Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Hesitant Fuzzy k-Nearest Neighbour (HFK-NN) Classifier for Document Classification and Numerical Result Analysis

Neeraj Sahu*
Singhania University
Rajasthan, India

G. S. Thakur
MANIT
Bhopal, India

Abstract— This paper presents new approach Hesitant Fuzzy K-nearest neighbour (HFK-nn) based document classification and numerical results analysis. The proposed classification Hesitant Fuzzy K-nearest neighbour (HFK-nn) approach is based on hesitant Fuzzy distance. In this paper we have used hesitant Fuzzy distance calculations for document classification results. The following steps are used for classification: data collection, data pre-processing, data selection, presentation, analysis, classification process and results. The experimental results are evaluated using MATLAB 7.14. The Experimental Results show proposed approach that is efficient and accurate compare to other classification approach.

Keywords—Hesitant Fuzzy k-nearest neighbour, hesitant Fuzzy distance, classification and data mining.

I. INTRODUCTION

Document classification is the recent issue in text mining. Document Classification areas are science, technology, social science, biology, economics, medicine and stock market etc. In last recent years lot of research work has been done decodes some best contributions on Document classification are as follows:

J.L. Castro, L.D. Flores-Hidalgo, C.J. Mantas, J.M. Puche, Extraction of fuzzy rules from support vector machines[2], T.M. Cover, P.E. Hart, Nearest neighbor pattern classification[4], B.A. Whitehead, T.D. Choate, Evolving space-filling curves to distribute radial basis functions over an input space[36], T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction[8], J.M. Keller, M.R. Gray, J.A. Givens Jr., A fuzzy K-nearest neighbor algorithm[12], C.C. Lee, Fuzzy logical in control systems: fuzzy logic controller[15], J.A. Leonard, M.A. Kramer, L.H. Ungar, Using radial basis functions to approximate a function and its bounds[16].

S. Destercke, S. Guillaume, B. Charnomordic, Building an interpretable fuzzy rule base from data using orthogonal least squares-application to a depollution problem[5], C.-F. Juang, C.-T. Lin, An on-line self-constructing neural fuzzy inference network and its applications [11], C.J. Lin, An efficient immune-based Symbiotic particle swarm optimization learning algorithm for TSK-type neuro-fuzzy networks design [17], C.J. Lin, Y.J. Xu, A self-adaptive neural fuzzy network with group-based symbiotic evolution and its prediction applications[19], J. Liu, M. Xu, Kernelized fuzzy attribute C-means clustering algorithm[21], J.M. Mendel, Fuzzy logic systems for engineering: a tutorial[22], W. Pedrycz, M. Reformat, Rule based modeling of nonlinear relationships[27], J. Moody, C.J. Darken, Fast learning in networks of locally tuned processing units[23], G.C. Mouzouris, J.M. Mendel, A single-value-QR decomposition based method for training fuzzy logic systems in uncertain environments[24], W. Pedrycz, Conditional fuzzy C-means[25], W. Pedrycz, M. Reformat, Evolutionary fuzzy modeling[28].

W. Pedrycz, Conditional fuzzy clustering in the design of radial basis function neural networks[26], G.B. Huang, P. Saratchandran, N. Sundararajan, An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks[9], T.M. Cover, Estimation by the nearest neighbor rule[3], J. Quinlan, Combining instance-based model and model-based learning[29], S. Elanayar, Y.C. Shin, Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems[7], M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification[30], Z.Q. Liu, F. Yan, Fuzzy neural network in case-based diagnostic system[20],A. Staiano, R. Tagliaferri, W. Pedrycz, Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering[31], C.M. Bishop, Neural Networks for Pattern Recognition[1], C.-T. Sun, Rule base structure identification in an adaptive network based fuzzy inference system[32], E.T. Kim, et al., A new approach to fuzzy modeling[14].

J.A. Dickerson, B. Kosko, Fuzzy function approximation with ellipsoidal rules[6], T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control[33], J.M. Keller, R. Krishnapuram, F.C.H. Rhee, Evidence aggregation networks for fuzzy logic inference[13],G. Tsekourasa, H. Sarimveisb, E. Kavaklia, G. Bafasb, A hierarchical fuzzy-clustering approach to fuzzy modeling[34], C.T. Lin, C.-S.G. Lee, Neural-network-based fuzzy logic control and decision system[18], M.G. Tsipouras, T.P. Exarchos, D.I. Fotiadis, A methodology for automated fuzzy model generation[35], J. Jang, C.T. Sun, Functional equivalence between radial basis function networks and fuzzy inference Systems [10]. The above mentioned work suffers from lack of efficiency and accuracy. The low accuracy is still issue and challenge in the Classification. This motivates us to construct the new method for

Classification. New Document Classification method we called Hesitant Fuzzy K-Nearest Neighbour. Hence we proposed new document classification approach HFK-nn.

