Abstract

k-Nearest Neighbor is a simple and effective classification method. The primary idea of this method is to calculate the distance from a query point to all of classified data points and make choice of a class which occurs maximum time in k closest neighbors. The Euclidean distance and cosine similarity are the common choice for similarity metric among all the similarity measures. Apart from Euclidean and Cosine there are various similarity measures available and being used to calculate similarity in n-dimensional vector space model for classification. Similarity calculation is a complex operation and computationally need high time if vector dimension increases. Hence this paper explores the usefulness of nine different similarity measures in kNN and presents their experimental results on agriculture dataset. We also compared the time required to finish the classification task and concluded that I-divergence is taking minimum time compared to these algorithms.

References
3. B.S. Charulatha, et. El, A Comparative study of different distance metrics that can be used in Fuzzy Clustering Algorithms, IJETTCS, National Conference on Architecture, Software systems and Green computing-2013(NCASG2013)

Index Terms

Computer Science

Information Sciences

Keywords

Euclidian, Manhattan, Minkowsaky, Canberra, Chebychev, Cosine, Correlation, Chi-square, I-divergence