Singaporean Journal of Scientific Research(SJSR)
International Journal of Software Engineering (IJSE)
Vol.8.No.4 2016,Pp.266-274
available at :www.iaaet.org/sjsr
Paper Received : 28-09-2016
Paper Accepted: 20-11-2016
Paper Reviewed by: 1.Prof. Sihen Liue 2. Dr.G.Ramesh
Editor : Dr. Ramnath Sigh

Comparison of Nearest Neighbor Technique

C.SenthilMurugan[1] and Dr. S. Prakasam[2]
[1]PhD Scholar ,[2]Associate Professor
[1,2]Dept of CSA, SCSVMV University, Kanchipuram
[1]Asst Professor, Dept of MCA, TCET, Vandavasi.
[1]dessenthil.c@gmail.com, [2]Prakasam_sp@yahoo.com,

Abstract

The Nearest Neighbor techniques are very simple techniques. The techniques highly effective and efficient in the many field of pattern recognition, text categorization and object recognition. Its prime main advantage, but the disadvantage cannot be ignored even. Typing to solve this problem several techniques presented as condensing techniques were proposed. The memory requirement and computing complexity for a large training sets. The nearest neighbor is a comparison to KNN algorithm of structure based on the model, weighted KNN model, model based KNN, condensed NN, reduced NN, and generalized NN are the structure less techniques. A structure algorithm developed to based on the overcome memory limitation and structure based techniques. Trying to solving the problem so many techniques presented as condensing techniques were proposed. A possibility is to hybridize them with other algorithm; called modern heuristics (or) Meta heuristics can improve the solution. This paper focus on some recent techniques of nearest neighbor algorithm.

Keywords-

Nearest neighbor (NN); kNN; Prototype selection; Condensed NN; Reduced NN; Condensing; Genetic algorithms; Tabu Search. Nearest neighbor (NN), kNN, Model based kNN, Weighted kNN, Condensed NN, Reduced NN.

I. Introduction

The Nearest Neighbor rule is a powerful classification method and identifies the category of compare the prototype set of pattern. This is the rule of pattern recognition, text categorization; ranking models object recognition and event recognition applications [10]. The K-Nearest Neighbor purpose by T.M Cover and P.E Hart. Weights KNN is improve by T [4]. Bailey A.K Jain. To NN algorithm is calculated on the base of value K [2]. To specified many nearest neighbor are to be considered to define the class of a sample data point. The research problem is an unavoidable importance of large set of the computational time [7]. The repeated patterns do not add extra information and training samples for eliminated. Gongde Guo selects the model based kNN [15]. This model automatically selects the value of K. Classical KNN using of ranking, false neighbor information, clustering. The KNN implementation on the using ball tree, k-d tree, nearest feature line, tunable metric, principal axis search tree and orthogonal search tree is a compare on the function. Basic KNN Algorithm Increase the speed of another algorithm [12]. The nearest neighbor is a comparison to KNN algorithm of structure based on the model, weighted KNN model, model based KNN, condensed NN, reduced NN, and generalized NN are the structure less techniques.

II. Software Engineering

Software engineering is the application of a systematic, disciplined, quantifiable approach to the development, operations and maintenance of software [12, 14]. Software engineering is that form of engineering that applies the principles of design, engineering mathematics, management, psychology and sociology and other disciplines. A software development is need to required that the user, design, customer and the capabilities and opportunities of software. A software development is importance of activities of domain that goals, pure customer activities, constructing activity and end point of activities [16]. Software engineering is the application of methods and scientific knowledge to create practical cost-effective solution for the design, construction, operation and maintenance of software. Software engineering is categorized at one of the domain in software engineering is which is concerned with goals, functions and constraints of software [5, 6].

III. Nearest Neighbor Techniques

Nearest Neighbor techniques are divided into two categories:

1. Structure less NN technique
2. Structure based NN technique

- Structure less NN technique

The K-nearest neighbor lies in first category in which whole data is classified into training data and sample data point [2]. This technique is very easy to implement but value of k affects the result in some cases. The structure less NN technique like as weighted k nearest neighbor (WkNN), condensed nearest neighbor (CNN), reduced nearest neighbor (RNN), model based k nearest neighbor (MkNN), rank nearest neighbor (kRNN), modified k nearest neighbor (MkNN), pseudo / generalized nearest neighbor (GNN), clustered k nearest neighbor (CkNN) [4]. These steps are repeated until all data tuples are grouped.

