

Improved Evidence Theoretic kNN Classifier based on Theory of Evidence

P.Umar Sathic Ali
 Research Scholar
 Bharathiar University, Coimbatore

Dr.C.Jothi Ventakeswaran
 Associate Professor & Head
 Presidency College, Chennai

ABSTRACT

The k-nearest neighbor rule is one of the simplest and most attractive pattern classification algorithms. However, it faces serious challenges when patterns of different classes overlap in some regions in the feature space. In the past, many researchers have developed various methods to improve its performance. In this paper, we propose an improved evidence theoretic kNN algorithm which combines Dempster Shafer theory of evidence and k nearest neighbouring rule with distance metric based neighborhood. It is shown that the proposed algorithm significantly improves the performance of the k-nearest neighbor rule. In experiments this algorithm performed better than voting, distance weighted and extended k nearest neighbours algorithms with best k, and it achieved highest performance when number of neighbours considered is seven.

Keywords

Dempster Shafer Theory, Nearest Neighbor Rule, Classification.

1. INTRODUCTION

Dempster Shafer Theory of Evidence [8] is widely accepted as a rich and flexible framework for representing and reasoning with imperfect information. Pattern Classification by the Distance Metric is one of the earliest concept in automatic pattern recognition. The k-Nearest Neighbour Rule is one of the most widely used pattern classification technique proposed by Fix and Hodges[6]. Cover and Hart[2] showed that under certain conditions k-NN method approaches the optimal Bayes error rate. Dudani[5] proposed a method to assign a weight to the nearest neighbors in the neighborhood.

Denoux[4] proposed an evidence theoretic k-NN method for classification based on Dempster Shafer Theory, in which each neighbor of a pattern to be classified is considered as a piece of evidence to support certain proposition concerning the class membership of a pattern. Based on the evidence, basic beliefs are assigned to the subset of all classes. Such basic belief assignments are obtained for each of k nearest neighbors and aggregated using the Dempster rule. It is well known that combining basic belief assignment is computationally expensive and it becomes impractical when the frame of discernment Ω has more than 15 to 20 elements.

Hai Hung and David Bell[11] proposed an alternative method for evidence theoretic classification to avoid the need of combination and the problem of choosing the best k for kNN. In this method, a single basic belief assignment is constructed for the neighbors. A classification rule was designed based on the basic belief assignment and it is known as extended kNN.

In extended kNN, the key issue was how neighborhoods were interpreted and selected. The Hypercube interpretation of neighborhood was adopted and neighbors were selected.

In this paper, we propose an evidence theoretic non-parametric algorithm for multivariate classification. Instead of using hypercube interpretation of neighborhood as used in

extended kNN, we adopt a distance metric based neighborhood. Each neighborhood is considered as a piece of evidence to support the class membership of the pattern to be classified. We call this improved evidence theoretic kNN. Experimental results are presented to show the competence of this algorithm.

2. DEMPSTER SHAFER THEORY

Dempster-Shafer Theory of Evidence is an extension of probability theory, which allows the representation of uncertainty and the combination of evidence. Dempster-Shafer Theory starts with the definition of all possible values, Ω , called a frame of discernment that a variable can take. An exact belief value is assigned to each subset of Ω and this represents the uncertainty that the value of the variable belongs to the set.

Let Ω be a finite set called frame of discernment. A mass function or basic belief assignment is a mapping $m : 2^\Omega \rightarrow [0,1]$ such that

$$\sum_{X \subseteq \Omega} m(X) = 1$$

$$m(\phi) = 0$$

The mass $m(X)$ measures the amount of belief that is exactly committed to X . $X \subseteq \Omega$ is called a focal element of m if $m(X) > 0$.

