Could be more critical.

Not very clear

Not well written, particularly the first page

Paper Review:

ST-DBSCAN: An algorithm for clustering spatio-temporal data

By Yongli Zhang, Lifeng Yan

detect clusters with varying densities

1 Summary:

Neutral Summary

This paper is a very good paper from all aspects. This paper proposes a novel algorithm called ST-DBSCAN, which makes some modifications and improvement on the DBSCAN algorithm. And in contrast to existing density-based clustering algorithm, the proposed algorithm can discover clusters with respect to non-spatial, spatial and temporal ~~values~~ attributs of the objects.

meaning what?

To be specific, the proposed algorithm improved DBSCAN in 3 aspects: firstly, ST-DBSCAN can cluster spatio-temporal data ~~which is not applicable to DBSCAN~~, secondly, ST-DBSCAN can detect noise points by assigning each cluster a density factor, even if when clusters of different densities ~~exist,~~ thirdly, ST-DBSCAN can identify adjacent clusters by comparing the average value of a cluster with new coming value.

What is a value of a cluster?

by requesting that the objects in the cluster are not too far away from what?

Moreover, in this paper, a spatial–temporal data warehouse system is designed for storing and clustering a wide range of spatial–temporal data, which provides user-friendly interfaces to allow relatively inexperienced users to operate. Also, special functions were developed for data integration, data conversion, visualization, analysis and management.

English - on what

2 Contributions:

how

What is the improvement over DBSCAN?

Novelty: The idea of modifying DBSCAN algorithm to make it applicable to analysing spatio-temporal data is novel and interesting. And this paper focuses on some restrictions of DBSCAN algorithm, and proposed an improved version, which overcomes some problems of DBSCAN, including the problem of identifying noise points and adjacent clusters in some cases, and the problem of clustering spatio-temporal data.

the case of the cluster can repeat to its non-spatial attribute

What makes it better? meaning what? differ

Somewhat vague ?

Technical Quality: The overall quality of this paper is very good. As is shown in the paper, the experiments are well performed with accuracy and proven to be justified. For example, in order to test the applicability of the proposed algorithm to real world problem, they apply the algorithm to the data warehouse, and the data mining result of the temperature database is really sound and convincing. The authors explain the experiment in details, and it should be easy for later researchers to replicate those results. However, the authors don't discuss the sensitivity of their algorithm to parameter settings. Also, the authors mention some weaknesses of their approach but they don't give detailed explanation or assessment to those weaknesses.

What makes them convincing

DBSCAN is already a spatial clustering algorithm.

What weakness?

Potential Impact and Significance: The knowledge discovery in spatial-temporal data is much more complex than non-spatial and temporal data. The algorithm proposed in this paper can extract useful knowledge effectively from spatio-temporal dataset, which is the state of the art. Therefore, this algorithm ST-DBSCAN can be used in many applications such as geographic information systems, medical imaging and weather forecasting, which is very promising.

Good!

Clarity of Writing: The paper is well organized and clearly written. For example, the use of temperature warehouse example gives a clear proof and idea how the proposed data warehouse system and the proposed algorithm ST-DBSCAN work. Also, by introducing the problems of DBSCAN, the authors explains clearly how the extension and modification is done.

3 Strongest points:

- The paper proposed a novel algorithm ST-DBSCAN as an extension of DBSCAN, which can cluster spatio-temporal data which is not applicable to DBSCAN.
- ST-DBSCAN can identify noise objects and adjacent clusters which exist in some special cases.
- In addition to a novel algorithm being proposed, this paper presents a spatial–temporal data warehouse system designed for storing and clustering a wide range of spatial–temporal data.

4 Weakest points:

- The input parameter of ST-DBSCAN are not generated automatically.
- The authors don't discuss the sensitivity of their algorithm to parameter settings.

5 Educational value:

What can students learn from it

In general, this paper is a good one with high educational value. It is well organized, starting with introduction of the basic algorithm DBSCAN they will later modify algorithm based on clear definitions and concepts aided with figures, then disadvantages of existing method are given, also in the form of definition and figures. These two parts together serve as a good introduction to the goals and objectives in the field of research. When it comes to their ST-DBSCAN algorithm, pseudocode is given and the authors also give out performance evaluation of the part. With this kind of structure, it is easy for those readers who are not familiar with this topic to understand what they are doing, and also why they are doing this. Furthermore, with detailed information and background knowledge about both the original and their modified algorithms provided, the paper might also inspire new thoughts, and thus can be a good starting point of new

mostly good

research. One more thing to mention is that the data they used for experiments are interesting and easy to understand, another proof that the paper is friendly to beginners.

