offre de transport. La principale difficulté dans la prévision est liée à la variabilité intrinsèque des séries temporelles des charges à prédire, induite par l’influence de plusieurs paramètres dont ceux liés à l’exploitation (horaire, retard, type de mission...) et au contexte (information calendaire, grand événement, météo,...). Nous proposons un modèle LSTM encodeur-prédicteur pour résoudre cette tâche de prévision. Plusieurs expérimentations sont menées sur des données réelles du réseau Transilien de la SNCF sur une durée d’un an et demi. Les résultats de prévision sont détaillés en vue de comparer les performances d’un tel modèle à plusieurs horizons temporels avec celles d’autres modèles plus classiques utilisés en prévision.

Mots-clef : Prévision, mobilité urbaine, séries temporelles, LSTM.

1 Introduction

La prévision de la demande de mobilité est un problème central dans l’organisation de tout système de transport. La capacité d’anticipation que confère un algorithme prédictif permet en effet de mieux planifier les déplacements et de spécifier à l’avance une offre adaptée, en s’aidant de critères usuels de type temps de parcours, ou de niveau de confort lié à l’affluence. A l’heure actuelle, l’offre pour les transports collectifs urbains fait déjà l’objet d’une adaptation à l’aide de méthodes de prévision de la demande, mais ces adaptations ont lieu longtemps en avance, en fonction d’informations calendaires, contextuelles et de fréquentations types estimées à partir d’enquêtes et de données de comptage collectés ponctuellement à des intervalles importants. L’introduction de nouvelles données temps réel ou quasi temps réel permet un suivi plus dynamique de la demande et elle ne fait qu’accroître l’intérêt pour des méthodes de prévision adaptées à ces nouvelles données.

Dans cet article, nous visons la prévision court-terme des charges (le nombre de passagers) des trains dans les transports collectifs. L’horizon temporel de prévision est court-terme, à savoir à 15mn, 30mn ou 1h. Cette prévision exploite deux sources de données hétérogènes : les données de comptage de passagers et les données de localisation automatique des trains.

Une grande majorité des travaux menés dans le domaine de la prévision de la demande en transport porte sur la prévision d’affluence dans les gares ou de flux avec un niveau agrégé (toutes les 15mn, 30mn, ...). L’originalité des travaux présentés réside dans la prise en compte du plan de transport réalisé pour la prévision, lequel peut s’éloigner assez fortement du plan de transport nominal. De fait, les séries tempo-

relles de charge que nous souhaitons prévoir possèdent un pas d'échantillonnage temporel variable, qui dépend de l'exploitation réelle prenant ainsi en compte les retards, les dessertes et missions des trains.

Une revue de littérature des travaux sur le sujet de la prévision d'affluence ou de charge conduit à distinguer plusieurs méthodologies selon les types de données utilisées et les objectifs visés en terme applicatif. Les premiers travaux sur la prévision du niveau de congestion dans les transports collectifs ont été proposés par [3] et [10]. En utilisant des techniques relativement simples basées sur des agrégations historiques des données billettiques, les auteurs de [3] ont proposé des modèles de prévision du niveau de congestion dans les transports publics. Dans l'article [10], les auteurs se sont intéressés à la recommandation tarifaire en y intégrant une étape de prévision des habitudes de déplacement. Ces dernières années, plusieurs travaux ont été menés sur le développement de modèles de prévision à base de méthodes d'apprentissage automatique. On citera les travaux de [12] où un réseau de neurone récurrent de type LSTM a été proposé pour la prévision court-terme de flux de passagers dans un réseau de transport. D'autres modèles plus complexes tenant compte des liens spatio-temporels entre les séries temporelles à prédire ont été proposés dans [9], [7] et [13]. Plus récemment, on notera les travaux de [8] qui s'attachent à la prévision de charges dans les tramways. Le problème est formalisé comme une tâche de classification où les différentes classes à inférer sont directement déduites du taux d'occupation des sièges dans les transports.

Dans cet article, la tâche de prévision est vue comme un problème de prévision à plusieurs horizons temporels sur des séries temporelles irrégulières impactées par plusieurs facteurs contextuels. Afin de prendre en compte ces spécificités et en s'appuyant sur les capacités d'abstraction des réseaux de neurones combinés à l'apprentissage de représentation [2], nous proposons un modèle LSTM encodeur-prédicteur combinée avec un apprentissage de représentation sur les facteurs contextuels. Le but est de fournir une prévision de la charge des trains au départ d'une station sur plusieurs pas de temps, tenant compte des valeurs de charges passées et de l'ensemble des facteurs contextuels caractérisant l'exploitation des trains.

