Download Ebook K Nearest Neighbor Algorithm For Classification Read Pdf Free

Lectures on the Nearest Neighbor Method *Data Algorithms* Dimensionality Reduction with Unsupervised Nearest Neighbors Weighted K-nearest Neighbor Algorithm as an Object Localization Technique Using Passive RFID Tags A Note on Rabin's Nearest-neighbor Algorithm **Nearest-neighbor Methods in Learning and Vision Comparison of Fuzzy and Crisp K-nearest Neighbor Algorithm for Linear Prediction Coefficient Based Speaker Dependent Isolated Word Recognition Systems** *Nearest Neighbor Search:* Data Mining in Agriculture Advances in Computational Science, Engineering and Information Technology **Adaptive Nearest Neighbor Algorithm on Hierarchical Weighted-index Graph** *Nearest Neighbor Classifiers with Improved Accuracy and Efficiency* **An Algorithm for Incremental Nearest Neighbor Search in K-dimensional Space** *Toby's Little Trees* **A Simple, Thread-safe, Approximate Nearest Neighbor Algorithm** Spherically Projected Nearest Neighbor Algorithm **Explaining the Success of Nearest Neighbor Methods in Prediction Efficient Expected-time Algorithms for Nearest-neighbor Problems A Tree Algorithm for Nearest Neighbor Searching in Document Retrieval Systems Machine Learning Algorithms From Scratch with Python Enhancement to Selective Incremental Approach for Transductive Nearest Neighbour Classification A Direct Algorithm for the K-nearest-neighbor Classifier Via Local Warping of the Distance Metric Machine Learning for Hackers License Plate Recognition System Using K Nearest Neighbor Algorithm Using a Nearest Neighbor Algorithm and Logistic Regression to Assess Hazard Identification in the U.S. Army Risk Management Process** A Direct Boosting Algorithm for the K-nearest Neighbor Classifier Via Local Warping of the Distance Metric Predicting image data with the SRS-12 algorithm. Applying random projection-based c-approximate nearest-neighbor queries to image data **Classic Works of the Dempster-Shafer Theory of Belief Functions A Randomized Approximate Nearest Neighbors Algorithm** *A Graphics Hardware Accelerated Algorithm for Nearest Neighbor Search* **Master Machine Learning Algorithms** K-Approximate Nearest Neighbor Search Using the GPU *Malani's Modified Kaplan-Meier Estimator Using the Kth Nearest Neighbor Algorithm Time Efficient and Quality Effective K Nearest Neighbor Search in High Dimension Space A Parallel Algorithm for Query Adaptive, Locality Sensitive Hash Search* Reducing the Computational Requirements of the Nearest Neighbor Classifier *A Study of Distance-based Machine Learning Algorithms* **Short-Term Load Forecasting by Artificial Intelligent Technologies** Advances in Databases and Information Systems **Nearest Neighbor (NN) Norms**

This is a collection of classic research papers on the Dempster-Shafer theory of belief functions. The book is the authoritative reference in the field of evidential reasoning and an important archival reference in a wide range of areas including uncertainty reasoning in artificial intelligence and decision making in economics, engineering, and management. The book includes a foreword reflecting the development of the theory in the last forty years. Distance-based algorithms are machine learning algorithms that classify queries by computing distances between these queries and a number of internally stored exemplars. Exemplars that are closest to the query have the largest in uence on the classi cation assigned to the query. Two speci c distance-based algorithms, the nearest neighbor algorithm and the nearest-hyperrectangle algorithm, are studied in detail. It is shown that the k-nearest neighbor algorithm (kNN) outperforms the rst- nearest neighbor algorithm only under certain conditions. Data sets must contain moderate amounts of noise. Training examples from the di erent classes must belong to clusters that allow an increase in the value of k without reaching into clusters of other classes. Methods for choosing the value of k for kNN are investigated. It is shown that one-fold cross-validation on a restricted number of values for k su ces for best performance. It is also shown that for best performance the votes of the k-nearest neighbors of a query should be weighted in inverse proportion to their distances from the query. Principal component analysis is shown to reduce the number of relevant dimen- sions substantially in several domains. Two methods for learning feature weights for a weighted Euclidean distance metric are proposed. These methods improve the performance of kNN and NN in a variety of domains. The nearest-hyperrectangle algorithm (NGE) is found to give predictions that are substantially inferior to those given by kNN in a variety of domains. Experiments performed to understand this inferior performance led to the

