Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Type-2 Fuzzy Sets for Pattern Classification: A Review

Jia Zeng and Zhi-Qiang Liu
School of Creative Media, City University of Hong Kong, China
E-mail: j.zeng@ieee.org

***Abstract*— This paper reviews the advances of type-2 fuzzy sets for pattern classification. The recent success of type-2 fuzzy sets has been largely attributed to their three-dimensional membership functions to handle more uncertainties in real-world problems. In pattern classification, both feature and hypothesis spaces have uncertainties, which motivate us of integrating type-2 fuzzy sets with traditional classifiers to achieve a better performance in terms of robustness, generalization ability, or classification rates. We describe recent type-2 fuzzy classifiers, from which we summarize a systematic approach to solve pattern classification problems. Finally, we discuss the trade-off between complexity and performance when using type-2 fuzzy classifiers, and explain the current difficulty of applying type-2 fuzzy sets to pattern classification.**

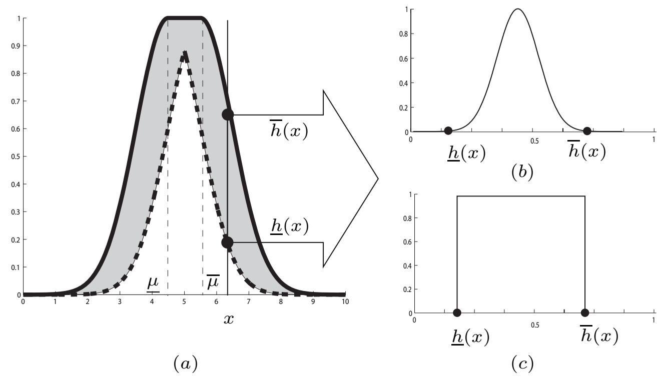

Fig. 1. The three-dimensional type-2 fuzzy membership function. (a) shows the primary membership with lower (thick dashed line) and upper (thick solid line) membership functions, where $\underline{h}(x)$ and $\overline{h}(x)$ are lower and upper bounds given the input x. The shaded region is the foot print of uncertainty. (b) shows the Gaussian secondary membership function. (c) shows the interval secondary membership function.

I. Introduction

As an extension of type-1 fuzzy sets (T1 FSs), type-2 fuzzy sets (T2 FSs) were initially introduced by Zadeh [1], and a subsequent investigation of properties of T2 FSs and higher types was done by Mizumoto and Tanaka [2], [3]. In [4] Klir and Folger explained that the T1 membership functions (MFs) might be problematical, because a representation of fuzziness is made using membership grades that are themselves precise real numbers. Thus it was natural to extend the concept of T1 FSs to T2 FSs and even higher types of FSs. In particular, they called interval type-2 fuzzy sets (IT2 FSs) as interval-valued FSs. Recently Mendel and John [5] introduced all new terminology to distinguish between T1 and T2 FSs, by which T2 FSs can be represented in vertical-slice and wavy-slice manners respectively. They also illustrated the concept of embedded FSs, which shows potential expressive power of T2 FSs for handling uncertainty. To rank T2 fuzzy numbers, Mitchell [6] ranked all embedded T1 fuzzy numbers associated with different weights. Set operations are foundations in the theory of T2 FSs, which were first studied by Mizumoto and Tanaka [2]. Their works were later extended by Karnik and Mendel [7] for practical algorithms to perform union, intersection, and complement between T2 FSs. In [5] Mendel and John reformulated all set operations in both vertical-slice and wavy-slice manners. They concluded that in practice general T2 FSs operations are too complex to implement, but operations on IT2 FSs involve only simple interval arithmetics so that they have been widely used. As the theoretical foundation of T2 FSs, Liu and Liu [8] established T2 fuzzy possibility theory. In [9] Mendel summarized developments and applications of T2 FSs before the year 2001. The recent advances in T2 FSs and systems since the year 2001 have been introduced in [10].

Recent advances of T2 FSs have been largely attributed to the three-dimensional type-2 fuzzy membership functions (MFs) as shown in Fig. 1. The T2 MF evaluates the uncertainty of the input x by the *fuzzy* primary membership, $[\underline{h}(x), \overline{h}(x)]$, as shown in Fig. 1 (a), which is further evaluated by the secondary MF in Fig. 1 (b) and (c). The footprint of uncertainty (FOU) is the shaded region bounded by lower MF $\underline{h}(x)$ and upper MF $\overline{h}(x)$. The FOU reflects the amount of uncertainty in the primary membership, i.e., the larger (smaller) the amount of uncertainty, the larger (smaller) will the FOU be. Fig. 1 (b) shows an example of Gaussian secondary MF. An IT2 FS has an interval set secondary MF in Fig. 1 (c). Because all the secondary grades are unity, we can represent the IT2 FS by the interval of upper and lower MFs, i.e., $[\underline{h}(x), \overline{h}(x)]$. Operations on general T2 FSs [5], [7], *meet* "$\sqcap$" and *join* "$\sqcup$", involve an intractable combinatorial problem of the primary membership, whereas IT2 FSs [11] use only interval arithmetic leading to very simple operations. Without loss of generality, we focus on IT2 FSs for pattern classification unless otherwise stated.