- Structure based NN technique

The K-nearest neighbor lies in second category is based on structure of data like as ball tree k nearest neighbor (KNS1), k-d tree nearest neighbor (kdNN), nearest feature line neighbor (NFL), local nearest neighbor, tunable nearest neighbor (TNN), center based nearest neighbor (CNN), principal axis tree nearest neighbor (PAT), orthogonal search tree nearest neighbor [11]. For this FL distance between query point and each pair of feature line is calculated for each class.

IV. Prototype Selection

Prototype selection is the process of representative patterns from the data, which can help in reducing these data [1, 25]. Like many other combinatorial problems, the prototype selection (PS) would require of optimization problems and use general techniques [22]. An exhaustive search to obtain optimal solutions in the general case. The problem to solve the proposed in Heuristics, especially the nearest neighbor algorithm, and met heuristics, especially genetic algorithms and tabu search [23].

V. Improving Prototype Selection

To correct this slight defect, the modern heuristics or met heuristics came to complete them [7]. Two types of met heuristics have been successful in their hybridization with the traditional KNN: Genetic algorithms and tabu search [23]. The tabu search is method of dynamic neighborhood, which selects, at each iteration, the best solution of the first local optimum by finding the best neighbor.

Table1. Comparison Of Nearest Neighbor Techniques

1	K Nearest Neighbor (KNN) [1]	Key Idea	Nearest Neighbor used for rules
		Advantage	1. Very fast to training 2. Easy and Simple to learn 3. Noisy training data secured for robust 4. Large training data is effective
		Disadvantage	1. Limited memory 2. Algorithm is runs slowly 3. Irrelevant attributes for easily fooled 4. Computation complexity 5. Biased by value of k
		Target Data	1. Large data samples 2. Massive data 3. Classification
2	Weighted k nearest neighbor (WkNN) [2]	Key Idea	Calculate for assign weights to neighbor as per distance
		Advantage	1. Overcomes limitations of Knn 2. Just k is not samples used all training 3. Makes the algorithm global one
		Disadvantage	1. Algorithm is runs slowly 2. Calculating weight is increases to computation complexity
		Target Data	1. Large sample data

			2. Massive data 3. Unbalanced data 4. Classification
3	Condensed nearest neighbor (CNN) [3, 4, 5]	Key Idea	Eliminate data sets which show similarity and do not add additional information to the classification
		Advantage	1. Training data size for reduce 2. Improve memory requirements and query time 3. Recognition rate to reduce
		Disadvantage	1. CNN is dependent order; it is not likely to pick up point on boundary. 2. Computation complexity
		Target Data	1. Memory requirement is main concern to data 2. Environment with limited memory 3. Duplicate data and pattern
4	Reduced nearest neighbor (RNN) [6]	Key Idea	Initially the resulting set is equal to the initial set, and then each instance that does not cause a wrong classification of another instance in the initial set is removed from the resulting set.
		Advantage	1. Improve memory requirements and search (query) time 2. Recognition rate to reduce 3. Reduces the training data and eliminates model
		Disadvantage	1. High cost of computational 2. Time consuming 3. Computational complexity 4. Its consistency depends on the consistency of the resulting set of CNN
		Target Data	Large data set and Null data set
5	Model based k nearest neighbor (MkNN) [7]	Key Idea	The model based kNN is constructed from data and classify new data using model
		Advantage	1. Automatically selected for value of k 2. Reduce number of data point is high efficiently 3. More classification accuracy
		Disadvantage	1. Do not consider the region of outside marginal data
		Target Data	Large repository is dynamic web mining
6	Rank nearest neighbor [18]	Key Idea	Each category of data training in assign ranks
		Advantage	1. There are too much variation for perform better to between features 2. Robust as based on rank
		Disadvantage	1. Multivariate kRNN depends on

