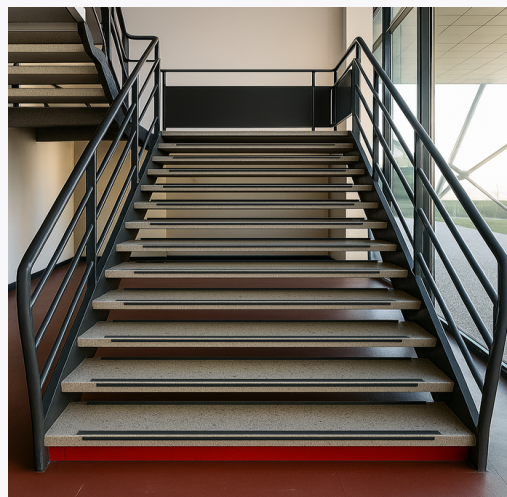


Portfolio

2019 – 2024



Unité de Recherche
Universités Lyon 2 et Lyon 1

Conception & mise en page

Textes

Crédits photos

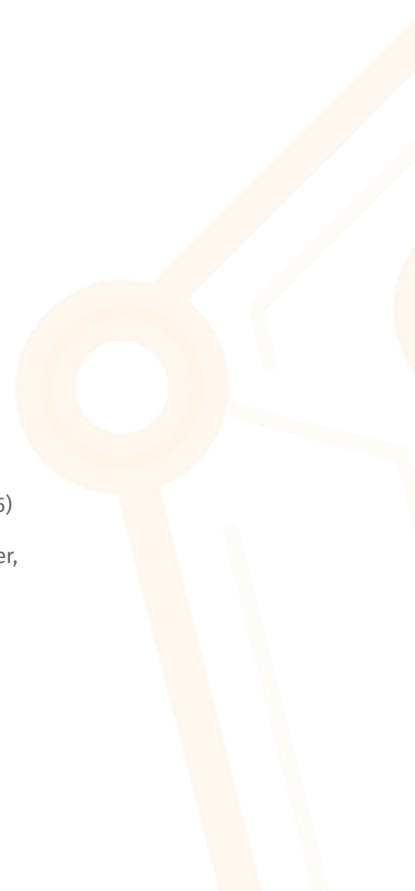
Portfolio du Laboratoire ERIC (2019 - 2024)

© **Laboratoire ERIC**, Unité de Recherche, Universités Lyon 2 et Lyon 1 (2025)

J. Cugliari

F. Bentayeb, J. Cugliari, J. Darmont, C. Favre, A. Guille, J. Jacques, S. Loudcher,
M.-L. Messai, A. Rolland et J. Velcin

J. Cugliari



Fondé en 1995, le laboratoire ERIC est une unité de recherche de l'Université Lumière Lyon 2 et de l'Université Claude Bernard Lyon 1, située sur le campus Porte des Alpes à Bron. Fort de 24 membres permanents et autant de doctorants, le laboratoire ERIC développe des activités de recherche en informatique et en mathématiques appliquées. Répartis en deux équipes, les chercheurs d'ERIC travaillent dans les domaines du machine learning, de l'aide à la décision, de la gestion et de l'analyse des données massives, avec une forte expertise dans le domaine des humanités numériques.

Ce portfolio présente des exemples de productions de notre laboratoire pour la période 2019-2024 : quelques publications scientifiques, ainsi que des organisations de manifestations, un exemple de projet de recherche, un dépôt de brevet... Ces éléments vous offriront un aperçu de notre activité au sein du laboratoire ERIC.

Je vous souhaite une agréable lecture.



Prof. Julien JACQUES
Directeur du Laboratoire ERIC.



Utilisateur.trices

- Problématiques de recherche
- Connaissances

Axe transversal Humanités numériques

Collaborations pluridisciplinaires

Autres terrains d'application :
industrie, énergie, santé...

Aide à
la décision

Enrichissement
des connaissances

Équipe
SID

Équipe
DMD

Big data management

Big data analytics

Analyse en ligne (OLAP)

Sécurité des données

Intelligence artificielle

Machine learning

Data mining

Agrégation multicritère

ENTREPÔT
OU LAC DE
DONNÉES

Intégration

volumineuses
variées
véloces

BIG DATA

imprécises
dynamiques
imparfaites



Brevet n° : FR3143115, 2024

Procédé et dispositif de pesée notamment pour la pesée de véhicules montés sur roues

Martial Amovin-Assagba, Julien Jacques, Irène Gannaz, Frédéric Fossi

Ce brevet est issu d'une collaboration entre l'entreprise Arpege Master K, acteur majeur dans le domaine du pesage industriel, le laboratoire ERIC et l'Institut Camille Jordan. Cette collaboration a pris la forme d'un premier contrat de recherche en 2018, suivi par la thèse CIFRE de Martial Amovin-Assagba de 2020 à 2023. Martial est à ce jour toujours dans l'entreprise sur un poste de data scientist.

La pesée de véhicules sur roues est effectuée à l'aide d'un pont-basculé comportant un plateau disposé sur un support, des capteurs de pesée étant fixés entre le plateau et le support pour mesurer le poids exercé par le plateau sur le support. Les signaux analogiques issus des capteurs peuvent alors être numérisés et constituent un ensemble de courbes, chaque courbe étant représentative des variations du signal issu d'un des capteurs.



L'analyse de données fonctionnelles est appropriée pour analyser de tels signaux issus des capteurs, notamment parce que ces données sont acquises à une fréquence élevée, de l'ordre d'une ou plusieurs centaines de Hertz.