The T2 MF can be viewed as a *union* of embedded T1 MFs with *fuzzy* parameters. Fig. 1 (a) can be viewed as the T1 Gaussian MF with uncertain mean μ, which is bounded by an interval $[\underline{\mu}, \overline{\mu}]$. We assume the mean vary anywhere in the interval, which results in the movement of the T1 MF to form the FOU. It is easy to see that if such movement is uniform, i.e., the mean has a uniform MF, then the FOU is also uniform with equal possibilities, so does the secondary MF in Fig. 1 (c). In other words, if the mean is a fuzzy variable [8] with uniform MF in Fig. 1 (a), the output $[\underline{h}(x), \overline{h}(x)]$ of the input x is also a fuzzy variable with uniform MF in Fig. 1 (c). However, if the mean is with Gaussian MF, the output is definitely not associated with Gaussian secondary MF in Fig. 1 (b). Therefore, in practice it is convenient to define the secondary MF directly without considering the MF of the fuzzy parameters of the original T1 MF, though we know that there is a complex relationship between MFs of fuzzy parameters and fuzzy outputs.

T2 FSs may be applicable when [5]:

1) The data-generating system is known to be time-varying but the mathematical description of the time-variability is unknown (e.g., as in mobile communications);
2) Measurement noise is non-stationary, and the mathematical description of the non-stationarity is unknown (e.g., as in a time-varying noise);
3) Features in a pattern recognition application have statistical attributes that are non-stationary, and the mathematical descriptions of the non-stationarity are unknown;
4) Knowledge is mined from a group of experts using questionnaires that involve uncertain words;
5) Linguistic terms are used that have a nonmeasurable domain.

Observe that pattern classification is concerned with all of them, which motivates us of using T2 FSs for handling uncertainties in pattern classification [12].

In the next section we discuss the types of uncertainty in pattern classification. In Section III we demonstrate that T2 FSs can provide additional information for pattern classification especially for outliers [13] using information theory. After integrating with other classifiers, T2 fuzzy classifiers may have the *potential* to outperform their counterparts. In Section IV we study recent T2 fuzzy classifiers for real-world problems, i.e., classification of MPEG VBR video traffic [14], evaluation of welded structures [15], speech recognition [16]–[18], handwritten Chinese character recognition [12], [19], [20], and classification of battlefield ground vehicles [21]. Based on these cases, we summarize a systematic method of applying T2 FSs to pattern classification. Section V discusses some implementation problems of T2 fuzzy classifier in terms of the complexity and performance trade-off.

II. Uncertainty in Pattern Classification

Pattern classification typically involves the partition of the unknown observation $\mathbf{X}$ (pattern) according to the class model (rule) $\lambda_\omega, 1 \leq \omega \leq C$. Fig. 2 shows a pattern classification system [13] including five basic components: sensing, segmentation, feature extraction (*feature space*), classification, and post-processing. This system reflects a functional relationship between the input and output decision. We shall choose a particular set or class of candidate functions known as *hypotheses* before we begin trying to determine the correct function. The ability of a hypothesis to correctly classify data

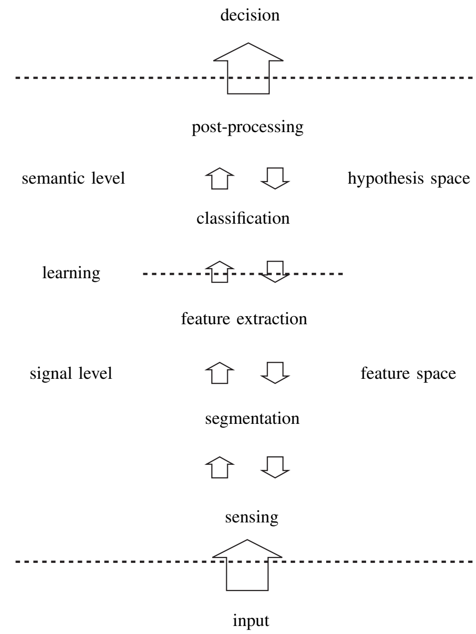

Fig. 2. The structure of the pattern classification system.

not in the training set is known as its *generalization*. The process of determining the correct function (often a number of adjustable parameters) on the basis of examples of input/output functionality is *learning*. Based on the above, we have three tasks in pattern classification:

1) Extract features that can be partitioned;
2) Choose a set of hypotheses that contains the correct representation of the decision function;
3) Design the learning algorithm that determines the *best* decision function from the feature and hypothesis spaces.

Inevitably uncertainties exist in both of the feature and hypothesis spaces. In statistical pattern recognition [13], we assume *randomness* in both spaces. In the feature space, random observations are generally expressed by the class-conditional probability density functions (PDFs). In the hypothesis space, the parameters of the decision function are random variables with some known prior distributions, and training data convert this distribution on the variables into posterior probability density. Whereas in T2 FSs we take all possibilities of uncertain parameters into account, Bayesian methods [13] select only the best precise parameters to maximize the posterior probability density. Thus classification is made by minimizing the probability of error. However, the insufficient and noisy training data often make the decision function not always the "best" in practice as shown in Fig. 3 (a) and (b). Furthermore, we find that randomness may be difficult to characterize the following uncertainties [12], [22], [23]:

1) Uncertain parameters of the decision function because of insufficient and noisy training data;
2) Non-stationary observation that has statistical attributes, and the mathematical description of the non-stationarity

