Data & Knowledge Engineering xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

Editorial

Efficient mining of platoon patterns in trajectory databases ☆

Yuxuan Li *, James Bailey, Lars Kulik

Department of Computing and Information Systems, The University of Melbourne, VIC 3010, Australia

ARTICLE INFO

Article history:
Received 13 July 2014
Received in revised form 27 January 2015
Accepted 17 February 2015
Available online xxxx

Keywords:
Spatial clustering
Trajectory database
Moving object cluster
Spatial pattern mining
Data mining

ABSTRACT

The widespread use of localization technologies produces increasing quantities of trajectory data. An important task in the analysis of trajectory data is the discovery of moving object clusters, i.e., moving objects that travel together for a period of time. Algorithms for the discovery of moving object clusters operate by applying constraints on the consecutiveness of timestamps. However, existing approaches either use a very strict timestamp constraint, which may result in the loss of interesting patterns, or a very relaxed timestamp constraint, which risks discovering noisy patterns. To address this challenge, we introduce a new type of moving object pattern called the *platoon pattern*.
We propose a novel algorithm to efficiently retrieve platoon patterns in large trajectory databases, using several pruning techniques. Our experiments on both real data and synthetic data evaluate the effectiveness and efficiency of our approach and demonstrate that our algorithm is able to achieve several orders of magnitude improvement in running time, compared to an existing method for retrieving moving object clusters.

1. Introduction

With the increasing availability of position-aware devices such as GPS receivers and mobile phones, it is now possible to collect and analyze large volumes of location databases that describe the trajectories of moving objects. Well known examples include taxi position data [1], animal movement data [2] and eye tracking data [3].

We address an important data mining challenge for trajectory data: discovering groups of spatial objects that move together for a certain period. We propose a new type of patterns, *platoon patterns*, that describes object clusters that stay together for time segments, each with some minimum consecutive duration of time. Fig. 1(a) shows an example of a platoon pattern. Wedding party vehicles o_2, o_3, o_4 and o_5 move together as a platoon at consecutive timestamps t_1, t_2, as well as consecutive timestamps t_4 and t_5.

The discovery of platoon patterns has a range of real-world applications. The identification of common routes among convoys may lead to more effective traffic control and the early discovery of truck platoons may assist traffic planning to avoid congestion. In eye tracking applications [3], the identification of common areas being viewed by a group of viewers can be used in advertising design and movie filming. In ecology, platoon patterns may provide a deeper understanding of animal migrations and in security may assist police to identify suspicious crowd movements.

1.1. Current techniques

Several recent approaches for discovering moving object clusters have been reported in the literature, but they are not directly applicable for mining platoon patterns. We use "moving object cluster" as a generic term in our paper.

☆ This research is supported under the Australian Research Council's Discovery Projects funding schema (project number DP110100757).
* Corresponding author.
E-mail addresses: yuxuan.li@unimelb.edu.au (Y. Li), baileyj@unimelb.edu.au (J. Bailey), lkulik@unimelb.edu.au (L. Kulik).

http://dx.doi.org/10.1016/j.datak.2015.02.001
0169-023X/

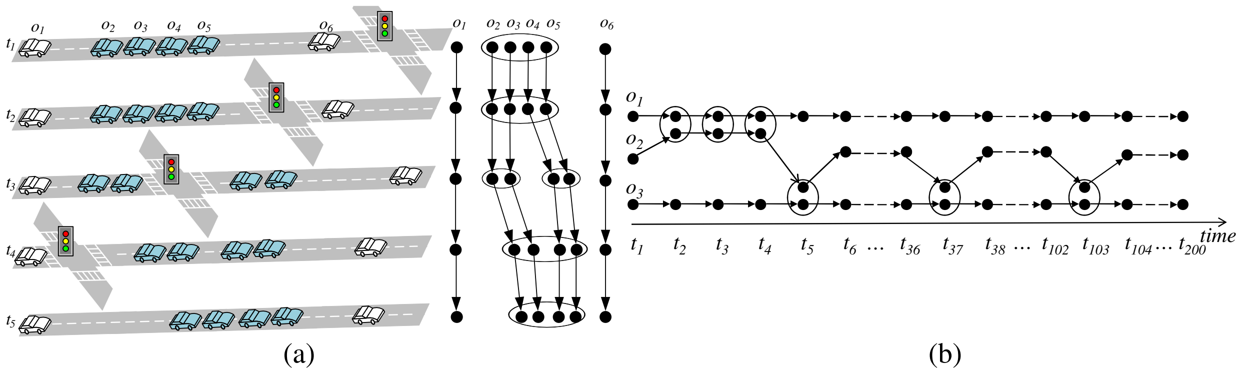

Fig. 1. (a) A platoon pattern example. Vehicles o_2, o_3, o_4 and o_5 travel together as a platoon at timestamps t_1, t_2, t_4 and t_5. Existing patterns such as flock and convoy fail to capture the co-location behavior of this pattern due to their strict constraint on timestamp consecutiveness. (b) The pattern that moving objects o_2 and o_3 travel together at isolated and non-consecutive timestamps t_5, t_{37} and t_{103} is a swarm when $k = 3$.