Given two mass functions m_1 and m_2 defined over the same Ω we can combine them using the Dempster Rule of Combination as follows:

$$m_1 \oplus m_2(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)}{1 - \sum_{X \cap Y = \phi} m_1(X) \cdot m_2(Y)} \quad (1)$$

The pignistic probability function [9] associated with m is $BetP: 2^\Omega \rightarrow [0,1]$ such that for any $E \subseteq \Omega$ is

$$BetP(E) = \sum_{X \subseteq \Omega} m(X) \frac{|E \cap X|}{|X|} \quad (2)$$

For $E_1, E_2 \subseteq \Omega$, we can define conditional pignistic probability $BetP(E_1|E_2)$ as follows, in a way similar to conditional (classical) probability:

$$BetP(E_1|E_2) = BetP(E_1 \cap E_2) / BetP(E_2) \quad (3)$$

3. K NEAREST NEIGHBOR RULE

The nearest neighbor (NN) rule, first proposed by Fix and Hodges [6], is one of the oldest and simplest pattern classification algorithms. Given a set of n labeled examples

$D_n = \{(\vec{X}_1, Y_1), (\vec{X}_2, Y_2), \dots, (\vec{X}_n, Y_n)\}$ with input vectors $X_i \in \mathbb{R}^d$ and class labels $Y_i = \{c_1, c_2, \dots, c_m\}$, the NN rule classifies an unseen pattern \vec{X} to the class of its nearest neighbor in the training data D_n . To identify the nearest neighbor of a query pattern, a distance function has to be defined to measure the similarity between two patterns. The basic rationale for the NN rule is both simple and intuitive: patterns close in the input space \mathbb{R}^d are likely to belong to the same class.

kNN is popular in pattern recognition community mainly due to its good performance and its simple-to-use feature. Since the inception of kNN some variations have been proposed in order to improve its performance.

3.1 Distance weighted kNN rule

In voting kNN the k neighbours are implicitly assumed to have equal weight in decision, regardless of their distances from the pattern x to be classified. It is intuitively appealing to give different weights to the k neighbours based on their distances from x , with closer neighbours having greater weights. Let d be a distance measure, and x_1, x_2, \dots, x_k be the k nearest neighbours of x arranged in increasing order of $d(x_i, x)$. So x_1 is the first nearest neighbour of x . [5] proposes to assign to the i^{th} nearest neighbour x_i a weight w_i defined as

$$w_i = \begin{cases} \frac{d(x_k, x) - d(x_i, x)}{d(x_k, x) - d(x_1, x)}, & \text{if } d(x_k, x) \neq d(x_1, x) \\ 1 & \text{if } d(x_k, x) = d(x_1, x) \end{cases} \quad (4)$$

Pattern x is assigned to the class for which the weights of the representatives among the k nearest neighbours sum to the greatest value. This rule was shown by Dudani to yield lower error rates than those obtained using the voting kNN rule. [5] provides an excellent and detailed review of distance weighted kNN.

3.2 Evidence theoretic kNN rule

The evidence theoretic k -nearest neighbour rule [4] is a pattern classification method based on the Dempster-Shafer theory of belief functions. In this approach, each neighbor of a pattern to be classified is considered as an item of evidence supporting certain hypotheses concerning the class membership of that pattern. Based on this evidence, basic belief masses are assigned to each subset of the set of classes. Such masses are obtained for each of the k nearest neighbours of the pattern under consideration and aggregated using the Dempster's rule of combination. In [12] Zouhal and Denoeux state "... in many situations, this method was found experimentally to yield lower error rates than other methods using the same information". They then proposed an optimization procedure to determine optimal or near-optimal parameter values from the data by minimizing an error function. This refinement of the original method is shown experimentally to result in substantial improvement of classification accuracy.