Afin d'assurer la fiabilité des mesures, la détection de mesures erronées est un problème crucial.

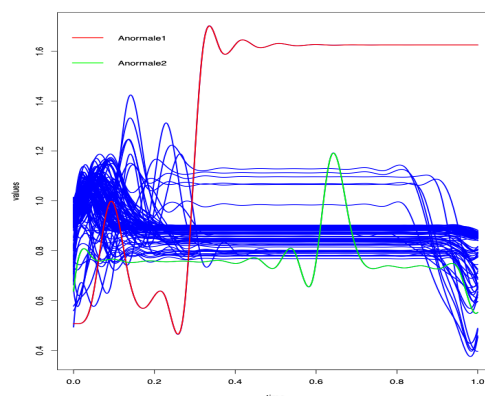
Pour ce faire, nous avons développé dans l'article [1] un modèle qui simultanément regroupe les mesures en clusters homogènes et détecte la présence d'anomalies. L'expérience a montré que ce modèle permet de détecter des anomalies dans les mesures, avec des taux de faux positifs et de faux négatifs très bas. En outre, il présente l'avantage de pouvoir s'affiner au fur et à mesure de son exécution, de ne pas nécessiter de fixer une proportion d'anomalies, ni de fixer un nombre de groupes.

Fig. 1: Camion en cours de pesée sur un pont développé par Arpege Master K. Lorsqu'un véhicule monte sur le pont, un premier train de roues du véhicule sollicite simultanément deux capteurs situés à l'entrée du pont, puis simultanément deux capteurs de sortie. Ces capteurs sont reliés à une unité de traitement. Le traitement des signaux mesurés (Figure 2) par les capteurs permet de déterminer le poids du véhicule.

Fig. 2: Cette figure représente un ensemble de mesures relevées sur un des 4 capteurs.

Les courbes bleues représentent des pesées normales : on constate la montée en charge des capteurs en début de pesées, lorsque le véhicule monte sur le pont, la stabilisation lorsqu'il s'arrête sur le pont, puis la descente en charge lorsque le véhicule descend du pont. On observe une grande variabilité dans ces courbes, du fait du tonnage du camion, du nombre d'essieux, ...

Les courbes rouges et vertes correspondent à deux pesées anormales, dues à deux défaillances différentes de capteurs. Ces courbes anormales, si elles ne sont pas détectées, peuvent induire une erreur de pesée.



[1] Outlier Detection in Multivariate Functional Data Through a Contaminated Mixture Model. *Computational Statistics and Data Analysis*, 2022, 174, 107496.



Fête de la science, 2 octobre 2020

Musée des moulages, université Lyon 2
Laboratoire ERIC

Le laboratoire ERIC est engagé dans une démarche active de médiation scientifique. À travers des actions individuelles et collectives, nous participons régulièrement à des événements et à des publications visant à diffuser les connaissances auprès du grand public.



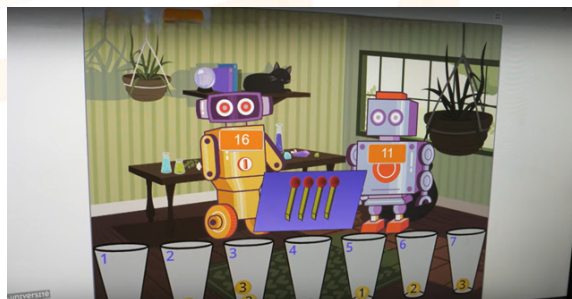
À l'occasion de la Fête de la Science 2020, nous avons souhaité mettre en lumière la richesse de nos travaux à l'intersection de l'intelligence artificielle et des sciences humaines et sociales. Le Musée des Moulages de l'Université Lumière Lyon 2 a offert un cadre idéal pour aller à la rencontre d'un public curieux et engagé.

Plusieurs activités ludiques, directement inspirées des recherches menées au sein du laboratoire, ont été proposées au public. Une série d'animations permettait de découvrir différents aspects de l'intelligence artificielle, du stockage de données à leur analyse décisionnelle. Le public pouvait ainsi se glisser dans

la peau d'un programmeur pour faire « penser » un robot, construire un chatbot (programme informatique capable de simuler une conversation vocale ou textuelle avec un humain), ou encore s'initier aux méthodes de décision collective.

Un temps d'échange avec des chercheurs était également proposé à travers un speed dating scientifique ou encore grâce au jeu des 7 familles de la recherche. L'un des stands présentait de manière accessible le fonctionnement des modèles de langage de grande taille (LLM — Large Language Models), illustré par une démonstration interactive sur ordinateur. Une dizaine d'autres ateliers ont permis de couvrir un large éventail des thématiques de recherche du laboratoire.

Enfin, une animation originale permettait aux plus jeunes de comprendre les bases de l'intelligence artificielle sans ordinateur, en rejouant plusieurs parties du jeu de Nim. Un simple dispositif de gobelets en plastique à remplir ou à laisser vides selon les résultats permettait d'illustrer les principes de l'apprentissage par renforcement, complété par une animation interactive sur écran pour renforcer l'apport théorique.





Article à la conférence The Web Conference (WWW 2019)

Global Vectors for Node Representations

Robin Brochier, Adrien Guille, Julien Velcin

The Web Conference est une des plus grandes conférences en informatique (notée A* dans le CORE), portant sur le futur du Web, organisée par l'ACM et dont l'édition 2019 s'est tenue à San Francisco. Les travaux présentés à cette conférence visent à analyser ou modéliser divers phénomènes observés sur le Web, la plupart du temps en ayant recours à l'apprentissage automatique.

