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 8 ماي 1945 بسيدي بلعباس
École Supérieure en Informatique
8 Mai 1945 Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Speciality: **Système d'Information et Web (SIW)**

Theme

Machine learning algorithms for geospatial data clustering

Presented by:

- Merzoug Manal
- Sahel Manar

Submission by : 02/07/2024

In front of jury composed by :

Dr. KAZITANI Yassine	President
Dr. SERHANE Oussama	Examiner
Dr. Chaib Souleyman	Supervisor

Academic Year : 2023/2024

ABSTRACT

Geospatial data, which includes information about locations, is very important in many fields like transportation, public health, and disaster management. Effective analysis of this data is paramount to unveil spatial relationships and extract valuable insights from this rich resource. Clustering is a key data analysis method that groups similar data points together, making it easier to find patterns and trends in large datasets.This well-established technique has been extensively employed in various research endeavors.

This thesis focuses on clustering algorithms for geospatial data. It provides an in-depth exploration of various clustering algorithm categories, examining and comparing their strengths and weaknesses and evaluating the application of these algorithms to diverse domains. Furthermore, it highlights recent algorithmic that address specific clustering problems related to geospatial data.

In this Master subject, we begin with an introduction to geospatial data, covering its concepts, characteristics, and sources. We include a detailed analysis of different clustering algorithm categories and their applications in various domains. Additionally, we present newer algorithms proposed by researchers. Finally, we introduce our proposed method for balanced customer clustering using geospatial data and incorporating visit frequency.

Key words: Geospatial Data,Data analysis,Clustering Algorithms.

RÉSUMÉ

Les données géospatiales, qui incluent des informations sur les emplacements, sont très importantes dans de nombreux domaines tels que le transport, la santé publique et la gestion des catastrophes. L'analyse efficace de ces données est essentielle pour révéler les relations spatiales et extraire des informations précieuses de cette riche ressource. Le clustering est une méthode clé d'analyse des données qui regroupe des points de données similaires, ce qui facilite la détection de modèles et de tendances dans de grands ensembles de données. Cette technique bien établie a été largement utilisée dans diverses recherches.

Cette thèse se concentre sur les algorithmes de regroupement pour les données géospatiales. Elle propose une exploration approfondie des différentes catégories d'algorithmes de regroupement, examinant et comparant leurs forces et faiblesses, et évaluant l'application de ces algorithmes à divers domaines. De plus, elle met en évidence des avancées algorithmiques récentes qui abordent des problèmes de regroupement spécifiques liés aux données géospatiales.

Dans ce sujet de Master, nous commençons par une introduction aux données géospatiales, en couvrant leurs concepts, caractéristiques et sources. Nous incluons une analyse détaillée des différentes catégories d'algorithmes de clustering et de leurs applications dans divers domaines. De plus, nous présentons de nouveaux algorithmes proposés par des chercheurs. Enfin, nous introduisons notre méthode proposée pour le clustering équilibré des clients en utilisant des données géospatiales et en intégrant la fréquence des visites.

Mots clés : Données géospatiales, analyse de données, algorithmes de clustering.

ملخص

البيانات الجغرافية المكانية، التي تشمل معلومات عن المواقع، مهمة جداً في العديد من المجالات مثل النقل، الصحة العامة، وإدارة الكوارث. إن التحليل الفعّال لهذه البيانات ضروري لكشف العلاقات المكانية واستخراج رؤى قيمة من هذا المورد الغني. تُعتبر التجميع طريقة أساسية لتحليل البيانات تجمع بين نقاط البيانات المتشابهة، مما يسهل اكتشاف الأنماط والاتجاهات في مجموعات البيانات الكبيرة. وقد تم استخدام هذه التقنية المعروفة على نطاق واسع في العديد من الأبحاث.

تركز هذه الأطروحة على خوارزميات التجميع للبيانات الجغافية المكانية. توفر استكشافًا متعمقًا لفئات خوارزميات التجميع المختلفة، مع فحص ومقارنة نقاط القوة والضعف فيها. من خلال تقييم تطبيق هذه الخوارزميات على مجالات متنوعة، تقدم الأطروحة مجموعة مختارة من الخوارزميات الحديثة التي تعالج مشكلات التجميع المحددة المتعلقة بالبيانات الجغرافية المكانية.

في هذا الموضوع الخاص بالماجستير، نبدأ بمقدمة عن البيانات الجغرافية المكانية، نغطي فيها مفاهيمها، وخصائصها، ومصادرها. ندرج تحليلًا تفصيليًا لفئات خوارزميات التجميع المختلفة وتطبيقاتها في مختلف المجالات. بالإضافة إلى ذلك، نقدم خوارزميات جديدة مقترحة من قبل الباحثين. وأخيرًا، نقدم طريقتنا المقترحة لتجميع العملاء بشكل متوازن باستخدام البيانات الجغرافية المكانية ودمج تكرار الزيارات.

الكلمات المفتاحية:البيانات الجغرافية المكانية، تحليل البيانات، خوارزميات التجميع.