ISSN: 0974-8571 Vol.13 No. 2 December, 2021

International Journal of Computational Intelligence in Control

AHP FOR FEATURE WEIGHTED KNN WITH GRANGER CAUSALITY

Mr. Dayaker P
Research scholar at CSE department of Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India

Dr. Neeraj Sharma
Professor at CSE department of Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India

Date of Submission: 15th August 2021 Revised: 27th October 2021 Accepted: 18th December 2021

Abstract: The kNN (k- Nearest Neighbour) algorithm is a popular algorithm for pattern classification due to its non-parametric nature, easeto way implement, and the fact that its classification error is bound with the Bayes error. In this paper, we show that the performance of the During the pairwise comparison of features in a given dataset, the employment of (training) class-wise group statistics based on two criteria enhances the kNN classifier significantly. Granger causality is employed to assign preferences to each criterion. Analytic Hierarchy Process (AHP) is applied to obtain weights for different features from the two criteria and their preferences. Finally, these weights are used to build a weighted distance function
for the kNN classification. Experimentation on fifteen UCI Machine Learning Repository benchmark datasets clearly demonstrates the superiority of the proposed Granger causality-driven model.AHP launched the kNN algorithm over the kNN method across a wide range of distance metrics and various element selection strategies. This proposed approach is also shown to function effectively on datasets with high layers of the different datasets like face and handwritten recognitions

Index Terms: kNN algorithm, Bayes error, Analytic Hierarchy Process (AHP) Granger causality, Machine Learning.

INTRODUCTION

Non-parametric, simple to construct, and limited to two times the Bayes error, the kNN calculation [1, 2] remains a well-known choice for classifying designs. Robotized web use information mining [3, 4], large-scale data categorization [5, 17, 18], and hyperspectral picture age

classification [6, 7] are only a few of the sectors where kNN computing and its modifications are being employed quite late. When it comes to determining the number of nearest neighbours, the kNN classifier relies heavily on distance and similitude work [8, 9, 16] as well as information pre-handling such as component selection [11, 12, 13, 14] to improve its accuracy. It is impossible to compare one distance measure to another while considering all possible difficulties with equal chance, according to the concept of speculating power. Distance metrics such as Euclidean distance, L1-standard distance, 2 distance, and Mahalanob is distance[10, 15, 19, 20] are often used for classification without the need for learning. For the sake of the

Xing distance, Information-Theoretic Metric Learning (ITML), Kernel Relevant Component Analysis (KRCA), Large Margin Nearest Neighbour (LMNN), Information Geometric Metric Learning (IGML), and Kernel Information Geometric Metric Learning (KIGML)-based nonlinear metrics have all been used to improve speculation capacity. For high speculating power, the distance work in kNN should boost valuable characteristics of the space. This necessitates giving priority to the illuminating components of the job above those presenting trivial or redundant information. Highlight determination for ideal classification is generally a very difficult assignment [20, 21]. This may be done in two unique ways: As a first step, it's best to prioritize the importance of each feature based on its importance and relevance to the overall

classification. As an example of the aforesaid system in action, see the RELIEF algorithm[22, 23, 24]. After then, the redundant and repeated dimensions will be turned off. Prior to any categorization calculation, a subset of highlights from the original datasets is picked in the vast majority of element determination techniques. Data mishaps may have a negative impact on these small-scale element determination processes, despite the fact that they may be successful in dealing with the dimensionality problem. A sparse component selection method that takes advantage of an irregular subset is RSFS.

RELIEF [10, 23], and other key expansions to the RELIEF algorithm family of weighting techniques are also included. LFDA is a useful calculation when dealing with many inputs and outputs. To handle noisy and fragmented information, the RELIEF algorithm [23] is an effective and efficient tool. I-guiding RELIEF's principle is to consider the personality of an example and the nearest neighbours as if they were covered up irregular elements. I-RELIEF algorithm uses the Expectation-Maximization (EM) algorithm to alter the component loads in different cycles are the two most advanced forms of I-RELIEF, Iterative RELIEF-1. Anomaly, mislabelling, and insignificant components may be better addressed with these two computations. Rather than the vast majority of the notable group including choice and online component determination calculations, some new internet-based highlight choice (OFS) techniques utilize just a little and fixed number of traits/elements of preparing cases, which is exceptionally suitable for costly high layered datasets as well with respect to the sequentially streaming datasets like web-based spam email identification framework. To deal with the test of precise expectation utilizing a predetermined number of fixed dynamic elements, these web-based calculations take the assistance of thorough learning. They use a variety of regularisation methods to avoid over and under fitting. Both first-request learning and second-request learning are important internet learning algorithms. Web-based learning utilizes the basic designs amongst highlights in an indisputably superior way to first-request calculations, regardless of the first-request learning calculations. Adaptive Regularization of Weight Vectors (AROW), and Soft Confidence Weighted algorithm (SCW) are famous examples of second request learning computations. To be successful, these tactics can only be used in online frameworks, where planning is more essential than the necessity for correct pre-linguistic authority Despite these benefits, the combination of high accuracy and decreased time complexity is more tempting than any of these individual needs.