However, the paper failed to discuss more about the big picture of the field of interest, ~~and no other~~ clustering methods are introduced~~, so~~ this could be a weakness when considering its educational value. ~~But~~ considering the fact that this paper is not a review paper, it is not strange to focus on the algorithm itself.

6 Comments and questions:

1). The concept of transaction time in the introduction part is not very clear. Although the paper focus on valid time aspect, I think it also should talk more about how the database store the temporal data when it mentioned building a data warehouse later and clarify the difference, otherwise it may be confusing since the temporal neighborhood might be "made up" by the database.

2). I think specific numbers should be added to show the distance between two classes that could be called "adjacent clusters" in part 3.3, and this kind of details should be helpful when choosing the parameters.

3). Further explanation about why the filters that reduced actually created paths could work in part 4.2 is needed.

4). I think the future work part in the conclusion needs to be enhanced to be more convincing.

7 Broader impact:

The authors have already discussed about what kind of aspects clustering could be used in, and since the paper is for a special kind of data, spatial-temporal data, the application of the proposed method could naturally be used in discussed aspects, such as geographic information systems, medical imaging, and weather forecasting given required spatial-temporal data. One real world application apart from the sea-surface temperature could be in video process and analyzing. And since they built a data warehouse, it is possible to use their algorithm on other kinds of data that could be put into the data warehouse.

There is a short paragraph about future work in the conclusion part, unfortunately it is not of much value. It however, points out that more attention should be paid on deciding the parameters in the algorithm, which requires both experiments and theoretical analyze. Since the paper talked little about other aspects apart from the clustering algorithms themselves and their experimental results, it is not likely that this paper could somehow connect some different, originally disconnected research communities. Good!

8 Numerical Scores

Criteria	Novelty	Technica l Quality	Potential Impact and Significance	Clarity of Writing	Educational Value	Broader Impact	Overall
Score	6	6	6	7	7	6	

9 Usefulness of Homework 2:

Writing this review not only makes us more familiar with the paper itself and the methods presented in it, but inspired us to think deeply and broadly about what is behind the paper as well. We could learn how to judge a paper from multiple angles, and it makes us consider important things we should care about while writing our own papers, not just contents and organization, but also things like educational value and application affairs. We can know that what kind of paper is helpful to the ordinary readers, and which points are vital for our papers to be accepted. The last thing is that it is of course very helpful when we are trying to officially review a paper for academic journals, which is very likely in our Ph.D career.

Homework 2: Reviewing of ST-DBSCAN Paper

Kunal Parmar - UHID: 1329834
Ha Le - UHID: 1344084

Not always convincing, but also some interesting observations
Could be more critical!

1. Summary

Clustering is the technique for grouping objects according to their similarity for knowledge discovery in large databases. Density-based clustering algorithms are based on the idea that objects which form a dense region should be grouped together into one cluster. They search for regions of high density in a feature space that are separated by regions of lower density.

Clustering of spatial-temporal data is one of the popular fields of knowledge discovery and data mining due to the increasing amount of spatial-temporal data being produced. The paper proposes a new density-based clustering algorithm, ST-DBSCAN, based on the DBSCAN algorithm. DBSCAN was chosen because of its ability to discover clusters of arbitrary shapes and its ability to process very large databases.

how does this contribute to that

ST-DBSCAN improves DBSCAN in three directions, namely, i) clustering based on spatial, non-spatial and temporal attributes by introducing similarity measures for spatial and non-spatial attributes, ii) detection of clusters of different densities by assigning a density factor which is the degree of density of a cluster iii) identification of adjacent clusters by setting a threshold value for the difference between the average value of cluster and the object.

with respect to non-spatial attributes

2. Reviewing Criteria

how does it improve OPTICS?