discovery of several improvements to NGE. Foremost of these is BNGE, a batch algorithm that avoids construction of overlapping hyperrectangles from di erent classes. Although it is generally superior to NGE, BNGE is still signi cantly inferior to kNN in a variety of domains. Hence, a hybrid algorithm (KBNGE), that uses BNGE in parts of the input space that can be represented by a single hyperrectangle and kNN otherwise, is introduced. The primary contributions of this dissertation are (a) several improvements to existing distance-based algorithms, (b) several new distance-based algorithms, and (c) an experimentally supported understanding of the conditions under which various distance-based algorithms are likely to give good performance. This text presents theoretical and practical discussions of nearest neighbour (NN) methods in machine learning and examines computer vision as an application domain in which the benefit of these advanced methods is often dramatic. Bachelor Thesis from the year 2015 in the subject Computer Science - Applied, grade: 1.0, University of Heidelberg (Computer Science), language: English, abstract: Determining point similarity is a cornerstone for a wide variety of applications. Whenever data is being compared, the problem can be stated as a matter of comparing vectors. Due to this fundamental importance, research has been conducted in this area for decades. For a long time, even for relatively low dimensions such as for example d = 10, complexity was far beyond of what was considered to be feasible in practice. A common approach is to soften the requirements for the accuracy of the neighbor search and look for points within a certain proximity to the query point instead of searching for exact neighbors. In this thesis we evaluate how well the SRS-12 algorithm works on real-world image data in order to lay the ground for future work such as image prediction. The SRS-12 algorithm is an approach to c-approximate nearest-neighbors and claims to have a tiny index and arbitrary approximation ratio while maintaining good theoretical guarantees. We first implement and verify the algorithm and subsequently examine the quality of its outputs when it is applied to image data by performing block matching. We find that the SRS-12 algorithm is indeed very suitable for processing image data as long as the input images are cut into patches that are sufficiently large. However the parameters of the algorithm have to be tuned carefully because they significantly affect not only the quality of the results, but also the computation time which can easily exceed an exact nearest-neighbor query if the parameters are not set properly. We conclude our experiment with a recommendation for well-working parameter settings and propose approaches to enhance the quality and speed of approximate nearest-neighbor queries on image data made with the SRS-12 algorithm such that it can be used in real-time. Modern applications are both data and computationally intensive and require the storage and manipulation of voluminous traditional (alphanumeric) and nontraditional data sets (images, text, geometric objects, time-series). Examples of such emerging application domains are: Geographical Information Systems (GIS), Multimedia Information Systems, CAD/CAM, Time-Series Analysis, Medical Information Sstems, On-Line Analytical Processing (OLAP), and Data Mining. These applications pose diverse requirements with respect to the information and the operations that need to be supported. From the database perspective, new techniques and tools therefore need to be developed towards increased processing efficiency. This monograph explores the way spatial database management systems aim at supporting queries that involve the space characteristics of the underlying data, and discusses query processing techniques for nearest neighbor queries. It provides both basic concepts and state-of-the-art results in spatial databases and parallel processing research, and studies numerous applications of nearest neighbor queries. Doctoral Thesis / Dissertation from the year 2018 in the subject Engineering - Computer Engineering, Jawaharlal Nehru University , language: English, abstract: The tremendous growth of data due to the Internet and electronic commerce has created serious challenges to the researches in pattern recognition. There is a need of processing and analysing data. Advances in data mining and knowledge discovery provide the requirement of new approaches to reduce the data. In this work, methods are proposed to overcome the computational requirements of the nearest neighbor classifiers. The work shows some of the possible remedies to overcome problems with nearest neighbor based classifiers. The work proposes a new method of reducing the data set size. The author reduces the data set size in terms of number of samples and also in terms of number of features. Therefore, the holistic goal of the work is to reduce the time and space requirements and at the same time not to degrade the performance of the nearest neighbor classifier. The nearest neighbor classifier is a popular non-parametric classifier used in many fields since the 1950s. It is conceptually a simple classifier and shows good performance. This book is a printed edition of the Special Issue "Short-Term Load Forecasting by Artificial Intelligent Technologies" that was published in Energies This text

