THE WRITER'S HANDBOOK

INTRODUCTIONS

Invest in Your Introduction

When they begin to read your paper, readers are trying to understand a complex text that is new to them. They want to know what it's about, to understand its background or context, and to see its goals or purpose. An explicit and detailed overview will help you show the richness and complexity of your work and set up the reader's expectations for your paper.

For a fuller version of this discussion, download this PDF.

When should you write the introduction?

You do not need to write your introduction first. Some writers write the introduction in the middle of the drafting process once they see the larger direction of the paper; others write the introduction last, once they know the exact content of their work. Try different approaches to see which one is best for you, but always check your introduction before you turn in your final draft to be sure it matches the paper you actually wrote.

√ Maybe

How can you start drafting an introduction?

Depending on the discipline you're writing in, an introduction can engage readers in many ways:

- Ask a question
- Identify a debate
- Give a comparison
- Explain a situation
- Describe a problem
- Quote an authority
- Cite an example
- Set up an intellectual problem
- Offer a hypothesis

√ What is unusual and exciting about the paper.

What should your introduction promise?

Introductions represent a promise the writer makes to the reader. Your introduction should announce your paper's topic and purpose, situate that purpose in relation to what you've discussed in your course or what has already been published on that topic, and offer your readers a preview of how you will satisfy that purpose.

Conference Papes: Give reason why the paper should be accepted!

To address a topic.

One of the important functions of an introduction is to **announce what you are writing about** to your readers.

To present a claim, finding, or argument.

You will often read them toward the end of an introduction: the thesis statement (which could also be a purpose statement or question). This is a promise that your paper is going to make a point, not just cover a topic.

To participate in a conversation.

If you are writing for an instructor, this promise might mean suggesting that your paper will touch on the main topic of the course. If you are writing as a scholar, this promise might meaning explaining how your research will fill an important gap in the existing research.

To excite and engage your reader's intellectual curiosity.

Vithdy Short

Khalil Introduction

for what?

Is capable to cluse

Clustering is an important problem in machine learning. It is the process of partitioning a database into set of clusters whose points have some similarity. There are different types of clustering one of which is density-based clustering which clusters areas that are dense. In fact, DBSCAN is a very well known density clustering approach which covers arbitrarily shaped clusters.

too general.

However, there are some problems that arise with clustering spatial-temporal data, identifying noisy elements and identifying adjacent clusters.

In this paper, ~~the authors~~ we present a new density-based clustering algorithm which is derived from DBSCAN. The algorithm has a minimum density parameter to decide if a region is a cluster. The algorithm starts by retrieving by the first point in the database and retrieves the neighbor points whose distance is closer than a given parameter. The number of points retrieved defines whether it is a cluster or not. It stops when all not noisy points are put in a cluster. ~~The authors~~ present an evaluation of the algorithm by evaluating the runtime complexity. The algorithm is an order of magnitude faster than similar algorithms such as CLARANS and DBCLASD.

The authors present three data-mining applications that use special-temporal data warehouses, with data-integration, data processing and transformation using special transformation functions and visualization of these clusters

The rest of the paper is presented as follows: Section 2 presents related work and basic concepts, section 3 tackles the problems that exist in the current approaches. Section 4 presents ST-DBSCAN and its functioning in addition to a thorough evaluation. Section 5 introduces data mining three applications of the algorithm. Conclusion and future work is discussed in section 6.

Typically, in an introduction you do not discuss alg. and other techn. contr. in detail. Reads like a review

ST view + Motivation - Establish Resch Area / Context first before introducing topics.

What is exciting/unusual/entertaining, of value for the reader of the paper

~~What is exciting unusual of the paper~~

Tianxiao

Should be write more like a preview not only

1 Introduction

Clustering is a technique that partition "similar" objects into groups (clusters), so that objects from the same group are more similar than to the other groups [1]. It is a main task of exploratory data mining and applied in many fields such as machine learning, pattern recognition, image analysis and bioinformatics [2]. Clustering belongs to unsupervised learning since no prior knowledge is known in advance. Traditional clustering techniques include k-means [3], Balanced Iterative Reducing and Clustering using Hierachies (BIRCH) [4], Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [5],Ordering Points To Identify the Clustering Structure (OPTICS) [6], and Statistical Information Grid-based method (STING) [7].

good: some context / motivation

Due to the recent advancement of geographic information science and tracking techniques (e.g. satellites),a large amount of data tagged with both location and timestamp ~~has been~~ obtained. As a consequence, spatial-temporal clustering techniques emerged for knowledge discovery in these situations. Spatial-temporal clustering ~~groups~~ objects based on both spatial and temporal similarities [8].

