Journal of Uncertain Systems
Vol.1, No.3, pp.163-177, 2007
Online at: www.jus.org.uk

Type-2 Fuzzy Sets for Pattern Recognition: The State-of-the-Art

Jia Zeng* and Zhi-Qiang Liu
School of Creative Media, City University of Hong Kong, China

Received 25 December 2006; Accepted 30 January 2007

Abstract

The 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 recognition, both feature and hypothesis spaces have uncertainties, which motivate us of integrating type-2 fuzzy sets with conventional classifiers to achieve a better performance in terms of the robustness, generalization ability, or recognition accuracy. In this state-of-the-art paper, we describe important advances of type-2 fuzzy sets for pattern recognition. Interests in type-2 fuzzy sets and systems is worldwide and touches on a broad range of applications and theoretical topics. The main focus of this paper is on the pattern recognition applications, with descriptions of how to design, what has been achieved, and what remains to be done.

1 Introduction

The advances of type-2 fuzzy sets (T2 FSs) and systems [1] have been largely attributed to their three-dimensional membership functions (MFs). As an extension of type-1 fuzzy sets (T1 FSs), T2 FSs were initially introduced by Zadeh [2], and a subsequent investigation of properties of T2 FSs and higher types was done by Mizumoto and Tanaka [3, 4]. Klir and Folger [5] explained that the T1 MFs might be problematical, because a representation of fuzziness is made using membership grades that are themselves precise real numbers. Thus it is 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 [6] 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 order T2 fuzzy numbers, Mitchell [7] 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 [3]. Their works were later extended by Karnik and Mendel [8] for practical algorithms to perform the union, intersection, and complement between T2 FSs. In [6] Mendel and John reformulated all set operations in both vertical-slice and wavy-slice manners. They concluded that practically general T2 FSs operations, *meet* "$\sqcap$" and *join* "$\sqcup$", are too complex to implement, whereas IT2 FSs [9] use only interval arithmetics leading to very simple operations. Without loss of generality, we focus on IT2 FSs for pattern recognition unless otherwise stated. As the theoretical foundation of T2 FSs, Liu and Liu [10] established T2 fuzzy possibility theory and introduced T2 fuzzy variables. In [11] Mendel summarized developments and applications of T2 FSs before the year 2001. The advances of theoretical and computational issues in T2 fuzzy sets and systems since the year 2001 can be found in [1].

The T2 MF evaluates the uncertainty of the input x by the *fuzzy* primary membership, which is bounded by the lower MF $\underline{h}(x)$ and the upper MF $\overline{h}(x)$ as shown in Fig. 1 (a). The fuzzy membership is further described by the secondary MF in Fig. 1 (b) or (c). The footprint of uncertainty (FOU) is the shaded region

*Corresponding author: J. Zeng (j.zeng@ieee.org)

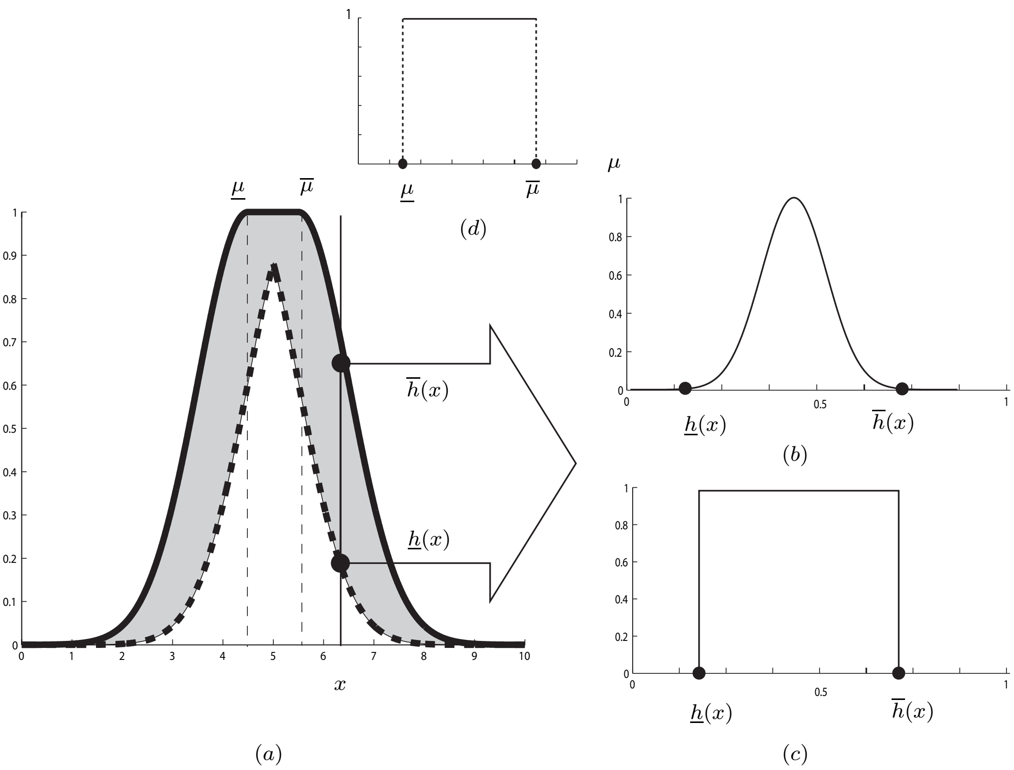

Figure 1: The three-dimensional type-2 fuzzy membership function. (a) shows the primary membership with the lower (thick dashed line) and upper (thick solid line) membership functions, where $\underline{h}(x)$ and $\overline{h}(x)$ are the lower and upper bounds given the input x. The shaded region is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

