Dr. Caetano Traina Jr.
Editor-in-Chief
Jornal of Information and Data Management

Dear Dr. Caetano Traina Jr.,

Please find attached a revised version of our manuscript “An Improved Base Algorithm for Online Discovery of Flock Patterns in Trajectories”. First and foremost we would like to apologize for the minor delay in submitting our work.

We thank the reviewers for analyzing our article. We also greatly appreciate your service and we have made our best effort addressing the comments. Please find our responses (in blue) below along with your comments.

We shall look forward to hearing from you at your earliest convenience.

Best regards,

Pedro Sena Tanaka, Marcos Rodrigues Vieira and Daniel S. Kaster

---

Responses to the comments

Reviewer A:

The authors propose an extension of the Basic Flock Evaluation (BFE) algorithm (Vieira et al, 2009) to detect flock pattern of moving objects trajectories. Using plane sweeping technique, binary signatures and inverted index, the proposed extension is faster than the original BFE algorithm. This paper is clear and well written. The authors cite related works and clearly emphasize the differences between the proposed method and the previous ones. This paper is an extension of the GeoInfo 2015 paper “Efficient algorithms to discover flock patterns in trajectories”. The authors improve the GeoInfo paper considering the reviewer comments and adding more results.

We thank you for your time.

Reviewer B:

The paper proposes some improvements over the first algorithm for on-line detection of flock patterns in trajectories of moving objects. The subject is relevant since the analysis of trajectories is nowadays a hot research topic. The text is well structured, but it has several typos and language problems that need to be improved before publication.

The paper is sometimes hard to read because of the excessive use of passive voice. Just a few examples: “From the analysis presented herein is possible to draw the conclusion” should be “From the analysis presented herein IT is possible to conclude”; “Based on Theorem 3.1 it was proposed..”, “should be “Based on Theorem 3.1 we proposed…”; “which require the entire historical dataset be loaded..”, should be “which require the entire historical dataset TO be loaded”; “depending on the hash functions chosen” should be “ depending on

the chosen hash functions"; "since we have to do perform these…" should be "since we have to perform…"

Thanks for pointing it out. We have made several corrections throughout the text, including those you cited.

In the last sentence of section 4.1 you mention figure 3(b) . Should it be Figure 3(c)? Because figure 3(b) does not have the MBRs.

You are right that Fig. 3(b) does not have MBRs. We removed this reference to the figure from the text to avoid confusing the reader.

The experimental section does not explain how the data are pre-processed in order to fit the requirements, i.e., to assure that in each time slice there is only one point for each trajectory. What happens if in a time slice there is no point of a trajectory?

We did not pre-process the datasets, except by some conversions between coordinate systems. Whenever a point of a trajectory is missing in a given timestamp, flock patterns will dismantle. This issue is not addressed in this work, however we included the following discussion regarding it (highlighted in BlueViolet in the manuscript):
Sec. 5: "All datasets are as provided by their owners, except by some conversions between coordinate systems. Some of the datasets have missing data, i.e., one entity appears in timestamp t, disappears in t+1 and reappears in t+2. Whenever this situation happens flock patterns will decompose."
Sec. 6: "Other possible future extension could be related to handling missing data. Currently, whenever an object helping to form a flock disappear for even *one* timestamp the flock is dismantled, therefore not reported. The base algorithm to merge candidate disks could be modified in order to keep disks from more previous timestamps instead of only one, and then check the new flocks against those disks."

Reviewer C:

This paper introduces an enhanced version of a flock patterns discovery algorithm, particularly, improving the online response time compared to baselines. The innovative idea is to use Plane Sweeping, binary signature and inverted indexes.

I found the paper interesting and quite well structured. The innovative contribution seems good for the target journal. The experimental results sound solid since authors show improvements compared with a baseline. The tests have been performed on 5 different datasets, including 4 real and one synthetic.

Therefore, I am overall positive with this paper and in favor of the acceptance.

However, I only have two main doubts:

1) The chosen baselines are previous versions of the algorithm of the same authors and incremental variants of the same adding features up to the final proposed version. I was wondering why authors do not compare with other state of the art flock patterns algorithms and only compare with themselves.

This was a good observation. For this next version of our work we included experiments that we conducted using an online version of the algorithm LCM_Flock which we developed (highlighted in red in the manuscript). The results presented in the article show that our method PSI outperforms LCM_Flock in most of the tested setups.

2) linked to the above observation is the second concern about the quality of the obtained flocks. It seems that the experimental evaluation is all about the time performance, do you have any hints on the quality of the results flocks?

Our goal was to improve the performance to obtain exact answers following the flock pattern definition employed. In an abstract view, the “quality” of a flock may be understood as the “suitableness” of the pattern for a given situation. Such problem is surely worth of further investigation for specific application fields, however it is out of the scope of this article.

Other minor points that I suggest authors should take into account is to motivate more the need of this online flock detection algorithm. Authors vaguely mention video surveillance but it could help to have more motivating examples on how having the online flock method would improve these applications.

The usefulness of group-based spatio-temporal patterns has been widely discussed in the literature. However, your suggestion to include examples that demand an online approach is interesting. Due to space limitations, we included the following discussion in the introduction:

“In order to fast act in response to observed activities in a monitored environment, we need algorithms that quickly detect flock patterns from applications that continuously consume large volumes of data from location-aware sensors. For example, a highway patrol can assemble an emergency task force to intercept and punish the participants of a street race after detecting a flock of vehicles moving in high speed.”