L'article est organisé de la manière suivante : la section 2 décrit les données et les spécificités applicatives. la section 3 présente le modèle proposé pour la prévision tandis que la section 4 détaille les expérimentations menées sur des données réelles et les résultats obtenus en les comparant aux performances

de plusieurs modèles pour la prévision à un et plusieurs pas de temps. La section 5 conclut cet article en fournissant des perspectives à ce travail.

2 Description des données

Les données que nous considérons sont collectées sur une ligne de chemin de fer desservant une cinquantaine de gares situées au nord de la banlieue parisienne. Environ 250 000 passagers par jour sont transportés sur cette ligne. L'ensemble des données couvre une période de 18 mois, allant de janvier 2015 à juin 2016, sur 40 stations exploitées pendant la journée de 5h00 à 2h00 du matin. La base de données inclut à la fois des informations sur les horaires réalisés et sur les données de comptage des passagers à l'embarquement et au débarquement de chaque train stationné à chaque gare. Ces sources de données hétérogènes enrichies d'informations de calendrier, permettent de reconstituer le nombre de passagers à bord de chaque train (la charge) au départ d'une gare.

Le principal objectif des travaux concerne la prévision de séries de charges univariées pour chaque gare. Afin d'illustrer l'allure temporelle des séries à prédire, la figure 1 présente deux profils de charge recueillies sur deux gares. Cette figure met en évidence un certain nombre de spécificités que le modèle de prévision devra intégrer, à savoir :

- Une période d'échantillonnage variable en raison des horaires de train et de l'exploitation ferroviaire. Chaque gare a sa propre fréquence de passage des trains.
- Un profil temporel spécifique de chaque série. Le profil est directement lié à l'usage de la station et en particulier à sa localisation spatiale et aux caractéristiques géographiques de la zone urbaine qui l'entoure (densité de population, densité d'emploi, loisirs, etc.).
- Les séries de charge de train sont influencées par des facteurs de calendrier tels que le type de jour (jour de semaine ou week-end), jour férié, vacances scolaires, etc.
- Les séries de charge de train sont également impactées par les caractéristiques des trains et par leurs missions (ligne à destinations multiples, services ferroviaires variés).

Outre ces facteurs contextuels, la demande en transports en commun est également impactée par des événements (sociaux, culturels, sportifs, etc.). Le modèle de prévision pour chaque gare doit faire face à tous ces facteurs temporels, spatiaux et exogènes

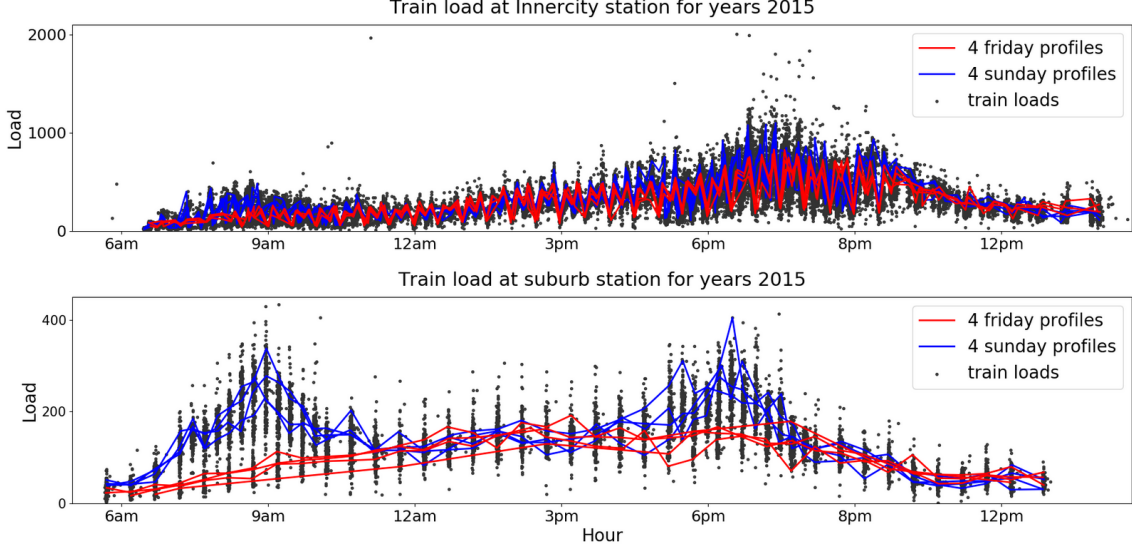


FIGURE 1 – Evolution des charges dans les trains sur l'année 2015 dans deux stations différentes