Previous work has proposed mining of moving objects that travel together for a minimum number of k consecutive timestamps such as flock [4–6] and convoy patterns [7,8]. These patterns commonly require that all timestamps are strictly (or *globally*) consecutive. As pointed out in [9], enforcing timestamp consecutiveness may lead to the loss of interesting patterns. For instance, in Fig. 1(a) with $k = 3$, there are no convoy or flock patterns, since the four objects split into two clusters at t_3 due to a red traffic light, before coming together again at t_4. In our opinion, these four objects are an interesting moving object cluster.

Secondly, swarm patterns [9], take an opposite approach and remove any consecutiveness constraint on timestamps. While this provides more latitude with regard to movement of clusters, it may also mine patterns that are overly "loose". Consider the example in Fig. 1(b) and assume we require at least $k = 3$ timestamps. Two vehicles (moving objects o_2 and o_3) might randomly encounter each other at some isolated and non-consecutive times (t_5, t_{37} and t_{103}), e.g. refilling fuel at the same petrol station, or stopping at the same car park. This does not imply that the drivers have a strong association with each other. Although one might avoid outputting this type of pattern by imposing a larger threshold value for the minimum number of timestamps (e.g. $k = 4$ timestamps), this would risk missing patterns with two objects that do move together over shorter consecutive durations (such as t_2, t_3 and t_4). Another alternative would be to first mine all swarm patterns and then filter the interesting ones. Such an approach is time consuming, however, since the postprocessing constraints are not pushed inside the swarm mining task. Indeed, our experiments will show that the number of swarm patterns can be extremely large but contain only a small proportion of platoon patterns.

1.2. Platoon patterns

Motivated by these issues, we propose a new definition for a moving object cluster called the *platoon pattern*, which allows the user to control the behavior of the consecutive time constraint to suit particular applications. Compared to the globally consecutive timestamp constraint of the convoy pattern [8], a platoon only requires that the timestamps are *locally* consecutive. Platoon patterns allow gap(s) in timestamps, but the consecutive time segments must have a minimum length (be locally consecutive). Given (1) a trajectory database with a timestamp-annotated history for moving objects, (2) a threshold for the minimum number of objects min_o that must appear in the platoon, (3) a threshold for the minimum number of timestamps min_t for which those objects travel together and (4) a threshold for the minimum number of consecutive timestamps min_c, a platoon pattern is an objectset and an associated timestamp sequence, denoted as $\{O : T\}$, such that $|O| \geq min_o$, $|T| \geq min_t$ and the timestamps in T are at least min_c locally consecutive. Intuitively, min_c denotes the minimum duration of a time segment in which objects stay together consecutively. In addition, platoon patterns do not rely on a particular clustering technique for deciding the spatial closeness of objects, which are instead modeled as preprocessing steps (c.f. Section 3 for our problem definition). The objects are required to be clustered.

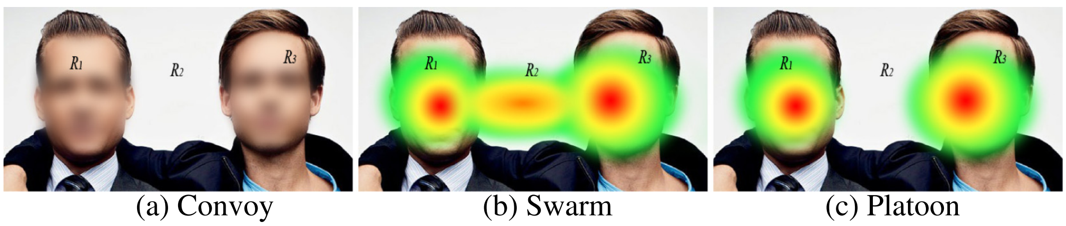

Fig. 2. Snapshot of a movie showing a dialog between two characters. Eye tracking data is represented as a heat map and eye movements of viewers focus on three dense regions: R_1, R_2 and R_3. (a) Convoy queries fail to identify interesting regions R_1 and R_3 due to the globally consecutive timestamp constraint. (b) Swarm queries erroneously consider R_2 to be interesting. (c) Platoon queries correctly identify R_1 and R_3 as interesting, using the locally consecutive timestamp constraint. Red color indicates high density of viewing, yellow indicates medium density and green indicates low density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Y. Li, et al., Efficient mining of platoon patterns in trajectory databases, Data Knowl. Eng. (2015), http://dx.doi.org/10.1016/j.datak.2015.02.001