bounded by lower and upper MFs. 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., $[\underline{h}(x), \overline{h}(x)]$. In this case, an IT2 FS can be completely described by its two-dimensional FOU in Fig. 1 (a). The result of the calculation between the fuzzy primary memberships of IT2 FSs is also an fuzzy variable with uniform possibilities according to the interval arithmetic [9]. For example, if the two fuzzy primary memberships are $[\underline{h}_1, \overline{h}_1]$ and $[\underline{h}_2, \overline{h}_2]$, then the sum and product are $[\underline{h}_1 + \underline{h}_2, \overline{h}_1 + \overline{h}_2]$ and $[\underline{h}_1\underline{h}_2, \overline{h}_1\overline{h}_2]$.

The T2 MF can be viewed as an *ensemble* of embedded T1 MFs with *fuzzy* parameters. Fig. 1 (a) is the T1 Gaussian MF with fuzzy mean μ, which is bounded by an interval $[\underline{\mu}, \overline{\mu}]$. We assume the mean vary anywhere in this interval, which results in the movement of the T1 MF to form the FOU in Fig 1 (a). We see that if such movement is uniform, i.e., the mean has a uniform MF in Fig. 1 (d), then the FOU is also uniform with equal possibilities, so does the secondary MF in Fig. 1 (c). More specifically, if the mean is a fuzzy variable [10] with the uniform MF in Fig. 1 (a), the output $[\underline{h}(x), \overline{h}(x)]$ of the input x is also a fuzzy variable with the uniform MF in Fig. 1 (c). However, if the mean is with the Gaussian MF, the output is definitely not associated with the 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 [6]:

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 recognition is concerned with all of situations, which motivates us of using T2 FSs for handling uncertainties in pattern recognition [12].

In the next section we discuss the types of uncertainty in pattern recognition. In Section 3 we demonstrate by information theory that T2 FSs can provide additional information for pattern recognition especially for outliers. After integrating with other classifiers, T2 fuzzy systems may have the *potential* to outperform their counterparts. In Section 4 we study the recent T2 fuzzy pattern recognition systems for real-world problems, i.e., classification of MPEG VBR video traffic [13], evaluation of welded structures [14], speech recognition [15–17], handwritten Chinese character recognition [12, 18, 19], and classification of battlefield ground vehicles [20]. Based on these systems, we summarize a systematic method of applying T2 FSs to pattern recognition. Section 5 discusses some implementation problems of T2 fuzzy systems in terms of the complexity and performance trade-offs.

2 Uncertainty in pattern recognition

Pattern recognition typically involves the partition of the unknown observation $\mathbf{X}$ (pattern) according to the class model (rule) $\lambda_\omega, 1 \leq \omega \leq C$, where C is the number of classes. Fig. 2 shows a pattern recognition system [21, Chapter 1.3] 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 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 recognition:

1. Extract features that can be partitioned;

2. Choose the 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 there are uncertainties in both of the feature and hypothesis spaces. In statistical pattern recognition, 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 in T1 FSs into account, Bayesian methods [21, Chapter 3.3] 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" as shown in Fig. 3 (a) and (b). Furthermore, we find that randomness may be difficult to characterize the following uncertainties [12, 19, 22]:

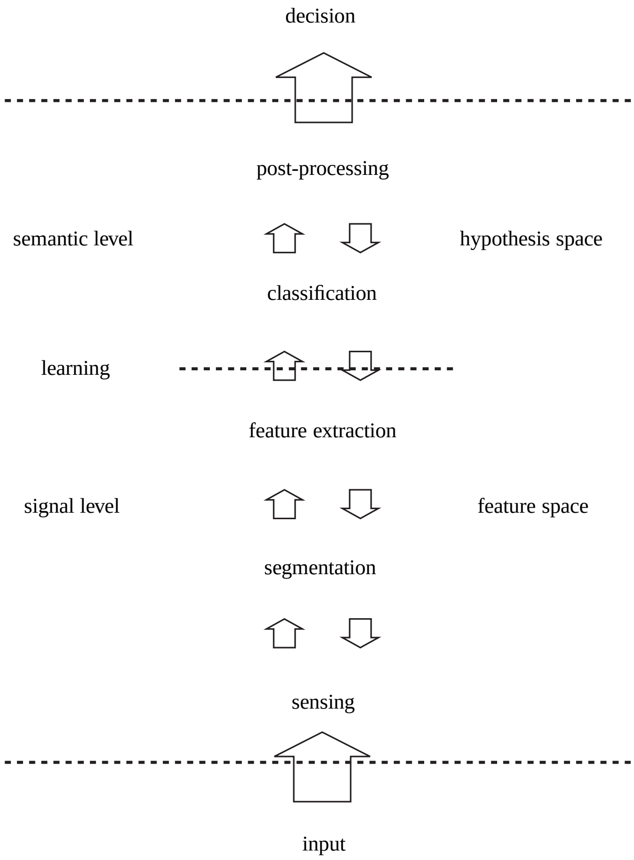

Figure 2: The structure of the pattern recognition system.