presents a wide-ranging and rigorous overview of nearest neighbor methods, one of the most important paradigms in machine learning. Now in one self-contained volume, this book systematically covers key statistical, probabilistic, combinatorial and geometric ideas for understanding, analyzing and developing nearest neighbor methods. Gérard Biau is a professor at Université Pierre et Marie Curie (Paris). Luc Devroye is a professor at the School of Computer Science at McGill University (Montreal). This book is devoted to a novel approach for dimensionality reduction based on the famous nearest neighbor method that is a powerful classification and regression approach. It starts with an introduction to machine learning concepts and a real-world application from the energy domain. Then, unsupervised nearest neighbors (UNN) is introduced as efficient iterative method for dimensionality reduction. Various UNN models are developed step by step, reaching from a simple iterative strategy for discrete latent spaces to a stochastic kernel-based algorithm for learning submanifolds with independent parameterizations. Extensions that allow the embedding of incomplete and noisy patterns are introduced. Various optimization approaches are compared, from evolutionary to swarm-based heuristics. Experimental comparisons to related methodologies taking into account artificial test data sets and also real-world data demonstrate the behavior of UNN in practical scenarios. The book contains numerous color figures to illustrate the introduced concepts and to highlight the experimental results. If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You ' ll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA Social network analysis (recommendation systems, counting triangles, sentiment analysis) We propose two methods for performing k-Approximate Nearest Neighbor search on the GPU. These methods are helpful in applications that require an estimate of a function in a region surrounding a point in space, for example, photon mapping. Our first approach employs a traditional method using an octree data structure. While the method can compute nearest neighbors at a rate as low as 80ms per query, the memory required by the algorithm restricts the maximum number ofqueries that can be processed simultaneously. The second approach gets over this problem by not using any data structure. This method is based on sorting, taking into account the fact that in most of the applications we know both the data and queries beforehand. With higher degrees of approximation this second approach can process up to 800 queries per millisecond. We also apply our nearest neighbor algorithms to photon mapping, showing that the approximations involved in thesealgorithms do not introduce any major artifacts. Machine learning or artificial intelligence is finally accessible to kids! While taking care of his gardens, Toby discovers that little trees are popping up everywhere! He wants to know if they are apple trees or orange trees. With Toby, you and your loved ones will learn the nearest neighbor method, a very important algorithm in machine learning. The classical nearest neighbors problem is formulated as follows: given a collection of N points in the Euclidean space R^d, for each point, find its k nearest neighbors (i.e. closest points). Obviously, for each point X, one can compute the distances from X to every other point, and then find k shortest distances in the resulting array. However, the computational cost of this naive approach is at least (d*N^2)/2 operations, which is prohibitively expensive in many applications. For example, "naively" solving the nearest neighbors problem with d=100, N=1,000,000 and k=30 on a modern laptop can take about as long as a day of CPU time. Fortunately, in such areas as data mining, image processing, machine learning etc., it often suffices to find "approximate" nearest neighbors instead of the "true" ones. In this work, a randomized approximate algorithm for the solution of the nearest neighbors problem is described. It has a considerably lower computational cost than the naive algorithm, and is fairly fast in practical applications. We provide a probabilistic analysis of this algorithm, and demonstrate its performance via several numerical experiments. Nearest Neighbor algorithms are non-

parametric algorithms that use distance measure techniques for classification and regressions. This thesis uses the method of pruning to improve accuracy and efficiency of a nearest neighbor classifier and also states the different stages the pruning algorithm can be applied and shows the best stage for pruning which gives the maximum accuracy. The performance of the classifier is shown to be better than other improved nearest neighbor classifiers. A fast method of finding the optimal k in a k-nearest neighbor classifier is proposed in the thesis. A method of optimizing the distance measure using a second order training algorithm in a k-nearest neighbor algorithm is also proposed in this thesis which results to better accuracy than the traditional k-nearest neighbor classifier. ABSTRACT: This thesis describes the implementation of a fast, dynamic, approximate, nearest-neighbor search algorithm that works well in fixed dimensions (d The k-nearest neighbor (k-NN) pattern classifier is a simple yet effective learner. However, it has a few