

Extraction de connaissances dans les Systèmes d'Information Pervasifs par l'Analyse Formelle de Concepts

Ali Jaffal, Bénédicte Le Grand, Manuele Kirsh-Pinheiro

Université Paris 1 Panthéon-Sorbonne (CRI)
90, rue de Tolbiac 75013 Paris
Ali.Jaffal@malix.univ-paris1.fr

Résumé. Nous présentons une méthode d'extraction de connaissances dans des systèmes d'information pervasifs. Nous étudions l'impact du contexte (environnement) d'un utilisateur sur les applications qu'il utilise sur son smartphone. Notre proposition pour gérer la complexité des données contextuelles repose sur l'Analyse Formelle de Concepts et les treillis de Galois. Nous nous focalisons sur l'automatisation du processus d'interprétation de ces treillis, pour généraliser l'extraction de connaissances et passer à l'échelle. Nous présentons des métriques originales illustrées sur des données réelles.

1 Introduction

Les Systèmes d'Information Pervasifs (SIP) étendent les SI traditionnels en tenant compte de l'environnement et de la mobilité des utilisateurs, rendue possible grâce aux smartphones (Kourouthanassis et Giaglis (2006)). Comprendre l'influence de ce «contexte» sur l'usage des terminaux constitue un défi (Fernandes et al. (2008)). Le projet ArtDeco (Anastasi et al. (2012)) se concentre sur l'extraction et la représentation de connaissances à partir d'informations hétérogènes provenant d'applications et de capteurs dans des entreprises. Le projet U-CAT (Lopez-Cozar et al. (2005)) a développé un système pour la réalisation d'activités éducatives via l'utilisation de différents appareils mobiles. Lors de travaux précédents (Jaffal et al. (2014)), nous avons utilisé l'Analyse Formelle de Concepts (AFC) et les treillis de Galois pour gérer les données contextuelles que nous appelons «*éléments de contexte*» et étudier leur impact sur le comportement des utilisateurs. L'AFC constitue un pont entre les mathématiques et l'analyse de données (Priss (2006)). L'AFC regroupe dans des *concepts formels* des éléments de données (appelés *objets*) en fonction des propriétés (appelées *attributs*) qu'ils ont en commun (Wille (2005)). Un *treillis* de concepts (ou de Galois) traduit la relation d'ordre entre les concepts formels, et peut être utilisé pour la classification ou la prédiction. Dans cet article, nous proposons des métriques pour automatiser l'interprétation de ces treillis. Plusieurs mesures comme la stabilité et le support (Jay et al. (2008), Klimushkin et al. (2010), Kuznetsov et al. (2007)) ont été proposées pour réduire les treillis (Ventos et Soldano (2005)). Notre contribution est complémentaire et vise à automatiser l'interprétation des treillis, réduits ou non. (Le Grand et al. (2009)) ont appliqué des mesures conceptuelles pour caractériser les membres d'un réseau social et leurs contacts. Ces mesures ne caractérisent ni les concepts, ni les relations du

treillis, elles se focalisent sur la caractérisation des objets et des attributs, et non sur le treillis (concepts et relations).

