



HAL
open science

Le Machine Learning pour un Context Mining Facility

Nourhène Ben Rabah, Manuele Kirsch Pinheiro, Bénédicte Le Grand, Ali Jaffal, Carine Souveyet

► **To cite this version:**

Nourhène Ben Rabah, Manuele Kirsch Pinheiro, Bénédicte Le Grand, Ali Jaffal, Carine Souveyet. Le Machine Learning pour un Context Mining Facility. INFORSID 2021, Jun 2021, Dijon, France. pp.107-110. hal-03877511

HAL Id: hal-03877511

<https://paris1.hal.science/hal-03877511v1>

Submitted on 29 Nov 2022

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.

Le Machine Learning pour un Context Mining Facility

Nourhène BEN RABAH^{1,2}, Manuele KIRSCH PINHEIRO¹,
Bénédicte LE GRAND¹, Ali JAFFAL², Carine SOUVEYET¹

1. Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne
90 rue de Tolbiac, 75013 Paris, France
{Manuele.Kirsch-Pinheiro|Benedicte.Le-Grand|Carine.Souveyet}@univ-paris1.fr
2. ESIEE-IT, 8 rue Pierre de Coubertin 1 95 300 Pontoise, France
{nbenrabah|ajaffal}@esiee-it.fr

RÉSUMÉ. Le Machine Learning est de plus en plus utilisé dans les applications sensibles au contexte pour la fouille des données de contexte. Dans l'article (Ben Rabah et al., 2020), nous posons la question de la généralisation de la fouille à l'échelle de tout un SI, en tant que « facilité » proposée à l'ensemble de ses composants. Ce changement d'échelle pose de nombreux défis à l'application des techniques de ML, notamment pour le prétraitement et le paramétrage. L'étude de littérature présentée dans cet article invite à une prise de conscience indispensable pour l'avènement d'un service de « context facility » dans les SI.

MOTS-CLÉS : Machine Learning, Fouille de contexte, Systèmes d'Information

1. Introduction

Les informations de contexte sont de plus en plus aisées à capturer, offrant de multiples opportunités au sein des SI (adaptation au contexte, recommandation de services ou de contenu, prédiction des besoins ou des actions des utilisateurs, etc.), surtout lorsqu'elles sont combinées à des techniques de *Machine Learning*. De nombreuses applications sensibles au contexte utilisent ainsi ces techniques (Ameyed, 2015). On peut alors envisager la généralisation de ces comportements « *intelligents* », en proposant la fouille de contexte comme un service propre à un SI.

Nous avons proposé dans (Ben Rabah et al., 2020) une vision d'une « *context mining facility* », *i.e.*, un service de fouille de contexte généralisé à tout un Système d'Information, ainsi utilisable par toutes ses applications. L'hétérogénéité des données de contexte apparaît comme un obstacle majeur pour leur analyse généralisée et totalement automatisée. Nous avons analysé l'utilisation du ML pour la fouille de contexte et les conditions nécessaires pour que cette analyse puisse tenir compte des particularités propres aux informations de contexte, ainsi que les défis pour sa généralisation à l'échelle d'un SI. Nous les résumons ici pour la communauté SI française.

2. Le *Machine Learning* pour une *context facility*

La généralisation de la fouille de contexte à l'échelle d'un Système d'Information représente la possibilité d'offrir, à l'échelle de tout un système, des mécanismes de raisonnement disponibles en continu pour toutes les applications et évoluant avec le système, avec un minimum d'intervention humaine. Ce support doit être pensé pour tout un système, avec son écosystème d'utilisateurs (décideurs, employés, clients...) et d'applications supportant les différents processus et corps de métiers qui cohabitent dans une organisation. Un même service de fouille de contexte devrait être capable de supporter les besoins de multiples applications (existantes ou futures), en intégrant à tout moment de nouvelles informations de contexte hétérogènes et en constante évolution. Dans ces conditions, on ne peut avoir de connaissance préalable de ces données.

Avec des mots clés tels que *context prediction / awareness*, *machine learning*, *data mining* et *IoT*, nous avons identifié 200 articles et étudié 30 en détail. Garcia et al. (2016) et Ramirez-Gallego et al. (2017) reconnaissent l'importance de la qualité des données. Or les données de contexte sont, par définition, hétérogènes et incertaines (Bettini et al., 2010) (Chalmers et al. 2004). Souvent issues de capteurs, ces données peuvent être erronées et incomplètes, ce qui a impact négatif sur le temps d'exécution, la complexité ou la qualité des modèles générés par le ML. De même, certaines techniques de ML requièrent une distribution équilibrée des données dans les classes disponibles, ce qui n'est pas toujours possible avec les données de contexte, en raison du caractère dynamique et incertain de ces données.

Des phases de prétraitement et une éventuelle intervention humaine sont alors nécessaires pour régler ce problème de qualité des données récoltées et pouvoir les exploiter. L'introduction de nouvelles données dans le système implique l'arrêt du service pour réexécuter le prétraitement et l'entraînement. Ceci peut impacter d'autres applications potentiellement critiques pour l'organisation. Il paraît alors clair qu'en état les techniques de ML actuelles ne semblent pas encore assez adaptées pour un usage l'échelle de tout un Système d'Information.

5. Conclusions

La réussite des techniques de ML repose encore beaucoup sur l'expertise humaine pour le choix de l'algorithme, des hyper-paramètres, la préparation des données, etc.). Il existe néanmoins de plus en plus de *framework* et d'outils, comme AutoML (Hutter et al., 2019) qui simplifient l'usage du ML, voire automatisent partiellement l'exécution de leur pipeline. On peut donc espérer que les difficultés identifiées dans (Ben Rabah et al., 2020) pourront s'amenuiser avec le temps, pour que cette vision d'un « *context mining facility* » puisse voir le jour.

Bibliographie

- Ameyed, D., Miraoui, M., Tadj, C. (2015). A survey of prediction approach in pervasive computing. *International Journal of Scientific & Engineering Research*, 6(5), p. 306-316.
- Ben Rabah, N., Kirsch Pinheiro, M., Le Grand, B., Jaffal, A., Souveyet, C. (2020). Machine Learning for a Context Mining Facility. *16th Workshop on Context and Activity Modeling and Recognition*, 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), p.678-684.
- Bettini, C., Brdiczka, O., Henriksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2), Apr 2010, p. 161-180.
- Chalmers, D., Dulay, N., Sloman, M. (2004). Towards Reasoning About Context in the Presence of Uncertainty. *1st Int. workshop on advanced context modelling, reasoning and management*. Nottingham, UK, September 2004.
- García, S., Luengo, J., Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98, 15 April 2016, p. 1-29.
- Hutter, F., Kotthoff, L., Vanschoren, J. (2019). *Automated Machine Learning: Methods, Systems, Challenges*, Springer, 2019.
- Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak M. , Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, 239, 24 May 2017, p. 39-57