			distribution of the data
		Target Data	1. Class distribution of nature the Gaussian 2. Data points with general information
7	Fast condensed nearest neighbor (FCNN) [8 ,9]	Key Idea	Decision boundary is very closed to the select points
		Advantage	1. It is independent of the order 2. Condensation is good rate 3. Computational complexity a smaller than CNN 4. Average efficiency of 96.01% for an average number of iterations of about 69 iterations
		Disadvantage	1. Large number of requires in an iterations 2. Low quadratic complexity.
		Target Data	S can be classified correctly without P
8	Modify k nearest neighbor (MkNN) [10]	Key Idea	Classify nearest neighbor is a used weights and validity of data point
		Advantage	1. MkNN is stable and robust 2. Overcome low accuracy of WkNN
		Disadvantage	1. Computation complexity
		Target Data	1. Classification 2. Methods facing outlets
9	Pseudo / Generalized nearest neighbor (GNN) [11]	Key Idea	Utilizes information of n-1 instead of only nearest neighbor
		Advantage	1. The whole training data set is used to consider of n-1 classes
		Disadvantage	1. Small data is does not hold good 2. Computational complexity
		Target Data	1. Large data set 2. Classification
10	Clustered k nearest neighbor [12]	Key Idea	Clusters are formed to the clustering data is select nearest neighbor
		Advantage	1. Nature the robust 2. Defect of uneven distribution overcome training samples
		Disadvantage	1. Threshold parameter is difficult before running algorithm selected 2. Value of k for clustering is biased
		Target Data	1. Text classification
11	Ball Tree k nearest neighbor (KNS1) [13]	Key Idea	kNN speed improve to ball tree structure
		Advantage	1. Easy to implement 2. Represent data structure is well tune
		Disadvantage	1. An insertion algorithm is costly 2. KNS1 degrades distance is increased
		Target Data	1. Robotic, vision, speed, graphics create on

			geometric learning tasks
12	k-d tree nearest neighbor (kdNN) [14]	Key Idea	The training data exactly divide into plane half
		Advantage	1. Simple and Fast 2. The perfectly balanced tree is produce
		Disadvantage	1. More computation 2. Intensive search is requirement 3. Slice point is blindly half may be miss data structure
		Target Data	1. The multi dimensional points is an organized
13	Nearest feature Line Neighbor (NFL) [21]	Key Idea	Lines for increasing data and accuracy of the mapping point
		Advantage	1. Improvement on the classification accuracy 2. Small size is a highly effective 3. Ignored in nearest neighbor of an utilizes information
		Disadvantage	1. Computational complexity 2. The features point describe on the straight line is hard task
		Target Data	1. Pattern recognition 2. Face recognition 3. Classification
14	Local Nearest Neighbor [22]	Key Idea	Query point is focused on nearest neighbor of local
		Advantage	1. NFL is a cover limitations
		Disadvantage	1. Computations for number
		Target Data	1. Pattern or Face recognition 2. Classification
15	Tunable Nearest Neighbor (TNN) [23]	Key Idea	Uses tunable metric
		Advantage	1. Small data sets is an efficient
		Disadvantage	1. Computations for large number
		Target Data	1. Classification 2. Bias or Discrimination problem
16	Center based Nearest Neighbor (CNN) [24]	Key Idea	The calculating is a center line
		Advantage	1. Small data sets is a highly efficient
		Disadvantage	1. Computations for large number
		Target Data	1. Pattern recognition 2. Classification
17	Principal Axis Tree Nearest Neighbor (PAT) [21]	Key Idea	Principal Axis Tree is used
		Advantage	1. The performance is good 2. Search for fast
		Disadvantage	1. Computation time low
		Target Data	3. Pattern recognition 4. Spatial data set

18	Orthogonal Search Tree Nearest Neighbor [22]	Key Idea	Orthogonal trees is used
		Advantage	1. Computation time is less 2. Large data sets is an effective
		Disadvantage	1. More query time
		Target Data	5. Pattern recognition 6. Spatial data set

VI. CONCLUSION

Through this study of comparison the two nearest neighbor techniques are discussed. One is structure less and another is structure base. Both techniques are improved over basic NN techniques. Proposed an improvement by researchers to space efficiency as well as gain speed efficiency. All techniques hold good in particular field under the particular circumstances. To ensure the minimality of this training set we presented some recent proposals using metaheuristics to check the optimality of the resulting set of some KNN reduction techniques. Note that each technique is very effective in a specific area and in special circumstances.

VII. REFERENCE

[1] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inform. Theory, Vol. IT-13, pp 21-27, Jan 1967.

[2] T. Bailey and A. K. Jain, "A note on Distance weighted k-nearest neighbor rules", IEEE Trans. Systems, Man Cybernetics, Vol.8, pp 311- 313, 1978.

