STSM Topic: Understanding Flock Patterns in Moving Objects (COST-STSM-ECOST--IC0903-230610-000000)

Grantee: Rebecca U. Ong (Department of Computer Science, University of Pisa, Italy)

Host: Prof. Monica Wachowicz (Centre for Geo-Information, Wageningen University and Research, the Netherlands)

Introduction

Understanding behaviour of moving objects is important in the management of different types of areas, such as recreational parks, and cities. These behaviours can be inferred from patterns automatically discovered using some data mining algorithms. In our ongoing work, we focus on a specific type of pattern called flock patterns in the context of pedestrian movement.

Purpose

The purpose of the two-week visit is to extend the work previously started with Prof. Monica Wachowicz on the topic of understanding flock patterns. We have previously developed an algorithm that extracts moving flock patterns and proposed an approach for understanding such patterns. The approach was tested on the Dwingelderveld National Park (DNP) dataset from which some interesting interactions were found.

Our initial plan for the visit was to test the approach on another dataset involving another type of moving object. However, after reviewing the developed methodology and the derived results, we found the need to perform more experiments on the DNP dataset to improve the methodology for inferring meanings from the discovered flock patterns.

Summary of Previous Work and Results

We have developed an approach for understanding moving flock patterns in the context of pedestrian movement. We define a moving flock pattern as a group of moving objects that moves spatially close together over some time period. As opposed to conventional definitions of flock patterns, we emphasize that a moving flock should not solely remain in one location (i.e., moving flock is different from a stationary flock).

Our approach consists of three main steps: flock discovery, semantic annotation, and flock interpretation. The flock discovery step involves running a selected flock detection algorithm in order to find flock patterns inherent in the dataset and extract the corresponding properties of these flocks. In our experiments, we used our own moving flock detection algorithm for this step. Once the set of flock patterns have been discovered, the next step involves the semantic annotation of individual trajectories and of the discovered flock patterns (i.e., semantic annotation at individual and pattern level). Finally, we propose a combination of correlation computation and hierarchical clustering on the set of individual attributes, the set of flock attributes, and the set of the discovered flock themselves in the flock interpretation step. Performing this last step aids in the explanation of interactions among discovered flocks. An interesting interaction found in the DNP dataset experiment is the inference of the path-following interaction among flock members, which refers to the tendency of the visitors to flock together in certain paths provided in the park.

Activities Performed during the Visit

In order to determine the enhancements that should be performed on the proposed approach for understanding flock patterns, we started by discussing and summarizing the approach and results we have at hand. We produced a table that summarizes the properties of the dataset we worked with. These properties include general description of the type of area where movements occur, the number of trajectories (total and per day), the number of days and the days when the observed points were collected, the number of visitors (total and per day), the average speed among the pedestrians, the sampling temporal rate (i.e., the frequency of collecting point observations in seconds), and synchronization rate (i.e., the regular time interval at which points are sampled for further mining and interpretation). These information are important in understanding the nature of the given datasets and in determining the most suitable parameters for the flock discovery algorithm.

We also reviewed the semantic annotation and the flock interpretation steps performed on the DNP dataset. All trajectories were annotated with semantic attributes explicitly available from the dataset and with flock membership attributes derived using the moving flock discovery algorithm. Interesting relations among these attributes were found but these types of relations may exist even when no flocking occurs. Thus, we decided to focus on relations present among flocking individuals by focusing only on the individual attributes of trajectories belonging to some discovered flock.

After focusing only the trajectories that belong to some flock for the semantic annotation phase, we performed the pattern interpretation that initially consisted of the correlation computation and hierarchical clustering as sub-steps. The experiments were encouraging as we have found interesting relations among trajectories belonging to flocks despite of the fact that there were only 25 trajectories belonging to some flock out of 370 trajectories available in the dataset.