i) a bunch of models to assess a few other options and

ii) a slew of specific loads for each of these requirements, in order to compare the different alternatives. Human weighting for alternatives compared to individual standards and manual calculation of measurement inclinations have both been shown to create inconsistent judgments from AHP on occasion.

In this work, we show how a dataset's alternatives compare to its high points. In order to eliminate the need to manually weigh alternatives based on specific measurements, we use (instructional course) astute collecting insights to automatically determine the weights. According to the mean and standard deviation of the collected data, we propose two separate models. To avoid the difficulty of manually selecting specific weights for the (two) models, we use Granger causality to relegate the loads. Priority was first recognised as a factor in determining the causal link between double cross series. Design-based data transit has been expanded by the developers of to discover the causal relationship. As a result, we're forced to consider the role of causation in our current endeavour. Causation here demonstrates that the model regarded as "because" is intended to supply further information on the assumption viewed as "effect" to deliver. A substantial measure-based cooperative approach administers Granger causality in today's context.

MODELS FOR AHP AND HYDROLOGICAL DATA GENERATION

AHP is a multi-criteria decision-making technique that was developed in the 1980s.Making a choice (MCDM). AHP has been widely utilizedin a variety

The intensity of importance (Scale)	Description
1	The equal importance of the two elements
3	Moderate importance of one element over the other
5	Strong importance of one element over the other
7	Very strong importance of one element over the other
9	The extreme importance of one element over the other
2,4,6,8 are intensities to express intermediates values	

of domains. AHP technique, in general, may be summed up as follows:

The problem is organized in a hierarchy: aim, options, and solutions. In addition, there are criteria

for evaluating the choices in the structure Sub-criteria can be created from criteria. Priorities are set among the aspects of the plan. The pair wise comparisons of the criteria are arranged in a hierarchy. Throughout these comparisons, the significance of the scale is used to establish the criterion. Individuals' judgments or experiences on a topic are used to assign significance intensities. A comparison matrix aggregates all information collected from pair wise comparisons, determining the hierarchy's overall priorities or preferences. The normalized primary priority vector (Eigen) is used to make the final choice. The computed consistency ratio (CR) should not be higher than 10% (in percentage form) or less than 1 (in unit form) for the judgment to be consistent (valid) during AHP implementation. Otherwise, the decision is void. Consistency ratio CR is calculated by dividing the constant index CI by the random index RI.

Table 1 Linkert scale for pair wise comparisons

Eq. (1) computes the constancy ratio (CR) from the judgment matrix:

$$CR = \frac{CI}{RI} \ldots\ldots (1)$$

$$CI = \frac{\lambda_{MAX} - n}{n-1} \ldots\ldots (2)$$

Table 2 shows the RI values, which are dependent on the comparison matrix's dimension (n). MAX is the comparison matrix's maximal Eigen value.

Table 2. Random index(RI) expressed as a function size (n) of the comparison matrix

N	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45

AHP's use has recently been expanded to the field of hydrology and water resources, with AHP being utilized for stream flow/rainfall data infilling difficulties.

However, there is no literature in which AHP is used to choose non-parametric stochastic models for the creation of hydrological data (such as rainfall and stream flow). The current research is an attempt to make this specific choice. AHP can be tackled at many levels for the current study since it can handle complicated issues in a hierarchy.

Level 1: The hierarchy's major purpose is to choose an appropriate category for a non-parametric stochastic technique/model in the process of generating hydrological data for a specific location in a certain catchment.