énumérés ci-dessus. La section suivante détaille la méthodologie développée pour réaliser la tâche de prévision.

3 Modélisation proposée

Nous proposons un modèle dédié à la prévision sur plusieurs horizons temporels des séries de charges avec les spécificités précitées. Ce modèle prend la forme d'un réseau de neurones récurrent auquel on associe un apprentissage de représentation du contexte, ce qui permet de capter les différents facteurs contextuels pouvant influencer la prévision. La Table 1 résume les notations utilisées dans cet article.

L'objectif est de fournir une prévision de la série temporelle (y_1, \dots, y_t) associée à une séquence d'observations S_t . Chaque élément de cette séquence est caractérisé par l'ensemble d'attributs contextuels e_t et de mesures passées m_t . On utilise la notation y_I, S_I, e_I pour désigner les sous-séquences $(y_t)_{t \in I}, (S_t)_{t \in I}, (e_t)_{t \in I}, (m_t)_{t \in I}$ avec $I \subset [1; T]$. Étant donné une fenêtre temporelle $W_i = [i - k, i + k']$ composée d'un horizon passé $P_i = [i - k, i]$, et d'un horizon futur $F_i = [i, i + k']$, le but de notre approche de prévision est d'inférer les réalisations y_{F_i} sur l'horizon futur à partir des informations disponibles sur la séquence d'observations $S_{W_i} = (e_{P_i}, e_{F_i}, m_{P_i})$ comme le montre la Figure 2. Le modèle doit être capable de capturer l'influence et les interactions des différents facteurs contextuelles sur les données historiques pour

inférer au mieux la dynamique future de la série (y_t) .

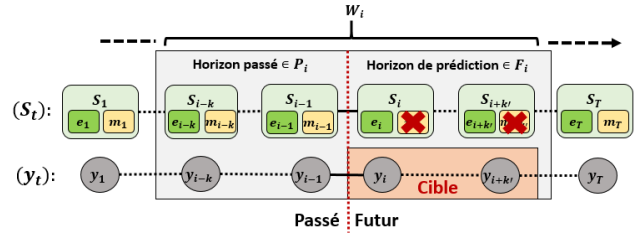


FIGURE 2 – Schéma général du modèle de prévision

S'inspirant des travaux de recherche sur les réseaux de neurones récurrents encodeur-décodeur proposés dans [5], nous proposons un modèle de réseau de neurones LSTM encodeur-prédicteur (LSTP EP) pour la prévision court-terme et multi-horizons incluant un apprentissage de représentation des facteurs contextuels. A partir d'observations sur une fenêtre temporelle S_{W_i} , le modèle reconstruit les k dernières réalisations \hat{y}_{P_i} et prévoit les k' prochaines réalisations \hat{y}_{F_i} en exploitant les informations contextuelles passées et futures e_{P_i}, e_{F_i} ainsi que les informations de mesure sur l'horizon passé.

$$\begin{aligned} LSTM_{EP}(X_i) &= LSTM_{EP}(m_{P_i}, e_{P_i}, e_{F_i}) \\ &= (\hat{y}_{P_i}, \hat{y}_{F_i}) \end{aligned} \quad (1)$$

