

DATA



hackAtech

Shake science. Shape innovation.

#decision

#pattern

#mining

#temporal

TENSE

Faites parler vos données horodatées !

CARACTÉRISTIQUES

Actuellement nous avons de plus en plus de données, qui peuvent être exploitées par des algorithmes de fouille de motifs (*pattern mining*), un motif est une séquence de données récurrente. Autrement dit, la fouille de motifs est une méthode d'apprentissage non supervisée qui permet de découvrir des connaissances structurées dans des données, tels que des événements, des corrélations ou des patterns.

La limite de ces algorithmes est qu'ils ne tiennent pas en compte les motifs pouvant comporter quelques données manquantes (motifs se retrouvant dans 90% des cas par exemple).

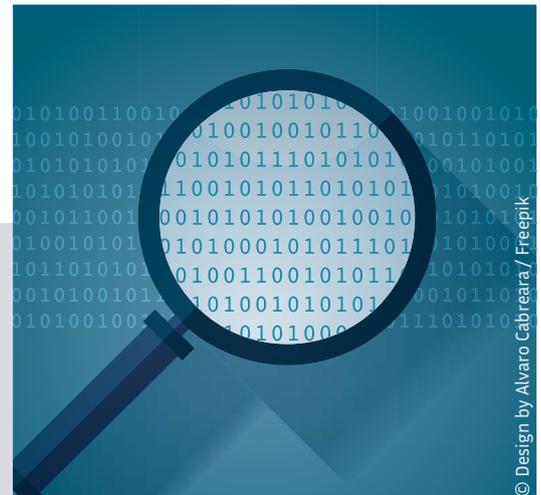
TENSE est un algorithme de fouille de motifs qui apporte une nouvelle approche au pattern mining « classique » puisqu'il considère l'absence de la survenue de certains événements pour décrire un comportement (négation dans les motifs).

L'utilisation de négation dans les motifs permet notamment d'identifier des événements indésirables qui sont à inhiber pour arriver dans des situations d'intérêt.

TRAITEMENT DES DONNÉES

Les données utilisées pour TENSE sont un dataset de séquences temporelles, où une séquence temporelle est une suite d'événements (*dits labels*) horodatés.

Il s'agit typiquement de logs d'événements ou logs d'activités, ou encore des données de type séries temporelles, soit des données volumineuses et datées.



© Design by Alvaro Cabreara / Freepik

USE CASES

Plusieurs domaines d'application :

- **Santé** : analyse de parcours de soin
- **Santé & Bien-être** : signaux d'un capteur physiologique
- **Retail** : analyse de séquence d'achats, trace d'utilisateurs sur le web
- **Industrie 4.0** : analyse des logs de fonctionnement de machine
- **Systemes de transport intelligents**

QUELS AVANTAGES ?

- Affinement des analyses de données
- Meilleure surveillance, gestion et optimisation des systèmes
- Prise en compte de la dimension temporelle dans la description des comportements
- Identification des caractéristiques cachées



FICHE IDENTITÉ

- Langage de programmation: : C++ et interface Python
- Propriété intellectuelle : Inria / IRISA
- Licence : APACHE
- Équipe projet : LACODAM

FONCTIONNALITÉS GÉNÉRIQUES

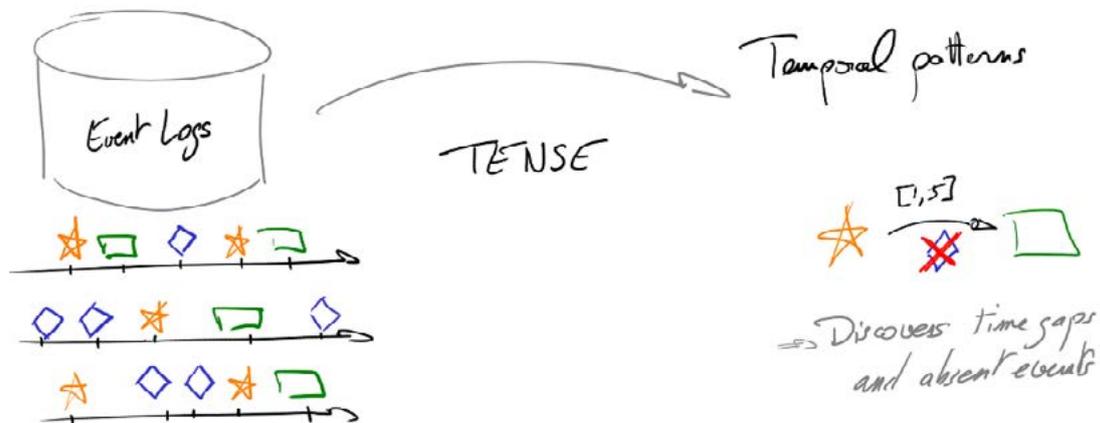
L'algorithme prend en entrée les séquences et en tire les motifs les plus fréquents. Plus particulièrement, à partir d'une suite d'événements datés, il découvre des connaissances structurées et détecte des séquences de motifs temporels avec négation.

Un motif temporel est un comportement typique qui se trouve dans les données en tenant compte des délais temporels entre les événements. De plus les motifs avec négation prennent en compte l'absence de certains événements.

TENSE écrit en C++, peut s'intégrer dans un logiciel comme une API donnant la possibilité d'exploiter des données temporelles. L'outil peut aussi s'utiliser par des appels extérieurs, dans le cas, le format d'entrée sous la forme de JSON (ou format plus brut). Le format de sortie est également sous la forme de JSON.

L'interface Python permet d'intégrer les fonctionnalités de la librairie compilée dans ce langage. Les librairies compilées ne sont disponibles que pour les plateformes de type Unix (compatibilité Windows non-finalisée).

Les méthodes et algorithmes correspondants ont été conçus pour « passer à l'échelle » et permettre l'analyse de données volumineuses. L'extraction des informations temporelles sur les motifs fait donc appel à des calculs multi-cœurs.



CONNAISSANCES MINIMUM REQUISES

- Manipulation de données
- Connaissances en data mining

READ ME

<http://people.irisa.fr/Thomas.Guyet/negativepatterns/>

<https://allgo.inria.fr/app/tensepatterns>

Référent : Thomas Guyet

* LACODAM est une équipe-projet commune à Inria, l'Université de Rennes 1, l'Agrocampus Ouest et l'INSA Rennes.



@Tguyet