3.3 Extended evidence theoretic kNN rule

Extended evidence theoretic kNN rule [11] is an alternative to evidence theoretic kNN. In this approach, multiple neighborhoods are computed and each of which provides a source of evidence supporting the proposition concerning the class membership of the given pattern. Here, the key issue is

how neighborhoods are interpreted and selected. Obviously there are many possible interpretations of neighborhood and neighborhood selection strategies. It adapts a hypercube representation of neighborhood along with simple selection strategy. It is assumed that the attributes in the dataset are all numerical. For a positive integer d , every attribute is partitioned into $d + 1$ equal-sized interval. This effectively gives equal weights to all attributes. Consider an attribute A , and let $\text{dom}(A)$ be its domain. The intervals are arranged in ascending order as $v_0, v_1, v_2, \dots, v_d$ such that for $x \in v_i, y \in v_j, x \leq y$ if $i \leq j$. Then every value of the attribute belongs to one and only one interval. For a value a of A , i.e., $a \in \text{dom}(A)$, let v_i be the interval that a belongs to. For a non-negative integer $q \leq d$ we call the following extended interval the q^{th} order interval of a , written as $V_A^q(a)$:

$$V_A^q(a) = \{x \in \text{dom}(A) : x \leq \max V_{[i+q]_0^d}, x \geq \min V_{[i-q]_0^d}\}$$

$$\text{Where } [n]_0^d = \begin{cases} 0, & \text{if } n < 0 \\ d, & \text{if } n > d \\ d, & \text{otherwise} \end{cases}$$

$$\text{Clearly } V_A^0(a) = V_1$$

For a data vector (tuple) t , its q^{th} order hypercube is

$$V^q(t) = \prod_A V_A^q(t(A))$$

where $t(A)$ is the projection of t to attribute A . Furthermore we let $\text{cov}(V^q(t))$ be the coverage of $V^q(t)$, i.e., the number of data instances in the hypercube.

We take each $V^q(t)$ as a neighbourhood of t , and so we have $d+1$ neighborhoods for $t: V^0(t), V^1(t), \dots, V^d(t)$.

Clearly $\text{cov}(V^i(t)) \subseteq \text{cov}(V^j(t))$ if $i \leq j$.

Neighbourhood selection strategy is as follows: for a given h , it consider h neighborhoods i NN for $i = 1, 2, \dots, h$ where 1NN is a non-empty $V^q(t)$ with the smallest q , and 2NN is $V^{q+1}(t)$, and so on.

4. DISTANCE METRICS

A non-negative function $d(x, y)$ describing the distance between neighboring points x and y , constitutes a metric [1]. A metric space is then a set possessing a metric. In general, a metric space is formed by a set of valid objects with a global distance function (the metric d) which, for every two point $x, y \in S$ gives the distance between them as a nonnegative real number $d(x, y)$. A finite subset of set that we could call X with size $n = |X|$ is the search of objects where we search. Then function $d(x, y)$ can also be expressed as $d: S \times S \rightarrow \mathbb{R}$. The smaller the distance $d(x, y)$, the closer x is from y . For a metric to be considered as such, it must satisfy:

$$(1) \forall x, y \in S, d(x, y) \geq 0$$

- (II) $\forall x, y \in S, d(x, y) = d(y, x)$
- (III) $\forall x, y \in S, d(x, x) = 0$
- (IV) $\forall x, y \in S, x \neq y \Rightarrow d(x, y) > 0$
- (V) $\forall x, y, z \in S, d(x, y) \leq d(x, z) + d(z, y)$

If the distance does not satisfy the strict positiveness property (IV), then the space is called a *pseudo-metric*. Also, in some cases property (II) does not hold. It then receives the name of *quasi-metric*. The above axioms express intuitive notions about the concept of distance: distances between different objects are positive and the distance between x and y is the same as the distance between y and x . The triangle inequality means roughly that the distance from x to z to y is never shorter than going directly from x to y . Some typical distance functions used in distance calculations are shown in the equations 5, 6 and 7. These most used metric belong to the Minkowski family of distances also known as the *L metric*. Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

Manhattan or City-block

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \tag{6}$$

Chebychev distance:

$$d(x, y) = \max_{i=1}^n |x_i - y_i| \tag{7}$$

5. PROPOSED ALGORITHM

We propose an algorithm which extends the standard majority voting kNN by evidence theory for multivariate S class ($S \geq 1$). There are two motivations behind this algorithm. First, each neighborhood of an unknown pattern t (to be classified) as well as the Pattern together provides some evidence supporting the class membership of that pattern. Hence the aggregation of such evidences using Dempster Shafer Theory is expected to result in a good performance of the algorithm. Second, we use the most popular Minkowski family of distance metrics for neighborhood computation. Hence it is expected that it will boost the accuracy of the algorithm substantially.

