



**HAL**  
open science

## Extraction de signaux comportementaux d'utilisateur et prédiction de conclusion de vente

Diana Nurbakova, Sylvie Calabretto, Timothée Saumet, Jialiang Wei

### ► To cite this version:

Diana Nurbakova, Sylvie Calabretto, Timothée Saumet, Jialiang Wei. Extraction de signaux comportementaux d'utilisateur et prédiction de conclusion de vente. Extraction et Gestion des connaissances, 2021, Jan 2021, Montpellier, France. pp.511-512. hal-03121968

**HAL Id: hal-03121968**

**<https://hal.archives-ouvertes.fr/hal-03121968>**

Submitted on 26 Jan 2021

**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.

# Extraction de signaux comportementaux d'utilisateur et prédiction de conclusion de vente

Diana Nurbakova\*, Sylvie Calabretto\*  
Timothée Saumet\*\*, Jialiang Wei\*\*\*

\*LIRIS UMR 5205 CNRS, INSA Lyon - Université de Lyon  
diana.nurbakova@insa-lyon.fr, sylvie.calabretto@insa-lyon.fr

\*\*Tilkee SAS

tim@tilkee.com

\*\*\*AKKA Technologies

jjaliang.wei@outlook.com

*Le problème de prédiction de conclusion de vente, PCV (deal closure prediction) consiste en l'estimation de la probabilité de convertir une opportunité (un client potentiel) en une vente. Ainsi, il est souvent formulé et abordé en tant que problème de classification binaire auquel sont appliqués des algorithmes d'apprentissage automatique (e.g. (Eitle et Buxmann, 2019)) et dont l'enjeu principal est l'extraction de features de différentes sources (e.g. (Aggour et Hoogs, 2013)). Dans le contexte industriel, ces features sont dans la majorité des cas directement issues de données du système de gestion de la relation client (GRC). Cependant, le problème de la qualité des données, en particulier les données manquantes et/ou erronées, persiste. Il est peut être lié au renseignement manuel par l'équipe commerciale mais aussi à l'absence de traces automatiques d'interactions des commerciaux avec les opportunités. Plusieurs travaux existants s'intéressent aux modèles de pertinence de document / résultats de recherche basés sur *les taux de clics* (CTR) (e.g. (Yu et al., 2019; Zheng et al., 2019)). Ainsi, ils considèrent peu l'engagement post-clic de l'utilisateur (Grinberg, 2018), en s'appuyant souvent sur le temps passé sur la page (dwell time) (Kim et al., 2014; Nurbakova et Saumet, 2020). Ceci est principalement lié à l'absence de données de log de traçage d'interactions d'utilisateur. Cependant, les interactions d'utilisateur avec le contenu des pages affichées (Huang et al., 2012; Grinberg, 2018) ont été utilisées afin d'étudier l'intention utilisateur ainsi que son attention lors de la lecture. Contrairement à ces travaux existant, nous cherchons à relier les interactions d'utilisateur lors de sa lecture de documents commerciaux avec son intention de conclure la vente. Dans (Nurbakova et Saumet, 2020), nous avons montré que la précision de la PCV basée uniquement sur le comportement de lecture de documents est relativement haute. Or, les features créées et utilisées par cette approche sont principalement de multiples agrégations de temps passé sur les pages de documents et des fréquences de visites. Ainsi, elles ne sont pas suffisamment discriminantes et n'exploitent pas le potentiel de tracking.*

**L'objectif principal** de ce projet consiste en l'élaboration d'un algorithme d'extraction de signaux comportementaux d'utilisateur caractéristiques de son intérêt commercial et issus de sa lecture de document numérique. Ceci va servir de base à l'approche de PCV. Pour atteindre l'objectif visé, nous proposons une étude en plusieurs **étapes** : (1) Analyse de l'impact

de signaux comportementaux sur la qualité de PCV en combinaison avec les données des opportunités issues de GRC ; (2) Extraction des régions d'intérêt de l'utilisateur et de ses motifs comportementaux de lecture (e.g. la séquence des pages consultées, événements de la souris, surlignement) à travers une étude utilisateur en faisant des parallèles avec les tâches de recherche d'information (e.g. (Grinberg, 2018)); (3) Proposition d'un algorithme de PCV tenant compte des résultats des étapes précédentes. En outre, il est envisageable d'effectuer l'étiquetage automatique des documents et d'en tenir compte dans l'algorithme de PCV. Ce dernier doit pouvoir passer à l'échelle et avoir un temps de réponse quasi-instantané ( $\leq 2s$ ).

Ce projet s'effectue en collaboration avec une entreprise spécialisée en tracking de documents numériques. Ceci permet la validation de la pertinence de notre algorithme dans un environnement industriel sous forme de test A/B et d'études utilisateur. Il sera comparée à (1) la baseline actuelle mise en production dans l'entreprise ; (2) l'approche proposée en (Nurbakova et Saumet, 2020); (3) RandomForest qui est l'algorithme de base utilisé par Einstein du GRC Salesforce appliqué aux données clients sans les données comportementales de lecture.

## Références

- Aggour, K. S. et B. Hoogs (2013). Financing lead triggers : Empowering sales reps through knowledge discovery and fusion. In *Proc. of the 19th ACM, KDD '13*, pp. 1141–1149.
- Eitle, V. et P. Buxmann (2019). Business analytics for sales pipeline management in the software industry : A machine learning perspective. In *52nd HICSS 2019, USA*, pp. 1–10.
- Grinberg, N. (2018). Identifying modes of user engagement with online news and their relationship to information gain in text. In *Proc. of WWW '18*, pp. 1745–1754.
- Huang, J., R. W. White, G. Buscher, et K. Wang (2012). Improving searcher models using mouse cursor activity. In *The 35th International ACM SIGIR '12*, pp. 195–204.
- Kim, Y., A. H. Awadallah, R. W. White, et I. Zitouni (2014). Modeling dwell time to predict click-level satisfaction. In *7th ACM WSDM 2014*, pp. 193–202.
- Nurbakova, D. et T. Saumet (2020). Deal closure prediction based on user's browsing behaviour of sales content. In *ICDM Workshops 2020*.
- Yu, H.-T., A. Jatowt, R. Blanco, J. M. Jose, et K. Zhou (2019). A rank-biased neural network model for click modeling. In *Proc. of the 2019, CHIIR '19*, pp. 183–191.
- Zheng, Y., J. Mao, Y. Liu, C. Luo, M. Zhang, et S. Ma (2019). Constructing click model for mobile search with viewport time. *ACM Trans. Inf. Syst.* 37(4).

## Summary

*The deal closure prediction problem* consists in the estimation of the probability of transforming an opportunity to a customer. This prediction is often performed based on the client data entered to a CRM system. However, these data and the methods built upon them lack consequential data about document reading behaviour of users. In this project in collaboration with a company specialised in document tracking, we aim at proposing an approach to extract user behaviour signals allowing to enhance the results of deal closure prediction.