B are important / desirable

In this paper

~~Here~~, we present ST-DBSCAN as a plausible implementation of spatial-temporal clustering technique. ST-DBSCAN is an extension of popular clustering technique: DBSCAN, which is a density-based clustering technique based on the idea that objects from a dense area should be grouped together. DBSCAN is capable of clustering dataset by arbitrary shapes. However, it ~~still~~ suffers from a problem of indentifying noise objects in a specific dataset that ~~has groups~~ of different density. In addition, traditonal density based clustering technique is inadequate to distinguish adjacent clusters. ST-DBSCAN solves the noise detection problem by introducing a metric called *density factor*. For adjacent clusters, we include a threshold $\Delta\epsilon$ to check the distance between the newcoming object and the cluster. ST-DBSCAN requires four parameters: Eps1 (spatial attribute distance), Eps2 (temporal attribute distance), MinPts and $\Delta\epsilon$. MinPts is determined heuristically as $ln(n)$, where n is the size of the datebase [5]. For a faster information extraction, we intruduced an improved R-Tree [9] indexing method during evaluation.

belongs to group

this seems to be idea of all cluster algorithm

too much details for introduc

ST-DBSCAN was tested in three data-mining applications of real-world problem: sea surface temperature, sea surface height residual, wave height. User-friendly interface was developed for inexperienced users to operate the system. Special functions were designed for data integration, data conversion, query, visulizaiton, analysis and management. All information were contained in a specifically designed spatial data warehouse system.

The rest of the paper is organized as follows. Section 2 reviews DBSCAN. Section 3 illustrates the existed problems of DBSCAN and the corresponding solution. Section 4 demonstrates the implementation of ST-DBSCAN algorithm. Section 5 states its application on real-world problem. Section 6 draws the conclusion of the paper and discusses the future work.

reads more like an abstract

References

[1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 3rd ed., 2011.

[2] L. V. Bijuraj, "Clustering and its Applications," in *Proceedings of National Conference on New Horizons in IT - NCNHIT 2013*, pp. 169–172, 2013.

[3] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, mar 1982.

[4] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," 1996.

[5] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996.

[6] M. Ankerst, M. M. Breunig, H.-p. Kriegel, and J. Sander, "OPTICS : Ordering Points To Identify the Clustering Structure," *SIGMOD '99 Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, vol. 28, no. 2, pp. 49–60, 1999.

[7] W. Wang and R. Muntz, "STING : A Statistical Information Grid Approach to Spatial Data Mining," *the 23rd VLDB conference*, pp. 1–10, 1997.

[8] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal clustering," in *Data Mining and Knowledge Discovery Handbook SE - 44* (O. Maimon and L. Rokach, eds.), pp. 855–874, Springer US, 2010.

[9] A. Guttman, "R-Trees: A Dynamic Index Structure For Spatial Searching," *84 Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, pp. 47–57, 1984.

Introduction Part of ST-DBSCAN paper

Lifeng Yan
1361158

Clustering is the task of grouping a set of objects in such a way that objects in the same group (cluster) are more similar to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields including machine learning, pattern recognition, image analysis, information retrieval and bioinformatics.

too general — but at least the context.

Density-based spatial clustering of applications with noise (DBSCAN) is a clustering method that is one of the most common clustering ones and also most cited in scientific literature. Two parameters are required for DBSCAN algorithm: neighborhood radius ε (eps) and the minimum number of points required to form a dense region (minPts), and the important concept of core point is defined as data points with number of samples in their ε neighborhood larger than the given threshold minPts. And it divides regions with enough density into clusters, where clusters are defined as the maximum set of core points that could be reached by another one through a chain of core points (density-connected). The process is to scan the whole database first, find out one core point and then extend it using cluster definition mentioned previously. The DBSCAN algorithm can find arbitrarily shaped clusters, is robust to outliers and is mostly insensitive to the ordering of the points in the database.

too much detail — as it is not a contribution of the paper

there are still other attributes

??