Dans l'article "Global Vectors for Node Representations", nous proposons une méthode (GVNR-t) pour apprendre à représenter des documents textuels organisés en réseau (cf. Figure 3) dans un espace vectoriel de faible dimension, et ce en tenant compte de l'information de dépendance donnée par les liens entre documents. Notre proposition s'appuie sur GloVe, une méthode conçue pour l'apprentissage de représentations de mots basée sur une décomposition matricielle. D'une part, nous modifions la formulation du modèle afin de représenter dans un même espace les mots (une représentation par mot) et les documents (deux représentations par document). D'autre part, nous proposons une démarche originale pour apprendre ces représentations. Nous construisons la matrice à décomposer à l'aide de marches aléatoires suivant la structure du réseau et nous affinons l'objectif guidant le calcul de la décomposition en incorporant notamment un terme sous-échantillonnant les valeurs d'après lesquelles l'erreur de reconstruction est mesurée. Cet échantillonnage vise à limiter le coût calculatoire mais

aussi à émuler le principe d'échantillonnage négatif employé dans d'autres types d'approches pour l'apprentissage de représentations.

Une part significative du Web est composée de pages interconnectées par des hyperliens, formant un réseau de documents. En appliquant notre méthode à un tel réseau, nous pouvons obtenir des représentations des documents pouvant servir à la résolution de tâches plus spécifiques nécessitant une représentation préalable dans un espace vectoriel de faible dimension. Nous pouvons aussi résoudre directement des tâches concrètes en nous basant sur des mesures dans l'espace de représentation. Par exemple, en mesurant des distances entre les représentations des mots et les représentations de documents, nous pouvons obtenir des recommandations de mots-clés. Chaque document étant associé à deux représentations, une orientée réseau et une orientée texte, on peut identifier les mots les plus proches de ces deux représentations et obtenir deux jeux de mots-clés complémentaires. La Table 1 présente un exemple, avec les mots-clés extraits pour une page présentant une revue de la littérature à propos du Web 2.0 et ses fonctionnalités. On remarque que les mots-clés liés à la représentation orientée réseau mettent l'accent sur des problématiques, tandis que les mots-clés liés à la représentation orientée texte mettent l'accent sur la nature des données à traiter.

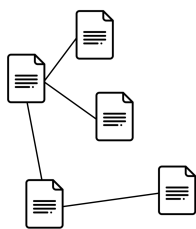


Fig. 3: La méthode GVNR-t permet d'apprendre des représentations de documents organisés en réseau. Cela nécessite la prise en compte à la fois du contenu textuel des documents et des relations entre ces documents.

Tab. 3: Mots-clés identifiés pour un document à propos du Web 2.0, à l'aide des représentations apprises avec la méthode GVNR-t. Les mots-clés orientés réseau ciblent les problématiques soulevées dans le document, alors que les mots-clés orientés texte ciblent la nature des données à traiter.

Document (extrait)	A brief survey of computational approaches in social computing Web 2.0 technologies have brought new ways of connecting people in social networks for collaboration in various on-line communities. Social Computing is a novel and emerging computing paradigm...
Mots-clés réseau	<i>cold start problem, computational humor, storylines, document titles</i>
Mots-clés texte	<i>social, social networks, extremely large datasets, sites blogs</i>

The World Wide Web Conference, San Francisco, CA, USA, 2019, pp. 2587-2593.

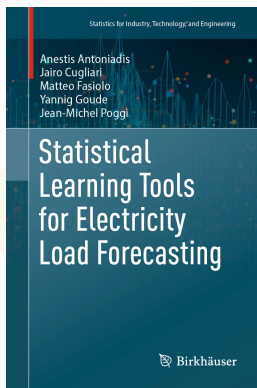


Livre publié en 2024

Statistical Learning Tools for Electricity Load Forecasting

Anestis Antoniadis, Jairo Cugliari, Matteo Fasiolo, Yannig Goude, Jean-Michel Poggi

La série Statistics for Industry, Technology, and Engineering series de l'éditeur Birkhäuser Cham est une collection de publications académiques éditée par Springer. Elle se concentre sur la présentation d'idées et de méthodes statistiques modernes pertinentes pour les chercheurs et les praticiens dans les domaines de l'industrie, de la technologie et de l'ingénierie.



Cette monographie explore un ensemble d'outils statistiques et d'apprentissage automatique pouvant être efficacement utilisés pour l'analyse de données appliquée dans le contexte de la prévision de la charge électrique. S'appuyant sur leur vaste expérience de la prévision de la demande en électricité dans des environnements industriels, les auteurs guident les lecteurs à

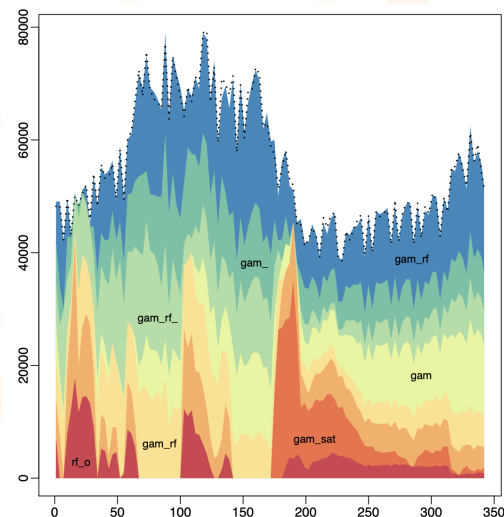
travers plusieurs méthodes et outils de prévision modernes, tant d'un point de vue industriel qu'appliqué : modèles additifs généralisés (GAMs), GAMs probabilistes, séries temporelles fonctionnelles et

ondelettes, forêts aléatoires, agrégation d'experts et modèles à effets mixtes.

