A Novel K-Nearest Neighbor Algorithm Based on I-Divergence with Application to Soil Moisture Estimation in Maize Field

Xiangyan Meng [1,2], Zhongxue Zhang [1,*], and Xinying Xu [2]
[1] College of Water Conservancy and Construction, Northeast Agricultural University, Harbin, China
[2] College of Science, Northeast Agricultural University, Harbin, China
*Email: zhongxuezh@163.com

Abstract— **This paper proposes a novel k-nearest neighbor algorithm to predict soil moisture in maize field. In order to estimate soil moisture in maize field accurately without any destruction to root and soil, this paper uses biological characteristics of maize to estimate soil moisture, including plant height, leaf area, stem diameter, dry weight and fresh weight, all the values of which are non-negative. So a novel k-nearest neighbor based on I-divergence (ID_KNN) is proposed. ID_KNN uses I-divergence as the distance metric instead of Euclidean distance, which is more effective when the data is positive. The proposed method is tested on datasets in six growth stages of maize, and the experimental results show that ID_KNN is more effective in accuracy and macro F_1 measure than traditional k-nearest neighbor algorithm.**

***Index Terms*—soil moisture, k-nearest neighbor, distance metric, I-divergence**

I. Introduction

Maize is one of the most important cereal crops, which is not only a kind of food for human and livestock, but also an industry material [1]. However, maize production is affected by diverse factors, such as sunshine, soil moisture, diseases, fertilizer etc., in which soil moisture is a key, indispensable and basic factor of the maize growth [2]. Water is an important factor of photosynthesis and transmission of, so only with enough water, can maize create, transport and absorb necessary nutrients for survival and growth. On the contrary, under drought conditions, water stress adversely affects maize plant in morphology, physiological and ecological, and decreases the yield of maize eventually [3]. Specifically, in morphology and ecology, soil moisture deficit results in incomplete development of morphological characteristics such as height, color, leaf area, stem diameter and root hair distribution etc.; in physiology, soil moisture deficit adversely affects photosynthetic rate, absorption and transmission of nutrition and minerals, as well as the growth speed [4]. Therefore, it is very important to measure accurately soil moisture for growth and yield of maize.

Nowadays, there are many methods to measure soil moisture, which can be divided into two classes: (i) direct methods and (ii) indirect methods [5]. Gravimetric method is the only direct method, and the others belong to indirect methods, such as neutron probe technology, Time Domain Reflectometry (TDR), gypsum block measurement, remote sensing etc.. Gravimetric method is widely used for estimation of soil moisture, but it costs much labor and time, it is also destructive for soil and root. Neutron probe technology requires a well trained operator and the equipment is harmful to health; furthermore, it lacks of depth resolution [6]. In TDR, the probe is sensitive to soil and salt, which can affect the measurement accuracy; the TDR equipment is very expensive. In gypsum block measurement, gypsum block will degrade gradually and need for recalibration with time, it is also sensitive to salt and temperature [7]. Remote sensing is expensive and complex with the restriction of brightness, temperature and soil moisture. It is suitable to large areas with appropriate brightness and temperature.

The methods above are restricted greatly during application and difficult to achieve the expected results, so a non-destructive method using discriminative learning methods to measure the soil moisture is proposed according to morphology and ecology characteristics of maize in this paper. In all the discriminative learning methods, K-Nearest Neighbor (KNN) is a simple, effective and nonparametric method, which is also a machine learning algorithm without building any generative models. Additionally, KNN has been used in many areas including text classification, pattern recognition and biometric etc.. In order to predict soil moisture accurately without any destruction to root and soil, biological characteristics (plant height, leaf area, stem diameter, dry weight, and fresh weight) are regarded as features of the sample, which constitute feature vectors, and discrete soil moisture contents are regarded as class labels. Therefore, KNN may be used to predict soil moisture according to biological characteristics.

The basic idea of KNN is to classify the test sample into some class according to its nearest neighbors. By this idea, number of its neighbors belonging to this class is maximum in all classes. In the estimation of soil moisture, each training sample is composed of biological features and soil moisture. The training dataset are composed of

doi:10.4304/jsw.9.4.841-846

the sample with known soil moisture. A test sample is classified, if the test sample is assigned to a certain class, the soil moisture of the test sample can be known. So we use KNN to estimate soil moisture.

