# Study of Continuous K-nearest Neighbor Algorithm

**Yidong Song**

School of North China Electric Power University, Hebei 071000, China

2261924775@qq.com

**Keywords:** k-nearest neighbor; continuous; prediction.

**Abstract:** In the machine learning algorithm, the K-nearest neighbor is mainly used to classify instances, and the selection of the value of K is especially critical. For the selection of the value of k, if the value is too large, the model is too simple to extract the law; if it is too small, the model is too complicated and over-fitting may occur. In general, we use cross-validation to get a good value of K. So that the model we built can better characterize the law of nature. This is only a fuzzy choice in the case of unclear data, and there is no guarantee that the model has strong practical value. Here, we introduce the "degree of structural law fluctuations", which can reflect the influence of different classifications on the overall law of things, and quantitatively express the objective laws of things for the first time, thus improving the accuracy and practical value of the K-nearest neighbor algorithm model. We call this algorithm as the improved algorithm, which is called continuous k-nearest neighbor algorithm.

## 1. Introduction

First, we divide the total data set into a training set and a test set. The training set is used to train the model, and the test set is used to test the effect of the model and select the optimal model.

In the K-nearest neighbor algorithm, we generally determine the category of an unknown point by K categories that are closest to the instance point. Here is to compare the distance between the points, select the K point with the smallest distance, and use the cross-validation method to optimize the K value.

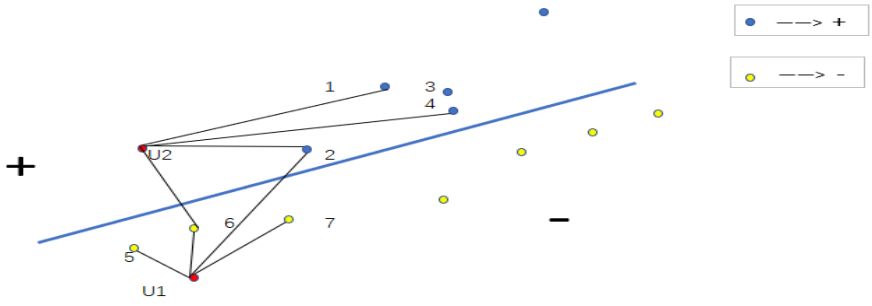


Figure 1. Wrong example

As shown in the Figure 1, we now need to classify the two unknown points U1 and U2; it is known that they are divided into two categories of "+" and "-" on both sides of the line. According to the K-nearest neighbor algorithm, we first calculate the distance between the unknown point U1 and the points of other known categories, and select the K known instance points that are closest to the unknown point (where the K value is 4, that is, 4 known Instance point);Among the four points, the category "-" has 5, 6, and 7; a total of three; the category "+" has 2; a total of one, we use a category with more categories as the category of unknown points, so unknown points The category of U1 is "-". Thus, the unknown point U1 is correctly classified.

However, such a selection method is relatively. straightforward; it does not reflect the overall structural law well and increases the error of model prediction.

For example, in the Figure 1, for the unknown point U2, the K known instance points closest to the distance are 2, 5, 6, and 7; wherein the category of 2 is "+", and the categories of 5, 6, and 7 are "-", so the category of the unknown point is "-".However, as can be seen from the image, the unknown point U2 is located on the side of the line type "+", that is, the actual category of the unknown point U2 is "+", which causes a problem of model classification error.

In order to solve the problem that the overall structural law of the algorithm appears insufficient during the training process, resulting in a large error, we introduce the “degree of structural law fluctuation” (SF).

First, let us introduce the basic idea of this article: correct classification always has less impact on known laws than incorrect classification (known laws are also filtered by this rule).In this way, we can judge the degree to which a certain point is classified correctly by comparing its influence on the known law.