The remaining paper is organized as follows: Section-I describe introduction and review of literatures. Section-II describes HFK-nn and K-nn. In Section-III, Methodology of document classification steps are described. In Section-IV, Experimental results are described. In Section-V, results Evaluation and measurement are described. Finally, we concluded and proposed some future directions in Conclusion Section.

II. CALCULATIONS FOR HESITANT FUZZY K-NN AND GENERAL K-NN CLASSIFIER

In this calculation we find k- nearest neighbour based on Hesitant Fuzzy distance (Hfd) and General distance (Gd). Hesitant Fuzzy distance and General distance of each pi to pj: Table 1 and Table 2. Represent all distance calculated by Hfd, Gd. Hesitant Fuzzy distance and General distance Cluster Point show in Table 3 and Table 4 with ascending order. This calculation shows hesitant Fuzzy distance based accuracy percentages and General distance based accuracy percentages Cluster Point show in Table 5 and Table 6.

For computational model we give tabulation form from Table 1 to Table 6.

TABLE I
HESITANT FUZZY DISTANCE FROM CLUSTER POINT

Clusters Points	**Hesitant fuzzy distance set from Cluster Point P1(7,4)**	**Selected Hesitant fuzzy distance from Cluster Point P_1(7,4)**
P_2(9,6)	{0.19,0.21,0.17,0.16,0.15,0.14}	0.19
P_3(11,4)	{0.29,0.30,0.27,0.31,0.33,0.32}	0.30
P_4(2,3)	{0.20,0.19,0.22,0.21,0.23,0.24}	0.22
P_5(4,5)	{0.13,0.12,0.11,0.14,0.15,0.16}	0.14
P_6(5,6)	{0.04,0.06,0.07,0.08,0.09,0.10}	0.09
P_7(7,9)	{0.12,0.13,0.15,0.17,0.14,0.16}	0.16
P_8(9,8)	{0.03,0.07,0.06,0.05,0.04,0.08}	0.04

TABLE III
GENERAL DISTANCE FROM CLUSTER POINT

Clusters Points	**General distance from Cluster Point P_1(7,4)**
P_2(9,6)	2.82
P_3(11,4)	4.00
P_4(2,3)	5.09
P_5(4,5)	3.16
P_6(5,6)	2.82
P_7(7,9)	5.00
P_8(9,8)	4.47

III. METHODOLOGY

In the Classification of document different the steps are used. The steps are as follows:

A. Data Collection: In this phase collect relevant documents like e-mail, news, web pages etc. from various heterogeneous sources. These text documents are stored in a variety of formats depending on the nature of the data. The datasets are downloaded from UCI KDD Archive. This is an online repository of large datasets and has wide variety of data types.

B. Classification Method: Initial step is to complete review of literature in the field of data mining. Next step is a detailed survey of data mining and existing Algorithms for Classification. In this area some work done by various researchers. After studying their work, it would be attempted to find the Classification algorithm.

C. Classification Process: Algorithms develop for Classification Process. Classification Process means transform documents into a suitable determined in classes for the Classification task. In Classification Process we performed Different tasks. Optimized classification will also be studied. The real data may be great source for the Classification.

D. Classification Results: In this Experiment we calculate k- nearest neighbour Based on Hesitant fuzzy distance and General distance. Hesitant fuzzy distance and General distance from Cluster Points Pi to Pj calculated and gives ascending order of the hesitant fuzzy distance and General distance for tabulation. Hesitant fuzzy Distance accuracy percentages and General distance accuracy percentages from Cluster Point show in tabulation. This Experiment show hesitant fuzzy distance based accuracy percentages is efficient and accurate compare General distance based accuracy percentages.

Algorithm 1: This Algorithm obtains hesitant fuzzy distance of a cluster from each cluster.

Step 1: Input eight clusters points.
Step 2: initialize x1, y1 for cluster point and x2, y2 for each clusters points.
Step 3: Produce and compare hesitant fuzzy distance one by one.
Step 4: find minimum Hesitant fuzzy distance Hfd from clusters points say first.
Step 5: arrange all hesitant fuzzy distance in ascending order.

Algorithm 2: This Algorithm obtains General distance of a cluster from each cluster.