Une collection d'études de cas basées sur des ensembles de données de grande taille et à haute résolution, accompagnée de packages R pertinents, illustre ensuite la mise en œuvre de ces techniques. Cinq ensembles de données réelles, à trois niveaux d'agrégation différents (national, régional ou individuel) et issus de quatre pays (Royaume-Uni, France, Irlande et États-Unis), sont exploités pour étudier cinq problèmes : prévision à court terme en un point donné, sélection des variables pertinentes pour la prédiction, construction d'intervalles de prédiction, prévision des pics de demande et utilisation des données de consommateurs individuels.

Ce texte s'adresse aux praticiens, chercheurs et étudiants en post-gradué travaillant sur la prévision de la charge électrique. Il peut également intéresser les universitaires appliqués ou les scientifiques souhaitant découvrir des outils de prévision de pointe pour les appliquer à d'autres domaines. Il est supposé que les lecteurs soient familiers avec les concepts statistiques de base, tels que les variables aléatoires, les fonctions de densité de probabilité et les valeurs espérées, et qu'ils possèdent une expérience minimale en modélisation.

Fig. 4: Prévision journalière de la demande en électricité en France durant l'année 2020 par un modèle de mélange d'experts. À partir du mois d'avril, l'expert modélisant l'ensemble des jours comme des samedis obtient un poids significatif. Cela s'explique par l'impact du confinement lié à la pandémie de COVID-19, qui a profondément modifié les habitudes de consommation d'électricité.





La Conférence internationale IEEE sur le Data Mining (ICDM) est l'une des conférences les plus reconnues (rang A*) de recherche mondiale dans le domaine du data mining.

Les modèles Stochastiques en Blocs (SBM) constituent une famille de modèles de clustering populaire, en vertu de leur interprétabilité et de la qualité de leurs résultats sur des tâches prédictives. Ces modèles sont perçus comme une alternative flexible aux techniques de décomposition tensorielle, notamment grâce à leur capacité à gérer des données étiquetées (par opposition à des données scalaires). Ces dernières années ont vu un regain d'intérêt pour leur utilisation, notamment dans les systèmes de recommandation.

Aussi, plusieurs travaux récents se sont attachés à leur amélioration, en particulier visant à étendre la taille du contexte fourni comme donnée d'entrée et à considérer des interactions entre *paires* d'éléments dudit contexte.

Dans cet article, nous montrons que ces modèles peuvent être formulés comme des cas particuliers d'un modèle plus général : le Serialized Interacting Mixed membership Stochastic Block Model

(SIMSBM). Ce modèle permet de représenter un contexte de taille arbitraire ainsi que des interactions d'un ordre arbitrairement élevé. Nous montrons comment retrouver plusieurs modèles de la littérature à partir du SIMSBM, et démontrons que notre formulation permet d'obtenir de meilleures prédictions que les cas particuliers qu'il généralise sur cinq jeux de données réelles traitant de tâches diverses – diagnostics médicaux, mécanismes de prise de décision humaine, et recommandation de musiques ou de films.

C'est en effet par sa flexibilité que le modèle présenté ici se distingue. En permettant la prise en compte de contextes et d'ordres d'interaction arbitrairement élevés, son champ d'application possible dépasse largement le cadre des systèmes de recommandation. Aussi, sa première vocation est de fournir à des chercheurs de tous horizons un cadre méthodologique formel à l'aide duquel étudier des corpus dont ils sont les spécialistes –ce qu'illustre une récente étude en archéologie (AFAM, 2024) faisant usage du SIMSBM dans le but d'inférer le sexe et l'âge de défunts d'époque mérovingienne à partir du seul petit mobilier funéraire qui les accompagne.

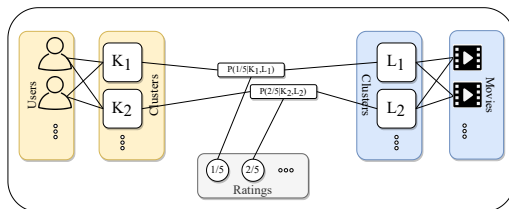
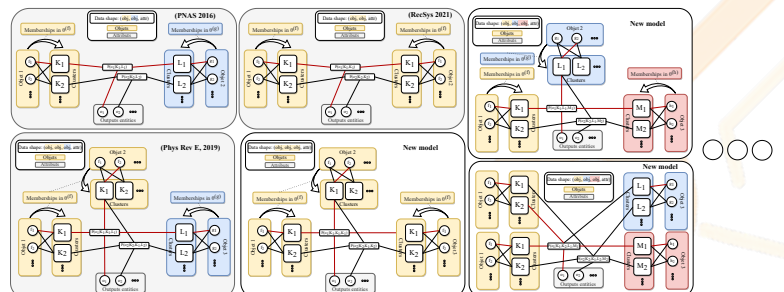


Fig. 6: Quelques itérations du SIMSBM –

Les couleurs représentent la nature des informations prises en compte, par exemple : rouge=film, jaune=utilisateur, bleu=acteur principal du film et gris (sortie)=note donnée au film. Lorsqu'une couleur est dupliquée, on parle d'interaction entre informations de même nature, par exemple si on choisit de considérer deux acteurs plutôt qu'un seul, l'effet de leur interaction sur les notes données au film sera également modélisé par le SIMSBM.

