International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 2
ISSN: 2454-4248
350 – 354

A Review on Advanced Decision Trees for Efficient & Effective k-NN Classification

Ms. Madhavi Pujari
Department of Technology,
Shivaji University,Kolhapur
Kolhapur ,India
email:madhavipujari9921@gmail.com

Mr. Chetan Awati
Department of Technology,
Shivaji University,Kolhapur
Kolhapur ,India
email:chetan.awati@gmail.com

Ms.Sonam Kharade
Delta Tech India,
Kolhapur,
Kolhapur ,India
email:skh9624@gmail.com

Abstract **-** K Nearest Neighbor (KNN) strategy is a notable classification strategy in data mining and estimations in light of its direct execution and colossal arrangement execution. In any case, it is outlandish for ordinary KNN strategies to select settled k esteem to all tests. Past courses of action assign different k esteems to different test tests by the cross endorsement strategy however are typically tedious. This work proposes new KNN strategies, first is a KTree strategy to learn unique k esteems for different test or new cases, by including a training arrange in the KNN classification. This work additionally proposes a change rendition of KTree technique called K*Tree to speed its test organize by putting additional data of the training tests in the leaf node of KTree, for example, the training tests situated in the leaf node, their KNNs, and the closest neighbor of these KNNs. K*Tree, which empowers to lead KNN arrangement utilizing a subset of the training tests in the leaf node instead of all training tests utilized in the recently KNN techniques. This really reduces the cost of test organize.

Keywords- *KNN, Classifier, KTree,Fuzzy*

*****

I. INTRODUCTION

KNN method is popular because of its simple implementation and works incredibly well in practice. KNN is considered a lazy learning algorithm that classifies the data sets based on their similarity with neighbors. But KNN have some limitations which affects the efficiency of result. The main problem with the KNN is that it is lazy learner as well as the KNN does not learn from the training data which affects the accuracy in result. Also KNN algorithm computation cost is quite high. So, these problems with KNN algorithm affect the accuracy in result and overall efficiency of algorithm. This work proposes the new KNN strategies KTree and K*Tree are more productive than the conventional KNN strategies. There are two recognized contrasts between the past KNN strategies and proposed KTree strategy. In the first place, the past KNN methods have no training stage, while KTree method has a sparse-based preparation stage, whose time complexity is O(n2). Second, the previous methods need at least O(n2) time complexity to obtain the ideal-k-values due to involving a sparse-based learning process, while KTree method just needs O(log(d) + n) to do that via the learned model. In this work, additionally stretch out proposed KTree technique to its change rendition called k*Tree strategy to speed test organize, by just putting additional data of training tests in the left node, for example, the training tests, their KNNs, and the closest neighbors of these closest neighbors. KTree methods learns different set samples and add a training stage in the traditional KNN classification. The K*Tree speed up its test stage. This reduces running cost of its stage.

II. LITERATURE SURVEY

Efficient kNN Classification With Different Numbers of Nearest Neighbors: In this paper[1] they proposes the new KNN technique KTree & K*Tree to conquer the impediments of customary KNN techniques. Accordingly, it is trying for all the while tending to these issues of KNN technique, i.e., ideal k-values learning for various examples, time cost lessening, and execution change. To address these issues of KNN techniques, in this paper, they initially propose a KTree technique for quick taking in an ideal k-esteem for each test, by including a training organize into the conventional KNN strategy. They additionally broaden proposed KTree strategy to its change form i.e K*Tree technique to speed test arrange. The key thought of proposed techniques is to outline a training stage for lessening the running expense of test arrange and enhancing the classification execution.

Block-Row Sparse Multiview Multilabel Learning for Image Classification: In this paper [2] they lead multiview picture order by proposing a piece push scanty MVML learning structure. They inserted a proposed blockrow regularizer into the MVML structure to lead the high level highlight choice to choose the instructive perspectives and furthermore lead the low-level element choice to choose the data highlights from the instructive perspectives. Their proposed strategy adequately led picture grouping by evading the unfriendly effect of both the excess perspectives and the boisterous highlights.

