The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Fast Approximate Nearest Neighbor Search via k-Diverse Nearest Neighbor Graph

Yan Xiao, Jiafeng Guo, Yanyan Lan, Jun Xu, Xueqi Cheng
University of Chinese Academy of Sciences
CAS Key Lab of Network Data Science and Technology,
Institute of Computing Technology, Chinese Academy of Sciences
xiaoyanict@foxmail.com,{guojiafeng, lanyanyan, junxu, cxq}@ict.ac.cn

Abstract

Approximate nearest neighbor search is a fundamental problem and has been studied for a few decades. Recently graph-based indexing methods have demonstrated their great efficiency, whose main idea is to construct neighborhood graph offline and perform a greedy search starting from some sampled points of the graph online. Most existing graph-based methods focus on either the precise k-nearest neighbor (k-NN) graph which has good exploitation ability, or the diverse graph which has good exploration ability. In this paper, we propose the k-diverse nearest neighbor (k-DNN) graph, which balances the precision and diversity of the graph, leading to good exploitation and exploration abilities simultaneously. We introduce an efficient indexing algorithm for the construction of the k-DNN graph inspired by a well-known diverse ranking algorithm in information retrieval (IR). Experimental results show that our method can outperform both state-of-the-art precise graph and diverse graph methods.

Introduction

Many efforts have been devoted to the approximate nearest neighbor search problem and graph-based methods have achieved the state-of-the-art performance in recent literature. The main idea is to connect each point with its several neighbors in the offline stage. In the online stage, some points are sampled as starting points and their neighbors are iteratively explored to approach the query point.

Most existing graph-based methods focus on building either the precise k-NN graph (Hajebi et al. 2011) or the direction-diverse graph (Harwood and Drummond 2016; Li et al. 2016). The precise k-NN graph has good exploitation ability, i.e., neighbors are exactly those nearest points, but the searching on the graph might be easily trapped in local optimums due to the lack of the exploration ability. On the contrary, the direction-diverse graph has good exploration ability, i.e., each point connects multidirectional neighbors so that different directions can be explored when traversing on the graph, but it might not be able to exploit the neighbors very well by focusing on exploration too much.

In this paper, we propose the k-DNN graph, where each point is connected to a set of neighbors that are close in distance while diverse in direction. In this way, we can balance the precision and diversity of the neighborhood graph to keep good exploitation and exploration abilities simultaneously. We take a novel view of the graph construction process as search result diversification in IR, which considers each point as the query and the neighbor candidates as documents, and re-ranks the neighbors based on an adaption of the maximal marginal relevance criterion. Experiments show that our method can outperform both state-of-the-art precise graph and diverse graph methods.

k-Diverse Nearest Neighbor Graph

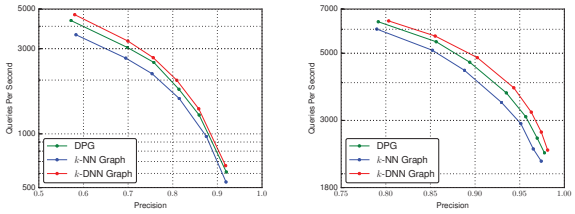
We propose to build a k-DNN graph for fast approximate nearest neighbor search, which takes into account both distance closeness and direction diversity simultaneously. This is quite similar to diverse ranking in IR and there is a simple but effective approach namely maximal marginal relevance (MMR) (Carbonell and Goldstein 1998) algorithm, which employs an iterative selection process and the document with the highest marginal relevance score is selected at each iteration.

Inspired by the MMR algorithm, here we consider the construction of the k-DNN graph as a two-step diverse ranking process for each data point. In the first step, we compute the initial neighbor candidates of each point using the state-of-the-art algorithm NN-Descent (Dong, Moses, and Li 2011). In the second step, we apply a new MMR-type algorithm to re-rank the candidates and obtain the k diverse nearest neighbors for each data point.

Specifically, for a target point v, a set of selected (i.e., ranked) neighbor points S, and a candidate point q, we define the relevance score as the negative distance between the target point and the candidate point, i.e., $-distance(q, v)$. Meanwhile, we define the margin score as the minimum angle between the candidate point and those selected neighbor points with respect to the target point, which is computed by the negative maximum cosine value $-max_{p \in S} cos(\vec{q} - \vec{v}, \vec{p} - \vec{v})$. To avoid the biases of different ranges, we re-scale these scores so that they are in the same range as follows.