drawbacks, one of which is the large model size. There are a number of algorithms that are able to condense the model size of the k-NN classifier at the expense of accuracy. Boosting is therefore desirable for increasing the accuracy of these condensed models. Unfortunately, there does not exist a boosting algorithm that works well with k-NN directly. We present a direct boosting algorithm for the k-NN classifier that creates an ensemble of models with locally modified distance weighting. An empirical study conducted on 10 standard databases from the UCI repository shows that this new Boosted k-NN algorithm has increased generalization accuracy in the majority of the datasets and never performs worse than standard k-NN. The k-nearest neighbor (k-NN) pattern classifier is a simple yet effective learner. However, it has a few drawbacks, one of which is the large model size. There are a number of algorithms that are able to condense the model size of the k-NN classifier at the expense of accuracy. Boosting is therefore desirable for increasing the accuracy of these condensed models. Unfortunately, there does not exist a boosting algorithm that works well with k-NN directly. We present a direct boosting algorithm for the k-NN classifier that creates an ensemble of models with locally modified distance weighting. An empirical study conducted on 10 standard databases from the UCI repository shows that this new Boosted k-NN algorithm has increased generalization accuracy in the majority of the datasets and never performs worse than standard k-NN. Rabin has proposed a probabilistic algorithm for finding the closest pair of a set of points in Euclidean space. His algorithm is asymptotically linear whereas the best of the known deterministic algorithms take order nlogn time. At least part of the speedup is due to the model rather than the probabilistic nature of the algorithm. If you ’ re an experienced programmer interested in crunching data, this book will get you started with machine learning—a toolkit of algorithms that enables computers to train themselves to automate useful tasks. Authors Drew Conway and John Myles White help you understand machine learning and statistics tools through a series of hands-on case studies, instead of a traditional math-heavy presentation. Each chapter focuses on a specific problem in machine learning, such as classification, prediction, optimization, and recommendation. Using the R programming language, you ’ ll learn how to analyze sample datasets and write simple machine learning algorithms. Machine Learning for Hackers is ideal for programmers from any background, including business, government, and academic research. Develop a na ï ve Bayesian classifier to determine if an email is spam, based only on its text Use linear regression to predict the number of page views for the top 1,000 websites Learn optimization techniques by attempting to break a simple letter cipher Compare and contrast U.S. Senators statistically, based on their voting records Build a “ whom to follow ” recommendation system from Twitter data Explains the success of Nearest Neighbor Methods in Prediction, both in theory and in practice. You must understand the algorithms to get good (and be recognized as being good) at machine learning. In this Ebook, finally cut through the math and learn exactly how machine learning algorithms work, then implement them from scratch, step-by-step. Master's Thesis from the year 2012 in the subject Computer Science - Didactics, , course: COMPUTER SCIENCE & ENGINEERING, language: English, abstract: During the last years, semi-supervised learning has emerged as an exciting new direction in machine learning research. It is closely related to profound issues of how to do inference from data, as witnessed by its overlap with transductive inference. Semi-Supervised learning is the half-way between Supervised and Unsupervised Learning. In this majority of the patterns are unlabelled, they are present in Test set and knowed labeled patterns are present in Training set. Using these training set, we assign the labels for test set. Here our Proposed method is using Nearest Neighbour Classifier for Semi-Supervised learning we can label the unlabelled patterns using the labeled patterns and then compare these method with the traditionally Existing methods as graph mincut, spectral graph partisan, ID3,Nearest

Neighbour Classifier and we are going to prove our Proposed method is more scalable than the Existing methods and reduce time complexity of SITNNC(Selective Incremental Approach for Transductive Nearest Neighbour Classifier) using Leaders Algorithm. Nearest neighbor search is a fundamental requirement of many machine learning algorithms and is essential to fuzzy information retrieval. The utility of efficient database search and construction has broad utility in a variety of computing fields. Applications such as coding theory and compression for electronic communication systems as well as use in artificial intelligence for pattern and object recognition. In this thesis, a particular subset of nearest neighbors is consider, referred to as c-approximate k-nearest neighbors search. This particular variation relaxes the