THEORETICAL FOUNDATIONS

The logical Hierarchy Process (AHP) gets a bunch of sources of info or options for picking the most ideal choice. Table 1 shows the results of applying several rules based on the Saaty-scale to these possibilities. Various sub-models may also be offered under the underlying organization of criteria for the formation of judgment. Additionally, a few relative loads given by a comparable Saaty-scale are used to compare the different standards. It's at this point that the AHP results in a genuine placement of the options. There are four rules that AHP adheres to in order to function:

Maxim 1. Examining two possibilities A(i, j) vs a rule/substandard on a complementary percentage scale may be done by the leader.

i.e., A(j, i)= 1/A(i, j). (3)

Saying 2. Never does the leader deem one option superior to another in terms of a metric., i.e., A(i, j) $*= \infty$.

Saying 3. An orderly sequence may be devised for the choice problem.

Maxim 4. All standards/sub-rules which solely affect the given issue, and every one of the pertinent other options, are addressed in the pecking order in one go.

AHP is carried out through the accompanying three significant stages:

Stage 1. Calculation of the element rules framework.

Stage 2.Calculation of the standards special loads.

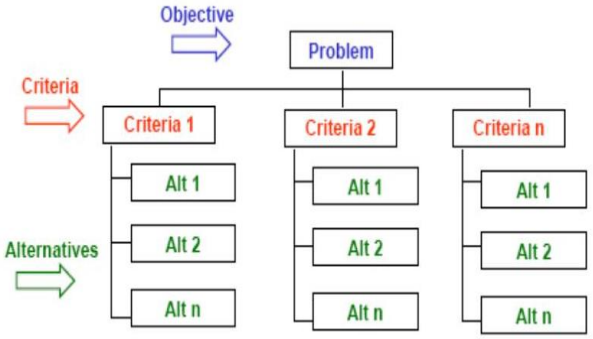
Stage 3.Positioning of the other options.

Weight vector calculation for a certain model in the model's special network P, NC NC is an element of the network. The inclination of the i^{th} model with respect to the jth basis is addressed in every section of P(i, j). This subsection's C(i, j) features are analogous to those of P(i, j). Our next step is to create the related standardized grid Pn by equating the number of segments' passes to 1.

PROPOSED METHOD

Classifying data using a collection of n well-defined and evenly dispersed sample points with d different features There are a total of N training points created from n samples, and these points are subsequently classified into M classes: CL1, CL2, •••,CLM. Points in the CLj class, where j is 1, 2,..., M and nj is the number of points in the class CLj. The aN d matrix X is used to represent these training data points.. x represents a collection of d-dimensional Nttest points, and the purpose is to categorize them. Our AHP-kNN technique may be shown with the help of two computations. Calculation 1 depicts the main phases of AHP-kNN and summons computation 2 for the calculation of the element loads W... Standardization of data occurs early on in computation 1. Calculation 2 takes care of the students' preparation and class information. To establish two standard networks, class-wise gathering measurements are taken for each component in C1 and C2.

Figure 4.1 Flow of AHP process



The components of the criteria grids address the pairwise correlation of any two elements for a particular measure. Two measures are intended to accomplish great Classification. The first criterion depends on bunch implies. Allow Xl to be the lth preparing point in the ith class, Ni. Xlj indicates the jth element of Xl. In this case, the ith class's jth-element mean may be calculated.

$$u_{ij} = \frac{1}{Ni} \sum_{l=1}^{Ni} X_{ij} \quad (4)$$

If the collection methods for different classes for a particular component are all around separated, it should be given high priority. In order to meet the first criteria, we aggregate the class-wise mean contrasts for each component in order to reduce burdens. The more valuable this aggregate is, the more attention it should get. The gathering mean for the ith class is μij, and the bunch mean for the lth class is μlj, as we will see in the jth highlight, respectively. It is thus possible to describe all M classes' total mean difference for the jth feature by using the following formula:

$$D_j = \sum_{i=1,l=1,i\neq 1}^{M} |u_{ij} - u_{lj}| \quad (5)$$

When the standard deviation of the collecting standard deviations is low (across all classes) and high (across all classes), the likelihood of a certain component is high. Let j be the standard deviation of the standard deviations gathered for the jth time. For the jth inclusion, the standard deviation-based burden is given by

$$S_j = \frac{\sigma_j}{\frac{\sum_{i=1}^{M} \sigma_{ij}}{M}} \quad (6)$$

Time-Complexity Analysis

B-fold partitioning has been used for cross-validation on n samples. Let N be the total number of training samples you'll be using. It's important to keep track of how many examples fall inside a certain area focus, and how many examples fall within a specific class. Now we have listed the worst-case scenario for our proposed computation in terms of time and complexity.

