NUST Journal of Engineering Sciences

Manuscript #: NJES-D-16-196

02-Mar-2017

Title: Using Open Source Tools for Spatial-Temporal Querying and Knowledge Discovery from Moving Object Data

Dear editor,

Thank you for your useful comments and suggestions on the language and structure of our manuscript.

We have modified the manuscript accordingly, and detailed corrections are listed below point by point while the revised manuscript is attached separately.

Reviewer 1

Comment 1: The paper describes using a query based DBMS Secondo on an existing map traversing mouse trajectory dataset. From a scientific point of view, it is very hard to infer any valid contribution in advancing the state-of-the-art in this area. Both the dataset as well as the chosen query tool are already existing and a mere application of querying this dataset seems like a very superficial task.

Response 1:

Our idea is related to personalizing map content using user trajectories in Human-Computer Interaction domain. At the same time these trajectories are acting as a proxy to physical trajectories in the real world. See the end references (2, 7, 11 & 17) highlighted in the final manuscript. The arena of Web especially with interactive mapping capability is on the rise. At the same time, Human-Computer Interaction especially with mapping has a lot of potential to understand user's behaviors based on a user's map interactions.

The data generated for this research is being intelligently exploited with highly specialized research oriented tool called Secondo. Although the tool exists, it is never used for map personalization. As described in the manuscript, Secondo is intelligently exploited for identifying specific movement patterns and behavior and ultimately extract knowledge which can be used in personalized web maps, spatial recommender systems, event detection and crime monitoring tasks. With the real data obtained from test subjects, the potential of this work has been highlighted throughout this research paper.

Comment 2: There are some clarity issues in the paper also suggesting some flaws in the author's chosen evaluation schemes. For example, the dataset used in this study refers to the UCD campus area (section 3), whereas in figure 2, the overlaid trajectories are covering almost the whole Dublin city and its surroundings. There seems to be a disconnect here.

Response 2:

This is true that the chosen study area is UCD campus. However, one of the aims of getting these spatial tasks on a Web map is to see how users navigate on UCD campus. As per Figure 2, few users did not focus on their tasks and were more interested in areas outside of the campus. This also indicates their intentions. For more clarity, we have added all the spatial tasks user performed on Page 3 & 4 of the manuscript.

Comment 3:

Similarly, in the author's chosen simulated use case of business intelligence in section 4.2, they state that 115 -118 users entered Bowl and stadium regions. This also seems like very strange given that the total number of users in their dataset are 27. It seems like the authors have confused here users with the multiple trajectories of the same users.

Response 3:

It is correct that there were total 27 users. Each user was asked to perform 10 tasks. Each task corresponds to a single trajectory. Therefore a total of 270 trajectories were formulated. However, few users did not perform task correctly and the final acceptable number of trajectories were 258. For this particular case, as explained in the manuscript, 115-118 users (trajectories) crossed over UCD Bowl and Stadium regions. This means out of 258 trajectories, 115-118 crossed over Bowl and Stadium. It was further segregated in the spatial database that which trajectories belong to a particular given user. The text in the manuscript has been updated. Instead of saying 115 users, it has been updated to 115 user trajectories for more clarity.

Comment 4: The simulated use cases also seem like very superficial. The main underlying query for all is the same i.e. given a specific time stamp span count the number of user trajectories. Which seems like a very trivial task given the data. Besides this, the queries and some of the keywords highlighted in bold in the paper do not really convey anything meaningful to readers not very familiar with the chosen tool. It is rather instructive

to explain these better in the text. For the simulated scenarios, given that all users are asked to perform these tasks i.e. go to the chosen, location, the relations to simulate this as the GPS coordinated to infer e.g. crime scene is also very weakly founded.

Response 4: There are four cases which are simulated in our research. Sensitive area analysis, business intelligence, social behavior analysis and crime event simulation. For all cases, a specific spatio-temporal query has been designed. This is definitely more than counting specific trajectory. For example, in sensitive area analysis, the points of interest (POIs) are marked and the user's movement have been recorded on those POIs based on their interactions. This provides a clear picture of events happening on and around specific locations in the town. The queries parameters can be tuned to get the filtered results. Furthermore, the spatio-temporal database is a complex matrix as opposed to simple Database Management System (DBMS). The spatio-temporal database is backed by a complete spatial and temporal algebra. This is why it was important to show the queries and the spatio-temporal operators they use. The queries are highlighted in the text.

The reviewers outlined that the crime simulation is weakly formulated. As explained in the text that Human-Computer Interaction data is used as a proxy for crime scenario. The crime scene scenario is simulated here to show the power of trajectory analysis for moving objects. The detailed justification is highlighted in Section 4.4 in the manuscript.

In our future work, we aim to apply these techniques that will play a great role in aiding the law enforcement agencies in apprehending the perpetrators. Keeping in view that cellular data is generated in huge sizes, the primary focus of the future research will be to develop scalable techniques which can process data on a cluster of machines in a distributed fashion. A brief direction of future work was provided in the conclusions section of the manuscript as highlighted.

Comment 5: Given above comments, In my honest opinion, the authors could probably do better by focusing on a more realistic user study by collecting realistic moving data e.g. from cars GPS and mobiles phones of a big study group over a longer period. They can then focus on the realistic use cases and highlight the problems in this realistic data together with some performance comparison of different search tools. This could at least be considered a valid contribution in showing the existing problems of the KDBMS on realistic moving data. A good further direction could be investigating the use of clustering techniques in comparison with such direct queries (e.g. clustering the data based on spatial coordinates and time stamps could also already reveal much of what you get from these direct queries). Investigating and comparing some new clustering methods against this could be one direction. If not this, then collecting and analyzing a realistic heterogeneous moving data using these existing tools should be considered.

Response 5:

We agree with the reviewers comment. However, coming from a mapping and Human-Computer Interaction perspective, our study was already a realistic user study which was conducted on interactive maps. The real essence is that user interactions with maps can reveal a lot more about their physical activities such as their likes and dislikes. This is relatively a unique contribution we have highlighted in this research as opposed to physical trajectories such as generated from Global Positing System (GPS). A lot of research has already been conducted on physical trajectories. For example, car, human and birds tracking. Nevertheless, we have highlighted future directions where we wish to scale up the moving data in particular user trajectories on a large number of users. We also aim to use the advanced version of Secondo called Parallel Secondo which supports massive data processing in a distributed environment. This is being referred in the conclusions in the manuscript.

The manuscript has been resubmitted to your journal. We look forward to your positive response.

Sincerely

Ali Tahir (corresponding author)
Muhammad Faisal Zia, Ali Tahir, Gavin McArdle, Michela Bertolotto