Novelty: (6/7)

The proposed algorithm performs clustering by taking into account spatial as well as non-spatial attributes. There are other density based algorithms, for example OPTICS, which are good at discovering clusters of varying densities. However, the algorithm proposed in this paper improves DBSCAN in three directions which includes discovering clusters of varying densities. Especially, the concept of density factor, which describes the degree of the density of a cluster, is very unique. The paper does propose a novel technique for clustering spatial temporal data.

but it is not clear how it is used by the clustering algorithm — see pseudo code

Technical Quality: (4/7)

The paper applies the proposed algorithm to three applications using a spatial temporal data warehouse. The runtime complexity of the algorithm is not changed compared to DBSCAN. The approach for the algorithm is very clearly explained and can be easily replicated for future research. The authors have considered prior work in the field and have given explanations about how their approach improves upon already existing methods. A potential weakness/vulnerability of the approach could occur when applying the algorithm to very large databases. This has been indicated as future work based on the paper. On the other hand, the proposed algorithm presents four parameters, but these parameters are picked based on a heuristic suggestion from another paper, which is insufficient to prove the validity of these parameters. There is not enough technique explanation on how good and bad of selecting parameters. ?

why

Significance and Impact: (6/7)

The paper provides an extension to the state of the art in density based clustering methods. The paper addresses the problem of applying clustering methods to spatial temporal data. Researchers who want to improve and develop some more techniques for density based clustering of spatial temporal data are likely to refer to this paper. Also, researchers trying to develop similar extensions for spatial temporal clustering using other methods, like hierarchical, model-based, etc., are likely to refer to the paper. People from different fields, not from machine learning, can find this paper useful. Examples can be in town planning, vegetation vigor, etc.

Clarity of writing of the paper: (5/7)

For the most part, the paper is very clearly written. It is well-organized, provides appropriate explanation of the approach taken and uses examples and figures in a nice manner to demonstrate results and flaws of the current approach. However, ~~the solution to the~~ problem of identifying adjacent clusters is not comprehensible.

3. Strongest points of the paper:

- The algorithm proposes a novel approach for clustering spatial temporal data using one of the better density-based clustering algorithms.
- The paper is very clearly written, along with appropriate examples and figures wherever needed.
- Three applications of the algorithm to real world problems and results for these applications have been presented.

4. Weakest points of the paper:

- There is not an appropriate explanation given for the section for solution of the algorithm to adjacent clusters (Sec 3.3).
- The selection of parameters is unclear.

5. Educational Value of the paper:

The paper is a good starting point for carrying out research in developing algorithms for clustering spatial temporal data. The paper does a good job in introducing the methods for developing algorithms for spatial temporal clustering. Students looking to do research in spatial temporal application such as geographic information systems, medical imaging can benefit from reading the paper.

The paper provides a good and clear understanding of the underlying approach and benefits for using the algorithm. It also introduces future direction of research in the field of applying the algorithm to very large databases and finding other useful heuristics for the algorithm. The results from the paper can be easily reproduced and work on extension of the current approach can be carried out.

6. Comments/Questions:

- Explanation of how the problem of identifying adjacent clusters is solved is not clear. (Sec 3.3)

- The definition of border point in DBSCAN is not extended in ST-DBSCAN.
- There is no comparison between the proposed method with other spatial-temporal data clustering methods.
- Fig. 5 is not necessary for the purpose of this paper.
- Fig. 8 and Fig. 9 is blurry.

reads more like a future work sec.

7. Broader Impact:

An algorithm like ST-DBSCAN designed for clustering spatial-temporal data can be used in applications such as geographical information systems, medical imaging, and weather forecasting. Some examples of applications can be detecting changes in vegetation vigor, town planning according to commuting patterns, etc. A similar application would be in trajectory clustering where you cluster data based on the trajectories of objects in the spatial-temporal domain.

Future research directions could be in developing spatial-temporal clustering algorithms that improve on other types of clustering methods, i.e. i) partition algorithms (k-means, k-medoids), ii) hierarchical algorithms (CURE, BIRCH), iii) grid-based algorithms (STING, WaveCluster) and iv) model-based algorithms (COBWEB). Other direction of research would be outlier/anomaly detection in large databases in spatial-temporal domain.

How does it contribute to solve broader problems!

8. Numerical Scores:

- Educational value: 6/7
- Broader Impact: 6/7
- Overall score: 6/7

9. Usefulness of Homework 2:

Homework 2 has been very helpful keeping the course in mind. The course focuses on introducing graduate students to research in machine learning and give them an idea about research in general. Reviewing a paper in an important aspect of research. Students need to look at a lot of different approaches to assess which one would be best for their research. They need to be able to evaluate, criticize and judge papers. Students are not asked for their opinion generally in school and hence never improve their ability to criticize. They need to be able to criticize to differentiate between good and bad which will help them in their research.

Could be more detailed!

Advanced Machine Learning
Prof. Dr. Eick

Homework 2
Date: 11.12.2015

Mostly understandable, and clearly presented although the English could be improved.

Some valid criticism! more than other case up while.