Fig. 5: Principe du SIMSBM – On suppose que des groupes de données étiquetées, ici d'utilisateurs et de films, auront tendance à interagir de façon similaire, ici en ce qui regarde la note donnée à un film par un utilisateur. Le SIMSBM permet d'inférer la note donnée à un groupe de films par un groupe d'utilisateurs, sans aucune information préalable quant à la composition des groupes.





Conférences organisées : ADBIS, TPDL & EDA 2020

Feedback from the participants of the ADBIS, TPDL and EDA 2020 joint conferences

Pegdwendé N. Sawadogo, Jérôme Darmont, Fabien Duchateau

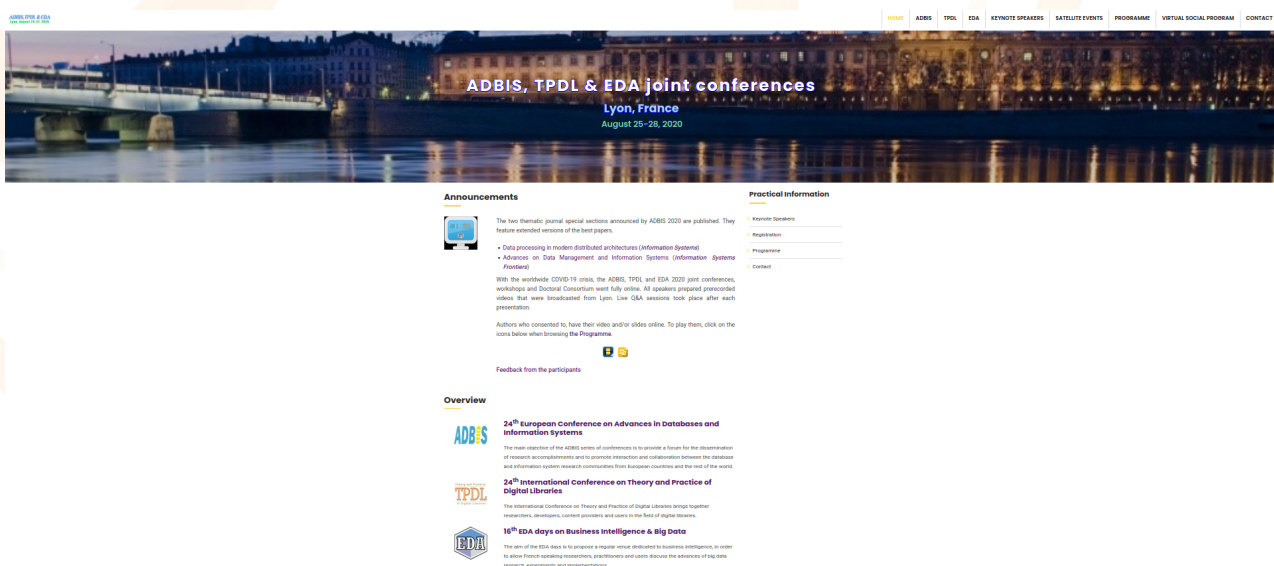
La 24th European Conference on Advances in Databases and Information Systems (ADBIS 2020), la 24th International Conference on Theory and Practice of Digital Libraries (TPDL 2020) et la 16th EDA days on Business Intelligence & Big Data (EDA 2020) ont été conjointement organisées du 25 au 27 août 2020. L'événement a également accueilli un Consortium Doctoral (DC) commun et six ateliers : 1st Workshop on Assessing Impact and Merit in Science (AIMinScience 2020) ; 2nd International Workshop on BI & Big Data Applications (BBIGAP 2020) ; International Workshop on Intelligent Data – From Data to Knowledge (DOING 2020) ; Modern Approaches in Data Engineering and Information System Design (MADEISD 2020) ; 10th International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA 2020) ; Scientific Knowledge Graphs (SKG 2020).

En raison du contexte de la pandémie mondiale, le comité d'organisation et les comités de pilotage respectifs ont décidé d'organiser l'événement en ligne. Ce choix a conduit à de grands défis dans le choix des bons outils pour garantir que les participants tirent le meilleur parti des conférences. Nous avons choisi de faire se dérouler les conférences à peu près aux mêmes dates que celles prévues pour la réunion physique à Lyon. Les journées de conférence et les présentations ont été raccourcies, car assister à des présentations en streaming est fatigant. Le programme a été centré autour de l'heure du

déjeuner pour permettre à autant de participants que possible d'assister, en particulier les chercheurs d'Amérique et d'Asie. Les frais d'inscription ont été réduits en conséquence et les participants sans article pouvaient s'inscrire gratuitement.

Le système de visioconférence que nous avons utilisé était Big Blue Button (BBB). La plateforme était hébergée à l'Université de Lyon 2, ce qui nous a permis d'avoir un meilleur contrôle sur l'infrastructure et le système. Nous avons demandé aux orateurs de préparer une présentation pré-enregistrée au format MP4. La durée des présentations était la suivante : présentations principales 20 minutes, articles longs 10 minutes et articles courts 5 minutes. Nous avons assisté les présidents de séance avec des « projectionnistes » chargés de diffuser les présentations pré-enregistrées. Les sessions de Questions et Réponses (Q&R) après les présentations étaient en direct, uniquement par chat, et modérées par le président de séance comme d'habitude. La durée des sessions de Q&R était la suivante : présentations principales 10 minutes, articles longs 5 minutes et articles courts 5 minutes.

Globalement, l'événement a attiré plus de 500 participants inscrits. Cependant, la participation n'était probablement pas aussi élevée, avec un peu plus de 100 connexions maximum au plus fort de l'événement.