constraints of exact nearest neighbors by introducing a probability of finding the correct nearest neighbor c, which offers considerable advantages to the computational complexity of the search algorithm and the database overhead requirements. Furthermore, it extends the original nearest neighbors algorithm by returning a set of k candidate nearest neighbors, from which expert or exact distance calculations can be considered. Furthermore thesis extends the implementation of c-approximate k-nearest neighbors search so that it is able to utilize the burgeoning GPGPU computing field. The specific form of c-approximate k-nearest neighbors search implemented is based on the locality sensitive hash search from the E2LSH package of Indyk and Andoni [1]. In this paper, the authors utilize the exceptional properties of the Leech Lattice [2], as a subspace quantizer for the locality sensitive hash families. The Leech Lattice is unique in that it provides the closest lattice packing of equal sized spheres in 24 dimensional space. In addition, advances from coding theory provide a very favorable decoding algorithm for finding the nearest lattice center to a query point in euclidean 24 dimensional space [3] [4]. The multilevel construction of the Leech Lattice provides an excellent opportunity for parallelization as it contains the minimization of many independent sub-lattice decodings resulting from the lattices exceptional symmetry among lattices. These decodings are additionally highly floating point computationally intensive, and because of which suggest a favorable implementation on GPGPU architectures such as NVIDIAs CUDA based framework. Furthermore, the overall construction of a locality sensitive hash based, nearest neighbors search algorithm, is able to be parallelized fairly efficiently as the hash decodings are completely independent of one another. The goal of this thesis is to present a CUDA optimized parallel implementation of a bounded distance Leech Lattice decoder [4] for use in query optimized c-approximate k-nearest neighbors using the locality sensitive hash framework of E2LSH. The system will be applied to the approximate image retrieval of SIFT transformed [5] image vectors.[1] A. Andoni, Nearest Neighbor Search: the Old, the New, and the Impossible. INSTITUTE OF [2] J. Leech, "Notes on sphere packings," [3] . J. Forney, G.D., "A bounded-distance decoding algorithm for the leech lattice, with generalizations, "[4] O. Amrani and Y. Beery, "Efficient bounded-distance decoding of the hexacode and associated decodersfor the leech lattice and the golay code," [5] D.G. Lowe, "Object recognition from local scale-invariant features," Data Mining in Agriculture represents a comprehensive effort to provide graduate students and researchers with an analytical text on data mining techniques applied to agriculture and environmental related fields. This book presents both theoretical and practical insights with a focus on presenting the context of each data mining technique rather intuitively with ample concrete examples represented graphically and with algorithms written in MATLAB ® . You must understand algorithms to get good at machine learning. The problem is that they are only ever explained using Math. No longer. In this Ebook, finally cut through the math and learn exactly how machine learning algorithms work. Using clear explanations, simple pure Python code (no libraries!) and step-by-step tutorials you will discover how to load and prepare data, evaluate model skill, and implement a suite of linear, nonlinear and ensemble machine learning algorithms from scratch. Technologies using identification by radio frequencies (RFID) are experiencing rapid development and healthcare is a major application area benefiting from it. Highly pervasive RFID enables remote identification, tracking and localization of the medical staff, patients, medications and equipment, thus increasing safety, optimizing in real-time management and providing support for new ambient-intelligent services. This thesis describes and evaluates an algorithm that enables object localization and tracking using passive RFID tags. This thesis also describes scenarios of how this technology can be used as a part of building a smart trauma resuscitation room by tracking the equipments. The main contribution of this thesis is the adaptation of the Weighted K-Nearest Neighbor Algorithm as a localization technique to track objects in a confined and crowded space by using passive RFID tags. The input parameter to the algorithm is the received signal strength indicator