However, traditional k-nearest neighbor (TR_KNN) is used to predict soil moisture. This easily leads to some defects as follows [8]:

(i) The complexity of computing is huge when the capacity of samples is large;

(ii) The performance is easily affected by feature weights;

(iii) Classification speed is low;

(iv) There is no certain way to determine the value of parameter k except repeatedly adjust it by experiment.

One of the most important reasons for the shortcomings of TR_KNN is the Euclidean distance used as the distance metric. So many researchers have improved the TR_KNN in distance metric, such as Han et.al.[9] and Jahromi et.al. [10]. Most of the improved methods are on the basis of Euclidean distance, which considers all the features have same weight in classification[11-12]. However, the importance of each feature is not the same with the others, so weight coefficients are introduced, which will be affected by subjective factors. Therefore, we introduce information divergence (I-divergence) as the distance metric instead of Euclidean distance in TR_KNN.

In this paper, we propose a novel KNN method by combining TR_KNN and I-divergence (ID_KNN), which can keep both the advantages of KNN and I-divergence. Some typical characteristics of maize are chosen as the classification features, such as maize height, fresh weight, dry weight, stem diameter and leaf area, the values of which are all non-negative, besides, I-divergence is effective when the values are positive, so I-divergence is suitable to be used in estimation of soil moisture.

The rest of the paper is arranged as follows: TR_KNN is introduced and a novel k-nearest neighbor based on I-divergence is proposed in Section **2**; experiments on estimation of soil moisture with proposed KNN method are carried out in Section **3**; the conclusion is shown in Section **4**.

II. Novel K-nearest Neighbor Algorithm Based on I-divergence

A. Traditional K-nearest Neighbor

KNN was proposed to text categorization by Cover and Hart in 1968 [13]. Nowadays, it has been shown that KNN is very effective to a variety of problems when related models are not known. The details of KNN algorithm are described as follows [14]:

Step 1: Given the training sample set and the test sample set. The training sample set is denoted by Ω such that $\{(x_i,c_i)|i=1,2,\cdots,n\}$, where $x_i=\left(x_i^1,x_i^2,\cdots,x_i^l\right)$ is a l-dimensional vector; l is the number of features, in which x_i^j denotes the jth feature component of the ith training sample; c_i denotes the corresponding class of the ith sample, and label c_i belongs to label set C such that $C=\{1,2,\cdots,t\}$, that is, the number of class is t. The test sample set is denoted by Φ such that $\Phi=\{y_j\,|\,j=1,2,\cdots,m\}$, where $y_j=\left(y_j^1,y_j^2,\cdots,y_j^l\right)$, in which y_j^i denotes the ith feature component of the jth test sample;

Step 2: Determine k value.

Step 3: Calculate the Euclidean distance between test sample and each training sample. Euclidean distance is denoted by:

$$d\left(x_i,y_j\right)=\sqrt{\sum_{k=1}^{l}\left(x_i^k\text{-}y_j^k\right)^2} \tag{1}$$

Step 4: Determine k nearest neighbors. Sort the distances in ascending order, and select k samples with relative minimum distances;

Step 5: Find the dominant class: let the k nearest neighbors be $x_1',x_2',\cdots,x_k'$, and the corresponding class labels be $c_1',c_2',\cdots,c_k'$ that belong to the label set C .The queried test sample is classified according to the classes of k nearest neighbors by means of maximum probability. The probability of each class appearing in k nearest neighbors is calculated as the number of the class appearing in k nearest neighbors divided by k. And the class with maximum probability is the dominant class. Let $S=\{s_1,s_2,s_3,\cdots,s_k\}$ be the set of number for each class in k nearest neighbors. The details are described as follows:

$$\tau^*=\arg\max_{\tau\in C}\left(s_\tau/k\right) \tag{2}$$

Step 6: Assign y_j to the class τ^* .

