**Introduction**

1. Classification is the problem of identifying to which a set if categories or sub-populations a new observation belongs, on the basis of a training set of data containing observations or instance whose category membership is known. In many domain, especially for big data and image recognition, classification has become a most hot and significant topic of research[1].In a training set, the information and features of a datum can be formally and numerically represented as a vector (x1,x2,...,xn). In such a n-dimensional features space, the main goal of classification is to estimate the label of a new unlabeled vector from a testing dataset through calculating the similarity among the labeled and unlabeled vectors. Generally, there is a common hypothesis that the training dataset and the testing dataset has a semblable distribution.
2. As a well known and widely used method, the k nearest neighbor (KNN) method, looked on as the most simple but effective method among many available classification method such as bayesian method, support vector machine and neural network. Due to its simplicity but good performance, the kNN algorithm has been considered as a top-10 data-mining method[2]. Known as a supervised learning algorithm, its working mechanism is fairly easy to grasp: for a given test sample, KNN algorithm try to get the nearest k samples in the training data set based on a specific distance metric, and then predict according to the information of these samples. In a classification problem, the vote method is usually applied to predict the class of one test samples. i.e. label this sample the most common class among the k nearest neighbors. In the past few decades, some researches are also carried on to replace the vote method with other case based reasoning method, for instance, [3][4] used evidence KNN based on the Dempster-Shafer Theory to extend the classical vote method.
3. Also, there are two issues often come into being discussed in the research: The First is about the similarity and distance compute between different vectors. recently many new algorithm has been proposed to solve this issue, Lin et al. [5] proposed a similarity compute method via fusing neighborhood information. In this proposed algorithm, both the euclidean distance and the neighbor information is considered simultaneously, the two metric is combined together to measure the similarity between samples. Yung-Kyun Noh et al.[6] proposed a generative metric learning method to enhance the performance of kNN algorithm. The second issue is about the selection of optimal value of k for a given data set. The general kNN algorithm simply assume the whole data set share a same value of k, which is actually not proper and inaccurate for a finite and non-uniformly distributed data set.Recently, many new method has been generate to find the k value adaptively or determine a k value for different sample respectively, Nicolás García-Pedrajas et al. [7] propose a algorithm to obtain local k value with a simple and fast procdeure, Zhang et al. [8] propose to learn a correlation matrix to reconstruct test data points by training data to assign different *k* values to different test data points.
4. The drawbacks of the foregoing methods is obvious, they ignore the uncertainty the training data sample itself. As we all know, the training dataset is composed of a group of labeled data samples, the label itself usually decided by some experts however, could be imprecise or even incorrect.While the label information of the k neighbors is combined to make the ultimate decision, the uncertainty of each point also accumulate. In this case, the accuracy of the kNN algorithm will decrease to some extent. Fortunately, a lot of additional and available information can used to quantify the uncertainty of the label of the sample itself. In a recommendation system, a rate of users or votes of experts of a history result could be considerated when quantify the uncertainties. For instance, Zhu et al.[9] described a a real-world decision support system for multi-disciplinary treatment called MdtDSS, the core recommend system predict a new patient's therapeutic regimen based on the historic patients. The ultimate therapy of one patient is voted by a group of doctor which reflect the doctors' opinion. The vote result of one doctor express if he/she approve the ultimate therapy of this person. Obviously, we can construct some connection between the vote result and the uncertainties.
5. In this paper, we define a method to quantify the uncertainties of samples in the training dataset with extra information, later, we fuse both the uncertainty and the label information as the evidence of that sample in the k nearest neighbors. Then, as to combine all these evidences, we use Dempster-shafer theory to accumulate all these evidences to decide which class a test sample belongs to. Later, we proposed two strategy to select the optimal k value for a specific test sample based on the combined evidences. At last, we carry on a series of experiments with the data in the MdtDSS mentioned in [9] to verify our method.
6. This paper is organized as follows: In section 2, we discuss related work on kNN algorithms. In section 3, we introduce the Dempster-shafer evidence theory and the MdtDSS. In section 4, we present the modified evidence-theory based on the uncertainty of data sample and the 2 algorithms upon optimal k value selection for this sample. Finally, in the section 5, we carry on several experiments and present results.