(RSSI), which gives a measure of back-scattered radio frequencies from passive tags. While using RFID technology special attention has to be given to the placement of antennas to get the optimum result. Therefore, we analyzed various antenna placement configurations with mean error and error consistency as the two performance parameters. The detection of multiple tags and human occlusion are two major concerns while tracking tags in a confined space with many team members collaborating on solving a problem. The RF signal can be interrupted by people walking around randomly and holding multiple (tagged) instruments at the same time. While the algorithm worked fine when tracking multiple tags, we had to modify the experimental set-up and attach an antenna onto the ceiling (which we call a vertical antenna), so that even if all the wall antennas are blocked we get at least one input parameter to base our localization decision on. We evaluated the algorithm for different combinations of configurations and number of neighbors, and achieved the following results. The best results were obtained for the 3 antennae (placed orthogonally) configuration considering the 4 nearest neighbors wherein a mean error rate of 15% of the maximum possible error was achieved under ideal conditions. We tested the algorithm for different human occlusion scenarios i.e. blocking 1 or 2 wall antennas, standing in random positions and then roaming in the field area randomly. The mean error rate for the standing scenario was measured as 20% of the maximum possible error and 18% in the case of roaming configuration. The error was found to be consistently within our defined maximum error for 100% of the recorded readings. The results obtained were found to be satisfactory for our application where, more than the exact location of the object, knowing whether the object is within a particular region is good enough for the users to know what task is being carried out in the trauma bay. Also the algorithm holds good in an indoor environment having a lot of factors and materials which affect the RF signal disrupting accurate calculation of the location co-ordinates. The algorithm does not require extensive data collection prior to implementation which makes it easily deployable in any environment. Apart from the problems mentioned there are some other factors like materials on which the tags are attached and orientation of tags which were found to be potential hindrances for accurate localization. Acceptable solutions to these problems form a part of our future work. This book constitutes the refereed proceedings of the 11th East European Conference on Advances in Databases and Information Systems, ADBIS 2007, held in Varna, Bulgaria, in September/October 2007. The 23 revised papers presented together with three invited lectures were carefully reviewed and selected from 77 submissions. The papers address current research on database theory, development of advanced DBMS technologies, and their advanced applications. This book is the proceedings of Third International Conference on Computational Science, Engineering and Information Technology (CCSEIT-2013) that was held in Konya, Turkey, on June 7-9. CCSEIT-2013 provided an excellent international forum for sharing knowledge and results in theory, methodology and applications of computational science, engineering and information technology. This book contains research results, projects, survey work and industrial experiences representing significant advances in the field. The different contributions collected in this book cover five main areas: algorithms, data structures and applications; wireless and mobile networks; computer networks and communications; natural language processing and information theory; cryptography and information security. K-Nearest-Neighbors (KNN) search is a fundamental problem in many application domains such as database and data mining, information retrieval, machine learning, pattern recognition and plagiarism detection. Locality sensitive hash (LSH) is so far the most practical approximate KNN search algorithm for high dimensional data. Algorithms such as Multi-Probe LSH and LSH-Forest improve upon the basic LSH algorithm by varying hash bucket size dynamically at query time, so these two algorithms can answer different KNN queries adaptively. However, these two algorithms need a data access post-processing step after candidates' collection in order to get the final answer to the KNN query. In this thesis, Multi-Probe LSH with data access post-processing (Multi-Probe LSH with DAPP) algorithm and LSH-Forest with data access post-processing (LSH-Forest with DAPP) algorithm are improved by replacing the costly data access post-processing (DAPP) step with a much faster histogram-based post-processing (HBPP). Two HBPP algorithms: LSH-Forest with HBPP and Multi- Probe LSH with HBPP are presented in this thesis, both of them achieve the three goals for KNN search in large scale high dimensional data set: high search quality, high time efficiency, high space efficiency. None of the previous KNN algorithms can achieve all three goals. More specifically, it is shown that HBPP algorithms can always achieve high search quality (as good as LSH-Forest with DAPP and Multi-Probe LSH with DAPP) with much less time cost (one to several orders of magnitude speedup) and same

memory usage. It is also shown that with almost same time cost and memory usage, HBPP algorithms can always achieve better search quality than LSH-Forest with random pick (LSH-Forest with RP) and Multi-Probe LSH with random pick (Multi-Probe LSH with RP). Moreover, to achieve a very high search quality, Multi-Probe with HBPP is always a better choice than LSH-Forest with HBPP, regardless of the distribution, size and dimension number of the data set.