 ISSN: 2454-4248
350 – 354

Biologically Inspired Features for Scene Classification in Video Surveillance: In this paper[3] they introduces a scene order technique in view of an enhanced standard model highlight., In this paper they recently proposed technique is more roboust more specific and of lower complexity. The moved forward models reliably beat as far as both power also, grouping exactness. Moreover, impediment and confusion issues in scene order in video observation are contemplated in this paper.

Learning Instance Correlation Functions for Multilabel Classification: In this paper[4], a powerful calculation is produced for multilabel order with using those information that are significant to the objectives. The proposes the development of a coefficient-based mapping amongst preparing and test examples, where the mapping relationship misuses the connections among the examples, instead of the unequivocal relationship between the factors and the class marks of information

Missing Value Estimation for Mixed-Attribute Data Sets: In this paper[5], they thinks about another setting of missing information attribution that is ascribing missing information in informational collections with heterogeneous traits, alluded to as crediting blended quality informational indexes. This paper proposes two predictable estimators for discrete what's more, constant missing target esteems. They additionally proposes a blend piece based iterative estimator is pushed to attribute blended characteristic informational indexes.

Feature Combination and the kNN Framework in Object Classification: In this paper[6], they take a shot at normal blend to investigate the fundamental instrument of highlight blend. They examine the practices of highlights in normal blend and weighted normal mix. Further they coordinate the practices of highlights in (weighted) normal blend into the kNN structure.

A Unified Learning Framework for Single Image Super-Resolution: In this paper[7], they propose another SR structure that flawlessly incorporates learning-and reconstruction based strategies for single picture SR to keep away from sudden relics presented by learning-based SR and reestablish the missing high-recurrence points of interest smoothed by recreation based SR. This incorporated structure takes in a solitary word reference from the LR contribution rather than from outside pictures to daydream points of interest, inserts nonlocal implies channel in the recreation based SR to improve edges and stifle ancient rarities, and step by step amplifies the LR contribution to the coveted top notch SR result

Single Image Super-Resolution With Multiscale Similarity Learning: In this paper[8] they propose a solitary picture SR approach by taking in multiscale self-likenesses from a LR picture itself to diminish the unfriendly impact brought by incompatible high-recurrence subtle elements in the preparation set, To incorporate the missing points of interest they proposes the HR-LR fix sets utilizing the underlying LR information and its down inspected form to catch the similitudes crosswise over various scales

Classification of incomplete data based on belief functions and K-nearest neighbors: In this paper[9] they proposes an option credal arrangement strategy for deficient examples (CCI) in light of the framewok of conviction capacities. In CCI, the K-closest neighbors (KNNs) of the articles are chosen to appraise the missing esteems. CCI manages K forms of the inadequate example with evaluated esteems drawn from the KNNs. The K variants of the fragmented example are separately arranged utilizing the traditional techniques, and the K bits of order are marked down with various measuring factors relying upon the separations between the protest and its KNNs. These reduced outcomes are all around combined for the credal grouping of the question.

Feature Learning for Image Classification via Multiobjective Genetic Programming: In this paper[10], they plan a developmental learning procedure to consequently create space versatile worldwide component descriptors for picture classification utilizing multiobjective hereditary programming (MOGP). In this design, an arrangement of crude 2-D administrators are haphazardly consolidated to develop include descriptors through the MOGP advancing and afterward assessed by two target wellness criteria, i.e., the grouping mistake and the tree many-sided quality. After the whole development system completes, the best-so-far arrangement chose by the MOGP is viewed as the(near-)ideal component descriptor got.

An Adaptable k-Nearest Neighbors Algorithm for MMSE Image Interpolation: In this paper[11] they propose a picture introduction calculation that is nonparametric and learning-based, principally utilizing a versatile k-closest neighbor algorithm with worldwide contemplations through Markov arbitrary fields. The proposed calculation guarantees picture comes about that are information driven and, subsequently reflect true pictures well, sufficiently given preparing information. The proposed calculation works on a nearby window utilizing a dynamic k-closest neighbor calculation, where varies from pixel to pixel.