B. Improvement Principles on K-nearest Neighbor

In TR_KNN, Euclidean distance is used as the distance metric between different samples, the result of which is that all the feature components have the same weight and contribution to classification. However, in fact, the weights of feature components are mostly different. So, many researchers set different weight for feature components according to their importance [15-16]. However, the principals of determining the weight coefficients are not uniform, which inevitably involve subjective factors in classification [17]. For selection of similarity measure, Csiszár identified all the similarity measures that are both consistent with the axioms he states and lead to good selection methods for use in solving divergence optimization problems [18]. He concludes that there are only two selection procedures of choice: if the functions involved in vector space are all real valued, having both positive and negative values, then minimizing the least squares measure is the only consistent choice; whereas, if all the functions are required to be nonnegative, then minimizing I-divergence is the only consistent choice. In estimation of soil moisture in maize field, all involved variables of biological characteristics (plant height, leaf area, stem diameter, dry weight, and fresh weight) are nonnegative. So, we introduce I-divergence as the distance metric in KNN to overcome this problem.

The initial I-divergence was proposed by Kullback and Leibler [19]. It was usually used in positive, linear inverse problem, and it is a kind of distance metric showing the difference between measured value and true value. In this paper, we introduce the I-divergence as the distance metric instead of Euclidean distance in KNN. For two nonnegative numbers p and q , the I-divergence is defined as follow [9]:

$$D(p,q) = p\log\frac{p}{q} - p + q \tag{3}$$

with the conventions $\frac{0}{0} = 0$, $0\log 0 = 0$ and $p/0 = \infty$ for $p \succ 0$. From the inequality $x\log x \geq x - 1$, it follows that $D(p,q) \geq 0$ with equality i.f.f. $p = q$. In (3), the value of $D(p,q)$ is very small if p and q is close each other. There are none quadratic terms in (3), which reduce calculation time and complexity. And it is unnecessary to introduce weight coefficient in I-divergence, so the value of $D(p,q)$ cannot be affected by subjective factors. For training sample x_i and test sample y_i , the I-divergence is defined as

$$d\_\mathrm{ID}(\mathrm{x}_i, y_j) = \sum_{k=1}^{l}\left(y_j^k \log\frac{y_j^k}{x_i^k} - y_j^k + x_i^k \right) \tag{4}$$

where it follows that $d\_\mathrm{ID}(\mathrm{x}_i, y_i) = 0$ with equality i.f.f. $x_i = y_i$. We use (4) as the distance metric instead of Euclidean distance used in TR_KNN. Using of I-divergence to solve the problem with all non-negative data can save calculation time and reduces calculation complexity.

C. Improved K-nearest Neighbor based on Information Divergence

According to the analyses above, we proposed ID_KNN with the combination of TR_KNN and I-divergence on the basis of TR_KNN. The details of ID_KNN are described as follows:

Step1: Given training sample set, test sample set;

Step2: Determine initial k value;

Step3: Calculate the distance between test sample and training samples with (4), sort the distances in ascending order, and then select the k sample with k relative minimum distances as k nearest neighbors;

Step4: Finding the dominant class according the **Step 5** in TR_KNN, and determine the class of the test sample;

Step5: Evaluation. If unsatisfied with the classification result, go back to **Step 2** and continue with **Step 2** to **Step 5**. Otherwise, go to the end.

The proposed KNN is based on TR_KNN, so it keeps the advantages of simplicity and easy implementation. Furthermore, its performance is better than TR_KNN.

III. Experiments

In this section, training dataset and test dataset at six stages (seeding, jointing, heading, grain filling, milky and mature) are tested to verify the performance of ID_KNN. All experiments are carried on Intel(R) Core(TM) i7 CPU 860@2.8GHz, memory with 4G and operating system with Windows XP by Matlab 2009b.

Matlab is developed by MathWorks, which is a high-level language and interactive environment for numerical computation, visualization, and programming. Using Matlab, you can analyze data, develop algorithm, and create models and applications. The language, tools and built-in math functions enable you to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. You can use Matlab for range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use Matlab, the language of technical computing.