Let $D = \{(t_i, c_i) : t_i \in V, c_i \in W, \text{ where } i = 1, 2, \dots, n\}$ be the training set, t_i is a vector of d attributes or feature whose domain is a relation $V = \text{dom}(x_1) \times \text{dom}(x_2) \dots \times \text{dom}(x_d)$, $W = \{c_1, c_2, \dots, c_m\}$ be the set of classes and t be an incoming sample to be classified based on D . Let $E = \{E_1, E_2, \dots, E_n\}$ be the set of neighborhoods and each neighborhood is a region in $\Omega(V)$ covering a set of neighbors of t . We take E_i as set of neighbors so $E_i \in 2^{\Omega}$. Figure 1 shows an input pattern t covered by three different neighborhoods. These neighborhoods are obtained by the distance metric

L_1, L_2 and L_∞ which are then taken as distinct source of evidence in classifying an input pattern t .

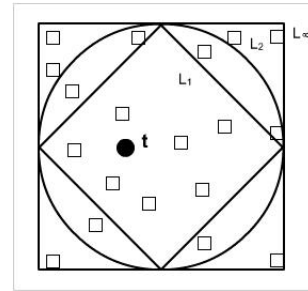


Figure 1. An example of pattern t in an overlapped neighborhoods L_1, L_2 and L_∞

The neighborhood provides a source of evidence supporting proposition concerning the class membership of t . Each neighborhood is taken as one part of source of evidence and all neighborhoods – together as a source of evidence – are used to generate a single mass function representing partial support by different neighborhoods.

Consider $E_i \in 2^{\Omega}$ and $c \in W$. We are interested in the joint probability $P(E_i, c)$ – the probability that a randomly selected element x of Ω belongs to E_i and is in class c , i.e., $x \in E_i$ and $f(x) = c$. Having no specific knowledge about the distribution p we can apply the principle of indifference to approximate $P(E_i, c)$ by

$$\bar{P}(E_i, c) = |E_i^c| / |D| \tag{8}$$

Where $E_i^c = \{x \in E_i : f(x) = c\}$

Then we define a function $m[t]$, induced for t from the h neighborhoods, as a mapping $m[t] : 2^{\Omega} \rightarrow [0, 1]$ such that, for $X \in 2^{\Omega}$ and $c \in W$,

$$m[t](X, c) = \begin{cases} \frac{\bar{P}(X, c)}{K}, & \text{if } X = E_i \text{ for some } i \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Here K is a normalizing factor. It follows that $K = \sum_{i=1}^n \sum_{c \in W} \bar{P}(E_i, c)$. Note that by $m[t](X, c)$ we mean $m[t](X \cap \{x \in \Omega : f(x) = c\})$, which is similar to the interpretation of joint probability $P(X, c)$. Clearly $m[t]$ is a mass function. In particular $\sum_{c \in W} \sum_{X \in 2^{\Omega}} m[t](X, c) = \sum_{c \in W} \sum_{i=1}^n m[t](E_i, c) = 1$

We propose to classify new patterns through conditional pignistic probability. For this we specify the joint pignistic probability as $\overline{\text{BetP}} : 2^{\Omega} \rightarrow [0, 1]$ such that, for $X \in 2^{\Omega}$ and $c \in W$,

$$\overline{\text{BetP}}(X, c) = \sum_{i=1}^n m[t](E_i, c) \times \frac{|X \cap E_i|}{|E_i|} \tag{10}$$