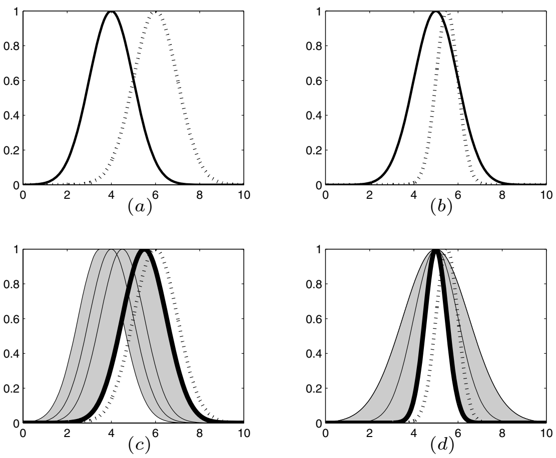

Fig. 3. In (a) and (b), the distribution of the training data and test data are the solid line and dotted line. Because of incomplete information and noise, the two distributions are not close. In (c) and (d), by incorporating uncertainty in the class model, i.e., letting the model move in a certain way, one of the models (the thick solid line) is probably to approximate the test data distribution. The shaded region is the "footprint" of the model uncertainty.

is unknown [5], [14];

3) Uncertain matching degree between the observation and class model.

One of the best sources of general discussion about uncertainty is Klir and Wierman [24]. Regarding the *nature of uncertainty*, they state that three types of uncertainty are now recognized:

1) *Fuzziness* (vagueness), which results from the imprecise boundaries of FSs;
2) *Non-specificity* (information-based imprecision), which is connected with sizes (cardinalities) of relevant sets of alternatives;
3) *Strife* (discord), which expresses conflicts among the various sets of alternatives.

Observe that the uncertainties in pattern classification may be certain fuzziness and non-specificity resulting from incomplete information, i.e., fuzzy decision functions (uncertain mapping), fuzzy observations (non-stationary data), and fuzzy match (uncertain matching degree).

For example, in Fig. 3 (a) and (b), the solid and dotted lines denote the distributions of the training and test data respectively. Because of incomplete information or noise, these two distributions are not close. In (c) and (d), if we assume that parameters of the distribution vary within an interval, one of the embedded distributions, denoted by the thick solid line, is probably to approximate the distribution of the test data. The "footprint" of the uncertainty reflects the degree of uncertainty in decision functions.

III. Motivation

In Section II we argue that some uncertainty is difficult to describe using randomness alone. Fuzziness is another important uncertainty that we have to handle in pattern classification.

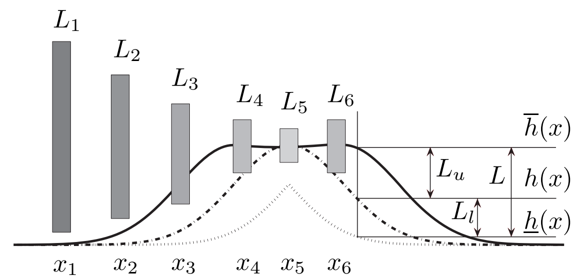

Fig. 4. The length $L = |\log \underline{h}(x) - \log \overline{h}(x)|$ describes the uncertainty of the input x to the T2 FS. The longer L the more uncertainty, which is marked by the darker gray. For example, x_1 deviates much from the mean, so it has not only a lower membership grade but a longer L_1 as well. Three intervals L, L_u, and L_l measure the uncertainty of the input x.

It is necessary to deal with both randomness and fuzziness within the same framework. Hence *fuzzy randomness* [25], [26], *fuzzy probability* [27], and *fuzzy statistics* [28] come into being. In contrast to these hybrid concepts, T2 FSs focus on the union of all possibilities of original T1 FSs simultaneously, which results in a fuzzy measurement of the primary membership grade called the secondary grade. The input of T2 MFs is the same with that of T1 MFs, but the output of T2 MFs is a fuzzy variable instead of a precise membership grade. We will explain later that such extension makes it possible to measure subtle distinction between patterns. Therefore, within T2 FSs framework, if we use the primary membership to describe the randomness in feature space, and use the secondary MF to describe the fuzziness of the primary membership, then *both kinds of uncertainties should be accounted for* [12], [16], [23]. Operations on T2 FSs propagate both randomness and fuzziness in the pattern classification system until final decision-making.

For analytical purpose, we often use the *log-likelihood* [13] in pattern classification. In the Gaussian case, the *maximum log-likelihood* estimation is equivalent to the *least squares* algorithm [29]. In Fig. 1 (a), the effect of fuzzy parameters of the Gaussian MF is that the likelihood becomes a fuzzy variable from a precise real number. This fuzzy variable contains more information of the input pattern x to the class model, which can be propagated by operations on T2 FSs. As shown in Fig. 4, T2 MFs measures each input x by a bounded interval set, $[\underline{h}(x), \overline{h}(x)]$, instead of a precise number $h(x)$ in T1 MFs or PDFs. In the case of Gaussian primary MF with uncertain mean (See Fig. 1 (a)), the upper boundary of the FOU is

$$\overline{h}(x) = \begin{cases} N(x; \underline{\mu}, \sigma), & x < \underline{\mu}; \\ 1, & \underline{\mu} \le x \le \overline{\mu}; \\ N(x; \overline{\mu}, \sigma), & x > \overline{\mu}, \end{cases} \tag{1}$$

and the lower boundary is

$$\underline{h}(x) = \begin{cases} N(x; \overline{\mu}, \sigma), & x \le \frac{\underline{\mu}+\overline{\mu}}{2}; \\ N(x; \underline{\mu}, \sigma), & x > \frac{\underline{\mu}+\overline{\mu}}{2}, \end{cases} \tag{2}$$