Ce modèle peut être vu comme un réseau de neu-

TABLE 1 – Notation et variable

Notation	
t	Horizon temporel $t \in [1, T]$
y_1, \dots, y_T	(y_t) Séries de réalisation
S_1, \dots, S_t	(S_t) Séquence d'observations
e_1, \dots, e_T	(e_t) Séquence d'attributs contextuels
m_1, \dots, m_T	(m_t) Séquence d'attributs de mesures
Fenêtre	
W_i	$[i - k, i + k']$: i ème Fenêtre temporelle.
P_i	$[i - k, i]$: Horizon passé de W_i
F_i	$[i, i + k']$: Horizon futur de W_i
X_i	$(m_{P_i}, e_{P_i}, e_{F_i})$ Attributs d'entrée
Espace latent	
u_1, \dots, u_T	(u_t) Représentation contextuelles
h_1, \dots, h_T	(h_t) Dynamique latente passée
r_1, \dots, r_T	(r_t) État latent de reconstruction
z_1, \dots, z_T	(z_t) État latent de prévision
Modèle et ses composants	
$LSTM_{EP}$	Modèle de réseau de neurones
$Fact$	Modèle contextuel
Enc	Encodeur récurrent
Dec	Décodeur récurrent
$Pred$	Prédicteur récurrent
$Reconst$	Sorties de reconstruction
$Predict$	Sorties de prévision

rones profond pouvant se décomposer en plusieurs éléments ayant un rôle spécifique. Une illustration générale est fournie dans la Figure 3.

Les composants du réseau de neurones sont décrits comme suit :

Fact : un modèle de contexte dédié à synthétiser les caractéristiques (e_t) en une représentation contextuelle (u_t) . Il s'agit d'un pré-traitement basé sur un perceptron multi-couches appliqué à chaque observation afin de régulariser les représentations contextuelles.

$$Fact(e_{P_i}, e_{F_i}) = \bigoplus_{t \in (P_i \cup F_i)} Fact(e_t) = \bigoplus_{t \in (P_i \cup F_i)} u_t = (u_{P_i}, u_{F_i}) \quad (2)$$

Enc : un LSTM encodeur de type "many-to-one" dédié à la capture de la dynamique latente passée (h_i) à partir des mesures m_{P_i} et des représentations contextuelles (u_{P_i}) de l'horizon passé.

$$Enc(m_{P_i}, u_{P_i}) = h_i \quad (3)$$

Dec : un LSTM décodeur de type "many-to-many" décodant récursivement les états latents de recons-

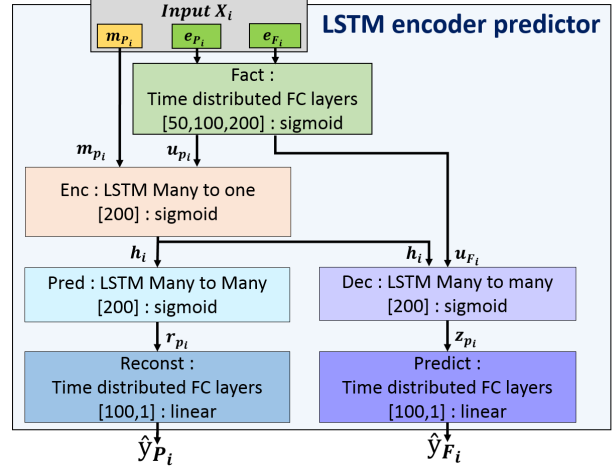


FIGURE 3 – Architecture générale du réseau LSTM encodeur-prédicteur

truction r_{P_i} des observations passées à partir de la dynamique latente de l'horizon passé (h_i) . Chaque état latent de reconstruction est ensuite interprété par des couches de reconstruction linéaire '**Reconst**' qui infèrent la réalisation des observations sur l'horizon passé.

A partir des sorties de '**Reconst**', on obtient \hat{y}_{P_i} qui est utilisé comme objectif intermédiaire durant la phase d'apprentissage pour faciliter la capture des dynamiques passées.

$$Dec(h_i) = r_{P_i} \quad (4)$$

$$Reconst(r_{P_i}) = \bigoplus_{t \in P_i} Reconst(r_t) = \hat{y}_{P_i} \quad (5)$$

Enc et **Dec** forment une structure encodeur-décodeur visant à synthétiser les dynamiques sur l'horizon passé à partir des attributs contextuels et des mesures sur cet horizon.

