Letter from the Special Issue Editor

High-dimensional vector similarity search is a key operation in several data analysis pipelines across many diverse domains and applications. It represents an important and challenging problem in modern data management, with a strong interest from users and industrial players alike. This interest has been steadily growing, especially thanks to the proliferation of a particular kind of vectors, the deep embedding representations (of all sort of different data objects), and the subsequent need for analyzing very large collections of such embeddings.

In our previous special issue on the topic in this Bulletin (High-Dimensional Similarity Search: from Time Series Management Systems to Vector Databases, IEEE Data Eng. Bull. 46(3), 2023), we explored various aspects of the high-dimensional similarity search problem. In particular, we examined the different kinds of vectors, ranging from time series to deep embeddings, we discussed different families of techniques that solve the problem, ranging from multidimensional trees and random projections to inverted files and k-nearest neighbor graphs, and we highlighted the work of different communities on this problem, ranging from the time series to the machine learning communities.

In this special issue, we focus on approximate similarity search, which is gaining popularity, because it provides very fast query answering times. We once again explore different aspects of the problem, including different forms of vector dimensionality reduction, methods designed for operation in memory and on solid state drives, and efficient solutions for sparse vectors. Moreover, we pay particular attention to the role of machine learning in this context, and how various machine learning techniques may help solve the vector similarity search problem more efficiently and effectively. The papers in this special issue (just like in the previous one) also include several details about real use cases of vector similarity search and their special characteristics and requirements in terms of data size, query latency and scalability, as well as discussions on (the several) open research directions.

In the first paper, Gao et al. discuss vector quantization, and present *RaBitQ*, a state-of-the-art binary and scalar vector quantization method that provides unbiased distance estimation, and achieves asymptotically optimal error bounds. In the second paper, Krishnaswamy et al. present the *DiskANN* library of k-nearest neighbor graph indices, which provides an efficient solution for both solid state drives and in memory, while at the same time supporting fast updates and predicate (filtered) search. In the third paper, Bruch et al. describe *SEISMIC*, an efficient solution for approximate similarity search on large collections of learned sparse representations (where the vast majority of the vector coordinates are 0), which finds important applications in information retrieval. In the fourth paper, Wang et al. examine the applicability and usefulness of dimensionality-reduction for speeding up approximate similarity search; they consider six dimensionality-reduction techniques, including traditional algorithms such as PCA and vector quantization, as well as algorithms based on deep learning approaches.

The last two contributions are opinion articles. Douze studies the interplay between vector similarity search and machine learning, comments on the reasons why machine learning does not currently play a big role in vector search, and discusses the role of vector search in machine learning. Finally, Papakonstantinou explores the engineering and machine learning aspects of vector search, and focuses on the emerging needs in this area that include the convergence of vector search with databases.

Overall, the above papers represent a multi-faceted view of vector similarity search, and the ongoing work in this area, in different domains, in both the academia and the industry. We hope that this special issue will further help and inspire the research community in its quest to solve this challenging problem. We would like to thank all the authors for their valuable contributions, as well as Haixun Wang for giving us the opportunity to put together this special issue, and Jieming Shi for helping in its publication.

Themis Palpanas
Université Paris Cité