



Research report

Determining shoal membership using affinity propagation

Vicenç Quera^{a,*}, Francesc S. Beltran^a, Inmar E. Givoni^b, Ruth Dolado^a^a Institute for Brain, Cognition and Behavior (IR3C), Adaptive Behavior and Interaction Research Group (GCAI), Department of Behavioral Science Methods, University of Barcelona, Campus Mundet, Passeig Vall d'Hebron 171, 08035 Barcelona, Spain^b Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada

HIGHLIGHTS

- We propose using the affinity propagation clustering for detecting multiple shoals.
- A soft temporal constraint is included in order to detect shoal fusion and fission.
- We explore how affinity propagation performs on agent-based simulated shoals.
- We compare affinity propagation clustering to human clustering of the same data.
- Affinity propagation is an appealing approach for detecting shoal dynamics.

ARTICLE INFO

Article history:

Received 31 August 2012

Received in revised form

16 November 2012

Accepted 20 November 2012

Available online 4 December 2012

Keywords:

Animal group membership

Shoal membership

Shoal fusion and fission

Affinity propagation clustering

Soft temporal constraint

Human clustering validation

ABSTRACT

We propose using the affinity propagation (AP) clustering algorithm for detecting multiple disjoint shoals, and we present an extension of AP, denoted by STAP, that can be applied to shoals that fusion and fission across time. STAP incorporates into AP a soft temporal constraint that takes cluster dynamics into account, encouraging partitions obtained at successive time steps to be consistent with each other. We explore how STAP performs under different settings of its parameters (strength of the temporal constraint, preferences, and distance metric) by applying the algorithm to simulated sequences of collective coordinated motion. We study the validity of STAP by comparing its results to partitioning of the same data obtained from human observers in a controlled experiment. We observe that, under specific circumstances, AP yields partitions that agree quite closely with the ones made by human observers. We conclude that using the STAP algorithm with appropriate parameter settings is an appealing approach for detecting shoal fusion–fission dynamics.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Determining group membership is a main concern of scientists studying animal collective behavior. Despite the growing body of literature devoted to the subject, criteria for determining shoal membership are still a matter of discussion. In particular, there is a need for methods that determine which fish are in a shoal and which are not. A common criterion to determine members of a shoal is based on body length based distances: animals within a criterion distance are considered to belong to the same shoal or group (in fish, four body lengths [1,2]; in dolphins, 100 m [3]). However, given the poor empirical evidence for validating that criterion, determining the limits of the shoal by eye is also common [4].

In this article, we follow the definition of a shoal given by [5], which permits quantification of this behavior, and is based on an earlier definition by [2]. According to [5, p. 614], a shoal is a group of individuals “presenting significant degree of cohesion, limited in a relatively small portion of space, a consequence of social interaction between these individuals”. However, defining a shoal, which is a particular kind of social interaction, as being a consequence of social interaction itself, is a circular definition; thus, “definitions based on the geometrical or statistical distributions of individuals (...) are useful in the study of aggregation behaviours because they are operationally objective and are independent of such behaviours” [6, p. 487]. Therefore, an appropriate way to define shoals is to provide an objective method to quantify cohesion, while characterizing them as consequences of social interaction seems unnecessary.

Our current work focuses on cohesion, or aggregation, and seeks to determine how to objectively characterize shoals. When synchronisation of displacements is measured, it is typically assumed that there exists a single group or shoal; however, when

* Corresponding author. Tel.: +34 933 125 088; fax: +34 934 021 359.
E-mail addresses: vquera@ub.edu (V. Quera), francesc.salvador@ub.edu (F.S. Beltran), inmar@psi.utoronto.ca (I.E. Givoni), ruth.dolado@ub.edu (R. Dolado).

several shoals are detected they can have different degrees of synchronisation and intra shoal aggregation. In such cases, it is more sensible to measure synchronisation and aggregation for each shoal separately. Therefore, a necessary first step in the analysis of collective motion is that of determining how many subgroups exist, and which individuals belong to each group.

When studying aggregation in groups of fish, three statistical methods have been proposed for determining which individuals belong to a shoal and which are outsiders. Such methods aim to detect a main cluster or shoal, and exclude outliers. These methods are based on momentary mean distances among individuals (Miller and Gerlai's 2008 method, MG08 for short, [7]), on trajectories of nearest neighbor distances during a session (Miller and Gerlai's 2011 method, MG11 for short, [8]), and on momentary nearest neighbor distances (Quera, Beltran and Dolado's 2011 method, QBD11 for short, [9]). When the MG08 is applied over several successive time units, the assignment of individuals to the main cluster tends to be unstable in certain cases: as individuals move, some of them may be considered members of the shoal at a certain time unit and excluded from it at the next, while for a human observer no substantial differences in membership can be distinguished. MG11 and QBD11, on the other hand, provide more stable results across time units, though MG11 can only be applied once all the data from a session have been gathered. None of these methods aims to segment a group into more than two disjoint subgroups or shoals. MG11 does segment a group into more than two subgroups, but the subgroups are not necessarily disjoint. A comparative summary of the three methods is shown in Table 1.