Note that E_i is a region covering some neighbours of t so we have $t \in E_i$. We can understand t as a singleton set, therefore $t \cap E_i^c = \emptyset$ and $|\{t\}| = 1$. Then we have the following joint, marginal and conditional pignistic probabilities for $t \in \Omega$,

$$\begin{aligned} \overline{\text{BetP}}(t, c) &= \sum_{i=1}^n m[t](E_i, c) / |E_i| \\ \overline{\text{BetP}}(t) &= \sum_{c \in W} \overline{\text{BetP}}(t, c) \end{aligned}$$

$$\begin{aligned} \overline{BetP}(c|t) &= \overline{BetP}(t, c) / \overline{BetP}(t) \\ &= \frac{\sum_{i=1}^h m[t](E_i, c) / |E_i|}{\overline{BetP}(t)} = \frac{\sum_{i=1}^h P(E_i, c) / |E_i|}{K \times \overline{BetP}(t)} \\ &= \frac{\sum_{i=1}^h |E_i^c| / |E_i|}{|D| \times K \times \overline{BetP}(t)} = \delta(t) \sum_{i=1}^h |E_i^c| / |E_i| \\ \text{where } \delta^{-1} &= |D| \times K \times \overline{BetP}(t). \end{aligned}$$

Classification then proceeds using the following rule:

Rule1.

For $x \in \Omega, f(x) = c, \text{ if } \overline{BetP}(c, |x) \geq \overline{BetP}(c|x) \quad \forall c \in W.$

If $h=1$ and we use Rule 1 for classification, then $\overline{BetP}(c/t) = \delta(t) |\{x \in E_1; f(x) = c\}| / |E_1|$ and we end up with a majority voting kNN. Therefore Rule 1 is an extended majority voting based kNN. The proposed algorithm is given in the following page.

6. EVALUATION

The Evaluation was done via experiments. The purpose of the evaluation is to show if and how the classification procedure improves upon the majority voting kNN, the distance weighted kNN and extended kNN. The data used in the experiments are public Dataset from UC Irvine Machine Learning Repository.

In our experiment, we set the number of neighborhood $h=4$, the neighborhoods E_1, E_2, E_3 and E_4 are obtained using the distance metrics Euclidean, Manhattan, Chebychev and Canberra respectively and recorded the classification accuracy. As a comparison, we implemented the voting based kNN classifier, the distance weighted kNN classifier and extended kNN and experimented with various values of k (from 1 to 10) and recorded the classification accuracy for each of the k values.

Throughout the experiment the validation method is 10 fold cross validation. The results for improved kNN, majority voting kNN, distance weighted kNN and extended kNN are shown in the Tables 1, 2, 3 and 4 respectively. From the results, we observe that on average improved kNN performed better than majority voting kNN, distance weighted kNN and extended kNN on these dataset. The performance of the improved kNN did not change much with different values of $k \geq 2$. Such a saturation property is useful since it relieves the designer of the kNN from the burden of searching for the optimal value of k . The highest performance is reported when the number of neighbors considered is 7. Figure 2 shows the average performance of our improved kNN with voting kNN, distance weighted kNN and extended kNN as a function of k over all datasets.

Algorithm 1 An Improved Evidence Theoretic KNN

Input:

$D = \{(t_i, c_i) : t_i \in V, c_i \in W, \text{ where } i = 1, 2, \dots, n\}$
be the training set, where t_i is a vector of d attributes or features whose domain is a relation
 $V = \text{dom}(x_1) \times \text{dom}(x_2) \dots \times \text{dom}(x_d)$ c_i is a class variable whose domain is finite set
 $W = \{c_1, c_2, \dots, c_m\}$

Unknown pattern $t = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$

Process:

for $i = 1$ to h do
Compute the neighborhood E_i with distinct distance metric for the unknown pattern t using the training set D .