Students:
- Pham Dinh Nguyen (1353607)
- Nikolaos Sarafianos (1392654)

Summarize what the research area and the topic of the paper is and what its contributions are 2-4 paragraphs); *write in a neutral or positive tone no matter how bad the paper is*!

The paper study is mostly related to clustering of spatiotemporal data and density-based clustering. The main contributions are: a new algorithm which extends DBSCAN, a data warehouse system specialized for analysis of spatiotemporal data.

good: Contribution are clearly described

The proposed algorithm is ST-DBSCAN, based on DBSCAN and having the same runtime complexity. Three important improvements: i) ST-DBSCAN can cluster spatiotemporal data according to its non-spatial, spatial and temporal attributes, overcoming the limitation of DBSCAN. ii) Unlike DBSCAN, which employs a fixed density, ST-DBSCAN can detect clusters with different densities, in the presence of noise. iii) ST-DBSCAN decides determines the borders of adjacent clusters ~~borders~~ by using an extra criteria: maintaining the cluster average value (such that the non-spatial attributes in a cluster ~~has~~ have lower variation).

The proposed data warehouse is briefly mentioned, but seems to covers all the essential tasks: data management, analysis, graphical interface, and result visualization.

Assess the contributions of the paper; follow the the KDD 2012 Reviewing Criteria (at the end of this slideshow). In particular, assess the novelty, technical quality, significance and impact, and clarity of writing of the paper (2-6 sentences for each criterium). If the paper makes contributions that do not fit into these 4 criteria, summarize those in an optional "*other contributions*" paragraph)

1. Novelty: 6/7
   The new algorithm is indeed a new contribution. Though it is based on DBSCAN, the three improvements are major ones. The application is extended to spatio temporal data, which is a remarkable contribution. meaning what

2. Technical Quality: 5/7
   The authors are informative on investigating previous works, and their new ideas are clearly introduced. However, there is a lack of throughout discussion on some parts: i)

The analysis of temporal attributes is quite simple. Ii) The heuristic choice of input parameters, and how the tuning of these parameters affects the results. iii) No numerical results to back the claim of their algorithm performance.

3. Significance and impact: 6/7
   The new algorithm has the potential to apply on a wide class of spatio temporal datasets. It addresses an important problem but within the Machine Learning and data mining fields. It proposes a significant advance on the spatiotemporal clustering problem and thus it will be cited in the following years. However new methods in domains such as clustering attract a lot of attention from the scientific community and thus existing methods have a certain lifespan which most of the time is not more than 10 years.

4. Clarity of Writing: 6/7
   In general the authors have a nice presentation of their ideas. The only part that is not clear is their discussion on clusters boundaries ~~decision~~ determined. Additionally, equations should be in the center of the line, Figures 8 and 9 are blurry and in a grayscale print the results are not clear in figure 9.

What are the 3 strongest points of the paper (just one sentence for each point)?

1. The 3 Improvements introduced on top of DBSCAN that address its limitations without increasing the runtime complexity
2. The experiment uses real and large dataset, and the results are visualized and verified by domain knowledge
3. Well structured, the basic concepts help build the necessary knowledge to understand the proposed method.

What are the 1-3 weakest points of the paper (just one sentence for each point)?

1. The evaluation of data mining results section is not clear and does not help the reader understand why the KDD is necessary to test the proposed method
2. There are no comparisons with other approaches (not DBSCAN, as DBSCAN cannot analyzes spatiotemporal data) in order to check if actually the proposed improvements result in a better performance.
3. There discussion on the heuristic choice of the parameters is insufficient and the analysis of temporal attributes is somewhat simple

Assess the educational value of the paper for graduate students (1-3 paragraphs)! Is the paper a good starting point to do work/research in the area? Does the paper do a good job in introducing the goals and objectives and the methods of the field of research? Does the paper do a good job in getting graduate students excited about working in the research field? What did you learn from reading the paper?

The paper is a good starting point for research in the area (clustering of spatial temporal data), assuming that students have been prepared with the related literature. Though the papers mention quite a lot of related clustering methods, only knowing some of them is enough to understand the paper. The others methods are mentioned, thus can serve as a guidance for further reading. The problem, that this paper tries to solve, is clearly presented, and is an important one. It is well structured in showing: the current methods, the current limitation, and the proposed solution. This paper is good in motivating the research. It also shows the amount of work required in similar research: ones must have done quite a study/experiment with the current methods, and then realize the limitation, before come up with a new idea. The implementation and experiment should also be convincing: using real dataset, and giving visualization. From reading this paper we realized that existing density-based clustering methods cannot handle effectively clusters with different densities. It was also interesting how you can use a state-of-the-art approach and by adding some extra characteristics to produce a high-quality paper.