A. Features Selection and Evaluation Indexes

The research objects are maize planted in Duerbote Mongolia Autonomous County in Heilongjiang Province in 2010. We observe six stages during maize grown life including seedling stage, jointing stage, heading stage, grain filling stage, milky stage, and mature stage. 180 maize plants in sixteen areas, with different sample capacity in different areas, are chosen as the whole sample dataset. We consider that soil moisture is mainly from irrigation, so irrigation quantity denotes soil moisture of corresponding area. At experimental fields, irrigation quantity of each area is fixed with five different irrigation levels. The data details are described in Table I.

Soil moisture is a critical element for the growth of maize, comparing normal level of soil moisture, the morphological characteristics show different state when the soil moisture is low. So, morphological characteristics, such as maize plant height, leaf area and stem diameter can be used as the features in estimation of soil moisture. Besides, water is the component with the largest percentage in weight of plants. Meanwhile, water mainly comes from soil moisture, so dry weight and fresh weight of plants can be chosen as the features in estimation. In summary, we choose maize plant height, leaf area, stem diameter, dry weight and fresh weight as the features in estimation of soil moisture in the six stage.

The different sample datasets in six stages (seeding, jointing, heading, grain filling, milky and mature) are denoted by

$$\Omega_k = \left\{ (x_i, y_i) \mid x_i = \left(x_i^1, x_i^2, x_i^3, x_i^4, x_i^5\right), y_i \in C \right\} \tag{5}$$
$$i = 1, 2, \cdots, 180, k = 1, 2, 3, 4, 5, 6$$

where $C=\{c_1,c_2,c_3,c_4,c_5\}$ represent 381.3Kg/Hm^2, 500 Kg/Hm^2, 618.7 Kg/Hm^2, 300 Kg/Hm^2, 700 Kg/Hm^2 respectively. $k=1,2,\cdots,6$ represent seeding, jointing, heading, grain filling, milky and mature stages respectively.

TABLE I
THE DETAILS OF DIFFERENT AREA

AREA NO.	SAMPLE NO.	FEATURES NO.	IRRIGATION QUANTITY(KG/HM2)
1	10	5	500
2	10	5	500
3	10	5	381.3
4	10	5	618.7
5	10	5	381.3
6	10	5	618.7
7	10	5	381.3
8	10	5	618.7
9	10	5	381.3
10	10	5	618.7
11	20	5	700
12	20	5	300
13	10	5	500
14	10	5	500
15	10	5	500
16	10	5	500

To evaluate the performance of ID_KNN used in estimation of soil moisture, we choose accuracy and macro-F$_1$ measure [8] as the evaluation indexes. Accuracy and macro-F$_1$ measure of TR_KNN and ID_KNN are denoted by A_{TR}, F_{TR}, A_{ID} and F_{ID}, respectively. The details of the indexes are described as follows:

$$A=\frac{Number\ of\ \ classified\ correctly}{Number\ of\ \ whole\ dataset}\times 100\% \qquad (6)$$

and

$$macro\text{-}F_1=\frac{\sum_{k=1}^{n}F_{1k}}{t} \qquad (7)$$

In (7), F_{1k} such that $F_{1k}=2r_kp_k/(r_k+p_k)$, where r_k and p_k represent recall and precision respectively. r_k such that $r_k=a_k/b_k$, where a_k denotes the number of the kth class test samples predicted correctly, b_k denotes the number of the kth class test samples; t is the number of the classes in test sample set. p_k such that $p_k=a_k/d_k$,where d_k is the number of test samples that predicted to be the kth class. F_{1k} combines the recall (r_k) with precision (p_k) into a single measure.

B. Experimental Results

In this experiment, the aim is to test the effects of irrigation quantity on biological characteristics, namely the biological characteristics have little differences between different areas with same irrigation quantity. We select heading stage as example and set the k to be 6 in TR_KNN and ID_KNN. The sixteen areas are divided into five classes according to irrigation quantity, that is area 3,5,7and 9 belong to class 1; area 1,2,13,14,15 and16 belong to class 2; area 4,6,8 and 10 belong to class 3; area 12 belongs to class 4 and area 11 belongs to class 5. We choose 20 samples in class 1 as the test sample set with equal quantity of 5 in each area, and the left 140 samples of the five classes form the training sample set. The results are described in Tab.II, in which Num.TR and Num.ID present the number of test samples classified into a class with TR_KNN and ID_KNN respectively.