1. Uncertain parameters of the decision function because of the insufficient and noisy training data;

2. Non-stationary observation that has statistical attributes, and the mathematical description of the non-stationarity is unknown [6, 13];

3. Uncertain measurement of the matching degree between the observation and class model.

One of the best sources of general discussion about uncertainty is Klir and Wierman [23]. 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 types of uncertainty in pattern recognition 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 similarity 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.

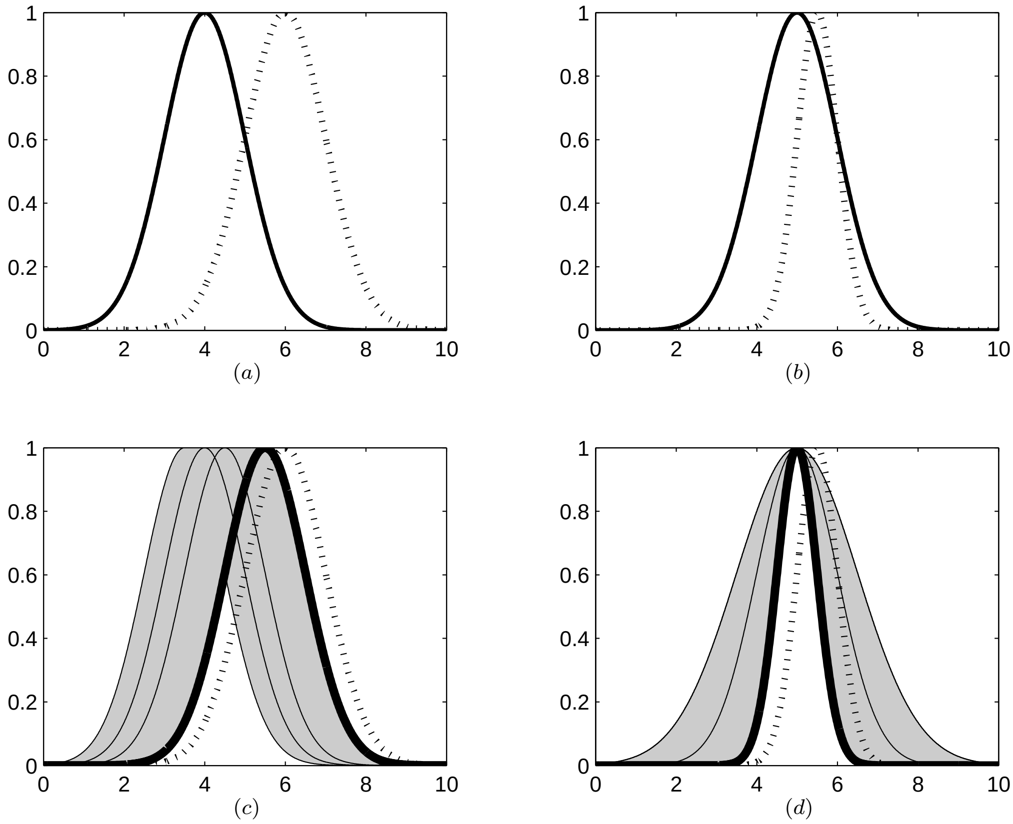

Figure 3: In (a) and (b), the distribution of the training data and test data are the solid line and dotted line. Because of incomplete information and noise, the two distributions are not close. In (c) and (d), by incorporating uncertainty in the class model, i.e., letting the model move in a certain 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 hypothesis uncertainty.

3 Motivation

In Section 2 we argue that some uncertainty is difficult to describe using randomness alone. Fuzziness is another important uncertainty that we have to handle in pattern recognition. It is necessary to deal with both randomness and fuzziness within the same framework. Hence *fuzzy randomness* [24, 25], *fuzzy probability* [26], and *fuzzy statistics* [27] come into being. In contrast to these hybrid concepts, T2 FSs focus on the ensemble of all possibilities of original T1 FSs simultaneously, which result in an additional measurement of the fuzzy 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 an ensemble representation makes it possible to measure subtle distinctions between patterns. Therefore, within the T2 FSs framework, if we use the primary membership to describe the randomness in the 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, 15, 22]. Furthermore, T2 FSs operations can propagate both randomness and fuzziness in the pattern recognition system until the final decision-making.

For analytical purpose, we often use the *log-likelihood* [21, pp. 86] in pattern recognition. In the case of Gaussian distributions, the *maximum log-likelihood* estimation is equivalent to the *least squares* algorithm [28]. In Fig. 4 (a) and (b), the effect of fuzzy parameters of the Gaussian primary MF is that the likelihood becomes a fuzzy variable from a precise real number. This fuzzy variable contains more informa-