Site web de la conférence jointe ADBIS, TPDL & EDA 2020.



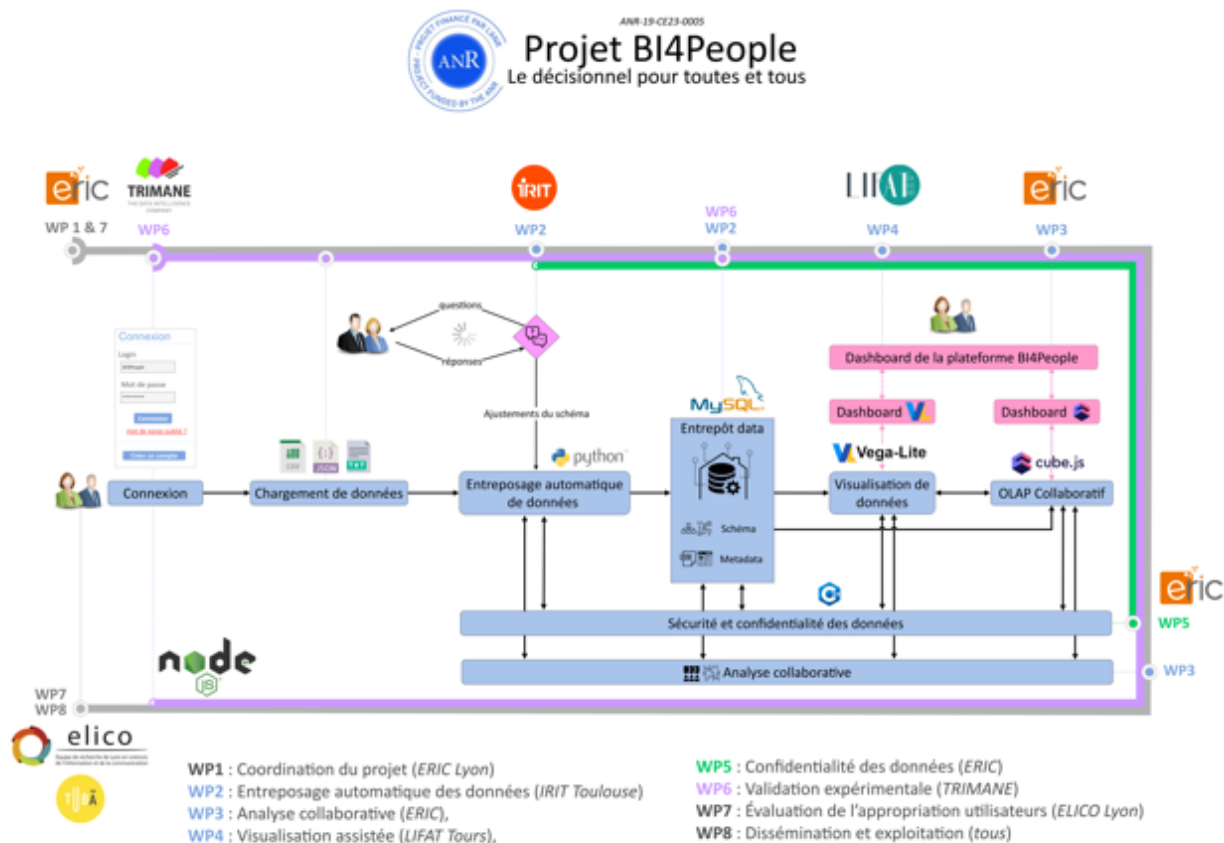
Projet BI4people : Le décisionnel pour tou·tes – Saison 2

Jérôme Darmont, Bilel Sdiri

Le projet BI4people a été présenté lors des journées EDA Business Intelligence & Big Data, qui sont une série de conférences nationales fondées en 2005.

Le projet BI4people vise à rendre l'informatique décisionnelle disponible en ligne à des usagers disposant de peu de ressources financières et de connaissances techniques, en automatisant des processus actuellement au mieux semi-automatiques. Plus précisément, il s'agit de rendre accessible la puissance de l'analyse interactive OLAP à la plus large audience possible, en mettant en œuvre le processus d'entreposage de données en mode *software-as-a-service*, de l'intégration de données multi-

sources, hétérogènes à une analyse OLAP et une visualisation très simples. Pour atteindre ce but, le service de BI doit inclure la *privacy by design*, être autonome, extrêmement simple, ergonomique et intelligible. Dans ce contexte, les étapes classiques de l'entreposage de données s'appliquent, mais doivent être complètement automatisées. À notre connaissance, BI4people est la première plateforme à atteindre complètement ce but. Par ailleurs, nous insistons sur l'importance de l'appropriation des visualisations fournies par l'outil par les usagers, ce qui implique une collaboration interdisciplinaire entre l'informatique et les sciences de l'information et de la communication.



Processus global du projet BI4people.

8^e journées Business Intelligence & Big Data (EDA 2022), Clermont-Ferrand. RNTI, B-18, pp. 107-108.



Article à la conférence Conceptual Modeling (ER 2019)

SkipSJoin: A New Physical Design for Distributed Big Data Warehouses in Hadoop

Yassine Ramdane, Nadia Kabachi, Omar Boussaid, Fadila Bentayeb

ER est une conférence de haut niveau (CORE A) dédié à la discussion et à l'enrichissement des fondements de la modélisation conceptuelle à la pointe de la technologie. Elle aborde également les défis émergents et futurs, ainsi que le rôle prépondérant que la modélisation conceptuelle occupe dans une multitude d'applications.