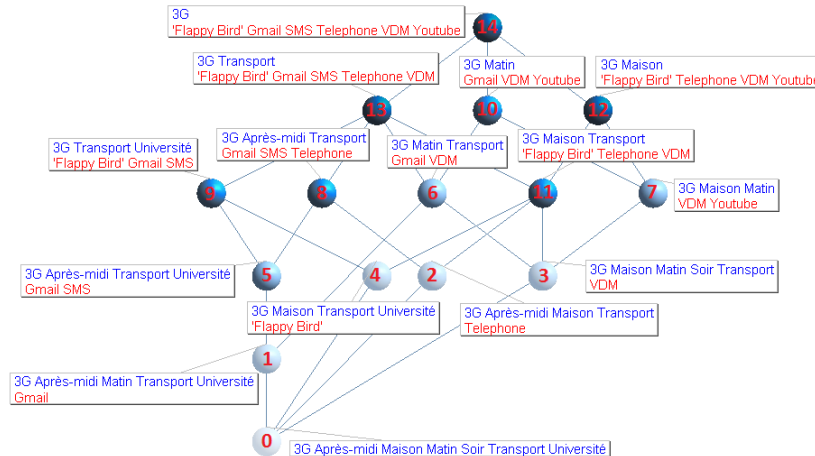


FIG. 1 – Exemple de Treillis de Galois

2 Proposition pour l'interprétation automatique des treillis

Notre proposition vise à automatiser l'interprétation des treillis de Galois via des mesures permettant d'analyser les objets, attributs, concepts et les relations dans ces treillis. L'analyse des regroupements effectués par un treillis permet d'extraire des informations importantes pour un SIP. Analysons par exemple le treillis de la fig. 1 qui correspond à un utilisateur. Dans chaque concept, les applications (objets) sont en bas et leurs éléments de contexte (attributs) communs sont en haut. Les données ont été collectées via un questionnaire soumis à 28 étudiants, où chacun a reporté, pendant une semaine, les applications utilisées et les éléments de contexte associés. Nous avons construit 282 treillis à partir de plusieurs stratégies. Le concept 14 contient l'élément de contexte *3G*, commun à toutes les applications, que l'on peut qualifier d'*universel* (pour cet utilisateur). Certaines applications sont *conceptuellement* proches, car fréquemment présentes simultanément dans des concepts formels, par ex. *Gmail* et *SMS*. De même, certains éléments de contexte sont *conceptuellement* proches comme *Transport*, *Après-midi* et *3G*. Nous proposons ici des métriques pour automatiser ces interprétations.

2.1 Poids des applications et des éléments de contexte

Le poids d'un élément (objet ou attribut) correspond à la proportion de concepts contenant cet élément. Le tableau 1 représente le poids des applications de la fig. 1. Il permet pour l'utilisateur étudié, d'identifier les applications (resp. éléments de contexte) les plus fréquents, donc ayant des contextes d'usage (resp. applications) communs.

	Gmail	SMS	Telephone	VDM	Flappy Bird	Youtube
Nbr Concepts	8	5	6	8	6	4
Poids	0,53	0,33	0,4	0,53	0,4	0,26

TAB. 1 – Poids des applications dans le treillis de la fig. 1

2.2 Impact mutuel

Dans un environnement ubiquitaire, les éléments de contexte ont un impact sur l'usage des applications, qui reflètent pour nous le comportement de l'utilisateur. Nous définissons l'impact mutuel entre un élément de contexte EC_i et une application A_j :

$$Impact\ mutuel_{(EC_i, A_j)} = \frac{Nbr_{concepts\ contenant\ EC_i\ et\ A_j}}{Nbr_{concepts\ contenant\ EC_i\ ou\ A_j}}$$

Le diagramme de la fig. 2 présente l'impact mutuel relatif entre les éléments de contexte et les applications dans le treillis de la fig. 1. Pour cet utilisateur, par exemple l'*Université* est majoritairement associée à *Gmail*, et le *matin* à l'application *VDM*.

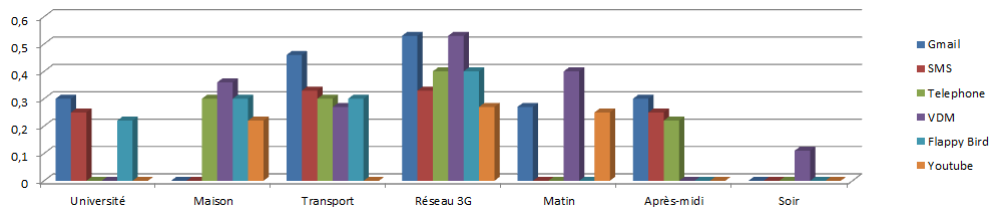


FIG. 2 – Impact mutuel relatif associé au treillis de la fig. 1

2.3 Similarité conceptuelle des éléments de contexte et des applications

Nous définissons la similarité entre deux éléments E_i et E_j (application ou contexte) :

$$Sim_{(E_i, E_j)} = \frac{Nbr_{concepts\ contenant\ E_i\ et\ E_j}}{Nbr_{concepts\ contenant\ E_i\ ou\ E_j}}$$

La fig. 3 représente la similarité entre les éléments de contexte du treillis de la fig. 1. Si la valeur de $Sim(EC_i, EC_j)$ est proche de 1, cela signifie que la plupart des applications utilisées avec l'élément de contexte EC_i sont aussi utilisées avec l'élément de contexte EC_j . Par exemple la plupart des applications utilisées par cet utilisateur à l'université sont aussi utilisées dans les transports, avec une connexion 3G et l'après-midi.

