**K-nearest Neighbors Search by Random Projection Forests**

**Abstract:**

K-nearest neighbors (kNN) search is an important problem in data mining and knowledge discovery. Inspired by the huge success of tree-based methodology and ensemble methods over the last decades, we propose a new method for kNN search, random projection forests (rpForests). rpForests ﬁnds nearest neighbors by combining multiple kNN-sensitive trees with each constructed recursively through a series of random projections. As demonstrated by experiments on a wide collection of real datasets, our method achieves a remarkable accuracy in terms of fast decaying missing rate of kNNs and that of discrepancy in the k-th nearest neighbor distances. rpForests has a very low computational complexity as a tree-based methodology. The ensemble nature of rpForests makes it easily parallelized to run on clustered or multicore computers; the running time is expected to be nearly inversely proportional to the number of cores or machines. We give theoretical insights on rpForests by showing the exponential decay of neighboring points being separated by ensemble random projection trees when the ensemble size increases. Our theory can also be used to reﬁne the choice of random projections in the growth of rpForests; experiments show that the effect is remarkable

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows XP/UBUNTU.
* Implementation : NS2
* NS2 Version : 2.28
* Front End : OTCL (Object Oriented Tool Command  Language)
* Tool : Cygwin (To simulate in Windows OS)