Numbered List of Specific Comments and Questions (e.g. *the claim stated in the second paragraph is not clear; I do not agree with the conclusion in the third paragraph..., symbol x was never defined, it is not clear to me what the purpose of Section 4.3.2 is; the author introduced formulas 2.4 that are never used in the remainder of the paper*, I do not understand what the term x means,...). Each review should have 3-7 specific questions/comments!

1. It is not clear how the problem of identifying adjacent clusters is solved and if the proposed solution is effective in solving this problem. Our guess is: trying to maintain the mean values of non-spatial attributes. For example: one cluster should not have some high temperature points on one side and low temperature ones on the other side.
2. The evaluation of the data mining results is very difficult to be reproduced since it requires a lot of parameters, data provided by the links in the footnotes in page 217 and also the data warehouse.
3. In the conclusions section the authors do not support why they decided to use this warehouse to test their method instead of another dataset.

Broader Impact (1-2 paragraphs); what real world applications will arise from this work? Assess how the paper will help society to make earth a better place! Does the paper foster new research/new approaches that could be investigated in future research? Does it establish new connections between different, originally disconnected research communities?

Density-based clustering algorithms do have an impact in real-world applications since the provide solutions that have been used and are currently used to many scientific and industrial domains. Thus, ST-DBSCAN is an additional tool that can be utilized by researchers to solve problems but it will definitely not make the world a better place since it is a publication that just proposes a few modifications and not something groundbreaking such as new feature descriptors (e.g., SIFT or SURF).

Give the paper a numerical score (1-7) using the KDD-2012 Criteria; 7 scores (add scores for educational value, broader impact and overall score!)!

Educational value: 6/7
Broader Impact: 6/7
Overall score: 6/7

Assess the usefulness of Homework2 (1 paragraph)!

Homework 2 was really interesting and helpful since I now have a more clear idea on how to assess papers that I will have to review in the future. The questions were nicely structured and their sequence helped me to first approach the publication based on them and then to search for its strong and weak points. For example, the fact that question 2 asked contributions based specifically on some aspects (e.g., novelty) guided me while reading the publication to search for the related parts.

Kajol ADVANCE MACHINE LEARNING Arthur Dunbar

REVIEW TASK 1

Valid

Some Criticism!

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

Not convincing in a few places

Derya Birant, Alp Kut

Department of Computer Engineering, Dokuz Eylul University.

I. Summary

"ST-DBSCAN: An algorithm for clustering spatial–temporal data" introduces a new algorithm for clustering spatial-temporal data, or, data grouped based on their spatial and temporal similarity. Spatial-temporal data plays a major role in many data analysis tasks including those related to geographic information systems. Most clustering algorithms, however, do not take Spatial temporal data in to account, and clustering algorithms that are not designed for spatial-temporal data are ill equipped to handle it. Before introducing their new algorithm though, the authors ~~describe~~ many things that are needed to understand it including density based clustering algorithms.

Spatial

The next sections of the paper contain a brief summary of the existing clustering algorithms followed by the importance of density based clustering algorithms. Various density based algorithms are listed by the authors, but their main focus is on the DBSCAN (Density based Spatial Clustering of applications with noise) algorithm which is used as the base for their new algorithm, ST-DBSCAN (Spatial-Temporal DBSCAN). The reason they chose DBSCAN was because it is good at discovering arbitrarily shaped clusters. Moreover it does not require apriori knowledge of the number of clusters, and is fast enough to be used on large databases. They then describe the basic terminology and concepts involved in using DBSCAN and they back up their explanations with examples that show its strengths and weaknesses. According to the paper, DBSCAN's weaknesses are, it cannot work with temporal data, it has trouble clustering data with different densities, and it can blur clusters together when there is a large amount of noise.

Is this the correct term?

After listing the weaknesses of DBSCAN, they explain how ST-DBSCAN overcomes them. First, ST-DBSCAN can cluster non spatial, spatial and temporal data because it adds a new parameter for non-spatial data. Second, using a 'density factor,' they are able to have clusters of different ~~densities~~ exist in the same data set. Finally, they remove the problem of having the clusters blurring together ~~under strong~~ noise by adding a new factor, Δε, which only allows new points to be added when they are within a certain distance of the current cluster mean. They then go on to explain how the new attributes are derived, and then they explain how the new algorithm works, and they provide pseudo code.