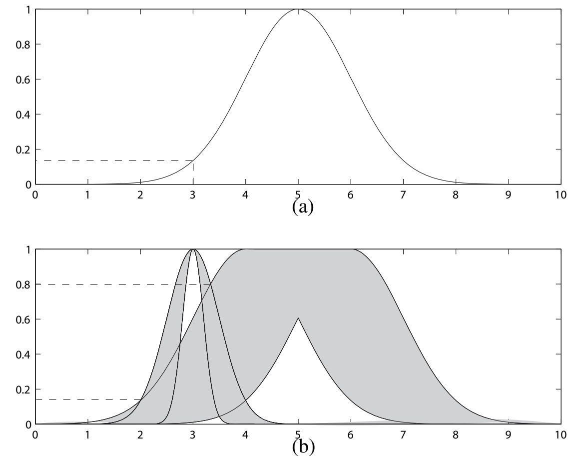

Fig. 5. (a) the singleton and (b) the T2 nonsingleton fuzzification.

where

$$N(x;\underline{\mu},\sigma) \triangleq \exp\left[-\frac{1}{2}\left(\frac{x-\underline{\mu}}{\sigma}\right)^2\right]. \tag{3}$$

The factor k [12], [16] controls the FOU,

$$\underline{\mu} = \mu - k\sigma, \quad \overline{\mu} = \mu + k\sigma, \quad k \in [0,3]. \tag{4}$$

Similar to the entropy of a uniform random variable, the uncertainty of the interval set is equal to the logarithm of the length of the interval [30]. Because we use the *log-likelihood* for classification, we are interested in lengths of three intervals, $L = |\log \overline{h} - \log \underline{h}|$, $L_l = |\log h - \log \underline{h}|$ and $L_u = |\log \overline{h} - \log h|$ as shown in Fig. 4. Given the factor k, we have the lengths (5)-(7). They are all increasing functions in terms of deviation $|x-\mu|$ and the factor k. For example, given a fixed k, the far the deviation of x from μ, the long the interval L in (5), which increases its entropy (uncertainty) [30]. This relationship accords with our prior knowledge. If the input x deviates far from the class model, so called *outlier* [13], [31], [32], it not only has a low membership grade $h(x)$, but also a long interval L reflecting its uncertainty to the class model. Indeed, we are often uncertain whether *outliers* belong to the class or not. From (5)-(7), we can see that k plays an important role in controlling uncertainty. If $k = 0$, then $L = L_l = L_u = 0$, which implies that no uncertainty exists so that the membership grade $h(x)$ is enough to make a classification decision. If k increases for a fixed deviation $|x-\mu|$, the length of interval increases representing more uncertainty of the input x. However, if k is large, L_l and L_u are long so that the two bounds $[\underline{h}, \overline{h}]$ will lose some information of the original $h(x)$.

In T2 fuzzy logic systems (FLSs) [11], [33], the nonsingleton fuzzification (NF) [34] is especially useful in cases where the available training data are corrupted by noise. Conceptually, the NF implies that the given input value is the most likely value to be the correct one from all the values in its immediate neighborhood; however, because the input is corrupted by noise, neighboring points are also likely to be the correct values. Fig. 5 compares the singleton fuzzification (SF) with the corresponding T2 NF. Besides uncertainty in data, T2 FSs are integrated with traditional classifiers to handle uncertainty in the hypothesis space [35]. For example, Liang and Mendel combined T2 FSs with T1 FLS-based classifiers for MPEG VBR video traffic classification [14]. Zeng and Liu integrated T2 FSs with hidden Markov model and Markov random fields for speech and handwritten Chinese character recognition [12], [16], [19], [20], [22], [23]. Wu and Mendel designed T2 FLS-based classifiers based on T1 counterparts for battlefield ground vehicles classification [21]. From these cases, we obtain a systematic design method in (8)-(9) to handle uncertain feature and hypothesis spaces in pattern classification.

TABLE I

CLASSIFICATION ERROR RATE COMPARISON (%) [15]

Dataset	T2 FSs	Benchmark
Welded Structures	5	6.8

IV. TYPE-2 FUZZY DATA AND CLASSIFIERS

This section reviews the state-of-the-art T2 fuzzy classification systems. We denote the class model with fuzzy parameters by the T2 FS, $\tilde{\lambda}_\omega, 1 \leq \omega \leq C$, where C is the number of classes. As discussed in Section III, the SF assumes no uncertainty in the feature space. The T2 (T1) NF models the observation as a T2 (T1) FS denoted by $\tilde{\mathbf{X}}$.