A Novel Template Reduction Approach for the k-Nearest Neighbor Method: In this paper [12]they propose another consolidating calculation. The proposed thought depends on characterizing the supposed chain. This is a succession of closest neighbors from substituting classes. They make the point that examples additionally down the tie are near the order limit and in light of that they set a cutoff for the examples keep in the preparation set.

A Sparse Embedding and Least Variance Encoding Approach to Hashing: In this paper[13],they propose an effective and proficient hashing approach by scantily implanting an example in the preparation test space and encoding the inadequate installing vector over a scholarly

 ISSN: 2454-4248
350 – 354

word reference. They segment the example space into bunches through a direct ghostly grouping strategy, and after that speak to each example as a scanty vector of standardized probabilities that it falls into its few nearest groups. At that point they propose a minimum difference encoding model, which takes in a word reference to encode the scanty implanting highlight, and therefore binarize the coding coefficients as the hash codes.

Ranking Graph Embedding for Learning to Rerank: In this paper[14], they demonstrate that bringing positioning data into dimensionality decrease altogether builds the execution of picture look reranking. The proposed technique changes chart inserting, a general system of dimensionality decrease, into positioning diagram implanting (RANGE) by demonstrating the worldwide structure and the nearby connections in and between various pertinence degree sets, separately. A novel essential parts investigation based closeness estimation strategy is introduced in the phase of worldwide chart development.

A Novel Locally Linear KNN Method With Applications to Visual Recognition: In this paper[15], a locally straight K Nearest Neighbor (LLK) strategy is given applications to strong visual acknowledgment. In the first place the idea of a perfect portrayal is displayed, which enhances the conventional inadequate portrayal from numerous points of view. The novel representation is handled by two classifiers, LLKbased classifier and a locally direct closest mean-based classifier, for visual acknowledgment. The proposed classifiers are appeared to interface with the Bayes choice run for least blunder. The new techniques are proposed for include extraction to additionally enhance visual acknowledgment execution.

Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects: In this work[16],they exhibited a study of fluffy closest neighbor classifiers. The utilization of FST and some of its expansions to the improvement of enhanced closest neighbor calculations have been checked on, from the principal recommendations to the latest methodologies. A few segregating attributes of the procedures has been de-scribed as the building pieces of a multi-level scientific classification, formulated to oblige introduce.

The Role of Hubness in Clustering High-Dimensional Data: In this paper[17], they take a novel point of view on the issue of bunching high-dimensional information. Rather than endeavoring to stay away from the scourge of dimensionality by watching a lower dimensional element subspace. They demonstrate that hubness, i.e., the propensity of high-dimensional information to contain focuses (center points) that much of the time happen in k closest neighbor arrangements of different focuses, can be effectively misused in grouping. They approve their theory by showing that hubness is a decent measure of point centrality inside a high-dimensional information bunch, and by proposing a few hubness-based grouping calculations.

Fuzzy similarity-based nearest-neighbour classification as alternatives to their fuzzy-rough parallels: In this paper[18], the hidden instruments of fluffy harsh closest neighbor (FRNN) and enigmatically evaluated unpleasant sets (VQNN) are in-vestigated and examined. The hypothetical confirmation and exact assessment demonstrate that the subsequent arrangement of FRNN and VQNN depends just upon the most noteworthy similitude and most noteworthy summation of the likenesses of each class, individually.

III. PROBLEM STATEMENT

To enhance the arrangement productivity of KNN algorithm by presenting new techniques KTree and K*Tree by outlining a training organize for ideal k esteems learning for various examples, reducing the cost of test organize, enhancing the exactness in result and enhancing the classification execution. Additionally we will outline and actualize framework which works and process high dimensional data to enhance the performance of proposed techniques and plan soft clustering classifier and compare this with KNN strategies.