Step 1: Input eight clusters points.
Step 2: initialize x1, y1 for cluster point and x2, y2 for each clusters points.
Step 3: Produce and compare General distance one by one.
Step 4: find minimum General distance Gd from clusters points say first.
Step 5: arrange all General distance in ascending order.

IV. EXPERIMENTAL RESULTS

In this Experiment we calculate k- nearest neighbour based on Hesitant fuzzy distance and General distance. Hesitant fuzzy distance and General distance from Cluster Points P_1 to P_8 calculated and gives ascending order of the hesitant fuzzy distance and General distance for tabulation describe in Table 3 and Table 4. Hesitant fuzzy Distance accuracy percentages and General distance accuracy percentages from Cluster Point show in Table 5 and Table 6. This Experiment show hesitant distance based accuracy percentages is efficient and accurate compare General distance based accuracy percentages.

TABLE IIIII

HESITANT FUZZY DISTANCE FROM CLUSTER POINT IN ASCENDING ORDER

Clusters Points	**Hesitant fuzzy distance from Cluster Point P_1(7,4)**
P_8(9,8)	0.04
P_6(5,6)	0.09
P_5(4,5)	0.14
P_7(7,9)	0.16
P_2(9,6)	0.19
P_4(2,3)	0.22
P_3(11,4)	0.30

TABLE IV

GENERAL DISTANCE FROM CLUSTER POINT IN ASCENDING ORDER

Clusters Points	**General distance from Cluster Point P_1 (7,4)**
P_2 (9,6)	2.82
P_6 (5,6)	2.82
P_5 (4,5)	3.16
P_3(11,4)	4.00
P_8(9,8)	4.47
P_7 (7,9)	5.00
P_4 (2,3)	5.09

TABLE V

HESITANT FUZZY DISTANCE ACCURACY PERCENTAGES FROM CLUSTER POINT

Clusters Points	**Accuracy percentages from Cluster Point %**
$P_2(9,6)$	27.63
$P_3(11,4)$	38.59
$P_4(2,3)$	52.37
$P_5(4,5)$	64.22
$P_6(5,6)$	76.29
$P_7(7,9)$	89.67
$P_8(9,8)$	98.99

TABLE VI
GENERAL DISTANCE ACCURACY PERCENTAGES FROM CLUSTER POINT

Clusters Points	**Accuracy percentages from Cluster Point %**
$P_2(9,6)$	23.34
$P_3(11,4)$	34.45
$P_4(2,3)$	49.45
$P_5(4,5)$	59.45
$P_6(5,6)$	72.34
$P_7(7,9)$	89.45
$P_8(9,8)$	95.45

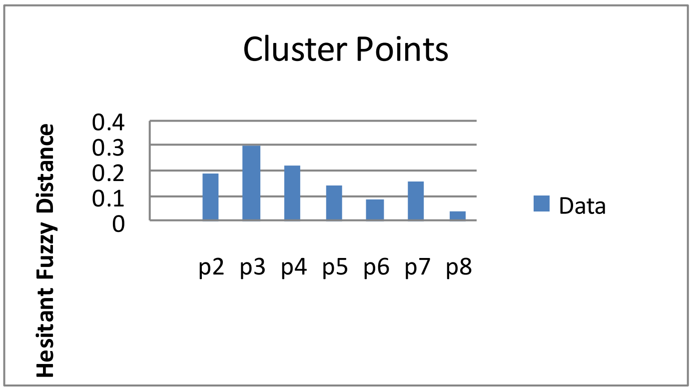

Fig 1: Hesitant fuzzy Distance from Cluster Point.