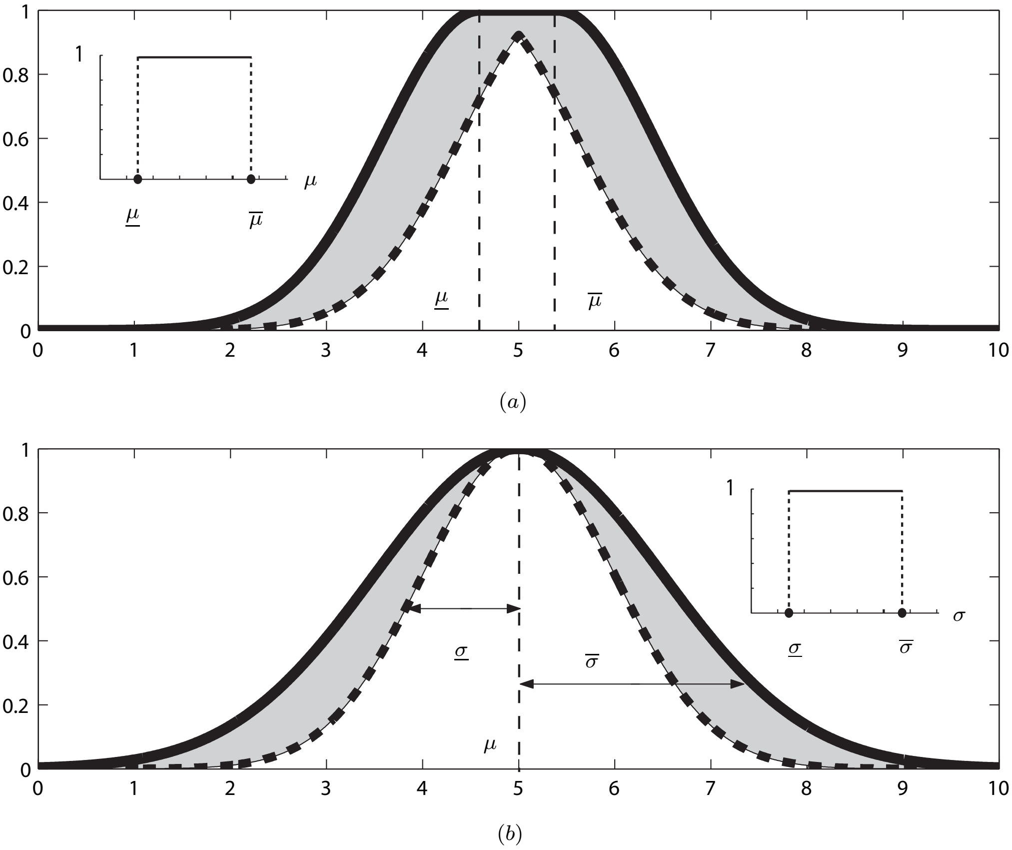

Figure 4: The Gaussian with uncertain mean (a) and std (b). The mean (a) and std (b) are fuzzy variables with uniform possibilities. The shaded region is the FOU. The thick solid and dashed lines denote the lower and upper boundaries of the FOU.

tion of the input pattern x to the class model, which can be propagated by operations on T2 FSs. In the case of Gaussian primary MF with uncertain mean (See Fig. 4 (a)) [9], the upper boundary of the FOU is

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

and the lower boundary is

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

where

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

In the case of the Gaussian with uncertain standard deviation (std) [9] (See Fig. 4 (b)), the upper MF is

$$\overline{h}(x) = N(x;\mu,\overline{\sigma}), \tag{4}$$

and the lower MF is

$$\underline{h}(x) = N(x;\mu,\underline{\sigma}). \tag{5}$$

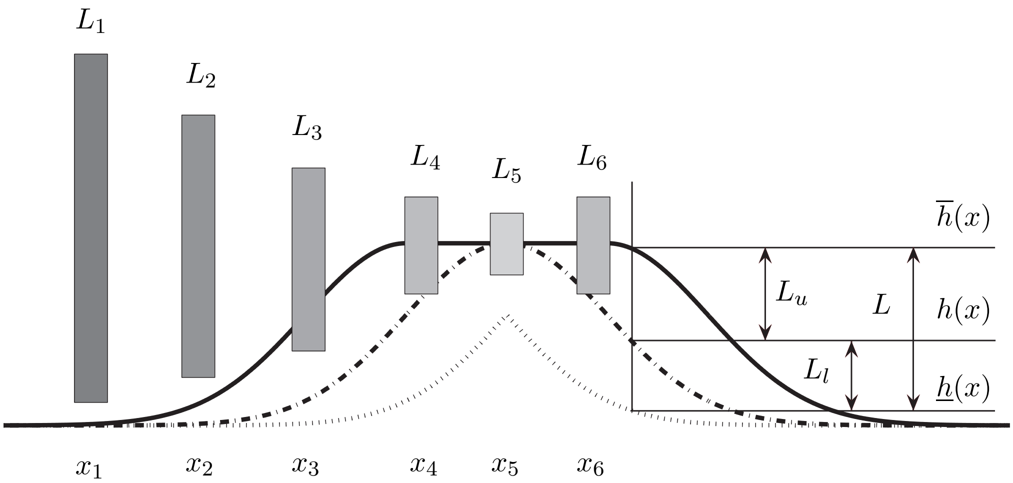

Figure 5: The length $L = |\ln \underline{h}(x) - \ln \overline{h}(x)|$ describes the uncertainty of the class model to the input x. The longer L the more uncertainty, which is marked by the darker gray. For example, x_1 deviates farther from the mean, so it has not only a lower membership grade but a longer L_1 as well. Three intervals L, L_u, and L_l measure the uncertainty of the class model.