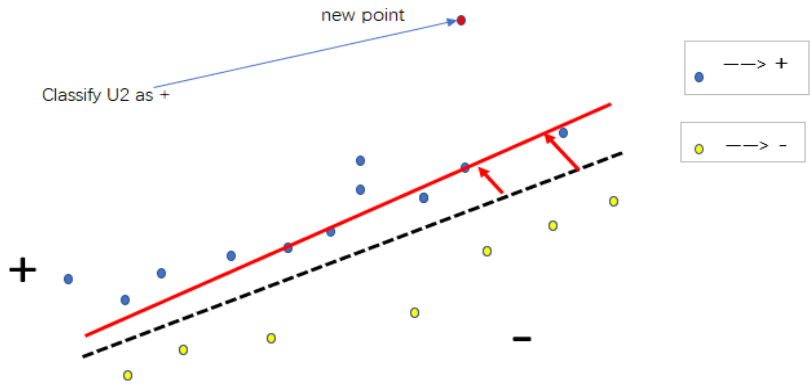


Figure 2. The effect of classifying the new point as +

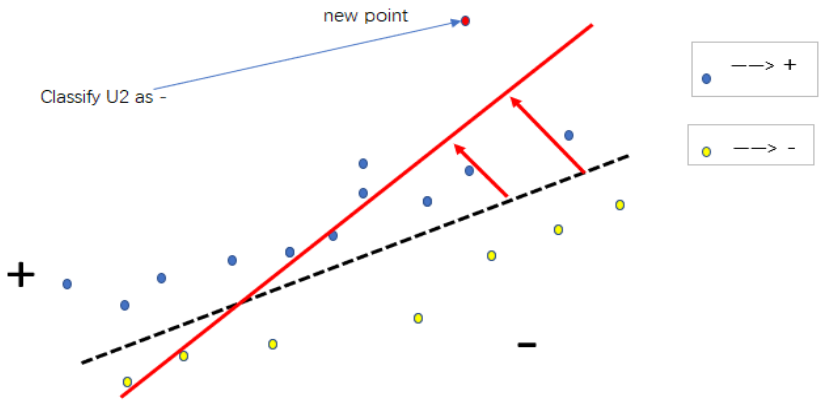


Figure 3. The effect of classifying the new point as –

Of course, we will definitely think that we can classify unknown points directly by comparing the “degree of structural law fluctuations” of various categories. That is to say, because the correct classification always has less impact on the original law than the wrong classification, then I select the classification with the smallest degree of structural law fluctuation, and the unknown points can be correctly classified. In fact, because we train the model through the training set, but the training set cannot fully reflect the inherent law of things, so here we only use the "degree of structural law fluctuations" as one of the parameters to train the model. The difference is that for different problems, we can assign different weights to the “structural regular variables”, making the model more realistic.

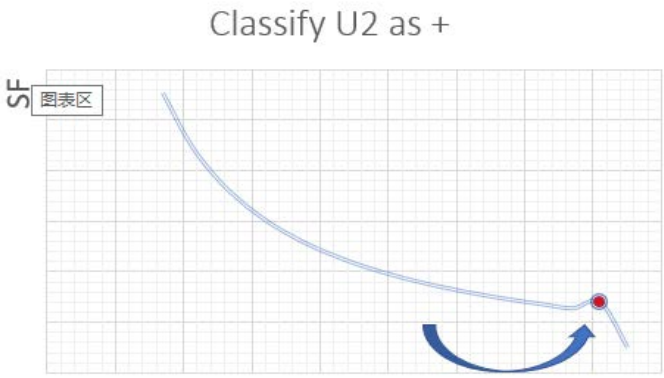


Figure 4. Classify U2 as +

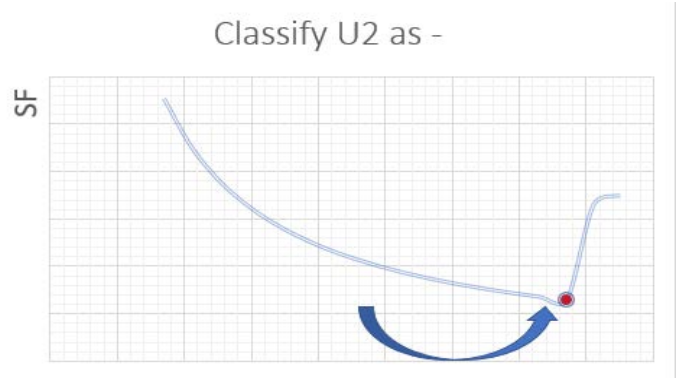


Figure 5. Classify U2 as +

The “degree of structural law fluctuation” SF can reflect the difference between the structural law of the newly obtained point set and the structural law of the original point set after classifying an unknown point. In the Picture 1, when the unknown point U2 is classified as "-", the "degree of structural law fluctuation" value is greater than the case where the unknown point U2 is classified as "+" (see Picture 2 and Picture 3).When we determine the value of K and select K instance points, we comprehensively consider the "distance" between the points and the value of the "degree of structural law fluctuation" in each classification case, so that the obtained K value is more reasonable, and the selected K instance points are also more reflective of the overall structural law, improving the predictive level and actual value of the model.