Ce papier aborde l'optimisation des entrepôts de données massifs sur Hadoop, un enjeu majeur pour améliorer les performances des requêtes décisionnelles, notamment face aux défis de gestion du trafic réseau lors des jointures complexes.

Le partitionnement et la distribution des données ont été largement utilisés dans les systèmes distribués qui utilisent le paradigme MapReduce, tels que Hadoop et Spark. Ils ont été utilisés pour l'équilibrage des charges de données, pour éviter le chargement des partitions inutiles et pour guider la conception physique des bases de données et des entrepôts de données distribuées.

Pour effectuer des analyses sur les données entreposées, nous utilisons généralement des requêtes

OLAP. Une requête OLAP est une requête complexe contenant plusieurs opérations coûteuses, telles que la jointure en étoile, la projection, le filtrage et les opérations d'agrégats.

Hadoop utilise le partitionnement horizontal pour optimiser la performance d'un entrepôt de données massives. Un défi majeur dans ce processus est de réduire le trafic réseau associé à une charge de travail spécifique. Pour remédier à ce problème lors de l'exécution d'une opération de jointure, une technique courante consiste à co-partitionner les tables de l'entrepôt en fonction de leur clé de jointure. Néanmoins, dans les schémas de partitionnement actuels, l'exécution d'une opération de jointure en étoile dans Hadoop nécessite encore un nombre considérable de cycles MapReduce.

Dans cet article, nous combinons un modèle basé sur les données "Data-Driven" et un modèle basé sur la charge de travail "Workload Driven" pour créer un nouveau design physique pour les entrepôts de données massives distribués sur Hadoop, appelé "SkipSJoin", et ainsi optimiser le schéma de partitionnement.

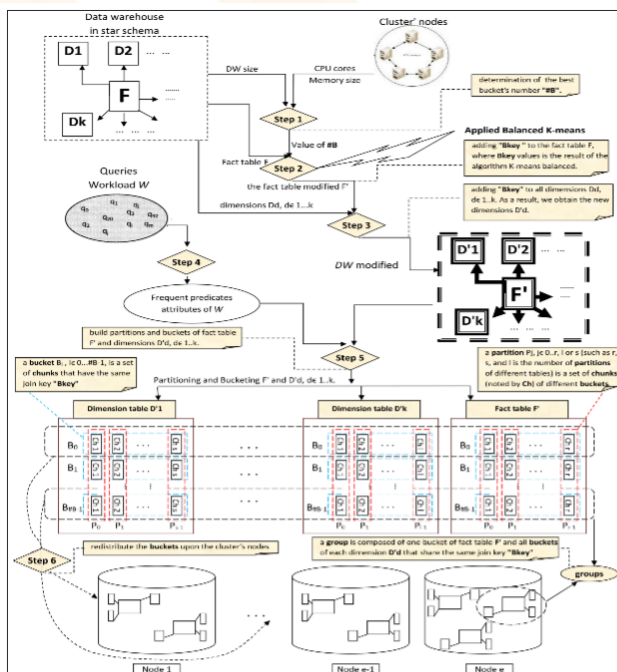


Fig. 7: "SkipSJoin" permet d'exécuter l'opération de jointure en étoile d'une requête OLAP en une seule étape Spark, sans la phase de "Shuffle", ce qui élimine le chargement de blocs HDFS inutiles et améliore considérablement le temps d'exécution des requêtes.



Article dans la revue *Information Systems Frontiers*, 2022 *Rumor Classification through a Multimodal Fusion Framework and Ensemble Learning*

Abderrazek Azri, Cécile Favre, Nouria Harbi, Jérôme Darmont, Camille Noûs

Le journal *Information Systems Frontiers* est un journal de l'éditeur Springer qui comprend des travaux à l'interface des systèmes d'information et des technologies de l'information sous des angles analytiques, comportementaux et technologiques. C'est une revue de très bonne qualité (Scimago Q1).

La plupart des méthodes de vérification de la véracité des messages de réseaux sociaux ne prennent pas en compte simultanément les contenus textuels/visuels et informations diverses. Dans cet article, nous présentons MONITOR (Multimodal Fusion Framework to Assess Message Veracity in Social Networks) qui permet de détecter la véracité des messages. Pour cela, l'originalité de l'approche réside dans l'usage de caractéristiques des images issues du domaine de l'évaluation de la qualité des images. Des expériences menées sur des jeux de données issus de Twitter montrent que MONI-

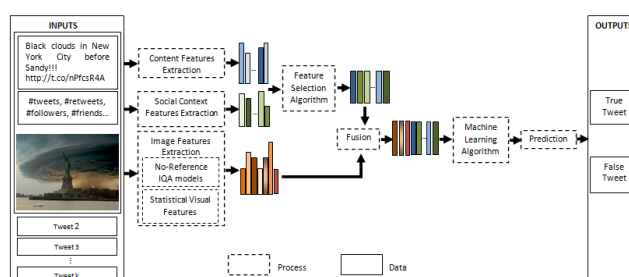
TOR surpasse les bases de référence de l'apprentissage automatique pour cette tâche et que le recours à l'apprentissage ensembliste pour combiner plusieurs algorithmes d'apprentissage afin d'obtenir de meilleures prédictions est concluant.