$$L = \begin{cases} 2k|x-\mu|/\sigma, & x \leq \mu - k\sigma,\ x \geq \mu + k\sigma; \\ |x-\mu|^2/2\sigma^2 + k|x-\mu|/\sigma + k^2/2, & \mu - k\sigma < x < \mu + k\sigma, \end{cases} \tag{8}$$

$$L_l = k|x-\mu|/\sigma + k^2/2, \tag{9}$$

$$L_u = \begin{cases} k|x-\mu|/\sigma + k^2/2, & x \leq \mu - k\sigma,\ x \geq \mu + k\sigma; \\ |x-\mu|^2/2\sigma^2, & \mu - k\sigma < x < \mu + k\sigma. \end{cases} \tag{10}$$

The factor k [12, 15] controls the FOU,

$$\underline{\mu} = \mu - k\sigma, \quad \overline{\mu} = \mu + k\sigma, \quad k \in [0, 3], \tag{6}$$

$$\underline{\sigma} = k\sigma, \quad \overline{\sigma} = \frac{1}{k}\sigma, \quad k \in [0.3, 1]. \tag{7}$$

Because a one-dimensional gaussian has 99.7% of its probability mass in the range of $[\mu - 3\sigma, \mu + 3\sigma]$, we constrain $k \in [0, 3]$ in (6) and $k \in [0.3, 1]$ in (7).

We take the Gaussian primary MF with uncertain mean as an example to explain why T2 FSs can handle uncertainties for outliers [19]. In Fig. 5 the T2 MF evaluates each input x by a bounded interval set, $[\underline{h}(x), \overline{h}(x)]$, rather than a precise number $h(x)$ in the T1 MF or PDF. Similar to the entropy of a uniform random variable, the uncertainty of the interval set is equal to the logarithm of the length of that interval [29]. Because we use the *log-likelihood* in pattern recognition, we are interested in the lengths of three intervals, $L = |\ln \overline{h} - \ln \underline{h}|$, $L_l = |\ln h - \ln \underline{h}|$ and $L_u = |\ln \overline{h} - \ln h|$ as shown in Fig. 5. Given the factor k, we have three lengths (8)-(10), which are all increasing functions in terms of the deviation $|x - \mu|$ and the factor k. For example, given a fixed k, the farther the deviation of x from μ, the longer the interval L in (8), which in the meantime increases the entropy (uncertainty). This relationship accords with our prior knowledge. If the input x deviates farther from the class model, so called *outlier* [21,30,31], it not only has a lower membership grade $h(x)$, but also a longer interval L reflecting its uncertainty to the class model. Indeed, we are often uncertain whether *outliers* belong to this class or not. From (8)-(10), we see that k plays an important role in controlling uncertainty of decision functions. If $k = 0$, then $L = L_l = L_u = 0$, which implies that there is

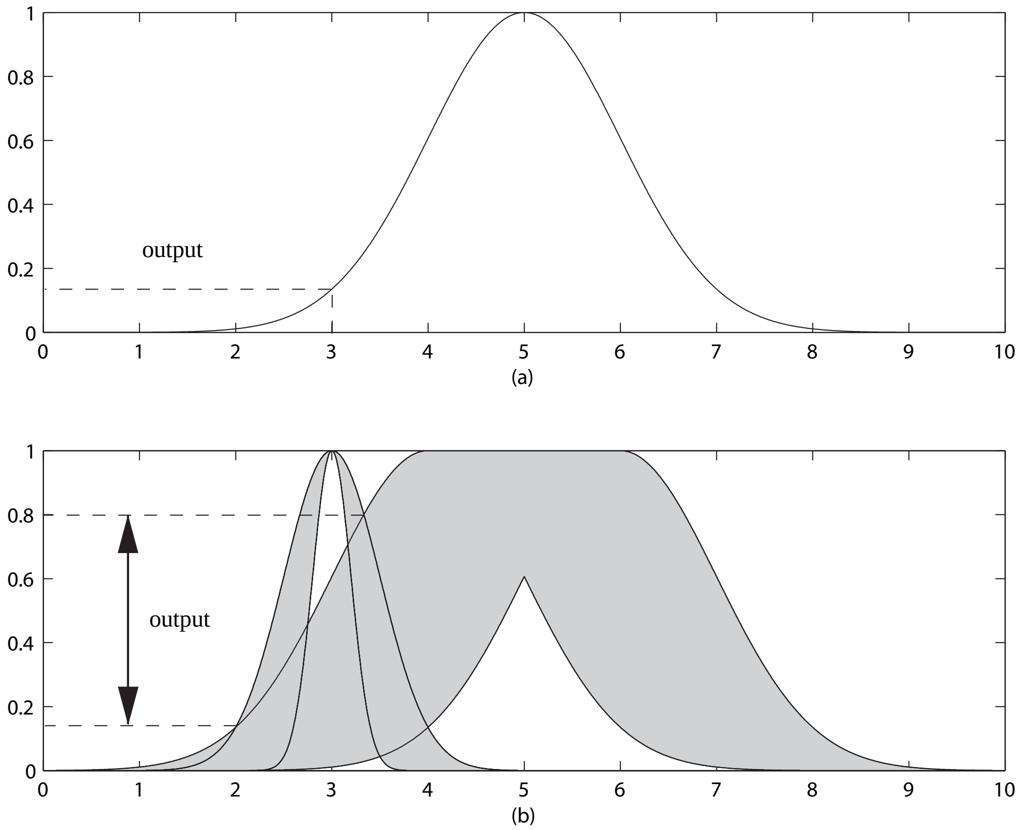

Figure 6: (a) the singleton and (b) the T2 nonsingleton fuzzification.

no uncertainty 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 the interval increases representing more uncertainty of the class model to the input x. However, if k is larger, L_l and L_u are longer so that the two bounds $[\underline{h}, \overline{h}]$ will lose some information of the original $h(x)$.