In Tab.II, by means of TR_KNN, seventeen test samples are classified into class 1, one is classified into class 2 and two are classified into class 4. In fact, the 20 test samples are all belong to class 1, so there are 3 samples incorrectly classified, the reason of which is that the irrigation quantities of class 1, 2 and 4 are close. The number of test samples classified correctly with ID_KNN is 19, and 1 is incorrectly classified in to class 4. The results show that the samples of area 3,5,7 and 9 in class 1 have little differences in biological characteristics with same irrigation quantity, and if the irrigation quantities of two area is close, the biological characteristics of maize plants is similar. It is pointed out that the performance of ID_KNN is superior to that of TR_KNN.

TABLE.II
CLASSIFICATION RESULTS WITH DIFFERENT IRRIGATION QUANTITIES

Area	3,5,7, 9	1,2,13,14, 15,16	4,6,8,10	12	11
Irrigation Qty.	381.3	500	618.7	300	700
Samples No.	40	60	40	20	20
Class	1	2	3	4	5
Num.TR	17	1	0	2	0
Num.ID	19	0	0	1	0

In this experiment, we set k value to be 6, and choose half samples of each class under different irrigation quantities to construct test samples set, the number of test sample is 90. The classification accuracy of each class is described in Tab.III, and the performances of TR_KNN and ID_KNN in six stages are described in Tab.IV.

Tab.III shows the classification accuracies of different irrigation levels with TR_KNN and ID_KNN. In overall, the accuracy is improved greatly with ID_KNN than TR_KNN at same dataset and irrigation quantity. When irrigation quantity is 700, the largest increments extends of A_{TR} and A_{ID} are 10% and 20% respectively at grain filling stage, and at heading stage dataset and mature stage dataset, the A_{ID} reaches 100%. The results denote that KNN can be used to estimate irrigation quantity, namely, soil moisture, and the method we proposed outperforms than TR_KNN.

Tab.IV shows the classification performances of TR_KNN and ID_KNN with mixing all the samples of six irrigation levels together, so the evaluation indexes shows the average classification state at six datasets. Both accuracy and macro-F_1 measure are increased in ID_KNN relative to TR_KNN. At grain filling dataset, the performance of ID_KNN is the best, with the increments of accuracy and macro-F_1 measure are 10.7% and 0.082 respectively. However, the accuracy of TR_KNN is the lowest at six stages. Accuracy and macro-F_1 measure reach peaks at mature dataset, the reason of which is the maize ear greatly different with different irrigation quantities during maize grown stages.

TABLE III
ACCURACIES OF DIFFERENT IRRIGATION QTY. IN EACH STAGES

Datasets	Irrigation Qty. (kg/hm^2)									
	381.3		500		618.7		300		700	
	A_{TR}/%	A_{ID}/%	A_{TR}/%	A_{ID}/%	A_{TR}/%	A_{ID}/%	A_{TR}/%	A_{ID}/%	A_{TR}/%	A_{ID}/%
Seedling	70	75	80	80	75	85	70	80	80	90
Jointing	80	80	83.3	90	80	80	70	80	80	80
Heading	75	80	86.7	90	85	90	60	70	90	100
Grain filling	75	85	90	93.3	80	95	60	70	70	90
Milky	85	90	90	93.3	80	85	80	80	90	90
Mature	85	90	93.3	93.3	85	90	90	100	90	100

TABLE. IV
THE AVERAGE ACCURACIES AND F_1MEASURE IN SIX STAGES

Datasets	A_{TR}/%	A_{ID}/%	F_{TR}	F_{ID}
Seedling	75	82	0.8257	0.882
Jointing	80	86.67	0.8683	0.9207
Heading	86	93.33	0.902	0.9529
Grain filling	74.33	85	0.8287	0.9109
Milky	83.67	86	0.8841	0.9053
Mature	88.67	96.43	0.9083	0.9627

Fig1. and Fig.2 show the change trends of accuracy and macro-F_1 measure, both of the two indexes increase from seedling stage to heading stage, then they decrease a little, and then increase to the top at mature stage. Soil moisture influences the biological characteristics of maize during their grown stages. Heading stage is affected most seriously by soil moisture deficit; jointing stage is the following one; and seedling stage is the one that maize has affected relatively less. Because of the influence of soil moisture deficit on the six stages of maize plants, the mature stage will show great difference between the areas under different irrigation levels, which provides good factors for classification. Thereby, classification performance is excellent when the influence of soil moisture deficit on maize plants is serious. So accuracy and macro-F_1 measure in heading, jointing and mature stages are higher than those of other stages.