Pred : un LSTM prédicteur de type "many-to-many" inférant récursivement les états latents de prévision (z_{F_i}) des futures observations à partir de leurs représentations contextuelles (u_{F_i}) en prenant en compte la dynamique latente sur l'horizon passé (h_i) . Chaque état latent de prévision est ensuite interprété par des couches de reconstruction linéaire '**Predict**' pour inférer la réalisation de l'observation. A partir des sorties de '**Predict**', on obtient \hat{y}_{F_i} qui correspond

à l’objectif de prévision multi-horizons.

$$Pred(h_i, u_{F_i}) = z_{F_i} \quad (6)$$

$$Predict(z_{F_i}) = \bigoplus_{t \in F_i} Predict(z_t) = \hat{y}_{F_i} \quad (7)$$

Notons que le modèle est intrinsèquement conçu pour prendre en compte des pas d’échantillonnage variables, ce qui le rend robuste à la suppression d’observations (données manquantes).

3.1 Apprentissage des hyper-paramètres

Le réseau de neurone profond est entraîné à travers une rétropropagation du gradient minimisant la fonction de coût suivante :

$$\mathcal{L}(\theta) = \alpha_p * \sum_{t \in P_i} \|y_t - \hat{y}_t\|^2 + \alpha_f * \sum_{t \in F_i} \|y_t - \hat{y}_t\|^2 \quad (8)$$

$$\theta = (\theta_{Fact}, \theta_{Enc}, \theta_{Dec}, \theta_{Pred}, \theta_{Reconst}, \theta_{Predict}).$$

Le premier terme mesure la capacité du modèle à reconstruire les observations antérieures à partir de la dynamique latente du passé. C’est un objectif intermédiaire qui vise à faciliter l’apprentissage des dynamiques passées. Le second terme mesure la capacité de prévision du modèle. Les hyper-paramètres α_p et α_f pondèrent les objectifs de reconstruction et de prévision.

Pour l’apprentissage, nous réalisons une optimisation par mini-batch grâce à la technique d’optimisation *Nadam* [11]. Les deux gradients de prévision et de reconstruction sont propagés depuis leurs couches de sorties (Predict and Reconst) vers les couches en amont en passant par le modèle de contexte à travers les couches du LSTM. Le modèle LSTM encodeur-prédicteur est implémenté dans l’environnement *TensorFlow* [1], et *Keras* [6].

Les hyper-paramètres ont été choisis empiriquement après plusieurs expériences sur la base des performances et de la convergence de l’apprentissage. *Fact* est composé de 3 couches denses de tailles [50, 100, 200] avec une fonction d’activation sigmoïde pour un total de 27000 paramètres. *Enc*, *Dec*, and *Pred* sont 3 couches LSTM de taille 200 avec une fonction d’activation sigmoïde pour un total de 880000 paramètres. *Reconst* et *Predict* sont composés de 2 couches denses de tailles [100, 1] avec une fonction d’activation linéaire pour un total de 40000 paramètres. Le réseau entier compte approximativement 900000 paramètres.

L’entraînement est réalisé de manière empirique avec des batchs de taille 128 sur plusieurs milliers d’itérations, ce qui prend quelques heures sur un GPU standard selon le jeu de données en entrée et la profondeur temporelle considérée. Des investigations complémentaires sont nécessaires pour accélérer la convergence du modèle.

4 Résultats expérimentaux et discussions

Pour la partie expérimentale, nous évaluons deux modèles de prévision de charge dans les trains sur la base de deux jeux de données relatifs respectivement à une station du centre-ville parisien, et une station de banlieue. Ces jeux de données couvrent une période d’exploitation comprise entre janvier 2015 et juin 2016, et sont répartis en un ensemble d’apprentissage représentant 66 % du jeu de données et un ensemble de test correspondant aux 33% restants. Les deux modèles utilisent plusieurs informations contextuelles ”long-terme” de type calendaires :

- Jour de l’année : il correspond à la position du jour dans l’année (365 valeurs possibles encodées dans un vecteur de dimension 8 par l’application des fonctions cosinus et sinus sur 2x4 fréquences).
- Type de jour : il correspond à la position du jour dans la semaine et est encodé sur 8 dimensions avec une information supplémentaire si le jour est férié ou non.
- Minutes : Encodage des minutes journalières (1440 valeurs possibles) sur 8 dimensions par cosinus et sinus sur (2x4) fréquences.
- Service : caractéristique relative au service des trains en termes de mission et de branche desservie.