In T2 fuzzy logic systems (FLSs) [9, 32], the nonsingleton fuzzification (NF) [33] 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. 6 compares the singleton fuzzification (SF) with the corresponding T2 NF. Besides handling uncertainty in data, T2 FSs have been integrated with conventional classifiers to handle uncertainty in the hypothesis space [34]. For example, Liang and Mendel have combined T2 FSs with T1 FLS-based classifiers for MPEG VBR video traffic classification [13]. Zeng and Liu have integrated hidden Markov model and Markov random fields with T2 FSs for speech and handwritten Chinese character recognition [12, 15, 18, 19, 22, 35]. Wu and Mendel have designed T2 FLS-based classifiers based on T1 counterparts for battlefield ground vehicles classification [20]. From these case studies, we obtain a systematic design method in (11) and (12) to handle uncertain feature and hypothesis spaces in pattern recognition.

4 Type-2 Fuzzy Data and Classifiers

This section reviews the state-of-the-art T2 fuzzy pattern recognition 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 3, the SF assumes no uncertainty in the feature space. The T2 (T1) NF models the observation as a

$$\text{uncertain feature space: data} + \text{T2 FSs} = \text{T2 fuzzy data} \quad (11)$$
$$\text{(noise or non-stationarity)}$$
$$\text{uncertain hypothesis space: classifier} + \text{T2 FSs} = \text{T2 fuzzy classifier} \quad (12)$$
$$\text{(unknown varieties of parameters)}$$

Table 1: Classification error rate comparison (%) [14]

Dataset	T2 FSs	Benchmark
Welded Structures	5	6.8

T2 (T1) FS denoted by $\tilde{\mathbf{X}}$.

Mitchell [14] has viewed pattern recognition as the similarity measure between two T2 FSs, in which one set accounts for the uncertain feature space in (11), and the other for the uncertain hypothesis space in (12). The task of pattern recognition is equivalent to finding the class model which has the largest similarity between these two T2 FSs:

$$\omega^* = \arg \max_{\omega=1}^{C} S(\tilde{\mathbf{X}}, \tilde{\lambda}_\omega). \quad (13)$$

In [7] Mitchell has defined the similarity measure by the weighted average of ordinary similarity measure of embedded T1 FSs,

$$S(\tilde{A}, \tilde{B}) = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{mn} S(A_e^m, B_e^n), \quad (14)$$

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. The classification error rate was 1.8% lower than the benchmark (See Table 1).

John *et al.* [36] have 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 (11). They demonstrated that T2 fuzzy preprocessing and MINMAX clustering produced least confusion in relation to consultants judgements.

Liang and Mendel [13] have classified video traffic by T2 FLS-based classifiers extended from T1 FLS-based classifiers as in (12), and showed better performance than the Bayesian classifiers when features have statistical attributes that are non-stationary. Firstly, they design 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, $R^l, 1 \leq l \leq M$, is

$$R^l: \text{ IF } x_1 \text{ is } F_1^l \text{ and } \ldots \text{and } x_d \text{ is } F_d^l, \text{ THEN}$$
$$\mathbf{x} \text{ is classified to } \lambda_1 \ (+1) \ [\text{or is classified to } \lambda_2 \ (-1)]. \quad (15)$$

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

$$h_{F_i^l}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]. \quad (16)$$

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

$$y = \sum_{l=1}^{M} (f^l_{\lambda_1} - f^l_{\lambda_2}), \tag{17}$$

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

$$\begin{aligned} R^l: \text{ IF } \tilde{x}_1 \text{ is } \tilde{F}^l_1 \text{ and } \ldots \text{and } \tilde{x}_d \text{ is } \tilde{F}^l_d, \text{ THEN} \\ \tilde{\mathbf{x}} \text{ is classified to } \tilde{\lambda}_1 \ (+1) \ [\text{or is classified to } \tilde{\lambda}_2 \ (-1)]. \end{aligned} \tag{18}$$

Suppose that the antecedents $\tilde{F}^l_i, 1 \leq i \leq d$ are described by a T2 Gaussian primary MF with uncertain mean or std. Similar to (17), the output of the T2 FLS,

$$\tilde{y} = \sqcup_{l=1}^{M} (\tilde{f}^l_{\lambda_1} - \tilde{f}^l_{\lambda_2}), \tag{19}$$

which is an interval rather than a precise number in (17). For comparison, they also design the Bayesian classifier as follows. If equal prior class probability is assumed, the Bayesian classifiers are

$$p(\mathbf{x}|\lambda_1) = \sum_{l=1}^{m} p(\mathbf{x}|\lambda^l_1), \tag{20}$$

$$p(\mathbf{x}|\lambda_2) = \sum_{l=1}^{n} p(\mathbf{x}|\lambda^l_2), \tag{21}$$

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

$$p(\mathbf{x}|\lambda) = \frac{1}{\sqrt{(2\pi)^d|\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{22}$$

where the covariance matrix is diagonal, $\mathbf{\Sigma} = \textit{diag}(\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_d)$. According to Bayesian decision theory [21, Chapter 2], the optimal decision rule is

$$\text{IF } p(\mathbf{x}|\lambda_1) - p(\mathbf{x}|\lambda_2) > 0, \text{ THEN } \mathbf{x} \text{ is classified to } \lambda_1, \tag{23}$$

$$\text{IF } p(\mathbf{x}|\lambda_1) - p(\mathbf{x}|\lambda_2) < 0, \text{ THEN } \mathbf{x} \text{ is classified to } \lambda_2. \tag{24}$$

Observe (17), (23), and (24) that the class model in the Bayesian classifier has a correspondence with each rule in the T1 fuzzy 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 (22), 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 the output interval in (19).

