Spatio-Temporal Modeling of the Topology of Swarm Behavior with Persistence Landscapes

Padraig Corcoran
School of Computer Science & Informatics
Cardiff University
Wales UK
corcoranp@cardiff.ac.uk

Christopher B. Jones
School of Computer Science & Informatics
Cardiff University
Wales UK
c.b.jones@cs.cardiff.ac.uk

ABSTRACT

We propose a method for modeling the topology of swarm behavior in a manner which facilitates the application of machine learning techniques such as clustering. This is achieved by modeling the persistence of topological features, such as connected components and holes, of the swarm with respect to time using *zig-zag persistent homology*. The output of this model is subsequently transformed into a representation known as a *persistence landscape*. This representation forms a vector space and therefore facilitates the application of machine learning techniques. The proposed model is validated using a real data set corresponding to a swarm of 300 fish. We demonstrate that it may be used to perform clustering of swarm behavior with respect to topological features.

CCS Concepts

•Mathematics of computing → Algebraic topology; •Information systems → Geographic information systems;

Keywords

Spatio-Temporal; Topology; Swarm; Persistence Landscape

1. INTRODUCTION

A swarm may be defined as a large set of agents moving in close spatial proximity to each other. The agents in question may correspond to animals, such as fish, birds or humans, robots or other environmental phenomena. Swarms can accomplish many complex tasks such as building complex structures [10]. As well as being able to accomplish such tasks, swarms can do so in a manner which is robust, scalable and flexible. For these reasons, the development of accurate models of swarm behavior has long been of interest to the research community.

Most existing models of swarm behavior model metric properties such as agent orientation. If one assumes the swarm to be samples lying on an unknown manifold, where a manifold is a space which locally looks like an open subset of $\mathbb{R}^n$, one can infer this manifold using methods such as kernel density estimation. One can then model the topology of this manifold and in turn that of the swarm. Computing the Betti numbers of a manifold is a commonly used approach to model its topology. Intuitively the k^{th} Betti number equals the number of k-dimensional holes in the manifold with the 0^{th} Betti number equaling the number of path-connected components [5]. Toward illustrating this approach to modeling swarm behavior consider Figure 1(a) which displays a swarm at a given time where the agents in question correspond to 300 Golden Shiners which are a type of fish [9]. The fish are swimming in a shallow pool and therefore their position may be accurately specified using x and y Cartesian coordinates. This manifold appears to have a single path-connected component which contains a single one dimensional hole. That is, the corresponding 0^{th} and 1^{st} Betti numbers are both equal to 1. Topological features which persist for a longer period of time are of greater significance than those which persist for a shorter period. Therefore when modeling the topology of a swarm it is important to model the persistence of Betti numbers with respect to time.

In this paper we propose a novel method for modeling swarm behavior. This method first computes the corresponding Betti numbers and their persistence with respect to time using *zig-zag persistent homology*. This information is subsequently transformed into a representation known as a *persistence landscape*. The latter representation forms a vector space that facilitates the application of vector based tools from machine learning. In this paper we specifically consider the task of clustering swarm behavior.

The layout of this paper is as follows. In section 2 we review related works on modeling swarm behavior. In section 3 we describe the model of swarm behavior proposed in this paper. In section 4 we demonstrate that it may be used to cluster swarm behavior. Finally in section 5 we draw conclusions.

2. RELATED WORKS

There exists a vast array of works which attempt to model the behavior of moving agents [7]. However in this section we only consider these works where the agents in question correspond to a swarm; that is, a large set of agents moving in close spatial proximity to each other. Two commonly used characteristics of swarm behavior are polarization and rotation order which provide measures of the alignment and

