

Scalable vector analytics: A story of twists and turns

Themis Palpanas*

* Université Paris Cite, IUF, équipe Dino

Résumé

Similarity search in high-dimensional data spaces was a relevant and challenging data management problem in the early 1970s, when the first solutions to this problem were proposed. Today, fifty years later, we can safely say that the exact same problem is more relevant (from Time Series Management Systems to Vector Databases) and challenging than ever. Very large amounts of high-dimensional data are now omnipresent (ranging from traditional multi-dimensional data to time series and deep embeddings), and the performance requirements (i.e., response-time and accuracy) of a variety of applications that need to process and analyze these data have become very stringent and demanding. In these past fifty years, high-dimensional similarity search has been studied in its many flavors. Similarity search algorithms for exact and approximate, one-off and progressive query answering. Approximate algorithms with and without (deterministic or probabilistic) quality guarantees. Solutions for on-disk and in-memory data, static and streaming data. Approaches based on multidimensional space-partitioning and metric trees, random projections and locality-sensitive hashing (LSH), product quantization (PQ) and inverted files, k-nearest neighbor graphs and optimized linear scans. Surprisingly, the work on data-series (or time-series) similarity search has recently been shown to achieve the state-of-the-art performance for several variations of the problem, on both time-series and general high-dimensional vector data. In this talk, we will touch upon the different aspects of this interesting story, present some of the state-of-the-art solutions, and discuss open research directions.

Biographie

Themis Palpanas is an elected Senior Member of the French University Institute (IUF), a distinction that recognizes excellence across all academic disciplines, and Distinguished Professor of computer science at the University Paris Cite (France), where he is director of the Data Intelligence Institute of Paris (diiP), and director of the data management group, diNo. He received the BS degree from the National Technical University of Athens, Greece, and the MSc and PhD degrees from the University of Toronto, Canada. He has previously held positions at the University of California at Riverside, University of Trento, and at IBM T.J. Watson Research Center, and visited Microsoft Research, and the IBM Almaden Research Center. His interests include problems related to data science (big data analytics and machine learning applications). He is the author of 15 patents. He is the recipient of 3 Best Paper awards, and

Scalable vector analytics: A story of twists and turns

the IBM Shared University Research (SUR) Award. His service includes the VLDB Endowment Board of Trustees (2018-2023), Editor-in-Chief for PVLDB Journal (2024-2025) and BDR Journal (2016-2021), PC Chair for IEEE BigData 2023 and ICDE 2023 Industry and Applications Track, General Chair for VLDB 2013, Associate Editor for the TKDE Journal (2014-2020), and Research PC Vice Chair for ICDE 2020.