In [15] Mitchell viewed pattern classification as the similarity measure between two T2 FSs, in which one set accounts for the uncertain feature space in (8), and the other for the uncertain hypothesis space in (9). The task of pattern classification is equivalent to finding the class model which has the largest similarity between these two T2 FSs:

$$\omega^* = \arg\max_{\omega=1}^{C} S(\tilde{\mathbf{X}}, \tilde{\lambda}_\omega). \tag{10}$$

Mitchell [15] defined the similarity measure by the weighted average of ordinary similarity measure of embedded T1 FSs,

$$S(\tilde{A},\tilde{B}) = \sum_{m=1}^{M}\sum_{n=1}^{N} w_{mn} S(A_e^m, B_e^n), \tag{11}$$

where w_{mn} is the weight (secondary grade) with mth and nth embedded T1 sets, and there are totally M and N embedded T1 sets in $\tilde{A}$ and $\tilde{B}$ respectively. Automatic evaluation of welded structures using radiographic testing was modeled by T2 FSs, and the classification error rate is 1.8% lower than the benchmark (See Table I).

In [36] John et al. represented consultant's interpretation of the input images by T2 FSs, and classified images of sports injuries by neuro-fuzzy clustering. They preprocessed the expertise of clinicians using T2 FSs to describe the imprecise data in (8). They showed T2 fuzzy preprocessing and MINMAX clustering produced least confusion in relation to consultants judgements.

In [14] Liang and Mendel classified video traffic by T2 FLS-based classifiers extended from T1 FLS-based classifiers

$$L = \begin{cases} 2k|x-\mu|/\sigma, & x \le \mu - k\sigma,\ x \ge \mu + k\sigma; \\ |x-\mu|^2/2\sigma^2 + k|x-\mu|/\sigma + k^2/2, & \mu - k\sigma < x < \mu + k\sigma, \end{cases} \tag{5}$$

$$L_l = k|x-\mu|/\sigma + k^2/2, \tag{6}$$

$$L_u = \begin{cases} k|x-\mu|/\sigma + k^2/2, & x \le \mu - k\sigma,\ x \ge \mu + k\sigma; \\ |x-\mu|^2/2\sigma^2, & \mu - k\sigma < x < \mu + k\sigma. \end{cases} \tag{7}$$

uncertain feature space: data + T2 FSs = T2 fuzzy data (noise or non-stationarity) (8)

uncertain hypothesis space: classifier + T2 FSs = T2 fuzzy classifier (unknown varieties of parameters) (9)

as in (9), and showed better performance than the Bayesian classifiers when features have statistical attributes that are non-stationary. Firstly, they designed the T1 FLS-based classifiers as follows. Consider the observation, $\mathbf{x} = [x_1, x_2, \ldots, x_d]'$, and two class models λ_1 and λ_2. For T1 fuzzy classifiers with a rule base of M rules, each having d antecedents, the lth rule, $R^l, 1 \le l \le M$, is

$$R^l:\ \text{IF } x_1 \text{ is } F_1^l \text{ and } \ldots \text{and } x_d \text{ is } F_d^l, \text{ THEN}$$
$$\mathbf{x} \text{ is classified to } \lambda_1\ (+1) \text{ [or is classified to } \lambda_2\ (-1)]. \tag{12}$$

Suppose that the antecedents $F_i^l, 1 \le i \le d$, are described by a T1 Gaussian MF,

$$h_{F_i^l}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]. \tag{13}$$

They used the unnormalized output in the T1 FLS (the firing strength of each rule is denoted by f^l), namely,

$$y = \sum_{l=1}^{M}(f_{\lambda_1}^l - f_{\lambda_2}^l), \tag{14}$$

and made a decision based on the sign of the output ($y > 0, \mathbf{x} \to \lambda_1$). Secondly, they extended T1 FLS-based classifiers to T2 FLS-based classifiers with a rule base of M rules, the lth rule, $R^l, 1 \le l \le M$, is

$$R^l:\ \text{IF } \tilde{x}_1 \text{ is } \tilde{F}_1^l \text{ and } \ldots \text{and } \tilde{x}_d \text{ is } \tilde{F}_d^l, \text{ THEN}$$
$$\tilde{\mathbf{x}} \text{ is classified to } \tilde{\lambda}_1\ (+1) \text{ [or is classified to } \tilde{\lambda}_2\ (-1)]. \tag{15}$$

Suppose that the antecedents $\tilde{F}_i^l, 1 \le i \le d$ are described by a T2 Gaussian primary MF with uncertain mean or standard deviation [11], [14], [21]. Similar to (14), the output of the T2 FLS,

$$\tilde{y} = \sqcup_{l=1}^{M}(\tilde{f}_{\lambda_1}^l - \tilde{f}_{\lambda_2}^l), \tag{16}$$

which is an interval rather than a precise number in (14). For comparison, they also designed the Bayesian classifier as follows. If equal prior class probability is assumed, the Bayesian classifiers are

$$p(\mathbf{x}|\lambda_1) = \sum_{l=1}^{m} p(\mathbf{x}|\lambda_1^l), \tag{17}$$

$$p(\mathbf{x}|\lambda_2) = \sum_{l=1}^{n} p(\mathbf{x}|\lambda_2^l), \tag{18}$$

where the number of prototypes of class λ_1 and λ_2 is m and n respectively. If the conditional probability of each prototype is described by the Gaussian distribution,

$$p(\mathbf{x}|\lambda) = \frac{1}{\sqrt{(2\pi)^d|\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{19}$$

where the mean vector, $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_d]'$, and the diagonal covariance matrix, $\mathbf{\Sigma} = \mathit{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2)$. According to the Bayesian decision theory [13], the optimal decision rule is

$$\text{IF } p(\mathbf{x}|\lambda_1) - p(\mathbf{x}|\lambda_2) > 0, \text{ THEN } \mathbf{x} \text{ is classified to } \lambda_1, \tag{20}$$

$$\text{IF } p(\mathbf{x}|\lambda_1) - p(\mathbf{x}|\lambda_2) < 0, \text{ THEN } \mathbf{x} \text{ is classified to } \lambda_2. \tag{21}$$