On the other hand, spatial-temporal data is consisted of two dimensions: spatial and temporal. Temporal dimension describes to which extent the evolution of the object is captured by the data, and the spatial dimension describes whether the objects considered are associated to a fixed location (e.g., the information collected by sensors fixed to the ground) or they can move, i.e., their location is dynamic and can change in time. A very basic example of spatial-temporal information are spatiotemporal events, each event is usually associated with the location where it was recorded and the corresponding timestamp. Both the spatial and the temporal information associated with the events are static, since no movement or any other kind of evolution is possible. Finding clusters among events means to discover groups that lie close both in time and in space, and possibly share other non-spatial properties.

Kind of redundant

Not very clear! What are the applications?

~~why is this the case~~ — what is the problem?

This paragraph is better!

However, DBSCAN and other density based algorithms could not be used directly on spatial-temporal data and 3 major extensions should be made, since following 3 problems exist. First is that the DBSCAN algorithm uses distance only valid for spatial data, so the distance measuring criteria should be expanded for non-spatial values. Secondly, existing density-based clustering algorithms, including DBSCAN could not give out good results when clusters of different densities exist, and the noise objects should be found to improve. Thus the concept of density factor which is the degree of the density of the

not true

??

Remove

other attributes
temporal attributes

is in the paper

They all have adjacent clus

cluster sh~~ould be~~ introduced, and each cluster should be assigned with such a density factor. Third problem is that existing density-based methods do not have satisfying results when dealing with datasets that contains adjacent clusters. And the solution should be comparing the difference between the average value of a cluster and new coming one with a threshold, since cluster objects should be within a certain distance from the cluster means. Motivated by these extension, a spatial-temporal version of DBSCAN, that is the ST-DBSCAN algorithm should be proposed.

??

is discussed in the remainder of the paper

the paper proposes a solu

The remaining part of the paper will be organized as follows: section 2 will talk about the related works and concepts in them; section 3 will have a detailed analysis of problems of existing methods; section 4 will introduce the ST-DBSCAN algorithm, and evaluate its performance; section 5 will discuss about the application of the algorithm and also practical experiment results; section 6 is about the conclusion and future work.

present a promise to the reader

to accomplish what?

objects near the border of adjacent clusters.

ST-DBSCAN: An algorithm for clustering Spatial–temporal data

Derya Birant, Alp Kut

Department of Computer Engineering, Dokuz Eylul University.

Stating the problem DBSCAN faces that are addressed by ST-DBSCAN more clearly and transparently could strengthen this introduc

good: identifies the task at hand

INTRODUCTION

too general!

Knowledge Discovery in Databases (KDD) is an interdisciplinary field, aimed at extracting valuable knowledge from large databases. At the core of the KDD process is the Data Mining step which embraces many data mining methods. One of the most important and influential step of data mining is Clustering. Clustering is an unsupervised learning process, it groups physical or abstract objects into classes of similar objects. It finds its application in many areas including weather report analysis, data mining, machine learning, and image analysis. It is classified into various types. Density based clustering approach is used in the paper. The idea behind them is that objects which form a dense region should be grouped together into one cluster. → true also true for other approaches

Spatial

Geographic and temporal properties are a key aspect of many data analysis problems in business, government, and science in today's world. Spatial-temporal clustering is a process of grouping objects based on their spatial and temporal similarity. Due to the evolution of all kinds of location-based or environmental devices that record position, time properties of set of objects in real time, Spatial- temporal clustering gained popularity especially in geographic information sciences. As a result novel approaches to knowledge discovery are required for large amounts of data. This papers main aim is to focus on clustering spatial-temporal data. A novel Spatial-temporal algorithm names ST-DBSCAN is presented in this paper which uses density based clustering methodology.

mostly good!

The density-based clustering methods use a density threshold around each object to distinguish the interesting data items from the noise. Its examples include DBSCAN, DENCLUE, OPTICS. All of them have certain disadvantages, firstly when clusters are adjacent to each other, and when clusters of different densities exist. DBSCAN (Density based Spatial Clustering of applications with noise) has been chosen as a base algorithm over which ST-DBSCAN has been developed because of its ability to process large databases, it can discover arbitrary shaped clusters such as linear, concave and predetermination of number of clusters is also not required.

Seem be two different problems

Seems to be unrelated.

KAJOL

ADVANCE MACHINE LEARNING
WRITING TASK 1

1358284

Obtaining clusters with diff densities

Does not do a good job to sell ST-DBSCAN

ST-DBSCAN (Spatial-Temporal DBSCAN) clusters data according to non-spatial, spatial and temporal neighbors. It uses the concept of density factor, degree of density of cluster to overcome problems related to different densities. Moreover it compares the average value of a cluster with new coming value with respect to the threshold to deal with adjacent clusters.