With respect to the non-spatial attributes.

REVIEW TASK 1

Next, they discus application and performance. The ST-DBSCAN algorithm has the same O(nlogn) time complexity of the standard DBSCAN algorithm. They also explain a new spatial-temporal data warehouse system that is used in conjunction with ST-DBSCAN. Finally, they show three applications of the ST-DBSCAN algorithm on data from the Mediterranean, Aegean, Marmara, and Black Seas that cluster based upon surface temperature, height of the water, and height of the waves.

Finally, they end the paper with a conclusion section.

II. Reviewing the paper in terms of the KDD 2012 criteria

Novelty

This paper introduces a new algorithm ST-DBSCAN, an improvement on DBSCAN, and it seems useful for clustering spatial-temporal data. It takes an old algorithm and updates it in a novel way to solve a new problem (spatial temporal clustering has other algorithms). This is novel, but it is not a paradigm shift.

Technical Quality

More than true —
Where does it occur in the pseudo-code

There are many places where it is very good like where it discusses the DBSCAN algorithm and its strengths and weaknesses, and where it uses clear diagrams to back up its arguments. The authors even discus the prior knowledge that is necessary to understand their new algorithm. There are, however, places where they are not so good. For example, they poorly explains some steps like how the density factor is used. Also, their demonstration of the algorithm is not the greatest because they apply it to a very simple task. They also do not go in to detail about where their algorithm might be weak. Finally, they did not compare the clustering results of ST-DBSCAN against other algorithms, so we are left without a strong sense of how well it performs relative to other algorithms.

Potential Impact or Significance

Very few people do that

Improving DBSCAN so that it applies to a new domain where it couldn't reasonably be used before is potentially significant, and the results it obtains seems useful. As a result, this paper could be very significant. Unfortunately though, they do not compare their results against others, so this is uncertain.

Clarity of Writing

Though mostly well written it does have a few places where ~~it's~~ grammar is iffy. It does not detract from the understanding to the paper however. ~~Also,~~ some ~~things are~~ explained in fragmentary parts across multiple paragraphs and diagrams (like $\Delta\varepsilon$). The paper would be much clearer if these things were all explained in a contiguous section. Finally a few concepts are never completely explained.

Which? List them!

Kajol | ADVANCE MACHINE LEARNING | Arthur Dunbar

REVIEW TASK 1

Additional Points

Some of the diagrams are great, especially towards the beginning when it is discussing DBSCAN and its problems.

III. Three strengths of the paper:

1. It explains well the main idea behind how ST DBSCAN works, and it gives good reasons for using algorithms like it that can work with spatial-temporal data.
2. It explains the things that need to be known prior to understanding ST DBSCAN (what density based clustering is, what DBSCAN is etc)
3. It gives pseudo-code for the algorithm and it explains it.

IV. Weaknesses of the paper:

1. The application of the algorithm seems very basic and does little for motivating its use. Perhaps they should have explained more information that they discovered than things like the area East of Crete has low waves and the area South of the Adriatic has high waves. It seems like this could have been done without a clustering algorithm (maybe a heat map for wave height).
2. Though many things are well explained, there are places where they skip over important details. For instance, they do a poor job explaining how the ~~density~~ factor is used. It explains why its needed and its impact, but not where in the equations ~~it goes.~~
3. It does not compare ST DBSCAN to other state of the art clustering algorithms of its type. This might be difficult to do for a clustering algorithm though.

V. Educational Value of the Paper: quite short

This paper has definite educational value. It not only introduces ST DBSCAN, but it goes through the trouble of explaining each idea that ST DBSCAN is based upon. This gives a broad overview of the topic and a practical application of the knowledge provided.

Another thing that it does well is it explains some of the pitfalls of DBSCAN and of density based clustering in general. This helps students to realize when it is appropriate to use these tools, and when it is not.

Does not occur in the pseudocode

VI. Specific Comments and Questions:

1. How is the density factor used? It says how they calculate it, and it says that they assign one to each cluster, and it says that it is for letting the algorithm recognize more sparse clusters when there are other denser clusters in the data. It however never says explicitly how it does this.

2. Similarly delta epsilon is not concisely and adequately explained. It manages to give enough information to find out what it is, but it is scattered between the times it is mentioned. It also never explains how to determine a good value for it.