Animal group membership can be highly variable, as groups fuse and fission [10]. Regarding the analysis of animal aggregations, it is accepted that determining objectively the number of groups present and which individuals are members of each group is a very difficult task [6]; consequently, a quantitative method is needed to detect multiple groups. Moreover, in order to determine the validity of such a method, its outcome should be compared to estimates made by human observers, which would “ensure consistency and objectivity across time and among different observers and studies” [6, p. 483].

In this article we propose using the affinity propagation (AP) clustering algorithm [11–14] for detecting multiple disjoint shoals. We explore how AP performs under a variety of circumstances, and we compare its output to partitioning results obtained from human observers asked to perform the same task in a controlled experiment. In addition, we propose a variant of the algorithm (denoted by STAP) that takes into account cluster dynamics so that the results are temporally consistent: that is, STAP can yield partitions at time t that are related to those obtained at time $t - 1$. This is in contrast to performing standard AP on each time step, yielding independent and not necessarily consistent, results. Unlike MG11, AP can be applied momentarily and not post hoc.

2. Affinity propagation

Cluster analysis, or clustering, is the task of partitioning data into disjoint subsets or groups. When groups also need to be associated with a label, the task is known as classification. There exist a wide variety of clustering and classification techniques, such as hierarchical cluster analysis, nearest neighbor classification, and techniques based on swarm intelligence algorithms (e.g., [15,16]). An iterative, k th nearest neighbor, hierarchical cluster analysis for detecting shoals was been proposed in [6]. However, results from hierarchical cluster analysis require a cutoff criteria in order to determine the actual clusters. k -Means clustering has been proposed for detecting clusters in animal social networks [17], but, similarly to many other methods, it requires specifying the number of clusters a priori. In comparison, affinity propagation partitions data into clusters without requiring a cutoff criteria or knowing the number of clusters to find. However, like other data clustering methods, AP requires that some parameters must be specified by the users; in this article we explore systematically how those parameters affect AP performance. Affinity propagation was developed by Delbert Dueck and Brendan J. Frey and has become a popular method in many research fields such as machine learning, bioinformatics, social networks analysis, computer vision, and neuroscience [14, pp. 6–7]. In this article we present an overview of AP; for details, we refer the reader to [11–14,18].

The affinity propagation algorithm takes as input a matrix of pairwise similarities for n points (s_{ij} , $i = 1 \dots n$, $j = 1 \dots n$). In our case, similarities are defined between individual fish represented by their spatial coordinates. The algorithm partitions the points into clusters so that each cluster contains exactly one prototypical data point, known as the *exemplar*, to which the other points in the cluster are associated. The similarity s_{ij} is a measure of how suitable point j is to serve as the exemplar for point i ; the similarity between a point and itself, s_{jj} , is known as the *preference*, and is a measure of the a priori suitability of point j to serve as an exemplar.

The algorithm operates by an iterative message-passing mechanism; each data point can be thought of as a node in a network. Nodes send and receive messages to and from other nodes along the edges of the network. Each node i transmits its *responsibility* (ρ_{ij}) for recognizing other nodes j as candidate exemplars, and its *availability* (α_{ij}) to be a candidate exemplar for other nodes and for itself (*self-availability*, α_{jj}). Respectively, these messages reflect the accumulated evidence “for how well-suited point j is to serve as the exemplar for point i , taking into account other potential exemplars for point i ” and “for how appropriate it would be for point i to choose point j as its exemplar, taking into account the support from other points that point j should be an exemplar” [13, p. 972]. Message passing is an iterative process in which responsibilities and availabilities are updated as functions of similarities and previous responsibilities and availabilities. After a certain number of iterations the process typically converges, and the messages no longer change between iterations. At that point, the messages can be used to compute the subset of points that are the exemplars, as well as

Table 1
Comparative summary of existing methods for detecting one single shoal vs. outliers. MG08, [7]; MG11, [8]; QBD11, [9].

	Methods		
	MG08	MG11	QBD11
Measure	Mean inter-individual distances	k th nearest neighbor distances	k th nearest neighbor distances
Analysis	Momentary	Post hoc	Momentary
Segmentation criterion	Arbitrary (square root of global mean distance)	Distribution-based and arbitrary (percentile)	Arbitrary (square root of global mean distance, and percentile)
Result	Main cluster/outliers	Main clusters/“ k -type excursions”	Main cluster/outliers
Size of main cluster	Liberal	More conservative as total group size increases	Conservative
Stability of segmentations	Fairly unstable	Fairly stable	Moderately stable
Computational cost	Low	High	Low

Download English Version:

<https://daneshyari.com/en/article/6259180>

Download Persian Version:

<https://daneshyari.com/article/6259180>

[Daneshyari.com](https://daneshyari.com)