SIGSPATIAL'16, October 31-November 03, 2016, Burlingame, CA, USA

DOI: http://dx.doi.org/10.1145/2996913.2996949

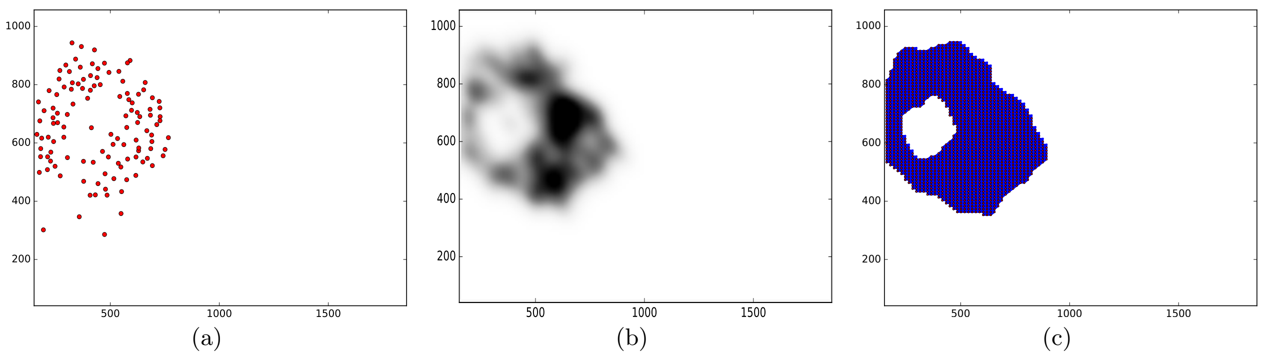

Figure 1: A swarm of 300 Golden Shiner fish, its KDE and simplicial complex are displayed.

angular momentum of the agents respectively [9]. These characteristics only model swarm behavior at a given instance in time. In order to model the temporal behaviour of a swarm [9] proposed to model how these characteristics vary as a function of time. [6] proposed a method for classifying swarm behavior as *flock*, *torus* or *disordered*. These three types of swarm behavior are considered to be the most common types exhibited by swarms. In a flock behavior the agents form a compact cluster with all agents moving in a common direction. In a torus behavior the agents move in a circular motion around a region in space. While in a disordered behavior the agents behave randomly.

[8] recently proposed a model of swarm behavior which computes the corresponding Betti numbers independently at each time step. The authors subsequently visualized the Betti numbers as a function of time. This approach was found to reveal characteristics of swarm behavior not captured by models which do not explicitly model topological features. A limitation of this method is that it does not compute the persistence with respect to time of the topological features corresponding to Betti numbers.

3. MODEL OF SWARM BEHAVIOR

The proposed method for modeling the topology of swarm behavior contains the following computational steps. Firstly we infer the manifold on which the agents lie. Next we use a methodology called *zig-zag persistence homology* for computing the persistence of topological features with respect to time [3]. We subsequently transform this information into a representation known as a *persistence landscape* which facilitates the application of tools from statistics and machine learning [2]. In order to achieve robustness with respect to noise we draw from recent advances in robust topological inference.

In this section we describe each of these steps in detail. Specifically in section 3.1 we describe how the manifold is inferred. In section 3.2 we briefly review background material on homology theory. In sections 3.3 and 3.4 we describe zig-zag persistence homology and persistence landscape respectively.

3.1 Inferring the Manifold

In this section we describe how the manifold on which the agents lie may be inferred using the combinatorial representation of a *simplicial complex* [5]. Briefly a simplicial complex $\mathcal{K}$ is a family of finite subsets of a universal set such that for each σ in $\mathcal{K}$ all subsets of σ are also in $\mathcal{K}$. A set σ in $\mathcal{K}$ is called a k-simplex if $|\sigma| = k + 1$ where $|.|$ represents the cardinality of the set in question. The faces of a simplex σ correspond to all simplices τ such that $\tau \subset \sigma$. This representation subsequently provides a platform for computing topological features of the manifold.

As stated in the introduction, we assume the agents to be samples drawn from an unknown manifold. When attempting to infer this manifold it is important to do so in a manner which is robust to noise. In this context noise corresponds to a minority of agents whose behavior differs from that of the majority and as a consequence their presence introduces topological artifacts.

Using a Gaussian kernel with bandwidth h we compute a *Kernel Density Estimation* (KDE) of the agent locations. Let f_h denote the kernel density estimator. The upper-level set $f_h^{-1}[a, \infty)$ of this estimator can be considered a robust estimate of the manifold provided the threshold a is appropriately chosen. This is justified by the fact that agents corresponding to noise will lie in regions of the space with low density and therefore will not be represented in the inferred manifold. We subsequently represent this super-level set using a simplicial complex $\mathcal{K}$ as follows. We first estimate the density for a grid of points over $\mathbb{R}^2$. For each point we include a corresponding 0-simplex in $\mathcal{K}$ if the density at that point is greater than a. For each pair of 0-simplices which are horizontally, vertically or main diagonally adjacent we include a corresponding 1-simplex in $\mathcal{K}$. For each triple of 0-simplices where all subsets of pairs are horizontally, vertically or main diagonally adjacent we include a corresponding 2-simplex.