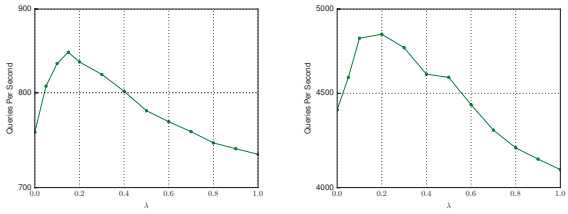
$$Relevance\ Score(q, v) = -\frac{distance(q, v)}{distance_{max} - distance_{min}}$$

$$Margin\ Score(q, v, S) = -\frac{max_{p \in S} cos(\vec{q} - \vec{v}, \vec{p} - \vec{v})}{1 - (-1)}$$



(a) GloVe (b) SIFT

Figure 1: The results for 10-NN search problem.



(a) GloVe (b) SIFT

Figure 2: The QPS with different λ under 90.0% precision.

The marginal relevance score (i.e., MR Score) is a linear combination of the relevance score and the margin score

$$MR\ Score(q, v, S) = \lambda * Relevance\ Score(q, v) + (1 - \lambda) * Margin\ Score(q, v, S)$$

where λ denotes the co-efficiency balancing the two scores. We then use the MR Score to iteratively select top-k neighbor points from the candidates for the target point v.

Experiments

We conduct experiments on GloVe[1] and SIFT[2] to evaluate the k-DNN graph. The GloVe dataset contains about 1.2M pre-trained 100 dimensional word vectors and we randomly sampled about 0.5% points to split the dataset into base set and query set, which contains 1,187,522 and 5,992 points respectively. The SIFT dataset contains 1,000,000 base and 10,000 query vectors, which are both 128 dimensional.

The evaluation metric is queries per second (QPS) against average precision in finding top 10 nearest neighbors. The precision of one query is described as follow.

$$Precision(R) = \frac{|R \cap T|}{|R|},$$

where T denotes the true 10-nearest neighbor set and R denotes the returned 10-nearest neighbor set of a given query.

We compared our k-DNN graph with existing k-NN graph (Hajebi et al. 2011) and diversified proximity graph (DPG)[3] (Li et al. 2016) methods. For fair comparison, we used bidirectional graph for all three methods (Li et al. 2016) and the same search algorithm with same starting points. We set $k = 20$ for all three methods. As our k-DNN graph, we set $\lambda = 0.15$ for GloVe and $\lambda = 0.2$ for SIFT dataset and select neighbors from 80 approximate nearest neighbors obtained by NN-Descent.

The experimental results are shown in Figure 1. As we can see, on both datasets, our method can outperform the two baselines consistently in terms of QPS under different precision criteria, and all the improvements are statistically significant on the t-test (p-value< 0.01). The results demonstrate that by balancing the precision and diversity, the k-DNN graph can obtain better search efficiency than existing methods. We further analyze the performance variance of the k-DNN graph with respect to the free parameter λ, and show the QPS results under 90.0% precision in Figure 2. We find that on both datasets, the performance of the k-DNN graph has the same trending, i.e., first increases then drops after certain point. The best performance is often obtained when λ is small, indicating that direction diversity should be more emphasized than distance closeness in building the graph.

[1]https://nlp.stanford.edu/projects/glove

[2]http://corpus-texmex.irisa.fr

[3]https://github.com/DBWangGroupUNSW/nns_benchmark/tree/master/algorithms/DPG

Conclusions and Future Work

In this paper, we propose the k-DNN graph for approximate nearest neighbor search and introduce an efficient indexing algorithm based on the well-known MMR criterion. We demonstrate through experiments that by balancing the precision and diversity of a graph, one can achieve better search efficiency in online stage. In the future, we want to test our idea with more complicated diverse ranking algorithms.

Acknowledgments

This work was funded by the 973 Program of China under Grant No. 2014CB340401, the National Natural Science Foundation of China (NSFC) under Grants No. 61232010, 61433014, 61425016, 61472401, 61203298 and 61722211, the Youth Innovation Promotion Association CAS under Grants No. 20144310 and 2016102, and the National Key R&D Program of China under Grants No. 2016QY02D0405.

References

Carbonell, J., and Goldstein, J. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, 335–336.

Dong, W.; Moses, C.; and Li, K. 2011. Efficient k-nearest neighbor graph construction for generic similarity measures. In *WWW*, 577–586.

Hajebi, K.; Abbasi-Yadkori, Y.; Shahbazi, H.; and Zhang, H. 2011. Fast approximate nearest-neighbor search with k-nearest neighbor graph. In *IJCAI*, volume 22, 1312.

Harwood, B., and Drummond, T. 2016. Fanng: Fast approximate nearest neighbour graphs. In *CVPR*, 5713–5722.

Li, W.; Zhang, Y.; Sun, Y.; Wang, W.; Zhang, W.; and Lin, X. 2016. Approximate nearest neighbor search on high dimensional data—experiments, analyses, and improvement (v1. 0). *arXiv preprint arXiv:1610.02455*.