??

, applied to accomplish what?

A Spatial data warehouse system has been developed to demonstrate our algorithm with respect to real world using data mining applications. Various steps in KDD like data integration and selection, data preprocessing ad transformation, data mining have been performed and obtained results have been compared and evaluated. Special functions like grids, tables, graphical illustration, maps and reports were developed along with easy interfaces so that system could be operated by any unexpert user. Thus it can be said that coupling in space and time has been well highlighted in this paper by looking at the obtained experiment results. Spatial-temporal data can also be used in other important application domains like traffic control, meteorology, Geophysics, Biology and ecology. Spatial-temporal datasets are highly auto correlated and are embedded in continuous space in contrast with the classical datasets needing much serious attention.

Seems to have not much to do with the data warehouse

Unused

misplaced and redundant

Each Section of the paper speaks about some information in support towards clustering spatial-temporal data. Firstly a brief description of the clustering algorithms and the dependence of other density based algorithms on DBSCAN and terms used in DBSCAN have been highlighted in Section 2. Secondly, how ST-DBSCAN overcomes the disadvantages in existing approaches is discussed in detail in Section 3. Logic behind the working of ST-DBSCAN and its performance measure is presented in Section 4 and this is followed by its applications and implementation of spatial data warehouse system is demonstrated in detail and results are shown using pictorial representations. The paper is concluded pretty well and scope of improvements in the future has been mentioned.

the proposed

??

Focus more on technical accomplishments, rather than getting lost into very technical details with respect to how ~~the~~ things are actually done

Writing style good!

Introduction

By YONGLI ZHANG 1269577

Nowadays, the spatio-temporal clustering has become an extremely popular research area of GIS-related knowledge discovery, which aims to group objects into meaningful clusters based on their spatial and temporal similarity. With the help of clustering, we can identify and extract patterns, trends, or evolutional changes with regard to the geographical phenomena. And practically speaking, spatio-temporal clusters may also serve as valuable input to other data mining techniques, such as association analysis.

Decent Motiv

paragraph should be merged Somewhat redundant

Also, due to the availability of cheap sensors and the advances in remote sensing and sensor networks, we have witnessed an exponentially growth of different types of spatio-temporal datasets. And extracting knowledge like spatial and temporal patterns from such dataset is very important in many applications, such as ~~geographic information system~~, weather forecasting, environment protection, etc. However, the combination of time domain makes the mining of spatio-temporal data more challenging and complex. And most traditional algorithms are inefficient in clustering spatio-temporal datasets as they do not incorporate the idiosyncrasies of the spatial domain and temporal domain.

with what?

In this paper, we propose a novel clustering method called ST-DBSCAN, which is able to discover clusters with respect to spatial, non-spatial, temporal values of the object. It is constructed by modifying DNSCAN, one of the most common clustering algorithms. Given a set of points in some space, DBSCAN can group together points that are closely packed together and mark those points that lie alone in low-density regions as outliers points. DBSCAN has several advantages: it is robust to noise, and it does not need to specify the number of clusters as priori, and it can find arbitrary shaped clusters. However, there are some issues ~~existing~~: firstly, DBSCAN needs modifications in order to be used for spatio-temporal data; secondly, problems with dealing with varying ~~in case of clusters with totally~~ different densities, some noise data may not be detected by DBSCAN; thirdly, adjacent clusters may make the objects around the border totally different from each other. And our proposed algorithm, ST-DBSCAN tries make some modifications and address solve those problems. ??

good motive what is novel with that gets lost in technical details.

???

with respect to DBSCAN

propose an extension of DBSCAN called

problems with

100% not clear

clusters objects on the boundaries of adjacent clusters mean is what?

Furthermore, we present a spatio-temporal data warehouse which could be used to handle and cluster a wide range of spatio-temporal dataset. And we also applied our algorithm to the data warehouse, which demonstrates the applicability of our proposed algorithm. And the experimental result shows that our proposed algorithm is effective in mining spatio-temporal data and overcomes some shortcomings of DBSCAN as well.

??

The rest of paper is organized as follows: Section II introduces the related works about existing algorithms and basic concepts about density-based algorithms. Section III discusses the problems of existing spatio-temporal clustering approaches. Section IV introduces the novel ST-DBSCAN algorithm in detail. Section V demonstrates the applicability of our algorithm to real world problems. Section VI concludes our study and discusses the potential future works.

demonstrats
the ~~te~~ proposed
extensions of
DBSCAN work