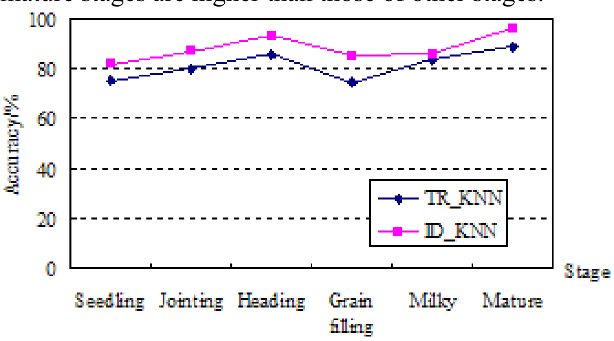

Figure 1. Classification accuracy with TR_KNN and ID_KNN

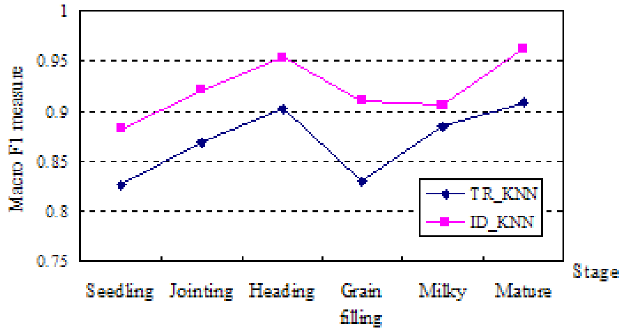

Figure 2. Macro F1 measure with TR_KNN and ID_KNN

IV. CONCLUSIONS

A novel k-nearest neighbor algorithm based on I-divergence is proposed to estimate the soil moisture in maize fields. Dataset with biological features (plant height, leaf area, stem diameter, dry weight and fresh weight) and irrigation quantity of maize plants in sixteen areas in six stages (seeding, jointing, heading, grain filling, milky and mature) are constitute to test the performance of ID_KNN. The experiment results show that ID_KNN is superior to TD_KNN in accuracy and macro-F_1 measure in estimation of soil moisture in maize fields. Therefore, ID_KNN can be used in estimation of soil moisture.

REFERENCES

[1] Z. R. Khan, J. A. Pickett, L. J. Wadhams, A. Hassanali, C. A. O. Midega, "Combined Control of Striga Hermonthica and Stemborers by Maize-Desmodium SPP. Intercrops", *Crop Protection,* vol.25, no.9, pp.989-995, Sep. 2006.

[2] C. Champagne, A. A. Berg, H. McNairn, G. Drewitt, T. Huffman, "Evaluation of Soil Moisure Extremes for Agricultural Productivity in the Canadian Prairies", *Agricultural and Forest Meteorology*, vol.165, pp.1-11, Nov.2012.

[3] I. M. Rahman, H. Hasegawa, "Water Stress", Intech, Jan.2012.

[4] U. R. Sangakkara, M. Frehner, J.Nösberger, "Effect of Soil Moisture and Potassium Fertilizer on Shoot Water Potential, Photosynthesis and Partitioning of Carbon