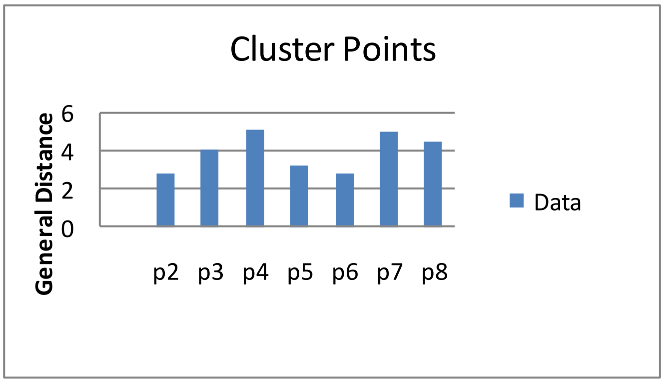

Fig 2: General Distance from Cluster Point

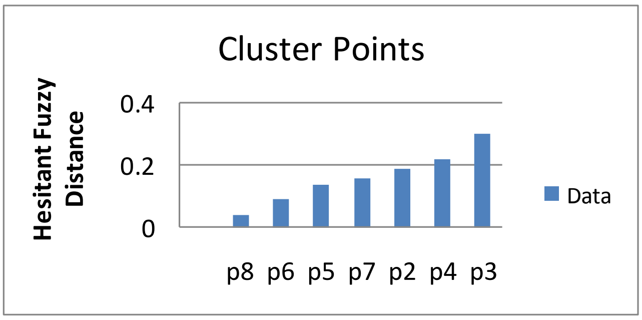

Fig 3: Hesitant fuzzy Distance from Cluster Point in ascending order

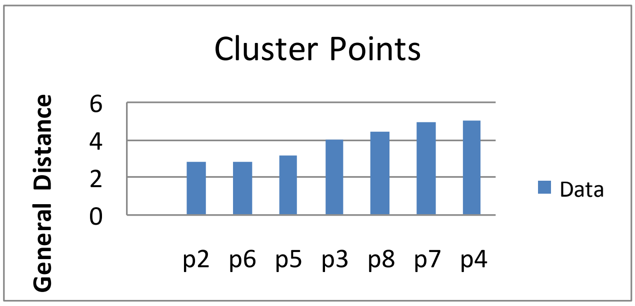

Fig 4: General Distance from Cluster Point in ascending order

The figures 5 and 6 describe document Classification results And Accuracy % of Classification process.

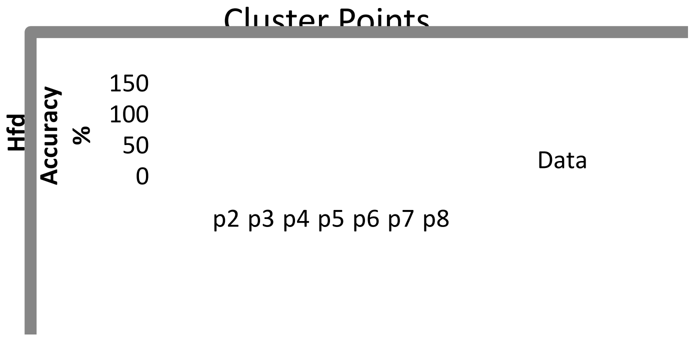

Fig 5: Accuracy % from Cluster Point for Hesitant fuzzy Distance

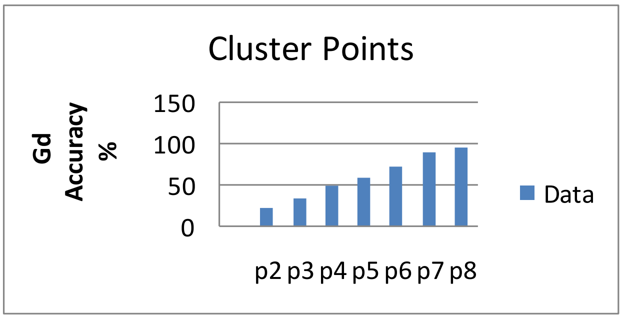

Fig 6: Accuracy % from Cluster Point for General Distance

ACKNOWLEDGMENT

This work is supported by research grant from MPCST, Bhopal, India Project ref. No. 1080/CST/R&D/2012, Dated 30/06/2012.

REFERENCES

[1] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, New York, USA, 1995.