IV. DIFFERENT CLASSIFICATION ALGORITHM COMPARISON

A. Decision Tree

A decision tree is a tree in which each branch hub speaks to a decision between various choices, and each leaf hub speaks to a choice. Decisions trees order occurrences by navigate from root hub to leaf hub [43]. We begin from root hub of choice tree, testing the characteristic indicated by this hub, at that point moving down the tree limb as per the quality incentive in the given set. This procedure is the rehashed at the sub-tree level. Decision tree learning calculation has been effectively utilized as a part of master frameworks in catching information. Decision tree is moderately quick contrasted with other order models. It additionally Obtain comparative and once in a while better exactness contrasted with different models

B. Decision stump

A decision stump is an extremely basic decision tree. A decision stump is a machine learning model comprising of a one-level choice tree. It is a decision tree with one inner hub (the root) which is quickly associated with the terminal hubs (its takes off). A decision stump makes a forecast in light of the estimation of only a solitary info include. At times they are additionally called 1-rules. It's a tree with just a single split, so it's a stump. decision stump calculation takes a gander at all conceivable incentive for each quality. It chooses best quality

 ISSN: 2454-4248
350 – 354

in view of least entropy. Entropy is measure of vulnerability. We measure entropy of dataset (S) concerning each trait. For each characteristic A, one level processes a score estimating how well trait An isolate the classes[44]

Table 1 discuss all about classification algorithm and comparison over different parameters

TABLE 1.DIFFERENT CLASSIFICATION ALGORITHM COMPARISON

Sr.No.	Algorithm	Features
1	C 4.5 Algorithm	1.Build model can be Effectively deciphered
		2. Easy to execute.
		3. Can use both discrete & continuous values.
		4. Deals with noise.
2	ID3 Algorithm	1. It delivers more accuracy result than C4.5 algorithm
		2. Detection rate is increment & space utilization is reduced.
3	Artificial Neural Network Algorithm	1. Need to parameter adjust
		2. Learning is required
4	Naive Bayes Algorithm	1. Easy to implement
		2. Great computational productivity & characterization rate
		3. Accuracy of result is high
5	Support Vector Machine Algorithm	1.High accuracy
		2. Work well even if data is not linearly separable in the base feature space.
6	K- Nearest Neighbor Algorithm	1.Classes need not be linearly distinct
		2. Zero cost of the learning process.
		3. Sometimes it is robust with regard to noisy training data
		4. Well suited for multimodal classes

V. CHOICE OF TOPIC WITH REASONING

K Nearest Neighbor is one of the best ten data mining algorithm on account of its simplicity of comprehend, basic execution and great characterization execution. Be that as it may, past shifted KNN strategies typically first take in an individual ideal k-esteem for each test or new example and after that utilize the conventional KNN order to anticipate test tests by the educated ideal k-esteem. In any case, either the way toward taking in an ideal k-esteem for each test or the way toward examining all training tests for finding closest neighbors of each test is take additional time. Along these lines, it is trying for at the same time conquer a few issues of KNN technique like optimal k-values learning for various examples, decreasing time cost, and enhancing execution proficiency. To overcome the restrictions of KNN techniques to enhance the effectiveness and exactness in comes about and controlling the time cost, this framework, to begin with propose a KTree strategy for quick taking in an optimal k-esteem for each test, by including a training arrange into the customary KNN technique. Additionally proposed framework outline the new form of KTree technique called K*Tree to speed test organize and diminishes the time cost of test arrange.

VI. CONCLUSION

In previous work, to conquer some issues of KNN technique, two new KNN classification algorithms, i.e., the KTree and the KTree strategies are proposed to choose ideal k-esteem for each test sample and successful KNN classification. The aim of proposed strategies is to plan training arrange for reducing the running expense of test organize and enhancing the classification execution. Additionally we will plan framework which works and process high dimensional data to increase the performance of proposed strategies and plan soft clustering classifier and compare this with KNN strategies.

VII. REFERENCES

[1] Shichao Zhang, Xuelong Li, Ming Zong, Xiaofeng Zhu, and Ruili Wang, Efficient kNN Classification With Different Numbers of Nearest Neighbors, Volume: PP, Issue: 99,April 2017

[2] X. Zhu, X. Li, and S. Zhang, Block-row sparse multiview multilabel learning for image classification, IEEE Trans. Cybern., vol. 46, no. 2, pp. 450461, Feb. 2015.