• Intricacy of standardization of n tests: O(n).

- Intricacy for weight change utilizing MOD-AHP:

There are M classes in the dataset, and Ni is the number of focuses within it.

- Intricacy for bunch mean estimation: O(MNi) ≈ O(N).

- Intricacy for bunch standard deviation estimation: O(MNilogNi)

Normalization and creation of criterion vectors from the matrix criteria takes time is O(d)+O(d) ≈ O(d)

– Granger-causality and the variation of criterion vectors make preference matrix generation more difficult.: O(d logd)+O(d2)

– Creating a vector of criterion preference preferences from a matrix of criteria preference preferences is difficult: O(NC)+O(NC) ≈ O(NC)

Now, for each of the samples tested,

– N-Sample Calculation of Distance Is Complex: O(N).

– The difficulty of sorting N training samples and estimating the similarity of k training samples: O(NlogN)+O(k).

FOR DATA GENERATION, AHP FORMULATIONAND IMPLEMENTATION

AHP is calculated as follows from Table 2:

1st level the objective is to find the finest nonparametric stochastic hydrological data generator possible (s).

Streamflow or rainfall is the data in this scenario. This choice will be made by the modeler, decision-maker, or water specialist. The AHP structure's highest level is Level 1.

2nd level As previously stated, there are five criteria: the ability to retain historic features (C1), ability to produce new hydrological data (C2), breadth of applicability (C3), existence of generated negative data (C4), and user friendliness (C5).

Level 3: When examining the criteria established at level 2, the decision maker or water expert/model user/developer should choose from among five kinds of non-parametric stochastic hydrological data generators or alternatives. Wavelet (A), reordering (B), closest neighbour (C), kernel density (D), and bootstrap (E) are the different types of models (E). The third rung of the hierarchy is the lowest.

Pairwise comparisons at level 2 and level 3 will be used to perform AHP for the purpose of generating hydrological data in order to meet the aim set at level 1.

EXPERIMENTAL RESULTS

Criteria weights

Table 3 shows the criterion weights obtained from Table 4 in the last column. Each row's weight is the average of the entries in that row. By dividing the sum in the final row of Table 6 by each unique entry, the table's six components are computed (in Table 4).The results in Table 6 demonstrate that criteria C2 and C5 have the highest (45.9%) and lowest (7.4%) preferences, respectively. It may be claimed that criterion C2 is 6 times more desired than criterion C5. As a result, the "capacity to create fresh hydrological data" will be more important to the water expert or modeler than the "scope of application." From a decision-perspective maker's

Table 3: Criteria weights

	C1	C2	C3	C4	C5	Criteria weight
C1	0.164	0.123	0.296	0.250	0.293	0.225
C2	0.658	0.493	0.370	0.333	0.439	0.459
C3	0.041	0.099	0.074	0.167	0.048	0.086
C4	0.054	0.123	0.037	0.083	0.073	0.074
C5	0.082	0.163	0.222	0.167	0.146	0.156
	1	1	1	1	1	1

Non-parametric stochastic hydrological data producers' relative weights

Alternative weights were calculated in the same way as criterion weights. Each element in the last column corresponds to the weights of options in each of these tables. Since non parametric stochastic models are weighted by dividing the entries (preferences) by the sum of the entries in each column, and then averaging each row, weights for criterion C1 (in Table) are derived as follows:

It has been tested on datasets from the UCI AI repository as well as datasets from. The LIBSVM website has provided us with some useful information. To the extent that it isn't too much bother, the datasets may be broken down into classes based on the aspect and number of instances they include. The classification execution of our technique is contrasted and

i) kNN utilizing different distance measurements,
ii) kNN utilizing unique include determination procedures and with

iii) some of the finest in class algorithms for selecting highlights. There are a number of large datasets that we've applied our method to as well.

The suggested method also works well on high layered face and handwriting acknowledgement datasets, as shown in this paper.We have confirmed that the Granger causality-based special weight determination for the two rules is realistic. This experiment involves shifting two distinct loads with a classification mistake, and the results are displayed in Figure 1 for two different datasets.As shown by the red dots, a modest scale mum blunder is achieved for both datasets using the same particular loads that were found in the Granger causality test.