Par ailleurs, nous considérons également des caractéristiques sur un horizon court-terme en prenant en compte les observations antérieures sur une fenêtre couvrant les six derniers passages de trains à la station étudiée. Ces caractéristiques se résument à :

- Retard : Différence entre le temps théorique et le temps réel de passage du train à la station en minutes.
- Charge : Nombre de passagers dans le train pour chacun des six derniers passages de trains à la station.
- Montants : Nombre de passagers qui sont montés à bord des trains lors des six derniers passages à la station.
- Descendants : Nombre de passagers qui sont des-

cendus des trains lors des six derniers passages à la station.

4.1 Modèles de prévision

Dans cette section, les performances du modèle LSTM encodeur-prédicteur entraîné sur les deux catégories de caractéristiques e_t et m_t sont comparées avec celles d'autres modèles classiques issus de la littérature et qui vont servir de référence à cette comparaison. Ces modèles se résument en :

- **Dernière Valeur (DV)** : un modèle de prévision simple qui consiste à renvoyer la dernière charge observée dans le train comme étant la prochaine charge à prévoir à la même station.
- **Moyenne Contextuelle (MC)** : cette approche considère la charge moyenne constatée sur le même type de jour et la même tranche horaire comme étant la charge à prévoir.
- **Gradient Boosting (XGB)** [4] : modèle de régression appartenant à la famille des méthodes ensemblistes et qui regroupe des arbres de décision dits faibles. Deux modèles sont proposés selon la nature des caractéristiques employées en entrée du modèle :
 - **XGB LT** : désigne le modèle XGB entraîné sur la base des caractéristiques long-terme e_t .
 - **XGB ST** : désigne quant à lui le modèle XGB entraîné sur la base des caractéristiques court-terme. m_t .

Pour l'ajustement des hyper-paramètres des modèles XGB, une recherche par grille ("search-grid") est réalisée en utilisant une validation croisée avec $k=3$ (3-fold). Enfin, nous évaluons les performances des modèles sur chaque pas de temps correspondant au passage d'un train en utilisant comme mesure de performances l'erreur quadratique moyenne (RMSE) et le pourcentage d'erreur absolue pondérée (WAPE). Ce dernier peut être interprété comme le pourcentage de l'erreur totale comparée à la valeur moyenne de l'observation actuelle, et est calculé comme suit :

$$WAPE\ score : \frac{\sum_t ||y_t - \hat{y}_t||}{\bar{y}} \quad (9)$$

4.2 Évaluation

L'évaluation des modèles prédictifs est basée sur une comparaison des performances obtenues par ces derniers, et est exprimée à l'aide des métriques décrites dans la section précédente. Ces mesures sont calculées

à la fois sur l'ensemble d'apprentissage et l'ensemble de test pour les deux stations étudiées, à savoir celle du centre ville et celle de banlieue. Les cinq modèles décrits dans la section 4.2 : Dernière valeur (DV), Moyenne Contextuelle (MC), le Gradient Boosting (XGB) dans ses deux variantes (ST) et (LT), ainsi que le LSTM EP sont comparés. Le Tableau 2 décrit les performances des différents modèles sur les stations étudiées.

TABLE 2 – Performances des modèles sur les stations étudiées

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

Comme attendu, les performances des modèles avancés (XGB, LSTM EP) surpassent celles des modèles DV et MC. Ces performances peuvent s'expliquer par le fait que les modèles XGB et LSTM EP ont une capacité de généralisation meilleure que les modèles de référence, ce qui permet d'éviter le problème de sur-apprentissage sur les données. Par ailleurs, les modèles XGB et LSTM EP sont plus complexes que les modèles de référence qui se contentent juste de renvoyer la dernière valeur observée ou la moyenne de la charge enregistrée sur un historique donné. Sur l'ensemble, le LSTM EP fournit de meilleurs résultats puisqu'à l'aide des caractéristiques court-terme, il est capable de capturer la dynamique du service des trains à la station.

En examinant l'erreur de prévision du LSTM EP sur la charge prédite (voir Figure 5), nous pouvons observer que le taux d'erreur croît avec l'augmentation de la charge à bord. Le modèle tend à surestimer légèrement les trains faiblement chargés, et à sous-estimer les trains surchargés. Ceci peut être expliqué par le fait que les trains fortement chargés sont peu nombreux, et présentent des informations contextuelles similaires à celles des trains faiblement chargés, ce qui rend difficile la prévision des charges élevées.