[2] J.L. Castro, L.D. Flores-Hidalgo, C.J. Mantas, J.M. Puche, Extraction of fuzzy rules from support vector machines, Fuzzy Sets and Systems 158 (18) (2007) 2057–2077.
[3] T.M. Cover, Estimation by the nearest neighbor rule, IEEE Trans. Inf. Theory IT-14 (1) (1968) 50–55.
[4] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory IT-13 (1) (1967) 21–27.
[5] S. Destercke, S. Guillaume, B. Charnomordic, Building an interpretable fuzzy rule base from data using orthogonal least squares-application to a depollution problem, Fuzzy Sets and Systems 158 (2007) 2078–2094.
[6] J.A. Dickerson, B. Kosko, Fuzzy function approximation with ellipsoidal rules, IEEE Trans. Syst. Man Cybern. 26 (4) (1996) 542–560.
[7] S. Elanayar, Y.C. Shin, Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems, IEEE Trans. Neural Networks 5 (1994) 594–603.
[8] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, Springer, Berlin, 2001.
[9] G.B. Huang, P. Saratchandran, N. Sundararajan, An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks, IEEE Trans. SMC-B 34 (6) (2004) 2284–2292.
[10] J. Jang, C.T. Sun, Functional equivalence between radial basis function networks and fuzzy inference systems, IEEE Trans. Neural Networks 4 (1993) 156–159.
[11] C.-F. Juang, C.-T. Lin, An on-line self-constructing neural fuzzy inference network and its applications, IEEE Trans. Fuzzy Syst. 6 (1) (1998) 12–32.
[12] J.M. Keller, M.R. Gray, J.A. Givens Jr., A fuzzy K-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. SMC-15 (4) (1985) 580–585.
[13] J.M. Keller, R. Krishnapuram, F.C.H. Rhee, Evidence aggregation networks for fuzzy logic inference, IEEE Trans. Neural Networks 3 (5) (1992) 761–769.
[14] E.T. Kim, et al., A new approach to fuzzy modeling, IEEE Trans. Fuzzy Syst. 5 (3) (1997) 328–337.
[15] C.C. Lee, Fuzzy logical in control systems: fuzzy logic controller, Part I, IEEE Trans. Syst. Man Cybern. 20 (2) (1990) 404–418.
[16] J.A. Leonard, M.A. Kramer, L.H. Ungar, Using radial basis functions to approximate a function and its bounds, IEEE Trans. Neural Networks 3 (1992) 624–627.
[17] C.J. Lin, An efficient immune-based symbiotic particle swarm optimization learning algorithm for TSK-type neuro-fuzzy networks design, Fuzzy Sets and Systems 159 (2008) 2890–2909.
[18] C.T. Lin, C.-S.G. Lee, Neural-network-based fuzzy logic control and decision system, IEEE Trans. Comput. 40 (12) (1991) 1320–1336.
[19] C.J. Lin, Y.J. Xu, A self-adaptive neural fuzzy network with group-based symbiotic evolution and its prediction applications, Fuzzy Sets and Systems 157 (8) (2006) 1036–1056.
[20] Z.Q. Liu, F. Yan, Fuzzy neural network in case-based diagnostic system, IEEE Trans. Fuzzy Syst. 5 (2) (1997) 209–222.
[21] J. Liu, M. Xu, Kernelized fuzzy attribute C-means clustering algorithm, Fuzzy Sets and Systems 159 (2008) 2428–2445.
[22] J.M. Mendel, Fuzzy logic systems for engineering: a tutorial, Proc. IEEE 83 (3) (1995) 345–377.
[23] J. Moody, C.J. Darken, Fast learning in networks of locally tuned processing units, Neural Comput. 1 (1989) 281–294.
[24] G.C. Mouzouris, J.M. Mendel, A single-value-QR decomposition based method for training fuzzy logic systems in uncertain environments, J. Intell. Fuzzy Syst. 55 (1997) 367–374.
[25] W. Pedrycz, Conditional fuzzy C-means, Pattern Recognition Lett. 17 (6) (1996) 625–632.
[26] W. Pedrycz, Conditional fuzzy clustering in the design of radial basis function neural networks, IEEE Trans. Neural Networks 9 (4) (1998) 601–612.
[27] W. Pedrycz, M. Reformat, Rule based modeling of nonlinear relationships, IEEE Trans. Fuzzy Syst. 5 (2) (1997) 256–269.
[28] W. Pedrycz, M. Reformat, Evolutionary fuzzy modeling, IEEE Trans. Fuzzy Syst. 11 (5) (2003) 652–665.
[29] J. Quinlan, Combining instance-based model and model-based learning, in: Proc. 10th ICML, San Matteo, CA, 1993, pp. 236–245.
[30] M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification, Fuzzy Sets and Systems 158 (2007) 2134–2152.
[31] A. Staiano, R. Tagliaferri, W. Pedrycz, Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering, Neurocomputing 69 (13–15) (2006) 1570–1581.
[32] C.-T. Sun, Rule base structure identification in an adaptive network based fuzzy inference system, IEEE Trans. Fuzzy Syst. 2 (1) (1994) 64–73.
[33] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Syst. Man Cybern. 15 (1985) 116–132.
[34] G. Tsekourasa, H. Sarimveisb, E. Kavaklia, G. Bafasb, A hierarchical fuzzy-clustering approach to fuzzy modeling, Fuzzy Sets and Systems 150 (2) (2005) 245–266.
[35] M.G. Tsipouras, T.P. Exarchos, D.I. Fotiadis, A methodology for automated fuzzy model generation, Fuzzy Sets and Systems 159 (2008) 3201–3220.

[36] B.A. Whitehead, T.D. Choate, Evolving space-filling curves to distribute radial basis functions over an input space, IEEE Trans. Neural Networks 5 (1994) 15–23.