Clustering, however, is still quite limited in terms of aiding the user in understanding the nature and occurrences of the flocks. Though clustering is able to pinpoint which attributes are correlated, it is not able to pinpoint how or why these attributes are correlated. Thus, we decided to extend the pattern interpretation step further by performing classification of flock membership based on individual attributes. We found interesting results that were based on the decision trees used in the classification process. Since the dataset mostly contains 0 (i.e., the dataset is quite sparse), we used the WEKA implementation of a cost-sensitive variation of J48, which is an extension of the C4.5 algorithm. We used cost-sensitivity to give more weight to misclassifying 1 values over 0 values (i.e., it is important to correctly classify positive instances of an attribute over that of its negative instances).The generated decision trees were consistent with the dendograms derived in the hierarchical clustering step.

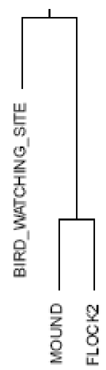
We also studied how association rules can aid in the pattern interpretation step. We used WEKA's implementation of the Apriori and the HotSpot algorithm. However, experiments with the DNP dataset gave association rules that focus on certain sets of attributes and were not very consistent with the clustering and classification results. Further experiments are needed to test the usefulness of association algorithms for the purpose of pattern interpretation.

Results

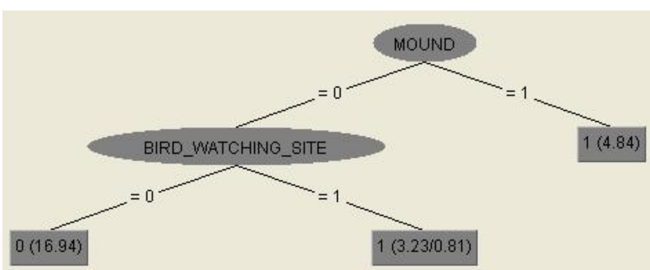
This section summarizes the clustering and classification results derived from the set of individual trajectories belonging to flocks.

For obtaining the hierarchical clustering results, the distance matrix should first be computed. As with our previous work, we chose to use different measures for computing distance matrices to accommodate different types of data. This also allows us to observe the consistency of clustering results obtained with different measures. The distance matrices are indirectly computed using correlation measures, which in our experiments include the symmetrical uncertainty coefficient (SUC), the correlation coefficient, and the absolute value of the correlation coefficient. We chose two versions of the correlation coefficient measure depending on whether negative correlations are important or not. In the case that negative correlations are important, the absolute value of the correlation score is generated. The correlation coefficient is a standard measure for computing correlations but it is only applicable for numerical data. Thus, we also chose to experiment with SUC, which is applicable to data containing nominal attributes as well.

The following is a sample relation found during the hierarchical clustering step using the standard correlation coefficient measure. This shows that belonging to flock2 is closely related to the member's preference of visiting mounds and bird watching sites.



In obtaining the classification results, we used a cost-sensitive implementation of J48 in WEKA. The following decision tree is obtained based on the dataset with focus on attributes affecting whether a flock is a member of flock2 or not. This tree further explains how flock2 is related to the pedestrians' preference of visiting mounds and of visiting bird watching sites. Note that these are relations also found in the hierarchical clustering step. The derived tree gives further details on how these attributes are related. It shows that members of flock2 are not interested in visiting mounds but are interested in visiting bird watching sites.



The experiments demonstrate that further application of a classification algorithm provides a good support for explaining how relations found using the hierarchical clustering step are correlated. Applying the described analysis steps to trajectories belonging to flock allow us to focus on relations that are present among visitors involved in flocking.

Future work

We plan to apply the approach on another dataset consisting of pedestrian movements in the city of Delft. We have already started with the execution of the flock discovery step and will proceed with the semantic annotation and the flock interpretation steps. We would also like to study the effect of using a taxonomy of semantic concepts for the flock pattern analysis step. We also plan to perform the classification sub-step for the pattern level attributes, as performed for the individual level attributes. Finally, we would like to provide a guideline for selecting parameters for the moving flock discovery algorithm.