Specifically, we convert the original single "distance" (D) into a "continuous distance" (CD) with weights; The CD is weighted by the physical distances D and SL between the unknown point and the actual point.

$$CD = A \times D + B \times SF, \quad A \text{ and } B \text{ are weights.}$$

## 2. Methods

For the K-nearest neighbor algorithm, we first determine an initial K value; first, we specify that the total number of instance points included in the training set is TO. Here we specify that the initial K value is half of TO.

$$K = \frac{TO}{2}$$

Then calculate the physical distance (D) between the unknown point and each instance point; the distance can be calculated in various forms, including Lp distance, Euclidean distance, Manhattan distance, etc.

Table 1: Definition of distance

| Number | Name |
|---|---|
| 1 | Lp Distance |
| 2 | Euclidean Distance |
| 3 | Manhattan Distance |
| 4 | Normalized Euclidian Distance |

Then we abstract the general law of things through the training set. We call it E. For the classification problem, we classify the categories according to the attributes of each instance. We can also abstract the overall law E through its attributes. For example, the classification problem of points on a two-dimensional plane, each point has two attributes x and y, that is, values on two axes. We fit the data by the coordinate values (xi, yi) of the known example points, and we can get the law of data like curve of the first degree and quadratic curve. This is the so-called overall development law of things E. Among them, the curve of the first degree is relatively simple, we also take this as an example to elaborate here.

After that, we need to calculate the weight of the ordinary physical distance D and the "degree of structural law fluctuation "(SF); this requires calculating the abstraction degree (Ad) of the training set we use on the overall law of the whole thing, that is, whether our training set can be very a good reflection of the nature of things. Here we obtain Ad by inverting the error value of the abstract whole law of the training set. The smaller the error, the more the training set can reflect the overall law of things, and the value of Ad is larger. When the value of Ad obtained by us is larger, it indicates that the data in the training set can better reflect the data law. Therefore, the law of abstraction of the training set is more powerful for the establishment of the overall model. Therefore, we should put the weight of the "degree of structural law fluctuation" Tune up; vice versa. Here, I give the specific calculation formula.

$$B = \begin{cases} 40 - \frac{\sqrt{40-B^1}}{e}, & B^1 \leq 40; \\ B^1, & 40 < B^1 < 70; \\ 70 + \frac{\sqrt{B^1-70}}{e}, & B^1 \geq 70; \end{cases}$$

$$B^1 = arctan(e^{|Ad|})$$

The role of $e^{|Ad|}$is to make the data scattered, so that different Ads are given different weights.

Then calculate the "continuous distance" between the unknown point and each instance point. Here, we take the classification problem of points on a two-dimensional plane as an example. For a point on a two-dimensional plane, we can initially fit the data according to its two attribute values and abstract the overall law of things; Then, when classifying the unknown points, calculate the change of the overall law of the things when the unknown points are divided into certain categories; that is, the degree of similarity (SIM) of the fitted curves obtained before and after the classification is calculated, and the higher the similarity degree is, the smaller the value of SF is, and vice versa.

$$SF = \begin{cases} SF^1, & 10^{-2} < \frac{D}{SF^1} < 10^2; \\ D + \frac{1}{\log_{SF^1} D}, & else; \end{cases}$$

$$SF^1 = SIM^2$$

The function of $\frac{D}{SF^1}$ is to prevent the difference between D and SF value from being too big, which will affect the weighted value.

After calculating the SF value, we add the SL and D values by weight to get the "continuous distance" CD.

To calculate the "continuity distance" (CD) between the unknown point and each instance point in the training set, we select K instance points with the smallest CD and use them to judge the category of the unknown point.

The specific judgment method is as follows. First, define a "category variable" CA. For category i, the corresponding category component is C$A_i$(i=1,2,3...). To begin with, we initialize each category component to 0; Then determine the category of the selected K instance points. If it belongs to category j, then add 1 to the corresponding C$A_j$, and finally get the value of each "category variable" C$A_i$ (i=1,2,3...). Select the largest "category variable" C$A_m$; and m is the category of the unknown point.

Through the above steps, we can train the model through a preset K value and determine the category of an unknown point. Below we need to determine the optimal K value by cross-validation to make the model more accurate.