In MPEG VBR video traffic classification (out-of-product testing) without parameter adjustment, Liang and Mendel [13] 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 2. 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 9.17% for T1 NF with T1 FLS-based classifiers (T1NFT1). So they concluded that T2 fuzzy

Table 2: False alarm rate comparison (%) [13]

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 3: Classification error rate comparison (%) [20]

Datasets	T2 fuzzy classifiers	T1 fuzzy classifiers
Battlefield ground vehicle	9.13	12.8

classifiers were substantially better than their T1 counterparts in terms of the robustness and classification error rate.

Similarly, Wu and Mendel [20] have designed 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 3). Besides, they showed that all FLS-based classifiers performed much better than the Bayesian classifiers.

In [12, 15–19, 22, 35] we view pattern recognition as the labeling problem, which is also a compound Bayesian decision problem [21]. 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 [21] guarantees the single best labeling configuration,

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} P(\mathcal{F}|\mathbf{X}, \lambda), \tag{25}$$

$$P(\mathcal{F}|\mathbf{X}, \lambda) \propto p(\mathbf{X}|\mathcal{F}, \lambda)P(\mathcal{F}|\lambda), \tag{26}$$

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 the fuzzy data and fuzzy class model, we incorporate T2 FSs into MAP (25)-(26) as follows,

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} h_{\tilde{\lambda}}(\mathcal{F}|\mathbf{X}), \tag{27}$$

$$h_{\tilde{\lambda}}(\mathcal{F}|\mathbf{X}) \propto h_{\tilde{\lambda}}(\mathbf{X}|\mathcal{F}) \sqcap h_{\tilde{\lambda}}(\mathcal{F}), \tag{28}$$

where $\tilde{\lambda}$ is the class model with fuzzy parameters. We use the NF to handle fuzzy observations due to noise. Set operations in (28) convey more information than (26) because we unite all possibilities of the class model due to fuzzy parameters into T2 FSs. Especially when $\tilde{\lambda}$ is certain, equation (28) will be reduced to (26). 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.

In [35] 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

Table 4: Classification rate (%) comparison [35]

Datasets	T2 FGMMs	GMMs
IONOSPHERE	77.7	75.3
PENDIGITS	91.9	88.3
WDBC	94.9	93.6
WINE	89.2	85.8

Table 5: Classification rate (%) comparison [35]

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

system, we use the generalized linear model (GLM) to make the final classification decision from fuzzy likelihoods. Extensive experiments on datasets from UCI repository [37] demonstrate that T2 FGMMs have an average 2.7% (the best results) higher classification rate than that of GMMs (See Table 4). Based on (27)-(28), we 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 the robustness, we also classified the phonemes corrupted by the multi-talker non-stationary babble noise with different signal-to-noise ratios (SNRs). Table 5 shows the best results of T2 FHMMs compared to HMMs. We see that on average T2 FHMMs outperform HMMs 1.85% in classification rate under babble noise with different SNRs.

In [12, 15–17] 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 adding the white Gaussian noise to the test data with different SNRs (See Table 6). Furthermore, a complete continuous phoneme recognition experiment demonstrate that T2 FHMMs outperform the competing HMMs 5.55% in dialect recognition accuracy (See Table 7).

Similarly, in [12, 18, 19] we have integrated T2 FSs with Markov random fields (MRFs) referred to as the T2 FMRFs for handwritten Chinese character modeling. From experiments on similar characters [19], we demonstrate that T2 FSs improve the performance of the MRFs for handwritten Chinese character recognition by 1.26% in classification rate on average (See Table 8). Furthermore, a generalization ability comparison (See Table 9) 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 (11) and (12) are effective in most pattern recognition problems. The T2 fuzzy data (11) and T2 fuzzy classifier (12) compose a T2 fuzzy pattern recognition system, which generally has a better performance than the competing T1 fuzzy and Bayesian classifiers. Though in some cases the T2 fuzzy system degrades a little than conventional methods, it is still a reliable approach to improve classification ability of the conventional methods in terms of the robustness, generalization ability, and recognition

Table 6: Classification rate (%) comparison [15]

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 7: Recognition accuracy comparison (%) [15]

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

Table 8: Classification error rate comparison (%) [19]

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

accuracy. However, note that, at present, *there is no theory that guarantees that a T2 fuzzy system will always do this* [1].

5 Discussions

T2 FSs can be viewed as an ensemble of T1 FSs or PDFs. Similarly, T2 fuzzy classifiers contain an ensemble of decision functions, which is definitely robust than the single best decision function in T1 fuzzy and Bayesian classifiers. More importantly, the ensemble T2 fuzzy classifiers keep all possibilities of decision functions until the final decision-making. In real-world applications, if we always make the correct classification decision from the FOU, the recognition accuracy *cannot be worse* than the original T1 fuzzy and Bayesian classifiers. Therefore, how to make the decision from the FOU poses the first problem of designing T2 fuzzy classification systems.