To illustrate this construction in the context of a swarm consider again the swarm displayed in Figure 1(a). The KDE corresponding to this swarm is displayed in Figure 1(b). The simplicial complex representation of the upper-level set of this KDE is displayed in Figure 1(c). It is evident that this simplicial complex contains a single path-connected component and a single one dimensional hole. As such it accurately models the topology of the original swarm in a robust manner.

3.2 Homology Theory

In this section we formally define Betti numbers and describe how the Betti numbers of a manifold may be inferred from a simplicial complex representation of that manifold. Let $\mathcal{K}$ be a simplicial complex. A k-chain on $\mathcal{K}$ is defined by

Equation 1 where each $\sigma_i \in \mathcal{K}$ is a k-simplex and each λ_i is an element from a given field. The set of k-chains forms a group known as the chain group $C_k(\mathcal{K})$. The boundary map is a map from a k-simplex to a sum of its $(k-1)$-simplex faces and is defined in Equation 2. Here $[v_1, \ldots, \hat{v_i}, \ldots, v_{k+1}]$ is the $(k-1)$-simplex obtained by removing the 0-simplex v_i from the k-simplex $\sigma = [v_1, \ldots, v_{k+1}]$. This map is distributive and therefore extends to the chain groups to give the sequence of chain groups defined in Equation 3. Such a sequence of groups is known as a chain complex C_*.

$$c = \sum \lambda_i \sigma_i \tag{1}$$

$$\partial_k \sigma = \sum_{i=1}^{k+1} [v_1, \ldots, \hat{v_i}, \ldots, v_{k+1}] \tag{2}$$

$$\ldots \rightarrow C_{k+1}(\mathcal{K}) \xrightarrow{\partial_{k+1}} C_k(\mathcal{K}) \xrightarrow{\partial_k} C_{k-1}(\mathcal{K}) \rightarrow \ldots \tag{3}$$

A k-chain $c \in C_k(\mathcal{K})$ is a k-boundary if there exists a $d \in C_{k+1}(\mathcal{K})$ such that $c = \partial d$. It is a k-cycle if $\partial c = 0$. The set of all k-boundaries and k-cycles form groups denoted by $B_k(\mathcal{K})$ and $Z_k(\mathcal{K})$ respectively. Both these groups are subgroups of $C_k(\mathcal{K})$. By virtue of the fact that $\partial_{k+1}\partial_k = 0$ it follows that $B_k(\mathcal{K}) \subseteq Z_k(\mathcal{K})$. The quotient group $H_k(\mathcal{K}) = Z_k(\mathcal{K})/B_k(\mathcal{K})$ is called the k-homology group of $\mathcal{K}$ and its rank is called the k^{th} Betti number. As discussed in the introduction to this paper, intuitively the k^{th} Betti number equals the number of k-dimensional holes in $\mathcal{K}$.

3.3 Zig-Zag Persistent Homology

For the purposes of this work we are not only interested in computing the Betti numbers but also the persistence with respect to time of the corresponding topological features. This is achieved using a methodology called *zig-zag persistence homology*. Consider the sequence of simplicial complexes $\mathcal{K}$ defined in Equation 4 which is called a *zig-zag diagram* [3]. Here each map $\leftrightarrow$ represents either a forward inclusion map $\rightarrow$ or a backward inclusion map $\leftarrow$. A forward inclusion map corresponds to the addition of simplices while a backward inclusion map corresponds to the removal of simplices. A zig-zag diagram induces a corresponding sequence of homology groups defined in Equation 5 which is called a *zig-zag module*.

$$\mathcal{K} : \mathcal{K}_1 \leftrightarrow \mathcal{K}_2 \leftrightarrow \cdots \leftrightarrow \mathcal{K}_n \tag{4}$$

$$H_k(\mathcal{K}) : H_k(\mathcal{K}_1) \leftrightarrow H_k(\mathcal{K}_2) \leftrightarrow \cdots \leftrightarrow H_k(\mathcal{K}_n) \tag{5}$$

This zig-zag module can be decomposed into a direct sum of interval modules. These modules have a form defined by Equation 6 where $I_i = k$ for $b \leq i \leq d$ and otherwise $I_i = 0$ and every $k \leftrightarrow k$ is the identity map.