[3] K. Huang, D. Tao, Y. Yuan, X. Li, and T. Tan, Biologically inspired features for scene classification in video surveillance, IEEE Trans. Syst., Man, Cybern., Part B, vol. 41, no. 1, pp. 307313, Feb. 2011.

[4] H. Liu, X. Li, and S. Zhang, Learning instance correlation functions for multi-label classification, IEEE Trans. Cybern., vol. 47, no. 2, pp. 499510, Feb. 2017.

[5] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, and Z. Xu, Missing value estimation for mixed attribute data sets, IEEE Trans. Knowl. Data Eng., vol. 23, no. 1, pp. 110121, Jan. 2011.

[6] J. Hou, H. Gao, Q. Xia, and N. Qi, Feature combination and the kNN framework in object classification, IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 6, pp. 13681378, Jun. 2016.

[7] J. Yu, X. Gao, D. Tao, X. Li, and K. Zhang, A unified learning framework for single image super-resolution, IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 4, pp. 780792, Apr. 2014

[8] K. Zhang, X. Gao, D. Tao, and X. Li, Single image super-resolution with multiscale similarity learning, IEEE Trans.

Neural Netw. Learn. Syst., vol. 24, no. 10, pp. 16481659, Oct. 2013.

[9] Zhun-ga Liu , Yong Liu , Jean Dezert , Quan Pan , "Classification of incomplete data based on belief functions and K-nearest neighbors, Knowledge-Based Systems, volume 89, November 2015.

[10] L. Shao, L. Liu, and X. Li, Feature learning for image classification via multiobjective genetic programming, IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 7, pp. 13591371, Jul. 2014.

[11] K. S. Ni and T. Q. Nguyen, An adaptable-nearest neighbors algorithm for MMSE image interpolation, IEEE Trans. Image Process., vol. 18, no. 9, pp. 19761987, Mar. 2009.

[12] H. A. Fayed and A. F. Atiya, A novel template reduction approach for the K-nearest neighbor method, IEEE Trans. Neural Netw., vol. 20, no. 5, pp. 890896, May 2009

[13] X. Zhu, L. Zhang, and Z. Huang, A sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process., vol. 23, no. 9, pp. 37373750, Sep. 2014.

[14] Y. Pang, Z. Ji, P. Jing, and X. Li, Ranking graph embedding for learning to rerank, IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 8, pp. 12921303, Aug. 2013

[15] Q. Liu and C. Liu, A novel locally linear KNN method with applications to visual recognition, IEEE Trans. Neural Netw. Learn. Syst., to be published.

[16] Joaqun Derrac,Salvador Garca,Francisco Herrera Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects Information Sciences, vol. 260, 1 March 2014,

[17] Nenad Toma sev, Milo s Radovanovi c, Dunja Mladeni c, and Mirjana Ivanovi c The Role of Hubness in Clustering High-Dimensional Data IEEE Trans. on Knowledge and Data Engineering, vol. 26, NO. 3, Match 2014.

[18] Yanpeng Qu a,b, Qiang Shenb, Neil Mac Parthalin b, Changjing Shangb, WeiWua Fuzzy similarity-based nearest-neighbour classification as al-ternatives to their fuzzy-rough parallels International Journal of Ap-proximate Reasoning, International Journal of Approximate Reasoning Volume 54, Issue 1, January 2013.

[19] Xiaofeng Zhu, Xuelong Li, Shichao Zhang,Chunhua Ju, and Xindong Wu, Robust Joint Graph Sparse Coding for UnsupervisedSpectral Fea-ture Selection IEEE Tranasactions on Neural Networks and Learning System, volume: 28 Issue: 6 June 2017.

[20] H. Wang, Nearest neighbors by neighborhood counting, IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 6, pp. 942953, Jun. 2006.

[21] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, Learning k for kNN classification, ACM Trans. Intell. Syst. Technol., vol. 8, no. 3, pp. 119, 2017.