First, for the initial K value, we use the instance points in the test set to test the accuracy of the model. and define the "optimization variable" OV. For each $K_i$ , the model corresponds to an optimization variable $OV_i$, which reflects the accuracy of the model when taking the $K_i$.At the beginning, we initialize each "optimal variable" $OV_i$ (i=1,2,3...) to 0, and test the model with the test set. When the model predicts the same category as the actual category, the OV value is increased by 1, and finally the value of K corresponding to the maximum OV value is obtained. We call it is $OV_k$.This is the optimal K value. Data analysis is shown in Figure 4.

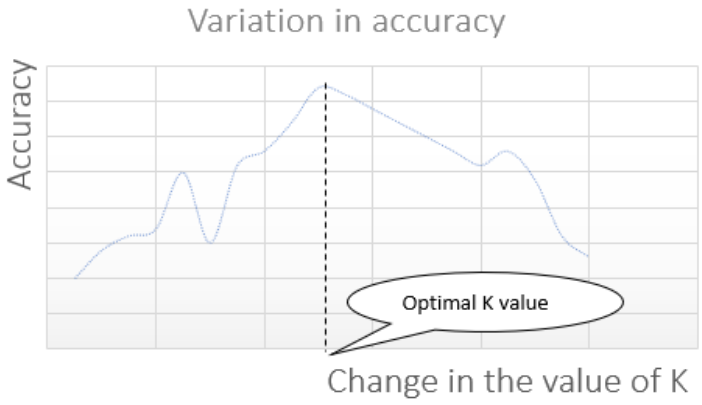


Figure 6. Variation in accuracy

At this point, we have built the model, which can achieve more accurate predictions.

similarly, in other aspects of machine learning, such as support vector machines, naive Bayesian methods, hidden Markov models, etc. we can also add "degree of structural law fluctuations" to improve the model to make it more optimized. When we add the “degree of structural law fluctuation” in different fields, the way to calculate the “degree of structural law fluctuation” is different. The following calculation methods are listed below.

Table 2: Definition of distance

| Model | Calculation standard |
|---|---|
| $SupportVectorMachine$ | Separation hyperplane |
| $HiddenMarkovModel$ | State sequence |
| $k-NearestNeighbor$ | Classification criteria |
| $Perceptron$ | Hyperplane used for classification |

## 3. Results

By introducing the “degree of structural law fluctuation” SF, the accuracy of the model and the actual application value are greatly improved.

Here we specify by an example given above. For the problem in Figure 1, we add the “degree of structural law fluctuation”, and the K known instance points of the model choose have changed.

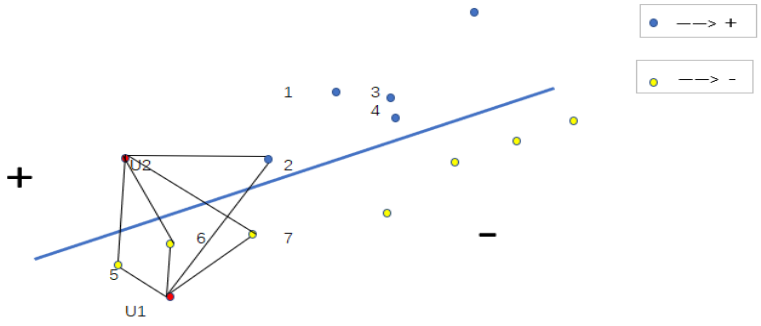


Figure 7. before adding degree of structural law fluctuation

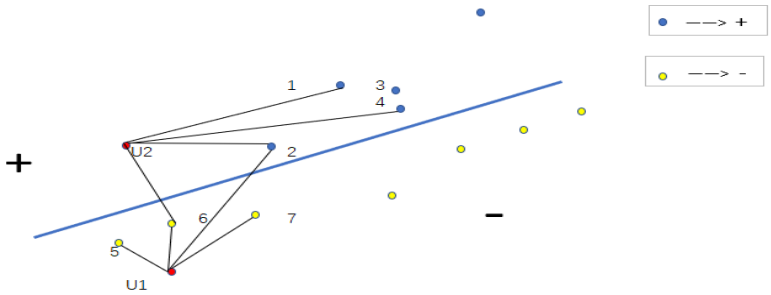


Figure 8. after adding degree of structural law fluctuation

In the figure, when judging the unknown point U2, the model calculates the selected K instance points from the original 1, 4, 5, and 6 points to 1, 2, 3, and 4 points, and the prediction result is also from the original " - "Changed to "+", which is in line with the actual situation.