$$\mathcal{I}_{[b,d]} : I_1 \leftrightarrow I_2 \leftrightarrow \cdots \leftrightarrow I_n \tag{6}$$

The *zig-zag persistent homology* of $\mathcal{K}$ for dimension p, which is denoted $\text{Pers}_p(\mathcal{K})$ and defined in Equation 7, is the multiset of intervals $[b, d]$ corresponding to the set of summands $\mathcal{I}_{[b,d]}$ of $H_p(\mathcal{K})$. Each interval $[b, d]$ corresponds to the persistence of a topological feature in $\mathcal{K}$ which exists from b to d inclusive. The total persistence $\text{Pers}(\mathcal{K})$ of the zigzag diagram $\mathcal{K}$ is the collection of $\text{Pers}_p(\mathcal{K})$ for each dimension p [4].

$$\text{Pers}_p(\mathcal{K}) = \{[b_j, d_j] | j \in J\} \tag{7}$$

When attempting to analyze the characteristics of swarm behavior one typically knows the locations of the agents in question at a sequence of discrete times. The corresponding sequence of simplicial complexes might not have the property that between each consecutive pair of simplicial complexes a forward or backward inclusion map exists. This is because the transformation between such a pair may include both the addition and removal of different simplices. Therefore one cannot directly compute the zig-zag persistent homology for such a sequence. To overcome this challenge for each consecutive pair of simplicial complexes we compute an intermediate simplicial complex corresponding to the union of the simplicial complexes in question. This gives the zig-zag diagram $\mathcal{K}$ of Equation 8 for which the zig-zag persistent homology can be computed. In this work we used the method of [4] to perform this computation.

$$\mathcal{K} : \mathcal{K}_1 \rightarrow (\mathcal{K}_1 \cup \mathcal{K}_2) \leftarrow \mathcal{K}_2 \rightarrow (\mathcal{K}_2 \cup \mathcal{K}_3) \leftarrow \mathcal{K}_3 \ldots \mathcal{K}_n \tag{8}$$

3.4 Persistence Landscape

Given the total persistence $\text{Pers}(\mathcal{K})$, corresponding to the collection of $\text{Pers}_p(\mathcal{K})$, we wish to transform this into a representation which facilitates the application of tools from statistics and machine learning. The most commonly used approach toward achieving this goal is to use a representation known as a *persistence diagram*. This representation is obtained by mapping the intervals of a given $\text{Pers}_p(\mathcal{K})$ to their endpoints [2]. It may be equipped with a metric, such as the bottleneck or Wasserstein metrics, to form a metric space [1]. However a metric space does not allow one to perform vector based machine learning tasks, such as computing averages and distances between averages. To overcome this limitation we use a representation known as a *persistence landscape* which forms a normed vector space [2].

4. EXPERIMENTS

This section presents experiments performed to evaluate the accuracy and usefulness of the proposed model of swarm behavior. It is structured as follows. In section 4.1 we describe the data used within the experiments. In section 4.2 we describe how the proposed model may be used to perform clustering of swarm behavior.

4.1 Data

The data used in our experiments corresponds to a swarm of 300 Golden Shiners. This data was described briefly in the introduction to this paper and was obtained from [9]. The fish in question were swimming in a small shallow tank (2.1 m $\times$ 1.2 m, water depth 5 cm). They were filmed for 56 minutes at 30 Hz. A vision based algorithm was used to track the position and orientation of individual fish. For the purposes of this paper we down-sampled the frame rate to 3 Hz. In all experiments swarm behavior is modeled over a temporal window of length equal to 10 time steps. An example of the swarm in question at a given time step is displayed in Figure 1(a).

