# 

# 

# 

# Extended Nearest Neighbour

# for Pattern Recognition

# Reading Report

## Team: 18

Tran Khanh Trung (e0146769)

Rahul Chauhan (e0146878)

Woo Yee Wan (e0146940)

# Summary of Extended Nearest Neighbour for Pattern Recognition Whitepaper

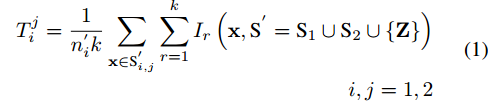
*Extended Nearest Neighbor Method for Pattern Recognition, IEEE Computational Intelligence Magazine, vol. 10, no. 3, Aug 2015*

The whitepaper proposed a new supervised classification method namely Extended nearest neighbour (ENN) which improvised the traditional K-nearest neighbour (KNN) method. K-nearest neighbors (KNN) algorithm is a popular nonparametric method used for supervised classification and regression. Even though the algorithm has gained many undeniable successes over other methods, there are some issues associated with KNN. This report summaries some definitions and mathematical functions of the new approach; Extended nearest neighbor (ENN) and how it overcomes the limitations of traditional KNN method.

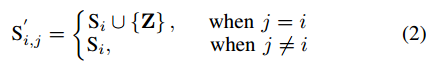
For many years, KNN has achieved its fame through the elegance and simplicity of the algorithm. It also has competitive performance and independent of the underlying data distribution. However, one of the weakest nature of the algorithm itself is that the method is sensitive to the scale or variance of distributions of the pre-defined classes. The nearest neighbors of an unknown test data will tend to be dominated by the class with the highest density.

To overcome the problem, ENN method which keeps all the advantages of the KNN and also introduces new idea is suggested. The most fundamental difference between KNN and ENN is that KNN technique involves only some nearest neighbours of a test sample to estimate a group membership. ENN goes further than that by taking into account to learn from all the training sample. This is often referred to as "two-way communication" style. Hence, ENN addressed the KNN issue of misclassification due to nearest neighbours dominated by other classes. ENN also defines some new terminologies and mathematical concept to achieve the objective such as generalized class-wise statistics, intra-class coherence, indicator function.

The whole working process of ENN method is described in Figure 1. The process considers binary classification problem. The first step works exactly like KNN method by applying the algorithm over entire training data set S. As a result of step 1, every training samples saves its k nearest neighbours and distances. The second step involves picking up test sample z from testing data Z. The third step is newly introduced which is an improvement over traditional KNN. The sample z is considered separately into two different classes (class 1 and class 2) for further investigation. Step four of the process is applying KNN to the union set S = S1 ∪ S2 ∪{z}. The fifth step, we apply generalized class-wise statistics Tij as follows to estimate the influence of sample z.



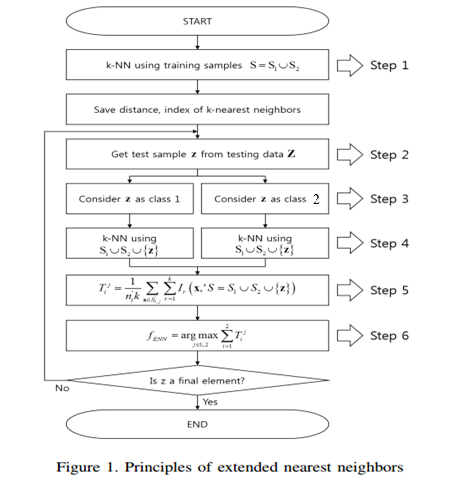
Where x denotes a sample in S’, k is user-defined number of nearest neighbor, ni’ is the size of Si,j’which is defined below. Ir is



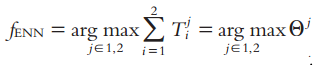
The indicator function indicates whether both the sample x and its r-th nearest neighbor belong to the same class



Where NNr (x,S) denotes the r-th nearest neighbor of x in S.

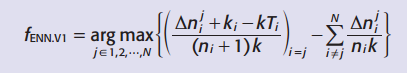


For the last step, we have already had four separate generalized class-wise statistics (for binary classification). ENN classifier predicts the outcome according to the following target function:



The final result proves that ENN method make the prediction based on j which has largest intra-class coherence.

**ENN.V1 classifier** is an improvement for ENN from the computational point of view as the function of ENN.V1 does not require recalculating the generalized class-wise statistic for every test sample. From the mathematical aspect, the classification outcome of fENN and fENN.V1 is the same, however, the implementation and calculation of fENN is easier for many practical applications.



Where,   
nji = change of the k nearest neighbour for class i when the test sample Z is assumed to be class j.

**ENN.V2 classifier** is an approximation version of ENN which simplifies the calculation dramatically under some assumptions. If the assumptions are satisfied, the ENN.V2 can provide competitive classification performance.   
The approximation is as such :-

1. All classes have the same number of data samples n (balanced classification problem).

2. For case i not equal to j, ( nji /((n+1)nk)) -> 0.

Using above approximation and the target function of ENN.V1, we can derive the target function for approximate version of ENN as below:-



Below table summarized the key features of each classifier:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | KNN classifier | ENN classifier | ENN.V1 classifier | ENN.V2 classifier |
| Training | Locate nearest neighbours, k and the distances. | Locate nearest neighbours, k and the distances. | Locate nearest neighbours, k and the distances. | Locate nearest neighbours, k and the distances. |
|  |  | Calculate class-wise statistics, T (for each class) | Calculate class-wise statistics, T (for each class) | Calculate class-wise statistics, T (for each class) |
| Prediction | Predict based on the k. | Consider each test sample z, for each class and obtain class-wise statistics. | Consider each test sample z, for each class and obtain class-wise statistics.  (Instead of re-calculating the class-wise statistics, calculate the change in k-nearest neighbours) | Approximate number of samples in class who consider test sample z as one of their k nearest neighbour.  Re-use the class-wise statistics T of the original class |
|  |  | Predict based on largest intra class coherence | Predict based on largest intra class coherence | Predict based on largest intra class coherence |

In conclusion, the ENN classification method based on the maximum gain of intra-class coherence provides a better performance than the classical KNN method. By utilizing “two-way communication” style, ENN can effectively eliminate the sensitivity to the scale or variance of the distributions of the predefined classes which has been a long-standing limitation of KNN method. There are three versions of ENN classifier: ENN, ENN.V1 and ENN.V2. Each classifier has its own advantages in practical application. As a classification algorithm, ENN is important and useful for many other machine learning and data mining problems, such as density estimation, clustering, regression.