La prolifération des rumeurs sur les réseaux sociaux est devenue une préoccupation majeure en raison de l'impact dévastateur qui peut en découler. La présence d'images dans les messages qui circulent accentue leur large diffusion. Or l'évaluation manuelle de la véracité des messages des réseaux sociaux est une tâche très chronophage qui peut être grandement facilitée par l'apprentissage automatique. Ainsi, ces travaux permettent d'apporter des pistes de solutions quant à la détection des rumeurs dans les réseaux sociaux, et plus particulièrement sur Twitter, ce qui représente un enjeu d'actualité important.



Fig. 8: L'image présentée ici était associée au texte "Black clouds in New York City before Sandy!!!" dans un message posté sur Twitter. Ce message est une rumeur qui a été lancée. L'enjeu est de pouvoir construire des caractéristiques de ces images issues du domaine de la qualité des images pour aider à la détection de fausses images. Ces caractéristiques peuvent alors être utilisées dans les algorithmes d'apprentissage pour la détection des faux messages qui contribuent à la diffusion de fausses informations.

Fig. 9: L'approche MONITOR prend en entrée un ensemble de tweets (avec les contenus textuels, les images, et diverses informations) pour finalement prédire en sortie si un tweet correspond à une fausse ou une vraie information. Pour ce faire, différentes caractéristiques sont extraites, à la fois par rapport au contenu du tweet, au contexte social et aux images, dont notamment les caractéristiques issues du domaine de la qualité des images (IQA : Image Quality Assessment). Ces caractéristiques sont ensuite fusionnées avant la phase d'apprentissage.





Article dans la revue The Ad Hoc Networks journal, 2023

A robust and efficient vector-based key management scheme for IoT networks

Sami Bettayeb, Mohamed-Lamine Messai, Sofiane Mounine Hemam

La revue Ad Hoc Networks, publiée par l'éditeur Elsevier, est une référence internationale dans le domaine des réseaux sans fil et des systèmes de communication autonomes (Scimago Q1). Elle se concentre sur les réseaux ad hoc dont les réseaux d'objets connectés. La revue publie des articles de recherche originaux, des études de cas et des analyses approfondies sur les protocoles, les architectures, la sécurité, la gestion des res de ces réseaux. Destinée aux chercheurs, ingénieurs et professionnels, elle vise à promouvoir l'innovation et les avancées technologiques dans ce domaine en pleine expansion.

L'article présente un nouveau schéma de gestion des clés, appelé EVKMS (Efficient Vectors-based Key

Management Scheme), conçu pour sécuriser la collection de données dans les réseaux IoT (Internet des Objets). Ce schéma utilise des vecteurs pré-distribués pour masquer les clés, tout en divisant la zone de déploiement en sous-zones pour améliorer la flexibilité, l'évolutivité et la résilience aux attaques. EVKMS minimise les communications pour l'établissement de clés cryptographiques tout en offrant une sécurité renforcée contre les attaques de capture de nœuds et autres menaces connues. Les résultats de simulations montrent que ce schéma est plus efficace que les solutions existantes en termes de stockage, de communication et de consommation d'énergie, le rendant adapté aux réseaux IoT à ressources limitées.

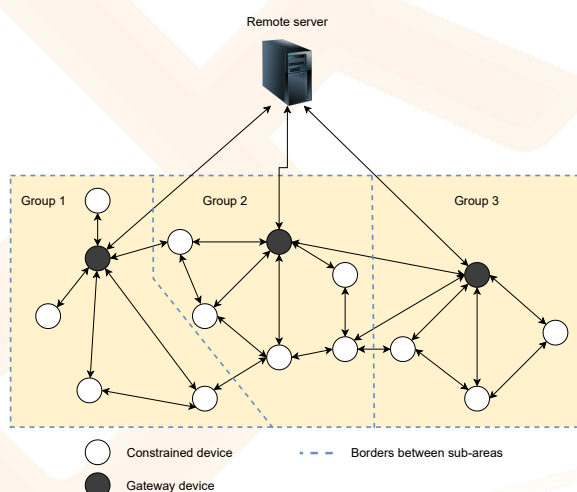
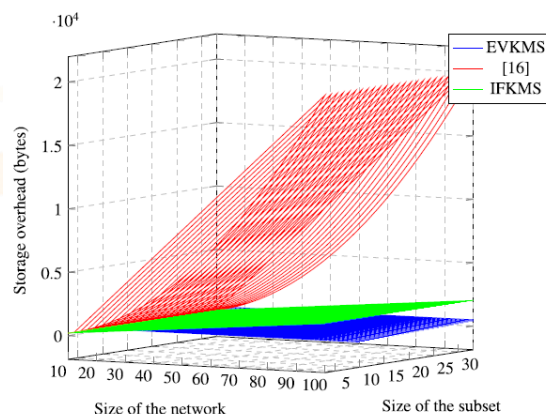


Fig. 11: Quelques résultats comparatifs sur la métrique de stockage mémoire nécessaire. L'analyse prend pour un nœud du réseau deux variables : la taille du réseau et la taille des sous-réseaux. Notre approche, comparée à deux autres de la littérature, consomme moins d'espace mémoire, ce qui la rend plus adaptée aux réseaux IoT.

Fig. 10: Cette figure présente le modèle réseau IoT. C'est une architecture réseau à trois niveaux. Il y a un serveur distant ainsi que trois groupes (Group 1, Group 2, Group 3), représentant des sous-réseaux divisés en zones. Le modèle peut être généralisé en N groupes. Dans chaque zone, un ensemble de nœuds IoT est déployé aléatoirement. Une passerelle par zone facilite la communication entre les nœuds IoT et le serveur distant.





Unité de Recherche

5 avenue Pierre Mendès France
69676 Bron Cedex, France
Tél. +33 (0)4 78 77 31 54