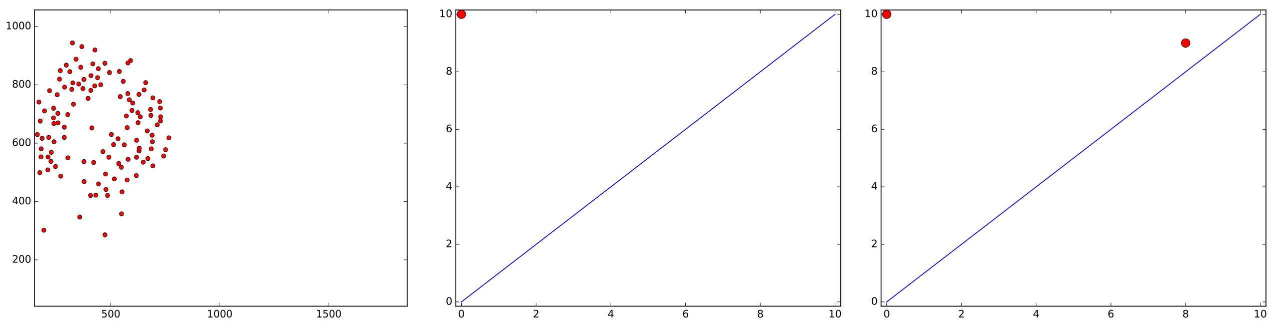

Figure 2: The swarm, persistence diagram $\mathbf{Pers}_0(\mathcal{K})$ and persistence diagram $\mathbf{Pers}_1(\mathcal{K})$ are displayed.

4.2 Clustering

In this section we describe an experiment performed to evaluate if the proposed model may be used to discover clusters of swarm behavior with distinct topological features. In order to perform clustering the K-medoids data clustering algorithm was used. The individual data points to be clustered correspond to the persistence landscape representation of swarm behavior in a given temporal window. K-medoids is an iterative clustering method which determines K clusters by assigning each cluster a corresponding cluster center represented by a data point such that the distance between each data point and its corresponding cluster center is minimized. We clustered swarm behavior using K-medoids for K=3. Figure 2 illustrates one of the clusters obtained. Recall that swarm behavior is modeled over a temporal window of length equal to 10 time steps. The left image of Figure 2 displays the swarm at the midpoint of this window for the cluster in question. The center and right images of Figure 2 display the corresponding persistence diagrams of $\text{Pers}_0(\mathcal{K})$ and $\text{Pers}_1(\mathcal{K})$ respectively. A persistence diagram is obtained by mapping the intervals in question to their endpoints. A point exists at coordinates (0,10) in the persistence diagram of $\text{Pers}_0(\mathcal{K})$ and this indicates that a path-connected component appeared at time 0 and disappeared at time 10. Likewise a point exists at coordinates (0,10) in the persistence diagram of $\text{Pers}_1(\mathcal{K})$ and this indicates that a one dimensional hole appeared at time 0 and disappeared at time 10. Points which lie closer to the diagonal of a persistence diagram, represented by blues lines in our figures, do not persist for a significant period and therefore are considered topological noise. This is the case for the point (8,9) in the persistence diagram of $\text{Pers}_1(\mathcal{K})$. This examination of the cluster of Figure 2 reveals that it corresponds to the swarm behavior *torus*. A similar examination of the two remaining clusters returned reveals that they correspond to the behaviors *flock* and *disordered*. As such, our model is able to discover these three behaviors in an unsupervised manner.

5. CONCLUSIONS

To the authors' knowledge the work presented in this paper represents the first attempt to model the topology of swarm behavior in a manner which facilitates the application of machine learning techniques. The experiment results presented demonstrate that the proposed model may be used to perform the machine learning task of clustering swarm behavior with respect to topological features. The model proposed has many potential applications beyond modeling swarm behavior. For example it could potentially be used to model topological events in sensor networks [11].

6. REFERENCES

[1] H. Adams, S. Chepushtanova, T. Emerson, E. Hanson, M. Kirby, F. Motta, R. Neville, C. Peterson, P. Shipman, and L. Ziegelmeier. Persistence images: An alternative persistent homology representation. *arXiv preprint arXiv:1507.06217*, 2015.

[2] P. Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

[3] G. Carlsson and V. De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.

[4] G. Carlsson, V. De Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256, 2009.

[5] H. Edelsbrunner and J. Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[6] M. Berger, L. M. Seversky, and D. S. Brown. Classifying Swarm Behavior via Compressive Subspace Learning. In *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.

[7] P. Corcoran, P. Mooney, and G. Huang. Unsupervised Trajectory Compression. In *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.

[8] C. M. Topaz, L. Ziegelmeier, and T. Halverson. Topological data analysis of biological aggregation models. *PloS one*, 10(5):e0126383, 2015.

[9] K. Tunstrøm, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin. Collective states, multistability and transitional behavior in schooling fish. *PLoS Comput Biol*, 9(2):e1002915, 2013.

[10] J. Werfel, K. Petersen, and R. Nagpal. Designing collective behavior in a termite-inspired robot construction team. *Science*, 343(6172):754–758, 2014.

[11] M. Worboys and M. Duckham. Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science*, 20(10):1087–1108, 2006.