

# Une nouvelle méthode de partitionnement de séquences avec motifs interprétables.

Sébastien Amoury\*, Karell Bertet\*, Damien Mondou\*

\* La Rochelle Université - 23 Avenue Albert Einstein, 17000 La Rochelle, France  
prenom.nom@univ-lr.fr

## 1 Contexte

À l'ère du « tout numérique », la génération et l'exploitation de données constituent un défi d'envergure pour la recherche et l'innovation. Le regroupement de données, ou clustering, repose sur de nombreuses approches (Ghosal et al., 2020), mais les solutions traditionnelles ne parviennent à partitionner les données complexes, comme les séquences, qu'après un pré-traitement consistant à effectuer un plongement numérique des données. Cette transformation rend difficile l'explication des partitions obtenues. Dans cet article, nous proposons une nouvelle méthode de partitionnement de données séquences à partir d'une hiérarchie de concepts, calculée par l'algorithme NEXTPRIORITYCONCEPT (Demko et al., 2020), issu de l'Analyse Formelle de Concept, avec motifs interprétables.

## 2 Travaux entrepris et résultats

L'explicabilité est un enjeu de taille lors de traitements automatique des données. Notre cas d'étude porte sur un Serious Game qui s'est déroulé en 2017 (Mondou, 2024). Les traces des joueurs peuvent être assimilées à des séquences d'activités et les comportements communs des joueurs peuvent être observés en calculant un treillis de concepts de leur séquences d'activités (Amoury et al., 2025). Nous proposons une solution automatisée permettant de partitionner les joueurs en sélectionnant des concepts à partir de la hiérarchie du treillis. Soit  $(A, \delta(A))$  un concept où  $A$  représente la partie objet du concept et  $\delta(A)$  un ensemble de prédicats décrivant le motif de  $A$ . L'ensemble des concepts dotés d'une relation de spécialisation/généralisation  $(A_1, \delta(A_1)) \leq (A_2, \delta(A_2)) \iff A_1 \subseteq A_2 (\iff \delta(A_2) \sqsubseteq \delta(A_1))$  forment un ensemble partiellement ordonné appelé treillis de concepts. Pour des données de type séquences, plusieurs descriptions sont possibles afin de calculer un treillis de concepts, notamment la description par préfixe maximal commun, qui permet d'obtenir un prédicat de la forme "sequence starts with", afin de grouper les joueurs possédant le même préfixe. Un concept  $(A, \delta(A))$  correspond donc à un groupe de joueurs  $A$  décrits par leur préfixe commun maximal  $\delta(A)$ . Notre algorithme va extraire un partitionnement sous forme d'une liste de cluster  $C$  et d'une liste de motifs  $M$  de chaque cluster, pour un seuil support donné en paramètre. Il parcourt le treillis de concepts en partant du concept top, qui contient tous les objets, puis pour chacun des pré-décresseurs  $(A', \delta(A'))$ , triés par ordre décroissant de support, d'un concept courant  $(A, \delta(A))$ ,