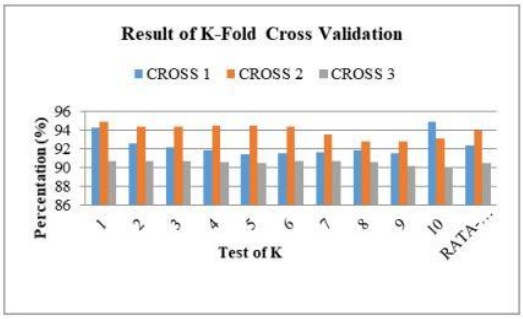


Fig 1:Accuracy Measurement Results

Table 1: *Comparison OfKnn And Mknn Accuracy*

C is a compromise between a standardization word and a calamity term if CW, AROW, SCW, SCW1, or SCW2 occur. C has a default value of 1 According to this group, a key boundary φ of the unfortunate work, i.e., the combined circulation function φ = φ−1(η) may be defined as the feature of the group of confidence-weighted learning calculations: In the second request computations, the boundary an is often used to introduce the covariance framework, i.e.∑ = a ∗ I, where I is a character network. Boundary an isn't very sensitive in most circumstances, and it's frequently anchored to 1. CW's default value for is 0.70, SCW's is 0.75, and SCW2's is 0.90, respectively. For M-CW a=1, η=0.75, for AROWC=1 and a=1, for M-SCW and M-SCW2 both C=1, a=1 and η=0.

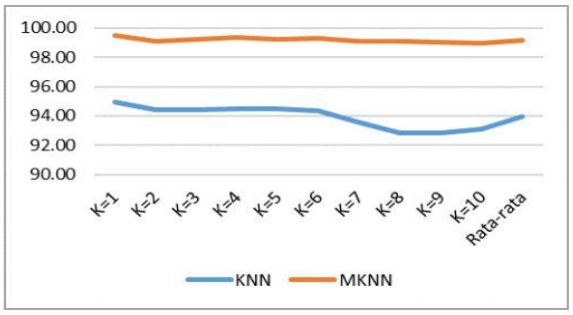


Fig 2: Accuracy Measurement Results

CONNCLUSION

AHP has shown to be a flexible tool since it can be used to a wide range of issues. For the first time, it has been used to the selection of non-parametric stochastic hydrological data production models. Pairwise comparisons were used to make the entire model selection process uniform and accessible. Bootstrap received the most support overall. There were no significant differences in the kernel density and wavelet categories from the bootstrap category. For the goal of generating hydrological data, the decision-maker, model user, or water expert might choose from a variety of methodologies. More work might be done to apply AHP to different types of stochastic models and additional criteria. To improve the presentation of traditional kNN, we use Granger causality and AHP. During the pairwise correlation of highlights, two models are used in light of instructional course-wise measures. Standards frameworks are given proper consideration using Granger causality. AHP is used to calculate the loads for each component. At last, these loads are utilized to fabricate a weighted dis-Similar examination of both MKNN and KNN was finished fully intent on knowing the precision capacity for order from the two calculations. And furthermore, to know the ideal information designs got from k-overlap cross approval into ideal information preparation and information testing the computation of exactness utilizes the guidelines of Confusion Matrix. A piece of decent informationdemonstrated in this paper was viewed as in cross 2 with an exactness of 93.945%. This displaying will be taken and utilized as preparing information and testing information to be tried in KNN and MKNNto break down the precision proportion with result that the most noteworthy precision of KNN was 94.95% with normal exactness during the test was 93.94% while MKNN's most noteworthy exactness was99.51% and normal exactness during the test was 99.20%. soit can be said the capacity of MKNN calculation is better as far as exactness with the distinction in precision by 5-7%.

Value K	KNN	MKNN
K=1	94.95	99.51
K=2	94.41	99.11
K=3	94.41	99.23
K=4	94.48	99.39
K=5	94.50	99.23
K=6	94.36	99.27
K=7	94.58	99.11
K=8	92.83	99.11
K=9	92.83	99.06
K=10	93.10	98.98
Average	93.94	99.20

REFERENCES

[1]T. M. Cover, P. E. Hart, Nearest neighbor pattern classification., IEEE Trans.Inf.Theory13(1967)21–27.

[2]Y.-C. Liaw, M.-L. Leou, C.-M. Wu, Fast exact k nearest neighborssearch using an orthogonal search tree, Pattern Recognition 43 (6) (2010) 2351 –2358.

[3]B. Yang, M. Xiang, Y. Zhang, Multi-manifold discriminant iso-map for visu-alizationandclassification,PatternRecognition55(2016)215–230.