The following is the prediction accuracy map of the model before and after adding SF.

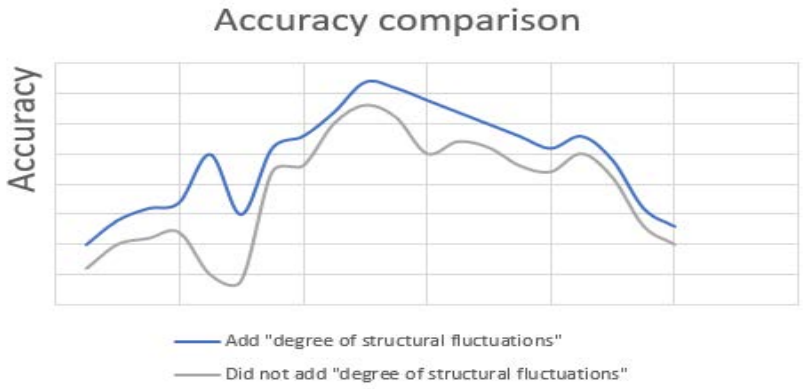


Figure 9. Accuracy comparison

Through the above analysis, after adding the “degree of structural law fluctuation”, the overall law of the data is quantified and involved in the process of model building, the accuracy of the model is improved, and we can better guide our lives.

## 4. Conclusion

In this article, we discuss the K-nearest neighbor algorithm after adding the “degree of structural law fluctuations”, that is, when training the model, the law of the essence of things is quantitatively expressed and added to the design of the model. Through actual calculations, we find that the continuity model obtained by this method is more optimized. At the same time, the model can have a lot of improvement. For example, when calculating the "continuous distance" between the unknown point and the instance point, the calculation is more complicated because the "continuous distance" from each instance point is calculated. Here, we can prioritize the instance points. The "continuity distance" between the instance points with high priority and the unknown point is less than the "continuous distance" between the instance point and the unknown point with lower priority, so that we don't need to Calculate the distance between all instance points and unknown points. Among them, one way to deal with data is to construct Kd-Tree.

At the same time, for the "singular training set", we avoid singular situations by special operations when selecting the training set. Here's a look at what is a "singular situation."

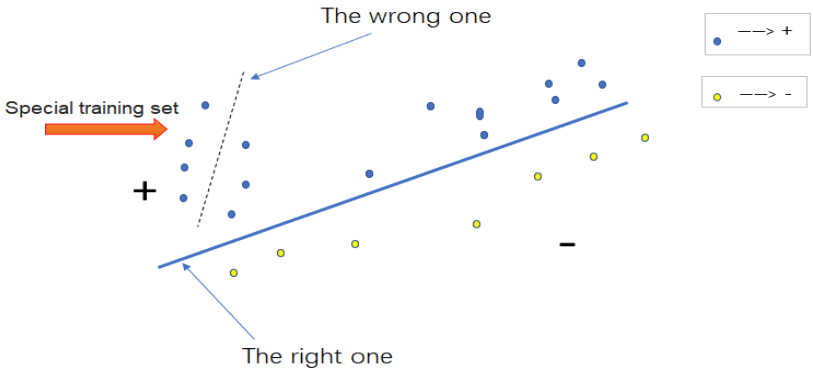


Figure 10. singular training set

In Figure 6, we see that the selection of the Special training set will make the overall development of the training set far from the correct development. This requires us to select the training set through a special method to make the selected instance points evenly distributed everywhere. Specifically, when selecting the training set, we make the distance H of any two instance points $P_i$ and $P_j$large enough to ensure uniform distribution.

$$H = |P_i - P_j| > S$$

S is a large enough number that will vary with specific problems.

## References

[1] Dalia M. Atallah, Mohammed Badawy, Ayman El-Sayed. Intelligent feature selection with modified K-nearest neighbor for kidney transplantation prediction[J]. SN Applied Sciences, 2019, 1 (10).

[2] R. B. Mcquistan, J. L. Hock. The shift operator matrix method applied to the two-dimensional nearest and next nearest neighbor problem[J]. Journal of Mathematical Chemistry, 1988, 2(1).

[3] Sahil Dhawan, Agnikumar G Vedeshwar, R P Tandon. Correlation of optical energy gap with the nearest neighbour short range order in amorphous V 2 O 5 films[J]. Journal of Physics D: Applied Physics, 2011, 44(21).