Mungbean and Cowpea", *Journal of Agronomy and Crop Science*, vol.185, no.3, pp.201-207, 2000.
[5] P. Dobriyal, A. Qureshi, R. Badola, S. A. Hussain, "A Review of The Methods Available for Estimating Soil Moisture And Its Implications for Water Resource Management", *Journal of Hydrology*, vol.458-459, pp.110-117, Aug.2012.
[6] V. Bavel, C. H. Underwood, "Soil Moisture Measurement by Neutron Moderation", *Soil Science*, Vol. 82, no. 1, pp.29-42, 1956.
[7] G. J. Bouyoucos , A. H. Mick. "An Electrical Resistance Method for The Continuous Measurement of Soil Moisture Under Field Conditions", Technical Bulletin., Michigan Agricultural Experiment Station, Vol. 172 pp. 38, 1940.
[8] S. Jiang, G. Pang, M. Wu, Limin Kuang, "An Improved K-Nearest-Neighbor Algorithm for Text Categorization", *Expert Systems with Applications*, vol.39, no.1, pp.1503-1509, 2012.
[9] E. H. Han, K. George, K. Vipin, "Text Categorization Using Weight Adjusted k Nearest Neighbor Classification", *Lecture Notes in Computer Science*, vol.2035, pp.53-65, 2001.
[10] M. Z. Jahromi, E. Parvinnia, R. John, "A Method of Learning Weighted Similarity Function to Improve the Performance of Nearest Neighbor", *Information Sciences*, vol.179, no.17, pp.2964-2973, 2009.
[11] R. Chang, Z. Pei, C. Zhang, "A Modified Editing k-nearest Neighbor Rule", *Journal of Computers*, vol.6, pp.1493-1500, 2011.
[12] J. Gou, T. Xiong, Y. Kuang, "A Novel Weighted Voting for K-Nearest Neighbor Rule", *Journal of Computers*, vol.6, pp.833-840, 2011.
[13] T.M. Cover, P. E. Hart, "Nearest Neighbor Pattern Classification", *IEEE. Transactions on Information Theory*, vol.13, no.1, pp.21-27, 1967.
[14] X. Xu, Z. Liu, Q. Wu. "A Novel K-Nearest Neighbor Classification Algorithm Based on Maximum Entropy. International Journal of Advancements in Computing Technology", vol.5, no.5, pp.966-973,2013, in press.
[15] E. H. Han, K. George, K. Vipin, "Text Categorization Using Weight Adjusted k Nearest Neighbor Classification", *Lecture Notes in Computer Science,* vol.2035, no.1, pp.53-65, 2001.
[16] S. Tan, "Neighbor-weighted K-Nearest Neighbor for Unbalanced Text Corpus", *Expert System with Application*, vol.28, no.4, pp.667-671, 2005.
[17] Xu, C. Li, S. Jiang, X. Liu, "Similarity measures for content-based image retrieval based on intuitionistic fuzzy set theory", *Journal of Computers*, vol. 7, pp.1733-1742, 2012
[18] I. Csiszár. "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problem", *The Annals of Statistics*, vol.19, no.4, pp.2032-2066, 1991.
[19] S. Kullback, R. A. Leibler. "On information and sufficiency", *The Annals of Mathematical Statistics*, vol.22, no.1, pp.79-86, 1951.

Xiangyan Meng was born at Heilongjiang Province, China, in 1977. She received the Master degree in ecology from Northeast Agricultural University in 2008. From September 2009 to the present, he is working in Ph. D in agricultural water and soil engineering from Northeast Agricultural University, Harbin, China. From July 1997 to the present, she is currently an Associate Professor in College of Science in Northeast Agricultural University. Her current research interests include irrigation and drainage theory and new technology, Principle of soil erosion and soil and water conservation technology, and ecological agriculture.

Zhongxue Zhang was born at Heilongjiang Province, China, in 1967. He received the Ph.D. degree in Ecology from China Agricultural University, Beijing, China, in 2000. From September 2000 to November 2002, he was a Post Doctor in the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science. From July 2002 to September 2002, he was a Visiting Professor in Hokkaido University of Japan. He is currently a Professor in College of Water Conservancy and Construction, Northeast Agricultural University. His current research interests include Irrigation and drainage theory and new technology, Principle of soil erosion and soil and water conservation technology, and ecological agriculture.

Xinying Xu (1985.04) born at Hebei province, received the bachelor degree in engineering from Harbin University of Commerce, Hei Longjiang province, China. Now she is a postgraduate in management science and engineering at Northeast Agricultural University, in Harbin, Hei Longjiang province, China. The fields of her research interests include pattern recognition and machine learning.