end for

for $i = 1$ to h do

for $j = 1$ to m do

Set $m[t](E_i, c_j) = \overline{P}(E_i, c_j) / K$

where $\overline{P}(E_i, c_j) = |E_i^{c_j}| / |D|$

$K = \sum_{i=1}^h \sum_{c \in W} P(E_i, c)$

end for

end for

for $j = 1$ to m do

Set $\overline{BetP}(t, c_j) = \sum_{i=1}^h m[t](E_i, c_j) / |E_i|$

end for

Set $\overline{BetP}(t) = \sum_{c \in W} \overline{BetP}(t, c)$

for $j = 1$ to m do

Set $\overline{BetP}(c_j|t) = \delta(t) \sum_{i=1}^h |E_i^{c_j}| / |E_i|$

where $\delta^{-1} =$

$|D| \times K \times \overline{BetP}(t)$

end for

Set $C = \underset{j}{\text{argmax}} \overline{BetP}(c_j|t)$

Output: Assign t to class C

7. CONCLUSION

Based on the conceptual framework of Dempster Shafer Theory and kNN rule, a new non parametric classification algorithm adopting a simple and efficient neighborhood selection strategy has been proposed. To classify the pattern, the algorithm considers several neighborhoods, each of which is a set of neighbors. The neighbors are taken as a single source of evidence supporting the propositions concerning the class membership of the pattern. This evidence is represented as single mass function in order to quantify the uncertainty attached to the class membership of that pattern.

Table 1. The classification accuracy of our improved kNN algorithm

Dataset	Min		Max		Avg.	Samples				
	Accu	K	Accu	K		Accu	K=1	K=3	K=5	K=7
Cancer	92.11	1	94.39	5	93.77	92.11	93.51	94.39	94.04	93.86
Diabetes	70.13	1	76.10	10	73.57	70.13	72.08	74.16	74.68	74.94
Ionosphere	87.14	10	90.29	3	89.17	89.14	90.29	89.43	89.43	87.71
Iris	94.66	4-5	95.33	1-3, 6-10	95.20	95.33	95.33	94.66	95.33	95.33
OCR	97.95	10	98.47	2	98.20	98.35	98.31	98.27	98.27	98.08

Table 2. The classification accuracy of majority voting kNN algorithm

Dataset	Min		Max		Avg.	Samples				
	Accu	K	Accu	K		Accu	K=1	K=3	K=5	K=7
Cancer	90.88	2	93.16	10	92.23	91.23	92.28	92.63	92.46	92.63

Diabetes	65.06	2,7,9,10	65.19	3-6	65.14	65.19	65.19	65.19	65.06	65.06
Ionosphere	80.57	2	84.57	1	82.14	84.57	83.43	82.86	82.00	82.86
Iris	92.66	10	96.00	1,3,5	94.80	96.00	96.00	96.00	95.33	94.00
OCR	98.06	2	98.49	1	98.24	98.49	98.42	98.29	98.33	98.26

Table 3. The classification accuracy of distance weighted kNN algorithm

Dataset	Min		Max		Avg. Accu	Samples				
	Accu.	K	Accu.	K		K=1	K=3	K=5	K=7	K=9
Cancer	91.23	1,2	92.28	4,5,10	91.96	91.23	91.93	92.28	91.93	91.93
Diabetes	67.79	1,2	73.90	10	71.48	67.79	70.65	72.47	73.38	72.73
Ionosphere	83.14	7,8	84.57	1,2,4	83.63	84.57	84.29	83.14	83.14	83.14
Iris	95.33	9,10	97.33	5,7	96.20	96.00	96.67	97.33	97.33	95.33
OCR	98.35	9	98.65	6	98.48	98.49	98.47	98.38	98.56	98.35

Table 4. The classification accuracy of extended kNN algorithm

Dataset	Min		Max		Avg. Accu	Samples				
	Accu.	K	Accu.	K		K=1	K=3	K=5	K=7	K=9
Cancer	92.50	1-1	92.50	1-1	92.50	92.50	92.50	92.50	92.50	92.50
Diabetes	71.43	10	74.16	5	72.84	72.60	73.25	74.16	72.73	71.82
Ionosphere	81.14	10	84.00	1-3	82.60	84.00	84.00	82.86	81.71	81.43
Iris	90.66	9,10	94.00	1,2	92.60	94.00	93.33	93.33	92.00	90.66
OCR	95.98	10	96.98	1	96.40	96.98	96.69	96.41	96.23	96.00