Comme l'illustre la Figure 4, l'erreur commise par

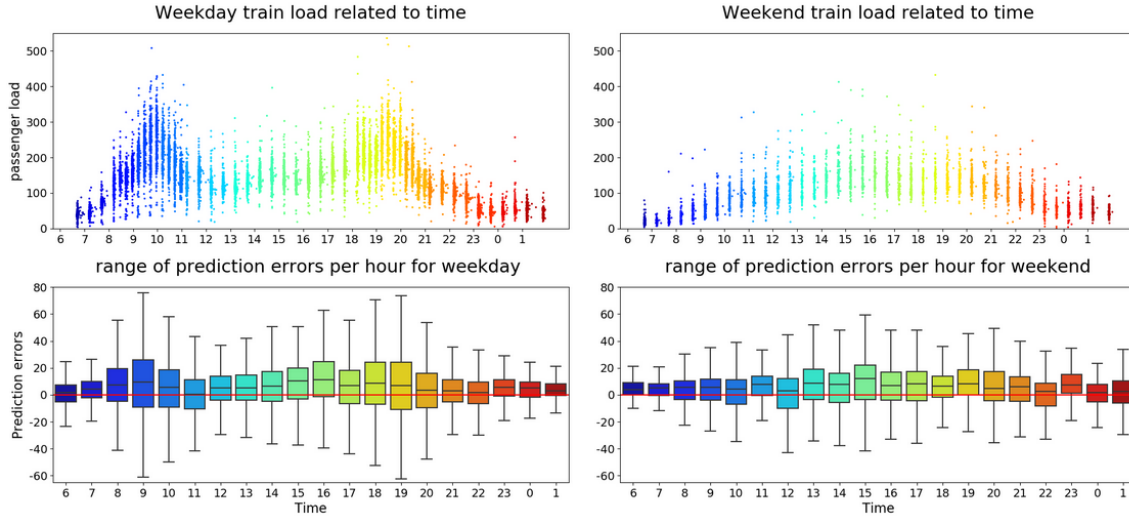


FIGURE 4 – L’erreur de prévision en fonction de la plage horaire pour la station de banlieue.

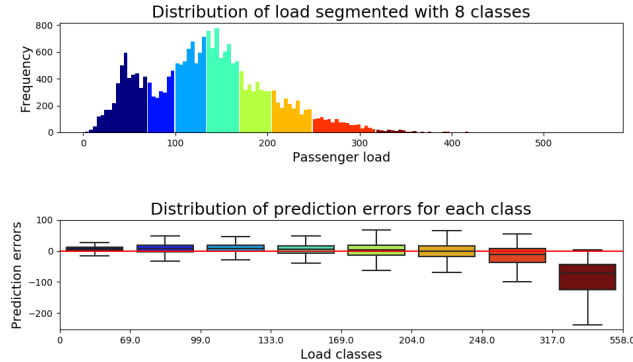


FIGURE 5 – L’erreur de prévision en fonction du segment de charge pour la station de banlieue.

le modèle est du même ordre de grandeur pour les jours de semaine que pour les jours de week-end avec néanmoins quelques différences. La variance de l’erreur calculée sur les créneaux horaires de la journée s’avère corrélée avec la charge à bord. Sur les jours de semaine, on constate des erreurs importantes durant les heures de pointe en lien avec une variance importante et une charge élevée. Le modèle commet des erreurs moyennes en milieu de journée et des erreurs faibles tôt le matin et en fin de soirée. Par ailleurs, durant les week-ends excepté en matinée, on observe une variance de l’erreur relativement stable avec une valeur maximale atteint à midi et au milieu de l’après-midi. On note également que le modèle a

plus de difficulté à prévoir les soirées de week-end que celles des jours de semaine.

D’autre part, lorsqu’on examine les performances des modèles de prévision sur des pas temporels successifs (multi-horizons), le LSTM EP surpasse XGB sur les six horizons (Tableau 3 et Tableau 4).

TABLE 3 – RMSE sur la station de banlieue pour les modèles de prévision multi-horizons.

Modèle	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

TABLE 4 – RMSE sur la station du centre-ville pour les modèles de prévision multi-horizons.