[3] K. Chidananda and G. Krishna, "The condensed nearest neighbor rule using the concept of mutual nearest neighbor", IEEE Trans. Information Theory, Vol IT- 25 pp. 488-490, 1979.

[4] F Angiulli, "Fast Condensed Nearest Neighbor", ACM International Conference Proceedings, Vol 119, pp 25-32. [5] E Alpaydin, "Voting Over Multiple Condensed Nearest Neighbors", Artificial Intelligent Review 11:115-132, 1997.

[5] E Alpaydin, "Voting Over Multiple Condensed Nearest Neighbors", Artificial Intelligent Review 11:115-132, 1997.

[6] Geoffrey W. Gates, "Reduced Nearest Neighbor Rule", IEEE Trans Information Theory, Vol. 18 No. 3, pp 431-433.

[7] G. Guo, H. Wang, D. Bell, "KNN Model based Approach in Classification", Springer Berlin Vol 2888.

[8] F. Bajramovic, Frank Mattern, Nicholas Butko, and Joachim Denzler "A comparison of nearest neighbor search algorithms for generic object recognition", LNCS 4179, ACIVS 2006, pp. 1186–1197.

[9] Hatem A. Fayed and Amir F. Atiya, "A novel template reduction approach for the K-nearest neighbor method", IEEE Transactions on Neural Networks, vol. 20 (5), pp. 890–896, May 2009.

[9] Y. Zeng, Y. Yang, L. Zhou, "Pseudo Nearest Neighbor Rule for Pattern Recognition", Expert Systems with Applications (36) pp 3587-3595, 2009.

[10] H. Parvin, H. Alizadeh and B. Minaei, "Modified k Nearest Neighbor", Proceedings of the world congress on Engg. and computer science 2008.

[11] Z. Yong, "An Improved kNN Text Classification Algorithm based on Clustering", Journal of Computers, Vol. 4, No. 3, March 2009.

[12] Djouadi and E. Bouktache, "A Fast Algorithm for Nearest Neighbor Classification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19. No. 3, 1997.

[13] V.Vaidehi, S. Vasuhi, "Person Authentication using Face Recognition", Proceedings of the world congress on engg and computer science, 2008.

[14] Shizen, Y. Wu, "An Algorithm for Remote Sensing Image Classification based on Artificial Immune b-cell Network", Springer Berlin, Vol 40.

[15] G. Toker, O. Kirmemis, "Text Categorization using k Nearest Neighbor Classification", Survey Paper, Middle East Technical University. [16] Y. Liao, V. R. Vemuri, "Using Text Categorization Technique for Intrusion detection", Survey Paper, University of California.

[17] E. M. Elnahrawy, "Log Based Chat Room Monitoring Using Text Categorization: A Comparative Study", University of Maryland.

[18] X. Geng et. al, "Query Dependent Ranking Using k Nearest Neighbor", SIGIR, 2008.

[19] Y. Yang and T. Ault, "Improving Text Categorization Methods for event tracking", Carnegie Mellon University.

[20] F. Bajramovie et. al "A Comparison of Nearest Neighbor Search Algorithms for Generic Object Recognition", ACIVS 2006, LNCS 4179, pp 1186-1197.

[21] S. Z Li, K. L. Chan, "Performance Evaluation of the NFL Method in Image Classification and Retrieval", IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 22, 2000.

[22] W. Zheng, L. Zhao, C. Zou, "Locally Nearest Neighbor Classifier for Pattern Classification", Pattern Recognition, 2004, pp 1307-1309.

[23] Y. Zhou, C. Zhang, "Tunable Nearest Neighbor Classifier", DAGM 2004, LNCS 3175, pp 204-211.

[24] Q. B. Gao, Z. Z. Wang, "Center Based Nearest Neighbor Class", Pattern Recognition, 2007, pp 346-349.

[25] F. Bajramovie et. al "A Comparison of Nearest Neighbor Search Algorithms for Generic Object Recognition", ACIVS 2006, LNCS 4179, pp 1186-1197.

[26] T. Liu, A. W. Moore, A. Gray, "New Algorithms for Efficient High Dimensional Non-Parametric Classification", Journal of Machine Learning Research, 2006, pp 1135-1158.

[27] S. N. Omohundro, “Five Ball Tree Construction Algorithms”, 1989, Technical Report.