Data has shape, one learns one day. Moreover, (some) data has beautiful shape, as one can infer, for example, by reading B. Mazur’s article in the Bulletin of the AMS (45, Number 2, pp. 185-228), where at the top of page 193 Mazur mentions the “What beautiful data!” exclamation of S. Holmes, a mathematician and statistician, when Mazur presented her some number-theoretical numerical data.

The natural question, then, is how one can see this (global) shape of data. As R. Ghrist reminds us, in another Bulletin of the AMS (45, Number 1, pp. 61-75) article, the answer is also natural: we should do it the same way we visualize a 2-dimensional object by its 1-dimensional level curves or projections. That is, in general we visualize a higher dimensional object from some of its low dimensional representations. Thus, we start visualizing the shape of a discrete data set by joining discrete points to make a global topological object.

Statisticians have, of course, for a long time extracted information from data sets that most of the times are huge and shapeless and, at least initially, very noisy and incomplete. The topological analysis of data sets, an area of interaction between applied (statistics) and pure (algebraic topology) mathematics, attempts to provide a rigorous treatment of the notion of data shape that is useful and practical for data analysis, with all its natural importance in several applications. This application would have been important by itself, but there is even more to it, because the interaction is bidirectional, and computational algebraic topology thrives nowadays.

How one can introduce shapes in clusters of data sets is a truly ingenious process, involving finding a right scale and organizing the data in hierarchical clusters. There are many fine points here: playing with forgetting or remembering patterns in or between clusters to obtain simplified collections or barcodes that
represent what *persists* after these manipulations. Persistence theory, a subject that just recently emerged and has reached a certain degree of maturity, attempts to produce these barcodes in all dimensions, providing a new tool for the analysis and visualization of data sets.

On the topological side, what is produced is a filtration

\[ X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n \]

of topological spaces. An algebraic topologist would then try to obtain information about these spaces by using some known invariants. In this case, the invariants that are easier to work with are the homology groups of the spaces and the morphisms induced between them by the inclusions of the filtration. Thus we are led to look at sequences of abelian groups, or vector spaces, or modules in general, of the form

\[ H_*(X_1) \to H_*(X_2) \to \cdots \to H_*(X_n), \]

a so-called *persistence module*. The problem of discerning which topological properties persist along the given filtration is translated into the algebraic problem of the compatibility of the sequence of modules, where compatibility amounts to showing that each allowable composition of morphisms can be represented by a diagonal matrix. Finding bases that solve this compatibility problem in general is a non-trivial matter, and here a big chunk of the theory of quiver representations is required.

The book under review, one of the few existing monographs in the subject of persistence theory, focuses mainly on the algebraic side of the theory, developing the quiver representation theory that is required and the theory of persistent modules. With these algebraic foundations, the second part of the book is devoted to topological applications. The third and last part of the book focuses on new trends in topological data analysis and possible generalizations on the algebraic side. The book is well organized, and although it is a monograph, it can be read by a graduate student who is well-grounded on representation theory or algebraic topology and with an interest in applications.

Felipe Zaldivar is Professor of Mathematics at the Universidad Autonoma Metropolitana-I, in Mexico City. His e-mail address is fz@xanum.uam.mx.

Tags:  
Applied Topology  
Computational Topology  
Data Analysis

Log in to post comments