الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي 08 ماي 1945 سيدي بلعباس
École Supérieure en Informatique
08 Mai 1945 Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Specialty: **Ingénierie des Systèmes Informatiques (ISI)**

Theme

Real-Time Extraction of Dynamic Closed Itemsets and Association Rules from Streaming Data

---

Presented by:
SAKOUTI Yasmine

Submission Date: **Sept, 2024**
In front of the jury composed of:

Dr. BENSENANE Hamdane	ESI Sba	President
Pr RAHMOUN Abdelatif	ESI Sba	Co-Supervisor
Dr BENABDERRAHMANE Sid Ahmed	NYU Abu Dhabi	Supervisor
Dr. BENDAOUD Fayssal	ESI Sba	Examiner

Academic Year : 2023/2024

Abstract

In data mining and machine learning, itemset search involves discovering patterns or associations within datasets by identifying frequent co-occurring item combinations.

While traditional algorithms, such as the Apriori algorithm, operate on static databases with a predetermined number of transactions, this project focuses on a novel approach involving a streaming-based algorithm.

Unlike classical methods, our approach is designed for dynamic transactional databases where new items are continuously inserted in real-time or through a sliding temporal window. The goal is to efficiently extract frequent itemsets and generate relevant association rules amidst the ongoing updates.

This project addresses the challenges of adapting itemset mining techniques to a streaming context, offering insights into managing and analyzing evolving data streams.

Keywords : Itemset Mining, Streaming Data, Machine Learning, Data Mining, Association Rules, Real-Time Data Processing, Sliding Temporal Window.

Résumé

En fouille de données et en apprentissage automatique, la recherche d'ensembles d'éléments consiste à découvrir des motifs ou des associations au sein de jeux de données en identifiant des combinaisons d'éléments qui apparaissent fréquemment ensemble.

Tandis que les algorithmes traditionnels, tels que l'algorithme Apriori, fonctionnent sur des bases de données statiques avec un nombre prédéterminé de transactions, ce projet se concentre sur une approche novatrice impliquant un algorithme basé sur le flux de données.

Contrairement aux méthodes classiques, notre approche est conçue pour des bases de données transactionnelles dynamiques où de nouveaux éléments sont continuellement insérés en temps réel ou via une fenêtre temporelle glissante. L'objectif est d'extraire efficacement les ensembles d'éléments fréquents et de générer des règles d'association pertinentes en dépit des mises à jour continues.

Ce projet aborde les défis de l'adaptation des techniques de fouille d'ensembles d'éléments à un contexte de flux de données, offrant des perspectives sur la gestion et l'analyse des flux de données évolutifs.

Mots clés : Fouille d'Ensembles d'Éléments, Données en Flux, Apprentissage Automatique, Fouille de Données, Règles d'Association, Traitement des Données en Temps Réel, Fenêtre Temporelle Glissante.

ملخص

في التنقيب عن البيانات وتعليم الآلة، تهدف عملية البحث عن مجموعات العناصر إلى اكتشاف الأنماط أو الارتباطات داخل مجموعات البيانات من خلال تحديد تركيبات العناصر التي تتكرر معًا بشكل متكرر.

بينما تعتمد الخوارزميات التقليدية، مثل خوارزمية Apriori على قواعد بيانات ثابتة تحتوي على عدد محدد مسبقًا من البيانات، يركز هذا المشروع على نهج مبتكر يتضمن خوارزمية تعتمد على تدفق البيانات.

بخلاف الأساليب الكلاسيكية، تم تصميم نهجنا للعمل مع قواعد بيانات المعاملات الديناميكية، حيث يتم إدخال عناصر جديدة بشكل مستمر في الوقت الفعلي أو من خلال نافذة زمنية متغيرة. الهدف هو استخراج مجموعات العناصر المتكررة بكفاءة وتوليد قواعد ارتباط ذات صلة مع التحديثات المستمرة.

يتناول هذا المشروع التحديات المتعلقة بتكييف تقنيات التنقيب عن مجموعات العناصر مع سياق تدفق البيانات، ويقدم رؤى حول إدارة وتحليل تدفقات البيانات المتغيرة.

كلمات مفتاحية :

التنقيب عن مجموعات العناصر، تدفق البيانات، تعليم الآلة، التنقيب عن البيانات، قواعد الارتباط، معالجة البيانات في الوقت الفعلي، النافذة الزمنية المتغيرة.