**3.PRELIMINARY**

**3.1 Dempster Shafer theory**

As a generalization of the Bayesian theory of subjective probability, Dempster Shafer theory was proposed in 1976 in [10]. It is a general framework for reasoning with uncertainty data. This theory provide a model to express *know and unknown* directly of different sources, and the ability to combine all the evidence. It mainly contains 2 crucial steps: Basic Belief Assignment(BBA) and evidence combination.

In the first step, let U represent a non-empty set of mutually exclusive and exhaustive propositions called the frame of decrement. The power set 2^U is all subsets of the set U, which includes both the empty set and the entire set U. For the frame of discernment U, the function   
 is a basic probability assignment (BPA) also called a mass function. This function satisfies the following two conditions:

(1)

m(A) indicates the degree of trust exactly in A. The subsets A of U where m(A) > 0 are called the focal elements of the belief function, and their union is called its core. *The function*  is the belief function over U, is defined as the sum of all the masses of subsets:

(2)

Belief (usually denoted *Bel*) measures the strength of the evidence in favor of a proposition but not any of its subsets. It ranges from 0 (indicating no evidence) to 1 (denoting certainty). An important difference with probability theory is the sum of belief of a proposition and its negation not necessarily equal to 1. Hence, the remaining belief of A, called total ignorance, is the belief of the whole frame of discernment.

In the second step, Dempster Shafer define a method to combine two or more mass assignments of a same frame of discernment in specific situations. To combine there assignments means to accumulate all evidence from all sources in one frame.This rule derives common shared beliefs among multiple sources and ignores *all* the conflicting (non-shared) beliefs through a normalization factor.

Given n mass functions m1,m2...mn on the same frame of discernment U, for arbitrary A included in U, the combination rule is as follow:

(3)

where:

(4)

the K called Normalization factor is a measure of the amount of conflict among mass functions.

**3.2 Dempster Shafer theory based kNN algorithm**

Also known as the evidence-theory based kNN (EKNN) ,the dempster shafer theory based kNN algorithm was proposed by Thierry Denœux [3] in 1995. He first established the connection between the multidimensional vector space in the KNN algorithm and the frame of discernment in DS evidence theory. In this approach, each neighbor of a pattern is considered as evidence supporting some hypotheses about the class membership of that pattern. The BPAs are calculated for each of the k nearest neighbors. The belief of each hypothese is obtained by aggregating BPAs using Dempster’s rule of combination. His contribution can be summarized as 2 points: (1)generalized a way to compute the BPA value of the each k nearest neighbor respectively based on the distance to the unlabeled data sample (2) applied the combination rule into the kNN algorithm with a special form. Our own research take both the DS evidence theory and the uncertainty of neighbors is a extension to Thierry's work which will be presented in this section.

In a classification problem, the training data set can be regarded as a collection on N P-dimensional training samples represented by X={xi =(x1i,...,xpi)|i=1,...,N}, every sample in the data set belongs to one and only one class from M classes C ={C1,...,CM}. In the very beginning, each sample in the training data is labeled a class in C with a certain degree of uncertainty, which will be discussed in detail in the following section.The labeled data set can be represented as a dualistic relationship (X,L), where L is the set of labels, that can be used to classify new patterns.

In the Dempster Shafer theory based kNN algorithm, all possible classes set C make up the frame of discernment, to predict the true label of a unlabeled pattern xs equals to assign it to one class in C.

For this unlabeled new pattern xs, the k neighbors based on a specific distance measurement make up a set . Each neighbor in provide a piece of evidence whether xs belongs to a specific class Cq in C, the negation of this piece of evidence is totally innocent, i.e. it doesn't refer to C itself rather than any subsets of C. If we use mass function to represent this piece of evidence, we get the following relationship:

(5)

where i=1,2,...,k.

**Definition 1: reflects how much the intensity is that the neighbor xi support that the unlabeled pattern xs should be classified as Cq.**

According definition 1, the should be a function of distance from xi to xs, because the smaller the distance between xi and xs is, the more crucial it is to decide wether the class of xs is the same with the class of xi. Moreover, as the distance between xs and xi gets infinitely large, the belief function given by ms, i becomes vacuous, which means that one’s belief concerning the class of xs is no longer affected by one’s knowledge of the class of xi.

we can replace with any reasonable decreasing function t, here we use the following one:

(6)

Then we can get all the k mass functions of the k nearest neighbors respectively now that the distance to xs is available. To make the ultimate decision which class xs belongs to, we should first combine all these pieces of evidence together based on the DS evidence combination rule, according to equation (3), for each class q, we get:

(7)

Notice that here the BPA function msq ({Cq}) measures the combining belief according to all the neighbors whose class is Cq.

Combining all the BPAs msq for each class, a global BPA is obtained as:

(8)

where K is the normalizing factor, q=1,2,..,M.

In the last step of the evidence rule based kNN algorithm, for each Cq in C we compute its BPA according to equation (8), and then assign xs with the optimal class.