Extraction de connaissances dans les Systèmes d'Information Pervasifs

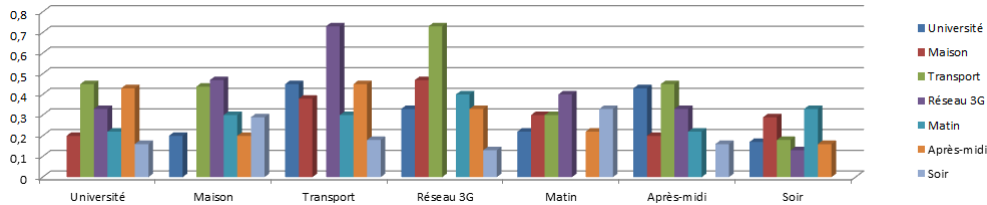


FIG. 3 – Similarité conceptuelle entre les éléments de contexte dans le treillis de la fig. 1

La fig. 4 montre si deux applications ont été utilisées avec les mêmes éléments de contexte. Par exemple l'application *Youtube* est très similaire à l'application *VDM*, car les éléments de contexte pour l'usage de *Youtube* sont presque les mêmes que pour l'application *VDM*, pour cet utilisateur. Jusqu'ici, les mesures présentées caractérisaient un unique treillis (utilisateur). Dans la suite, nous comparons les résultats obtenus pour les treillis de tous les utilisateurs.

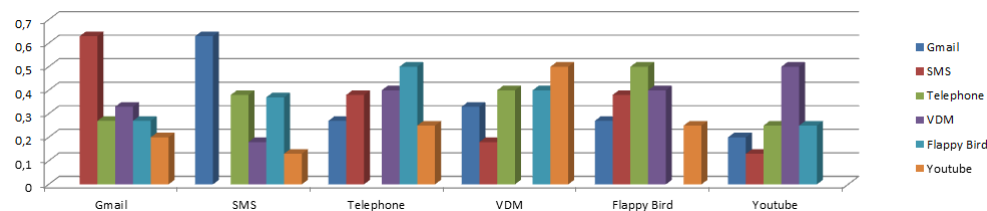


FIG. 4 – Similarité conceptuelle entre les applications dans le treillis de la fig. 1

2.4 Comparaison des treillis

Les treillis de Galois que nous avons générés représentent les comportements des utilisateurs ; dans cette partie nous comparons le comportement des utilisateurs avec les mesures d'impact et de similarité.

Nous avons comparé l'impact entre un élément de contexte EC_i et une application A_j dans 282 treillis. La fig. 5 représente l'impact mutuel absolu entre l'élément de contexte *Université* et l'application *Gmail*. La fig. 5 montre que même si la plupart des utilisateurs utilisent l'application *Gmail* à l'*université*, l'impact de l'*université* sur *Gmail* n'est pas très fort (max 0,46). La mesure d'impact nous a permis d'extraire des connaissances sur l'usage des applications en fonction des éléments de contexte. Elle permet aussi de classer les utilisateurs.

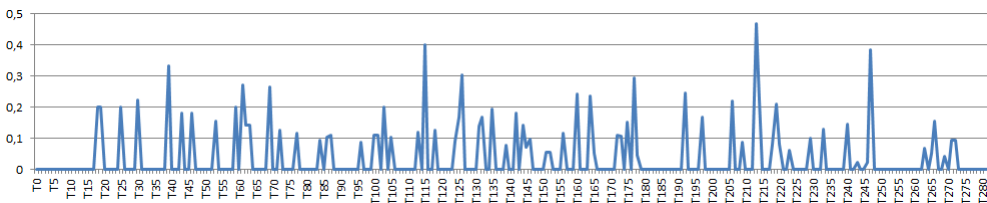


FIG. 5 – Diagramme d'impact entre l'Université et l'application Gmail

Nous avons calculé la similarité entre les éléments de contexte et les applications. La fig. 6 montre la similarité entre les deux éléments de contexte *Transport* et *Matin* dans les différent treillis. Les deux éléments de contexte sont très similaires *i.e.*, les applications qui sont été utilisées dans les transports ont aussi été utilisées le matin (et vice versa). Le diagramme montre aussi que ces deux éléments de contexte apparaissent pour la plupart des utilisateurs.

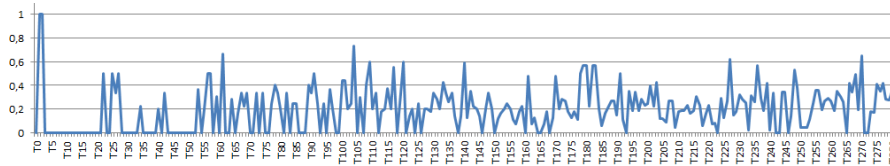


FIG. 6 – Similarité entre *Transport* et *Matin*

La fig. 7 représente la similarité entre les deux applications *WhatsApp* et *SMS*. Ces deux applications sont peu utilisées ensemble (valeur nulle fréquente) : les utilisateurs choisissent en général l'une ou l'autre pour communiquer. On constate cependant que dans les cas où l'utilisateur utilise les deux applications, elles apparaissent comme très similaires en termes de contexte (similarité élevée).