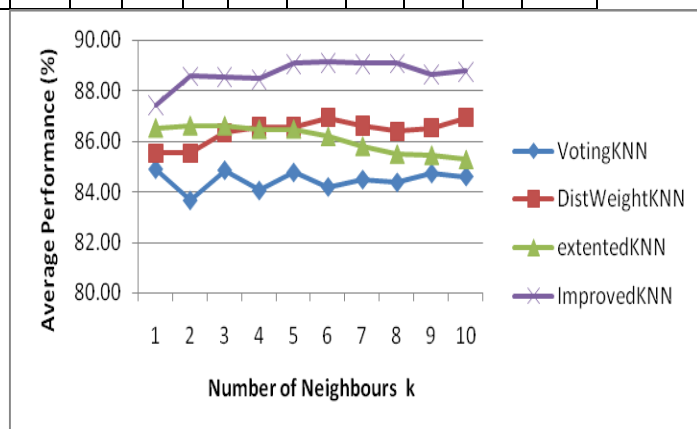


Figure 2. Average Performance of all four kNN over all datasets as a function of k.

This Classifier is different from the extended kNN in that it doesn't adopt the hypercube representation of neighborhood and also different from TBM classifier in that it doesn't use the time consuming Dempster rule of combination to aggregate mass functions for classification. *Transactions on Systems, Man and Cybernetics*, 28,263–271.

In experiments using real world datasets, the classifier outperformed on average the majority voting kNN, distance weighted kNN and extended kNN.

In this investigation, we have used the most popular distance metrics for the neighborhood computation. It will be more interesting to use more distance metrics and their combination so that it can deal with uncertainty about the class membership of the data. We leave it for future investigation.

8. REFERENCES

- [1] Chavez, E., Navarro, G., Baeza-Yates, R., Marroquin, J. L. 2001 Searching in Metric Spaces, *ACM Computer Surveys*, Vol. 33:3, pp. 273-321.
- [2] Cover, T. M. and Hart, P. E. 1967. Nearest neighbour pattern classification. *IEEE Trans. Inform. Theory*, IT-13, 21–27
- [3] Cunningham, P., and Sarah Jane Delany, 2007: k-Nearest Neighbour Classifiers, Technical Report UCD-CSI-2007-4.
- [4] Denoeux, T. 1995: A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25, 804–813.
- [5] Dudani, S. A. 1976. The distance-weighted k-nearest-neighbor rule. *IEEE Trans.Syst. Man Cyber.*, 6, 325–327.
- [6] Fix, E., and Hodges, J.L. 1951. Nonparametric discrimination: consistency properties, USAF School Aviation Medicine, Randolph Field, TX, Tech. Rep. 4.
- [7] Pal, N.R., and Susmita Ghosh. 2001. Some Classification Algorithms Integrating Dempster–Shafer Theory of Evidence with the Rank Nearest Neighbor Rules, *IEEE Transactions on Systems, Man, and Cybernetics— Part A: Systems And Applications*, Vol. 31, PP 59-66.
- [8] Shafer, G. 1976. *A mathematical theory of evidence*. Princeton University Press, Princeton, New Jersey.
- [9] Smets, P. and Kennes, R. 1994. The transferable belief model. *Artificial Intelligence*, 66,191–234
- [10] Wang, J., Neskovic, P., Cooper, L.N. 2007: Improving the Nearest Neighbor rule with a simple adaptive distance measure. *Pattern Recognition* 28, 207–213.
- [11] Wang, H., and David Bell. 2004. Extended k-Nearest Neighbours based on Evidence Theory, *The Computer Journal*, vol 47, pp 662–672
- [12] Zouhal, L. M. and Denoeux, T. 1998. An evidence-theoretic k-nn rule with Parameter optimization. *IEEE*