[22] D. Tao, J. Cheng, X. Gao, X. Li, and C. Deng, Robust sparse coding for mobile image labeling on the cloud, IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 1, pp. 6272, Jan. 2017.

[23] H. Wang, Nearest neighbors by neighborhood counting, IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 6, pp. 942953, Jun. 2006.

[24] X. Zhu, S. Zhang, J. Zhang, and C. Zhang, Cost-sensitive imputing missing values with ordering, in Proc. AAAI, 2007, pp. 19221923

[25] S. Zhang, M. Zong, K. Sun, Y. Liu, and D. Cheng, Efficient kNN algorithm based on graph sparse reconstruction, in Proc. ADMA, 2014, pp. 356369.

[26] B. Li, S. Yu, and Q. Lu. (2003). An improved k-nearest neighbor algorithm for text categorization. [Online]. Available: https://arxiv.org/abs/cs/0306099

[27] F. Sahigara, D. Ballabio, R. Todeschini, and V. Consonni, Assessing the validity of QSARS for ready biodegradability of chemicals: An applicability domain perspective, Current Comput.-Aided Drug Design, vol. 10, no. 2, pp. 137147, 2013.

[28] X. Zhu, Z. Huang, Y. Yang, H. T. Shen, C. Xu, and J. Luo, Selftaught dimensionality reduction on the high-dimensional small-sized data, Pattern Recognit., vol. 46, no. 1, pp. 215229, 2013.

[29] E. Blanzieri and F. Melgani, Nearest neighbor classification of remote sensing images with the maximal margin principle, IEEE Trans. Geosci. Remote Sens., vol. 46, no. 6, pp. 18041811, Jun. 2008.

[30] H. Liu and S. Zhang, Noisy data elimination using mutual k-nearest neighbor for classification mining, J. Syst. Softw., vol. 85, no. 5, pp. 10671074, 2012.

[31] X. Zhu, H.-I. Suk, and D. Shen, A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis, NeuroImage, vol. 100, pp. 91105, Oct. 2014.

[32] Y. Pang, Y. Yuan, and X. Li, Effective Feature Extraction in High-Dimensional Space, IEEE Trans. Syst., Man, B, vol. 38, no. 6, pp. 16521656, Dec. 2008.

[33] S. Zhang and X. Wu, Large scale data mining based on data partitioning, Appl. Artif. Intell., vol. 15, no. 2, pp. 129139,2001

[34] X. Li, Y. Pang, and Y. Yuan, L1-norm-based 2DPCA, IEEE Trans. Syst., Man, Cybern. B, vol. 40, no. 4, pp. 11701175, Aug. 2010.

[35] X. Zhu, Z. Huang, H. Cheng, J. Cui, and H. T. Shen, Sparse hashing for fast multimedia search, ACM Trans. Inf. Syst., vol. 31, no. 2, p. 9, 2013.

[36] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 3, pp. 328340, Mar. 2005.

[37] X. He and P. Niyogi, Locality preserving projections, in Proc. Neural Inf. Process. Syst., vol. 16. 2004, p. 153.

[38] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] H. Li and J. Sun, Majority voting combination of multiple case-based reasoning for financial distress prediction, Expert Syst. Appl., vol. 36, no. 3, pp. 43634373, 2009.

[40] A. Bahety, Extension and Evaluation of ID3Decision Tree Algorithm, Entropy, vol. 2, no. 1, p. 1, 2014.

[41] Z. H. Zhou and Y. Yu, Ensembling local learners throughmultimodal perturbation, IEEE Trans. Syst. Man, B, vol. 35, no. 4, pp. 725735,Apr. 2005.

[42] Z. H. Zhou, Ensemble Methods: Foundations and Algorithms. London,U.K.: Chapman & Hall, 2012.

[43] XindongWu Vipin Kumar J. Ross Quinlan Joydeep Ghosh Qiang Yang Hiroshi Motoda Geoffrey J. McLachlan Angus Ng Bing Liu Philip S. Yu Zhi-Hua Zhou Michael Steinbach David J. Hand Dan

[44] Wyne Iba, Pat Langley, Induction of One Level Decision Trees, Machine Learning 1992.