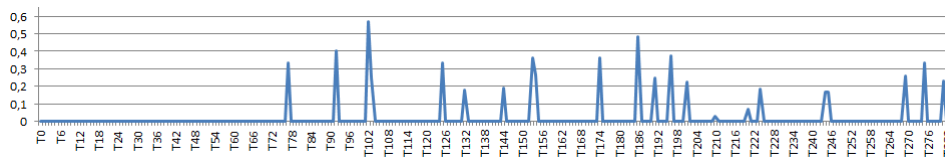


FIG. 7 – Similarité entre *WhatsApp* et *SMS*

3 Conclusion

Nous avons présenté dans cet article des métriques originales pour l'interprétation de treillis de Galois dans le cadre de l'analyse formelle de concepts. Dans notre étude de cas dédiée aux SIP, un treillis de Galois regroupe les applications utilisées en fonction des éléments de contexte associés. Ces métriques ont été appliquées à un ensemble de données réelles. Nous avons donné des exemples d'interprétations automatiques rendues possibles par ces mesures, qui aident à comprendre le comportement des utilisateurs en termes d'usage d'applications en fonction du contexte. D'autres métriques sont actuellement en cours de développement ; leur évaluation constitue l'une de nos perspectives, ainsi que la conception de métriques complémentaires pour aller encore plus loin dans l'interprétation automatique des treillis de Galois.

Références

Anastasi, G., E. Bellini, E. Di Nitto, C. Ghezzi, L. Tanca, E. Zimeo, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Ran-

- gan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, et G. Weikum (Eds.) (2012). *Methodologies and Technologies for Networked Enterprises*, Volume 7200 of LNCS. Berlin, Heidelberg : Springer.
- Fernandes, J. E., R. J. Machado, et J. I. Carvalho (2008). Model-Driven Development for Pervasive Information Systems. In *Advances in Ubiquitous Computing : Future Paradigms and Directions*, 45–82.
- Jaffal, A., M. Kirsch-Pinheiro, et B. Le-Grand (2014). Unified and Conceptual Context Analysis in Ubiquitous Environments. *8th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 48–55.
- Jay, N., F. Kohler, et A. Napoli (2008). Analysis of social communities with iceberg and stability-based concept lattices. In *6th International Conference on Formal Concept Analysis - ICFCA 2008*, Volume 4933 of LNAI, Canada, pp. 258–272. Springer.
- Klimushkin, M., S. Obiedkov, et C. Roth (2010). Approaches to the selection of relevant concepts in the case of noisy data. In *Formal concept analysis*, pp. 255–266. Springer.
- Kourouthanassis, P. E. et G. M. Giaglis (2006). A Design Theory for Pervasive Information Systems. In *in Proc. 3rd Int. IWUC'06*, pp. 62–70.
- Kuznetsov, S., S. Obiedkov, et C. Roth (2007). Reducing the Representation Complexity of Lattice-Based Taxonomies. In U. Priss, S. Polovina, et R. Hill (Eds.), *Conceptual Structures : Knowledge Architectures for Smart Applications*, Number 4604 in LNCS, pp. 241–254. Springer.
- Le Grand, B., M.-A. Aufaure, et M. Soto (2009). Empreintes conceptuelles et spatiales pour la caractérisation des réseaux sociaux. In *EGC*, pp. 349–354.
- Lopez-Cozar, R., Z. Callejas, M. Gea, et G. Montoro (2005). Proceedings of the Int. Workshop ISCA Tutorial and Research Workshop on ASIDE. In *COST278 Final Workshop and ITRW on Applied Spoken Language Interaction in Distributed Environments*.
- Priss, U. (2006). Formal Concept Analysis in Information Science. In : *Blaise, C. (ed.) Annual Review of Information Science and Technology, ASIST 40*.
- Ventos, V. et H. Soldano (2005). Les treillis de Galois Alpha. In *conférence d'apprentissage*, 175–190.
- Wille, R. (2005). Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies. *Formal Concept Analysis, B.Ganter et al., eds.*, 1–33.

Summary

This paper presents a new method for extracting knowledge in pervasive information systems. We study the impact of the context of a user on the applications he runs on his smartphone. Our proposal relies on FCA and Galois lattices to manage the heterogeneity and complexity of contextual data. Manual analysis of *Galois lattices* in previous work has given promising results. We focus here on automating the process of interpretation of the *Galois lattice*, in order to generalize the knowledge extraction. We present original metrics illustrated on real data.