Observe (14), (20), and (21) that the class model in the Bayesian classifier has a correspondence with each rule in the T1 classifier. We find that the T1 FLS-based classifier is mathematically the same with the Bayesian classifier except the normalization factor $1/\sqrt{(2\pi)^d|\mathbf{\Sigma}|}$ in (19), which generally does not affect the classification results so that there is no essential distinction between T1 fuzzy classifiers and Bayesian classifiers. However, T2 FLS-based classifiers may make a quite different decision from output interval in (16).

In MPEG VBR video traffic classification (out-of-product testing) without parameter adjustment, Liang and Mendel [14] reported the lowest average false alarm rate 14.11% for T1 NF data with T2 FLS-based classifiers (T1NFT2), which was slightly lower than the average 15.07% for SF data with T1 fuzzy classifiers (SFT1) as well as the average 14.29% for Bayesain classifiers (BC) as shown in Table II. Furthermore, they adjusted parameters of fuzzy classifiers by the steepest-descent algorithm, and obtained the lowest average false alarm rate 8.03% for T2 NF data with T2 FLS-based classifiers (T2NFT2) , which was also slightly lower than the average

TABLE II

FALSE ALARM RATE COMPARISON (%) [14]

Classifiers	Without parameter adjustment	Parameter adjustment
BC	14.29	-
SFT1	15.07	9.41
T1NFT1	14.35	9.17
SFT2	14.24	13.65
T1NFT2	14.11	8.43
T2NFT2	14.35	8.03

TABLE III

CLASSIFICATION ERROR RATE COMPARISON (%) [21]

Datasets	T2 classifiers	T1 classifiers
Battlefield ground vehicle	9.13	12.8

9.17% for T1 NF with T1 FLS-based classifiers (T1NFT1). So they concluded that T2 fuzzy classifiers were substantially better than their T1 counterparts in terms of robustness and classification error rate.

Similarly, Wu and Mendel [21] used T2 FLS-based classifiers to classify multi-category battlefield ground vehicles, and demonstrated that T2 FSs can model unknown varieties of features. They reduced the average classification error rates of T1 FLS-based classifiers by T2 FLS-based classifiers from 12.8% to 9.13% over more than 800 experiments (See Table III). Besides, they showed that all FLS-based classifiers performed much better than the Bayesian classifiers.

In [12], [16]–[20], [22], [23] we view pattern classification as the labeling problem, which is in fact a compound Bayesian decision problem [13]. The solution is a set of linguistic labels, $1 \leq j \leq J$, assigned to a set of sites, $1 \leq i \leq I$, to explain the observation, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_I\}$, at all sites. The label j at site i is a random variable, so that the labeling configuration at all sites, $\mathcal{F} = \{f_1, f_2, \ldots, f_I\}$, is a stochastic process. Given the model λ, the *maximum a posteriori* (MAP) estimation [13] guarantees the single best labeling configuration,

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} P(\mathcal{F}|\mathbf{X}, \lambda), \tag{22}$$

$$P(\mathcal{F}|\mathbf{X}, \lambda) \propto p(\mathbf{X}|\mathcal{F}, \lambda)P(\mathcal{F}|\lambda), \tag{23}$$

where $p(\mathbf{X}|\mathcal{F}, \lambda)$ is the likelihood function for $\mathcal{F}$ given $\mathbf{X}$, and $P(\mathcal{F}|\lambda)$ is the prior probability of $\mathcal{F}$. However, because of fuzzy data and class model, we incorporate T2 FSs into MAP (22)-(23) as follows,

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} h_{\tilde{\lambda}}(\mathcal{F}|\mathbf{X}), \tag{24}$$

$$h_{\tilde{\lambda}}(\mathcal{F}|\mathbf{X}) \propto h_{\tilde{\lambda}}(\mathbf{X}|\mathcal{F}) \sqcap h_{\tilde{\lambda}}(\mathcal{F}), \tag{25}$$

where $\tilde{\lambda}$ is the class model with fuzzy parameters. We use NF to handle fuzzy observations due to noise. Set operations in (25) convey more information than (23) because we unite all possibilities of the class model due to fuzzy parameters into T2 FSs. Especially when $\tilde{\lambda}$ is certain, equation (25) will be reduced to (23). The T2 FSs $h_{\tilde{\lambda}}(\mathbf{X}|\mathcal{F})$ and $h_{\tilde{\lambda}}(\mathcal{F})$ describe *fuzziness* of the likelihood and prior respectively within the Bayesian framework.