[4]D.Adeniyi,Z.Wei,Y.Yongquan,Automatedwebusagedatamin ingand recommendation system using k-nearest neighbor (knn) classificationmethod,AppliedComputingandInformatics12(1)(20 16)90–108.

[5]J. Maillo, I. Triguero, F. Herrera, A MapReduce-based k-nearest neighbor approach for big data classification, in: IEEE TrustCom/BigDataSE/ISPA,Helsinki,Finland,Volume2,2015,pp. 167–172.

[6] J. Zou, W. Li, Q. Du, Sparse representation-based nearest neighbor class-fiersforhyperspectralimagery,IEEEGeosci.RemoteSensingLett.12 (12)(2015)2418–2422.

[7] W.Yang,C.Sun,L.Zhang, A multi-manifolddiscriminantanalysismethodforimagefeatureextraction,PatternRecognition44(8)(2011)1649–1657.

[8] P.Mitra,C.A.Murthy,S.K.Pal,Unsupervisedfeatureselectionusingfeaturesimilarity,IEEETrans.Patt.Anal.Mach.Intell.24(3)(2002)301–312.

[9] G. Bhattacharya,K.Ghosh,A. S. Chowdhury,An affinity-based new localdistancefunctionandsimilaritymeasureforknnalgorithm,Patt.Recog.Lett.33(3)(2012)356–363.

[10] W. Zheng, L. Zhao, C. Zou, Locally nearest neighbor classifiers for patternclassification,PatternRecognition37(6)(2004)1307–1309.

[11] G.Bhattacharya,K.Ghosh,A.Chowdhury,Testpointspecifickestimationfor knn classifier, in: Patt. Recog. (ICPR), 2014 22nd Int. Conf. on, 2014,pp.1478–1483.

[12] M. A. Tahir, J. E. Smith, Creating diverse nearest-neighbour ensembles us-ing simultaneous metaheuristic feature selection, Patt. Recog. Lett. 31 (11)(2010)1470–1480.

[13] C. Schaffer, A conservation law for generalization performance., in: W. W.Cohen,H.Hirsh(Eds.),ICML,MorganKaufmann,1994,pp.259–265.

[14] D.R.Wilson,T.R.Martinez,Improvedheterogeneousdistancefunctions,J.ofArtif.Intell.Research6(1)(1997)1–34.

[15] M.Liu,B.C.Vemuri,Arobustandefficientdoubly regularizedmetriclearn-ing approach, in: Proc. of the 12th European Conf. on Comput. Vis. - VolumePartIV,Springer-Verlag,Berlin,Heidelberg,2012,pp.646–659.

[16] P. Xie, E. P. Xing, Large scale distributed distance metric learning, CoRRabs/1412.5949.

[17] J. V. Davis, B. Kulis, P. Jain, S. Sra, I. S. Dhillon, Information-theoreticmetric learning, in:Proc. of 24th Int. Conf. on Mach. Learning (ICML),2007,pp.209–216.

[18] I.W.Tsang,P.mingCheung,J.T.Kwok,Kernelrelevantcomponentanalysis for distance metric learning, in: IEEE Int. Joint Conf. on Neural Networks(IJCNN,2005,pp.954–959.

[19] K. Q. Weinberger, J. Blitzer, L. K. Saul, Distance metric learning for largemarginnearestneighborclassification,in:NIPS,MITPress,2006.

[20] S. Wang, R. Jin, An information geometry approach for distance metriclearning, in: D. V. Dyk, M. Welling (Eds.), Proc. of the Twelfth Int. Conf.onArtif. Intell. and Statistics (AISTATS-09), Vol. 5, J. of Mach. LearningResearch-Proc.Track,2009,pp.591–598.

[21] J.M.Peña,R.Nilsson, On the complexity of discrete features election for optimal classification, IEEE Trans. Patt. Anal. Mach. Intell. 32 (8) (2010)1517–1522.

[22] K. Kira, L. A. Rendell, A practical approach to feature selection, MorganKaufmannPublishersInc.,pp.249–256.

[23] I.Kononenko,Estimatingattributes:Analysisandextensionsorelief,in:L.DeRaedt, and F. Bergadano (Eds.), Machine Learning: ECML-94, Springer-Verlag,1994,pp.171–182.

[24] I. Kononenko, E. Simec, M. Robnik-Sikonja, Overcoming the myopia of inductive learning algorithms with relief,AppliedIntell.7(1997)39–55.