Occam's razor [21, Chapter 9.2.5] has come to be interpreted in pattern recognition 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 classifiers have more parameters with a higher computational complexity than their counterparts such as T1 fuzzy and Bayesian classifiers [13–15, 20, 34, 36]. In most cases, at least twice computations (interval type-2 fuzzy sets) have to be done in T2 fuzzy classifiers than conventional methods. Therefore, when we design T2 fuzzy systems to solve the real-world problems, we have to consider carefully if the problem at hand is needed to pay more complexity. Based on comprehensive experiments, we say that T2 fuzzy classifiers have the *potential* to outperform their counterparts, but in the meantime they add more complexity to the system leading to the performance-complexity trade-offs.

No Free Lunch Theorem [21, Chapter 9.2.1] tells us that there are no context-independent or usage-independent reasons to favor one learning or classification method over another. Looking back at strategies (11) and (12), T2 fuzzy systems are natural extensions of the original pattern recognition systems, which

Table 9: Classification error rate comparison (%) [19]

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

means the performance has been already ensured, and T2 FSs just improve it. We should also note that T2 fuzzy systems do not always outperform their counterparts in all pattern recognition problems, and T2 fuzzy systems are not always effective for modeling uncertainties [20]. The major reason may be that the designed FOU covers too much or too little uncertainty that the system does not have. Another reason may be that we use ineffective methods for the final classification decision-making.

The great success of statistical pattern recognition as well as Bayesian decision theory has been attributed to the recognition of *randomness* in both 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 3 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. Through many case studies, we obtain the design methods in (11) and (12) for the pattern recognition system, and further extend them within the Bayesian framework in (27) and (28). Based on encouraging experimental results, we are optimistic about the future of T2 FSs for pattern recognition applications.

References

[1] J. M. Mendel, "Advances in type-2 fuzzy sets and systems," *Information Sciences*, vol. 177, pp. 84–110, 2007.

[2] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Sciences*, vol. 8, pp. 199–249, 1975.

[3] M. Mizumoto and K. Tanaka, "Some properties of fuzzy sets of type-2," *Information and Control*, vol. 31, pp. 312–340, 1976.

[4] ——, "Fuzzy sets of type-2 under algebraic product and algebraic sum," *Fuzzy Sets and Systems*, vol. 5, pp. 277–290, 1981.

[5] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty and Information*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[6] 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.

[7] H. Mitchell, "Ranking type-2 fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 287–294, Apr. 2006.

[8] N. N. Karnik and J. M. Mendel, "Operations on type-2 fuzzy sets," *Fuzzy Sets and Systems*, vol. 122, pp. 327–348, 2001.

[9] 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.

[10] Z.-Q. Liu and Y.-K. Liu, "Type-2 fuzzy variables," *IEEE Trans. Fuzzy Syst.*, to be published.

[11] J. M. Mendel, *Uncertain Rule-based Fuzzy Logic Systems : Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[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] 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.

[14] H. Mitchell, "Pattern recognition using type-II fuzzy sets," *Information Sciences*, vol. 170, pp. 409–418, 2005.

[15] 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.

[16] ——, "Interval type-2 fuzzy hidden Markov models," *Proc. FUZZ-IEEE*, pp. 1123–1128, 2004.

[17] ——, "Type-2 fuzzy hidden Markov models to phoneme recognition," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 192–195, 2004.

[18] ——, "Type-2 fuzzy Markov random fields to handwritten character recognition," *Proc. Int'l Conf. Pattern Recognition*, pp. 1162–1165, 2006.

[19] ——, "Type-2 fuzzy Markov random fields and their application to handwritten Chinese character recognition," *IEEE Trans. Fuzzy Syst.*, to be published.

[20] H. Wu and J. M. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 56–72, Feb. 2007.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, 2001.

[22] J. Zeng, "Type-2 fuzzy sets, Markov processes for pattern recognition," Ph.D. dissertation, City University of Hong Kong, Hong Kong, Oct. 2006.

[23] G. J. Klir and M. J. Wierman, *Uncertainty-based information*. New York: Physica-Verlag, 1998.

[24] B. Möller and M. Beer, *Fuzzy Randomness : Uncertainty in Civil Engineering and Computational Mechanics*. New York: Springer, 2004.

[25] B. Liu, *Uncertainty Theory : An Introduction to Its Axiomatic Foundations*. Berlin: Springer, 2004.

[26] J. J. Buckley, *Fuzzy Probabilities : New Approach and Applications*. New York: Physica-Verlag, 2003.

[27] ——, *Fuzzy Statistics*. New York: Springer, 2004.

[28] T. M. Mitchell, *Machine Learning*. Boston, Mass.: McGraw-Hill, 1997.

[29] 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.

[30] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford University Press, 1995.

[31] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[32] 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.

[33] 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.

[34] J. Zeng and Z.-Q. Liu, "Stroke segmentation of Chinese characters using Markov random fields," *Proc. Int'l Conf. Pattern Recognition*, pp. 868–871, 2006.

[35] J. Zeng, L. Xie, and Z.-Q. Liu, "Type-2 fuzzy Gaussian mixture models," *Pattern Recognition*, submitted for publication.

[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/~mlearn/MLRepository.html