TABLE IV

CLASSIFICATION RATE (%) COMPARISON [22]

Datasets	T2 FGMMs	GMMs
IONOSPHERE	77.7	75.3
PENDIGITS	93.3	90.2
WDBC	94.9	93.6
WINE	89.2	85.8

TABLE V

CLASSIFICATION RATE (%) COMPARISON [22]

Classifiers	clean	20db	10db	50db	0db	-5db	-10db
T2 FHMMs	58.1	47.5	32.4	24.2	16.9	11.4	7.0
HMMs	54.9	45.1	30.7	22.6	15.4	10.0	5.9

In [22] we have integrated T2 FSs with Gaussian mixture models (GMMs) referred to as the T2 FGMMs, which describes fuzzy likelihoods by lower and upper boundaries of the FOU. In the proposed classification system, we use the generalized linear model (GLM) to make the final decision from fuzzy likelihoods. Extensive experiments on datasets from UCI repository [37] demonstrate that T2 FGMMs have an average 2.5% (the best results) higher classification rate than that of GMMs (See Table IV). Based on (24)-(25), we also extend the T2 FGMMs-based hidden Markov model (HMM) referred to as the T2 FHMM. Forty-six-category phonemes were classified using T2 FHMMs. To test robustness, we also classified the phonemes corrupted by multi-talker non-stationary babble noise with different signal-noise-ratios (SNRs). Table V shows the best results of T2 FHMMs compared to HMMs. We can see that on average T2 FHMMs outperform HMMs 1.85% in classification rate under babble noise with different SNRs.

In [12], [16]–[18] we have used T2 NF to describe fuzzy observations, and modeled the fuzzy transition probability by fuzzy numbers in T2 FHMMs. In this classification system, we propose a heuristic ranking of output fuzzy likelihoods. A broad-five-category phoneme classification shows that a significant improvement (7.03% on average) in classification rate when the white Gaussian noise is added to test data with different SNRs (See Table VI). Furthermore, a complete continuous phoneme recognition experiment demonstrates that T2 FHMMs outperform the competing HMMs 5.55% in dialect recognition accuracy (See Table VII).

Similarly, in [12], [19], [20] we have integrated T2 FSs with Markov random fields (MRFs) referred to as the T2 FMRFs for Chinese character modeling. From experiments on similar characters [20], we demonstrate that T2 FSs improve the

TABLE VI

CLASSIFICATION RATE (%) COMPARISON [16]

Classifiers	5dB	10dB	15dB	20dB	25dB	30dB
T2NF FHMMs	50.6	59.9	65.4	71.3	75.1	79.3
HMMs	38.7	48.0	58.2	66.0	72.3	76.2

TABLE VII
RECOGNITION ACCURACY COMPARISON (%) [16]

Datasets	T2NF FHMMs	HMMs
TIMIT phoneme	62.94	62.59
TIMIT dialect	56.94	51.39

TABLE VIII
CLASSIFICATION ERROR RATE COMPARISON (%) [20]

Datasets	T2 FMRFs	MRFs
ETL-9B / ETL-9B	3.11	4.25
Hanja1 / Hanja1	3.29	4.67

performance of the MRFs for Chinese character recognition by 1.26% in classification rate on average (See Table VIII). Furthermore, a generalization ability comparison (See Table IX) shows that T2 FMRFs have a better performance (2.63% on average) in classifying unknown Chinese character patterns from different datasets.

In conclusion, the strategies (8) and (9) are effective in most pattern classification problems. The T2 fuzzy data (8) and T2 fuzzy classifier (9) compose a T2 fuzzy pattern classification system which generally has a better performance than the competing T1 and Bayesian classifiers. Though in some cases the T2 fuzzy system degrades a little than traditional methods, it is still a reliable approach to improve classification ability of the traditional methods in terms of robustness, generalization ability, and classification error rates. Note that, at present, *there is no theory that guarantees that a T2 fuzzy system will always do this* [10].

V. DISCUSSIONS

Occam's razor [13] has come to be interpreted in pattern classification as counseling that one should not use classifiers that are more complicated than are necessary, where "necessary" is determined by the quality of fit to the training data. Indeed, T2 fuzzy systems have more parameters with higher computational complexity than their counterparts such as T1 fuzzy and Bayesian systems [14]–[16], [21], [35], [36]. In most cases, at least twice computations [12] have to be done in T2 fuzzy systems than traditional methods. Therefore, when we apply T2 fuzzy classifiers to real-world problems, we should consider if the problem at hand is needed to pay more complexity. Currently, we say that T2 fuzzy systems have the *potential* to outperform the traditional ones, but in the meantime they add more complexity to the system leading to the performance-complexity trade-off.

TABLE IX
CLASSIFICATION ERROR RATE COMPARISON (%) [20]

Datasets	T2 FMRFs	MRFs
ETL-9B / Hanja1	4.44	6.78
Hanja1 / ETL-9B	4.16	7.08

No Free Lunch Theorem [13] tells us that there are no context-independent or usage-independent reasons to favor one learning or classification method over another. Looking back to the strategies (8) and (9), T2 fuzzy systems are natural extensions of the original classification systems, which means the classification performance has been already ensured, and T2 FSs just improve it. More importantly, we should note that T2 fuzzy systems do not always outperform their counterparts in all pattern classification problems. Also T2 fuzzy systems are not always effective for modeling uncertainties [21]. The major reason is that the FOU may cover too much or too little uncertainty that the system does not have. Another reason is that we may use the ineffective method for final decision-making.