3. How useful is the obtained data? The paper could have done better explaining a use of the discovered information, or some sort of important fact discovered by the clustering algorithm. Perhaps it was not the point of the paper to explain the Mediterranean sea's surface conditions, but highlighting some important discovery would motivate the use of the algorithm.

4. Does spatial-temporal data apply to more than 2 spatial dimensions across time? Can it be used for 3 spatial dimensional data?

5. They split the data clusters into height, heat, and wave height diagrams, but could they combine them to find clusters on the combined data?

VII. Broader Impact:

The paper provides a novel algorithm that is an improvement of an older algorithm that allows it to work in a new domain. This is useful and has impact, but it is not some sort of revolutionary change.

Extending DBSCAN to explicitly work with spatial temporal data is a good idea, and doing it without making it asymptotically slower is good too. This paper seems like it could have significant impact within spatial-temporal clustering.

The actual data they applied it to probably does not have much of an impact though.

VIII. Score from 1 to 7:

Novelty: 6

Technical Quality: 5

Potential Impact or Significance: 5

Clarity of Writing: 5

Additional Points: 5

REVIEW TASK 1

Educational Value: 6

Broader Impact: 6

Overall Score: 5.5

Should be moved to the fre

Without having direct access to the data used or the algorithm being covered, this paper seems technically sound, impactful, and novel. It is decently well written, and it takes the time to explain concepts that are necessary to know to understand the algorithm before explaining the algorithm.

The biggest criticisms of the paper is ~~that~~ it does have a few key concepts that are poorly explained (density factor) and a few things that are explained are scattered piecemeal across several paragraphs. Also ~~the application~~ it runs the algorithm on seems weak and not very useful for gauging it's worth.

IX. Usefulness of Homework 2:

We think it is a good idea to give an assignment where students have to review papers. Reviewing papers is an important skill for any person who is in a scientific field because it allows them to determine the quality of the ideas that papers they are reading give. It also helps those who work as researchers learn how to communicate their findings by learning what a good paper looks like.

If you know the criteria that a reviewer will use to judge your paper, you are in a much better position to tailor your paper so that it has a higher chance of being accepted, and you are more likely to do research that makes for a good paper. Theory only gets you so far in situations like this, so practice is necessary.

Could be more critical! Not convincing in a few places

Too broad for a paper summary

Not always clear, particularly the first page

Summary

Clustering is a technique that partitions "similar" objects into groups (clusters), so that objects from the same group are more similar than to the other groups. It is a main task of exploratory data mining and applied in many fields. Clustering belongs to unsupervised learning since no prior knowledge is known in advance.

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.

varying density

This paper presented a new algorithm for spatial-temporal clustering: ST-DBSCAN, which is an extension of DBSCAN. DBSCAN is a density-based clustering technique which targets arbitrary shaped clusters. The author proposed to use DBSCAN to cluster spatial temporal data by integrating both spatial and temporal attributes into distance measurement. Other than that, ST-DBSCAN solved the noise detection problem of DBSCAN in different-density dataset by introducing a new metric called density factor. In the case of adjacent clusters, ST-DBSCAN uses a threshold to decide whether to include the new point or not during cluster expanding. To make it applicable for real-world application, the author proposed an improved spatial indexing method based on R-Tree. The algorithm evaluation on real-world problems given at the end of the paper demonstrated its effectiveness.

?! non-spatial attributes

restricts clusters with respect to

Novelty

Should say more clearly what capabilities these extensions add

The authors propose three extensions to DBSCAN:

- Extension to identify distance metrics to support both spacial and non spacial objects
- Extension to identifying noisy objects by introducing a density factor
- Extension to identify adjacent clusters

What does this buy us?

The authors also present a fast indexing method in a spatial-temporal data warehouse application.

reference not clear.

The extensions are very novel and make DBSCAN perform better on spatial-temporal datasets

What is the message here? Is the evidence for this claim convincing or not

Technical quality

The author did a great job explaining the problems with DBSCAN and the reason why an extension may solve a set of problems namely distance between spatial and non spatial objects, adjacent clusters and noisy objects. The author extended the algorithm by adding 2 distance function instead of one. However, the author did not give a guideline on how the choice of these parameters can be performed. The author claims that ST-DBSCAN performance is similar to DBSCAN which is much faster than competing algorithms such as CLARANS and DBCLASD. However, there does not seem to be surprising as ST-DBSCAN still have to scan through all points and using the similar R-Tree indexing method as in DBSCAN. Detailed complexity comparison between DBSCAN and other methods has been given in the original paper.

