Decentralized Scalable Algorithm for Approximate Nearest Neighbor Search[1]

Alexander Ponomarenko

National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, 136 Rodionova Ulitsa, Nizhny Novgorod 603093, Russia
e-mail: aponomaremko@hse.ru

The nearest neighbor search problem is an important task of many fields such as pattern recognition [1], machine learning [2], document semantic search [3].

The problem formulated as follows. Given a domain U, a metric function $\sigma : U \times U \to R$ and a finite set X. Need to preprocess X into a data structure S, so the search of nearest neighbor of q from X can be performed with a minimal numbers of metric calculation σ .

Many methods have been proposed for exact nearest neighbor search and for approximate nearest neighbor search. The search complexity of algorithms for exact nearest neighbor search has an exponential dependence from the number of dimensions. The cause lies in the "curse" of dimensionality [4]. Algorithms for an approximate version of the problem are less dependent on the number of dimensions.

In this talk we present an algorithm for approximate nearest neighbor search. We suggest to build a data structure S as graph $G(V, E)$, where $V = X$, with a small world properties and having an approximation of graph Delone as sub graph. The search algorithm based on a greedy walks and has ability to boost the accuracy without rebuilding of the structure.

Computation experiments on published datasets demonstrate performance that outperforms all existing methods.

Presented algorithms do not use the coordinate representation and do not presume the properties of Euclidean spaces, because they are based only on comparing distances between the objects and the query, and therefore are applicable to data from general metric (or even non-metric) spaces.

References

1. T. M. Cover and P. E. Hart *Nearest neighbor pattern classification* Information Theory, IEEE Transactions on, vol. 13, no. 1, pp. 21-27, Jan. 1967
2. Salzberg, S. Cost, and Steven *Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features*, Machine Learning, vol. 10, no. 1, pp. 57-78, 1993
3. S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman *Indexing by Latent Semantic Analysis*, J. Amer. Soc. Inform. Sci., vol. 41, pp. 391-407, 1990
4. E. Chavez, G. Navarro, R. Baeza-Yates, and J. L. Marroquin, *Searching in metric space*, Journal ACM Computing Surveys (CSUR), vol. 33, no. 3, pp. 273-321, Sep. 2001

[1]The author is partially supported by LATNA Laboratory, NRU HSE, RF government grant, ag. 11.G34.31.0057