Modèle	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

Ces horizons correspondent aux prochains passages des trains à la station et varient entre 14 et 182 minutes pour la station de banlieue, et entre 2 et 61 minutes pour la station du centre ville. Cette différence

est due à une fréquence de trains plus élevée en centre ville qu'en banlieue. Il est important de souligner que ces prévisions sont obtenues avec un modèle LSTM EP unique qui prévoit simultanément la charge dans les trains sur l'ensemble des horizons temporels, tandis que pour le modèle XGB, nous avons autant de modèles que d'horizons temporels. Les prévisions fournies par le modèle XGB LT restent invariantes au fil des horizons temporels. Cependant, les performances de la variante XGB ST se dégradent avec l'avancement dans le temps. Le LSTM EP maintient des prévisions compétitives et robustes sur l'ensemble des horizons temporels aussi bien en station du centre-ville qu'en banlieue. Ceci peut être expliqué par une meilleure compréhension des facteurs contextuels à travers la représentation latente qui contribue à capturer la dynamique sous-jacente du service lié aux trains qui desservent la station.

5 Conclusion

Dans cet article, un modèle de prévision à base d'un réseau de neurones LSTM encodeur-prédicteur (LSTM EP) associé à un apprentissage de représentation des facteurs contextuels a été proposé. Ce modèle vise à prévoir la charge dans les trains à plusieurs horizons temporels en tenant compte des facteurs contextuels et des horaires réels de passage des trains dans les stations. Les difficultés posées par ce problème de prévision sont induites par une variabilité de l'exploitation ferroviaire et par plusieurs facteurs contextuels. Le réseau de neurones que nous proposons a la particularité de pouvoir apprendre une représentation à partir de caractéristiques contextuelles, de capturer la dynamique passée latente à travers la sous-structure sous-jacente de l'encodeur-décodeur, puis de prévoir la dynamique à venir à l'aide de la couche prédictive.

Au travers des résultats obtenus sur un jeu de données réelles couvrant une année et demi d'exploitation, nous avons pu montrer le potentiel du modèle proposé à traiter la prévision à court terme de la charge des trains en comparaison à d'autres modèles tels que le modèle gradient boosting. Les performances du modèle proposé ont été évaluées sur deux gares avec des profils temporels différents et pour des horizons de prévision à un et plusieurs pas de temps. Sur les deux configurations, le LSTM-EP présente une meilleure robustesse de la qualité des prévisions fournies.

Les futurs travaux devront s'attacher à l'exploration de la représentation apprise, et en particulier, la capacité de l'espace latent prédictif à caractériser des situations anormales telles que des perturbations et des anomalies

de trafic.

Références

- [1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow : A system for large-scale machine learning. *OSDI Symposium on Operating Systems Design and Implementation*, pages 265–283, 2016.
- [2] Y. Bengio, A. Courville, and P. Vincent. Representation learning : A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8) :1798–1828, 2013.
- [3] I. Ceapa, C. Smith, and L. Capra. Avoiding the crowds : Understanding tube station congestion patterns from trip data. *ACM SIGKDD International Workshop on Urban Computing*, pages 134–141, 2012.
- [4] T. Chen and C. Guestrin. Xgboost : A scalable tree boosting system. *SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.
- [5] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *EMNLP Empirical Methods in Natural Language Processing*, page 1724–1734, 2017.
- [6] F. Chollet et al. Keras. 2015.
- [7] C. Ding, D. Wang, X. Ma, and H. Li. Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8(11) :1100, 2016.
- [8] L. Heydenrijk-Ottens, V. Degeler, D. Luo, N. van Oort, and J. van Lint. Supervised learning : Predicting passenger load in public transport. 2018.
- [9] J. Ke, H. Zheng, H. Yang, and X. M. Chen. Short-term forecasting of passenger demand under on-demand ride services : A spatio-temporal deep learning approach. *Transportation Research Part C : Emerging Technologies*, 85 :591–608, 2017.
- [10] N. Lathia and L. Capra. Mining mobility data to minimise travellers' spending on public transport. *17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1181–1189, 2011.
- [11] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton. On the importance of initialization and

- momentum in deep learning. *ICML International Conference on Machine Learning*, 28(1139-1147) :5, 2013.
- [12] F. Toqué, E. Côme, M. K. El Mahrsi, and L. Oukhellou. Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks. *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1071–1076, 2016.
- [13] A. Ziat, E. Delasalles, L. Denoyer, and P. Gallinari. Spatio-temporal neural networks for space-time series forecasting and relations discovery. *IEEE International Conference on Data Mining ICDM*, pages 705–714, 2017.