?? reference not clear.

Significance and impact

explain the approach

objects whose values for their non-spatial attributes are too far away from the cluster

These extensions make DBSCAN perform better for spatial-temporal dataset and could be more robust with regards to noise. Short!

Clarity of writing of the paper

The paper is overall very well written. It is clear and concise. However, certain parts of the paper leave questions unanswered because they are short. For instance when introducing the extension to adjacent clusters. It is not clear how it is performed and how does one choose this value. be specific — e.g. list all!

Strongest points

1, Extend DBSCAN by assigning two separate threshold values for spatial and temporal distances. to accomplish what!

2, Enhance DBSCAN for noise detection in dataset of different densities.

3, Enhance DBSCAN for adjacent cluster separation by adding threshold during cluster expansion.

Weakest points

- The algorithm has more parameters than the original DBSCAN. As a result, the choice of these parameters will affect the performance. The authors did not provide hint or heuristics on how to choose these parameters

- The performance evaluation is not very strong. The authors claim that the modifications do not change the complexity of the algorithm but no algorithm analysis is performed

- The application part only shows the use of ST-DBSCAN and how it performs but it does not give an indication of how better it performs compared to other algorithms.

Educational Value

Starting from the beginning, the author clarified that ST-DBSCAN is an extension of DBSCAN and it solved two problems from DBSCAN. These two modifications do not have to be restrained only in spatial-temporal clustering. To accommodate DBSCAN into spatial-temporal dataset, the author came up with a very intuitive solution. ST-DBSCAN is definitely more like an enhanced algorithm rather than a completely new method. Its modifications deserve careful testing for ~~those~~ who are interested in using DBSCAN somewhere else.

Interesting. Although DBSCAN is explained in detail in original paper and many of its extensions, the author still replicated its basic definitions. For most graduate students who are familiar with DBSCAN, this is unnecessary. However, thanks to its review on these basic definitions, the author could clarify his modifications more clearly. Moreover, this paper pointed out a very practical problem coming with the large spatial-temporal real-world dataset: spatial-temporal indexing method. Generally speaking, this paper is good start point for students who are interested in using DBSCAN for spatial-temporal clustering and more generally for those who are interested in density-based clustering (DBSCAN) method. The work it has done is simple, diverse and encouraging. A good lesson from this paper is that there is no obsolete algorithm, just depending on how you use and improve it.

not true for the
assignment of borders pts.
I guess it is
not retrieved.

Specific comments

In section 4.1, does retrieve_Neighbors returns points that are both spatially and temporally close. What happens when one is spatially close but not temporally close?

In section 4.1, If a point belongs to 2 clusters, then ST-DBSCAN chooses the cluster discovered first. This is the same effect with original DBSCAN. But its reason is not clarified here. What if we made a medication on this by assigning the border points to the nearest cluster? DBSCAN is insensitive to the ordering of the points. Does shuffling the dataset affect the outcome and therefore the performance in ST-DBSCAN?

In section 4.2, DBSCAN is said to be as much as 3 times faster, and the current modification do not change the performance order. However, 2 algorithms with the same performance order say O(n log n) may have a constant multiplier in difference. That said, it is possible that ST-DBSCAN may be 2, 3 or even 10 slower that DBSCAN and still be of the same order.

Broad impact

The author already showed a successful application of ST-DBSCAN on real-world problem. However, this approach can generally be used in many other applications with similar spatial-temporal dataset structure. A good example would be another geological science problem with recordings at consecutive time (days, years).

By adjusting two neighborhood parameters, ST-DBSCAN can tune the algorithm to put different weight on spatial and temporal similarity. This is very intriguing if one of them is more important than the other in real application. Moreover, its modification can be applied wherever density-based method (DBSCAN) is used for clustering. Probably other density-based methods can benefit from these two modifications.

Numerical score

6: A very good paper, should be accepted.

Usefulness of homework2

Writing a review on the paper is often not required for graduate student. Through this unique task, we learned how to form our own opinions. When we read a paper, we tend to get stranded by its technical details. Most of time, we ignore the overall structure of the paper and its general idea. It is impossible to remember all the details for even one or two months, but it is possible to remember the generalized 'main' idea of the paper for enough long period. Reviewing a paper is based on your understanding of the paper, questions can be asked to help you understand deeper. Writing your own comment is an indirect way of generating a higher-level concept. However, limited by the time and effort, we cannot write the review in a very professional way unless we investigated the related research field. But it definitely helps for a better understanding and memorizing.