Abstract

Locality Sensitive Hashing (LSH) is an index-based data structure that allows spatial item retrieval over a large dataset. The performance measure, $\gamma$, has significant effect on the computational complexity and memory space requirement to create and store items in this data structure respectively. The minimization of $\gamma$ at a specific approximation factor $c$, is dependent on the load factor, $\beta$. Over the years, $\gamma=4$ has been used by researchers. In this paper, we demonstrate that the choice of $\gamma=4$ 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 $\gamma=5$. Experiments on the Defense Meteorological Satellite Program imagery dataset have shown that $\gamma=5$ 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 $\gamma=4$. 
Optimal Load Factor for Approximate Nearest Neighbor Search under Exact Euclidean Locality Sensitive Hashing


---

2 / 5

Index Terms
Computer Science Algorithms

Keywords
Approximate Nearest Neighbor Exact Nearest Neighbor Approximation Factor Performance Measure
Optimal Load Factor