The great success of statistical pattern classification as well as Bayesian decision theory [13] has been attributed to the recognition of *randomness* in the feature and hypothesis spaces. Now we realize that it is necessary to incorporate *fuzziness* into the same framework to solve real-world problems. In Section III we have explained the mechanism of T2 FSs to handle both randomness and fuzziness and demonstrated that T2 FSs have more expressive power to tackle more difficult problems. From many case studies, we obtain the design methods for classification systems in (8) and (9), and further implement them within the Bayesian framework in (24) and (25). Based on recent encouraging experimental results, we are optimistic about the future of T2 FSs for pattern classification applications.

REFERENCES

[1] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Sciences*, vol. 8, pp. 199–249, 1975.
[2] M. Mizumoto and K. Tanaka, "Some properties of fuzzy sets of type-2," *Information and Control*, vol. 31, pp. 312–340, 1976.
[3] ——, "Fuzzy sets of type-2 under algebraic product and algebraic sum," *Fuzzy Sets and Systems*, vol. 5, pp. 277–290, 1981.
[4] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty and Information*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[5] J. M. Mendel and R. I. B. John, "Type-2 fuzzy sets made simple," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 117–127, Apr. 2002.
[6] H. Mitchell, "Ranking type-2 fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 287–294, Apr. 2006.
[7] N. N. Karnik and J. M. Mendel, "Operations on type-2 fuzzy sets," *Fuzzy Sets and Systems*, vol. 122, pp. 327–348, 2001.
[8] Z.-Q. Liu and Y.-K. Liu, "Type-2 fuzzy variables," *IEEE Trans. Fuzzy Syst.*, submitted for publication, 2006.
[9] J. M. Mendel, *Uncertain Rule-based Fuzzy Logic Systems : Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.
[10] ——, "Advances in type-2 fuzzy sets and systems," *Information Sciences*, vol. 177, pp. 84–110, 2007.
[11] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: Theory and design," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 535–549, Oct. 2000.
[12] J. Zeng and Z.-Q. Liu, "Type-2 fuzzy sets for handling uncertainty in pattern recognition," *Proc. FUZZ-IEEE*, pp. 6597–6602, 2006.
[13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, 2001.
[14] Q. Liang and J. M. Mendel, "MPEG VBR video traffic modeling and classification using fuzzy technique," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 1, pp. 183–193, Oct. 2001.
[15] H. Mitchell, "Pattern recognition using type-II fuzzy sets," *Information Sciences*, vol. 170, pp. 409–418, 2005.

[16] J. Zeng and Z.-Q. Liu, "Type-2 fuzzy hidden Markov models and their application to speech recognition," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 3, pp. 454–467, June 2006.
[17] ——, "Interval type-2 fuzzy hidden Markov models," *Proc. FUZZ-IEEE*, pp. 1123–1128, 2004.
[18] ——, "Type-2 fuzzy hidden Markov models to phoneme recognition," *17th Int'l Conf. on Pattern Recognition*, vol. 1, pp. 192–195, 2004.
[19] ——, "Type-2 fuzzy Markov random fields to handwritten character recognition," *18th Int'l Conf. on Pattern Recognition*, pp. 1162–1165, 2006.
[20] ——, "Type-2 fuzzy Markov random fields and their application to handwritten Chinese character recognition," *IEEE Trans. Fuzzy Syst.*, submitted for publication, 2006.
[21] H. Wu and J. M. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Trans. Fuzzy Syst.*, accepted by special issue on Extensions to Type-1 Fuzzy Sets, 2006.
[22] J. Zeng, L. Xie, and Z.-Q. Liu, "Gaussian mixture models with uncertain parameters," *Pattern Recognition, submitted for publication*, 2007.
[23] J. Zeng, "Type-2 fuzzy sets, markov processes for pattern recognition," Ph.D. dissertation, City University of Hong Kong, Hong Kong, 2006.
[24] G. J. Klir and M. J. Wierman, *Uncertainty-based information*. New York: Physica-Verlag, 1998.
[25] B. Möller and M. Beer, *Fuzzy Randomness : Uncertainty in Civil Engineering and Computational Mechanics*. New York: Springer, 2004.
[26] B. Liu, *Uncertainty Theory : An Introduction to Its Axiomatic Foundations*. Berlin: Springer, 2004.
[27] J. J. Buckley, *Fuzzy Probabilities : New Approach and Applications*. New York: Physica-Verlag, 2003.
[28] ——, *Fuzzy Statistics*. New York: Springer, 2004.
[29] T. M. Mitchell, *Machine Learning*. Boston, Mass.: McGraw-Hill, 1997.
[30] H. Wu and J. M. Mendel, "Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 5, pp. 622–639, Oct. 2002.
[31] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford University Press, 1995.
[32] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[33] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 643–658, Dec. 1999.
[34] G. C. Mouzouris and J. M. Mendel, "Nonsingleton fuzzy logic systems: Theory and application," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 1, pp. 56–71, Feb. 1997.
[35] J. Zeng and Z.-Q. Liu, "Stroke segmentation of Chinese characters using Markov random fields," *18th Int'l Conf. on Pattern Recognition*, pp. 868–871, 2006.
[36] R. I. John, P. R. Innocent, and M. R. Barnes, "Neuro-fuzzy clustering of radiographic tibia image data using type-2 fuzzy sets," *Information Sciences*, vol. 125, pp. 65–82, June 2000.
[37] D. Newman, S. Hettich, C. Blake, and C. Merz, "UCI repository of machine learning databases," 1998. [Online]. Available: http://www.ics.uci.edu/$\sim$mlearn/MLRepository.html