Abstract

Locality Sensitive Hashing (LSH) is an index-based data structure that allows spatial item retrieval over a large dataset. The performance measure, $\beta$, has significant effect on the computational complexity and memory space requirement to create and store items in this data structure respectively. The minimization of $\beta$ at a specific approximation factor $c$, is dependent on the load factor, $\gamma$. Over the years, $\beta=4$ has been used by researchers. In this paper, we demonstrate that the choice of $\beta=4$ does not guarantee low computational complexity and low memory space of the data structure under the LSH scheme. To guarantee low computational complexity and low memory space, we propose $\beta=5$. Experiments on the Defense Meteorological Satellite Program imagery dataset have shown that $\beta=5$ saves more than 75% on memory space; cuts the computational complexity by more than 70% and answers query two times faster on the average compared to that of $\beta=4$.
Optimal Load Factor for Approximate Nearest Neighbor Search under Exact Euclidean Locality Sensitive Hashing


Optimal Load Factor for Approximate Nearest Neighbor Search under Exact Euclidean Locality Sensitive Hashing


Index Terms
Computer Science Algorithms

Keywords
Approximate Nearest Neighbor Exact Nearest Neighbor Approximation Factor Performance Measure

Optimal Load Factor
Optimal Load Factor for Approximate Nearest Neighbor Search under Exact Euclidean Locality Sensitive Hashing