

Fuzzy Techniques for Modelling Uncertainty in Medical Data and Knowledge

Jonathan M. Garibaldi¹

School of Computer Science, e-mail: jmg@cs.nott.ac.uk

Abstract: Fuzzy techniques may be used to represent uncertainty in both data and knowledge, and fuzzy inference systems (FISs) may be used to reason with such uncertain data and knowledge. In this paper, definitions are first provided for conventional type-1 fuzzy sets, more complex type-2 fuzzy sets and the recently introduced non-stationary fuzzy sets. Two medical applications in which such fuzzy sets have been utilised to model various forms of uncertainty are then presented.

Keywords: Fuzzy Sets, Fuzzy Inference, Umbilical Acid-Base Analysis, Breast Cancer

1. Definitions

Definition 1 A type-1 fuzzy set A of the universe of discourse X is characterised by a membership function $\mu_A : X \rightarrow [0, 1]$ which associates with each element x of X a number $\mu_A(x)$ in the interval $[0, 1]$, with $\mu_A(x)$ representing the grade of membership of x in A . The type-1 fuzzy set A is denoted by:

$$A = \int_{x \in X} \mu_A(x)/x. \quad \blacksquare$$

Definition 2 A fuzzy set is of type-2 if its membership function ranges over fuzzy sets of type-1. That is, a type-2 fuzzy set $\tilde{A}$ of the universe of discourse X is characterised by a type-2 membership function $\mu_{\tilde{A}}(x, u)$ where $x \in X$ and $u \in J_x \subseteq [0, 1]$, $\forall x \in X$. The type-2 fuzzy set $\tilde{A}$ is denoted by:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u)/(x, u).$$

where J_x is called the primary membership of x . ■

Definition 3 The union of all primary memberships of a fuzzy set $\tilde{A}$ is termed the footprint of uncertainty (FOU). This is:

$$FOU(\tilde{A}) = \bigcup_{x \in X} J_x. \quad \blacksquare$$

Essentially, a type-2 fuzzy set is a fuzzy set in which the membership function has been ‘blurred’ so that, for a specific value of x , the membership function is no longer a single value but, rather, takes a (fuzzy) set of values.

¹Address of correspondence: School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, UK.

Definition 4 Let A denote a fuzzy set of a universe of discourse X characterised by a membership function μ_A , T be a set of time points t_i (possibly infinite) and $f : T \rightarrow \mathbb{R}$ denote a perturbation function. A non-stationary fuzzy set $\dot{A}$ of the universe of discourse X is characterised by a non-stationary membership function $\mu_{\dot{A}} : T \times X \rightarrow [0, 1]$ which associates with each element (t, x) of $T \times X$ a time-specific variation of $\mu_A(x)$. The non-stationary fuzzy set $\dot{A}$ is denoted by:

$$\dot{A} = \int_{t \in T} \int_{x \in X} \mu_{\dot{A}}(t, x) / x / t.$$

However, an additional restriction is imposed on $\mu_{\dot{A}}$. Given that $\mu_A(x)$ can be expressed as $\mu_A(x, p_1, \dots, p_m)$, where $p_1, \dots, p_m$ denote the parameters of $\mu_A(x)$:

$$\mu_{\dot{A}}(t, x) = \mu_A(x, p_1(t), \dots, p_m(t)),$$

where $p_i(t) = p_i + k_i f_i(t)$ and $i = 1, \dots, m$. In this way, each parameter is varied in time by a perturbation function multiplied by a constant. ■

Hence, for a given standard fuzzy set A and a set of time points T , a non-stationary fuzzy set $\dot{A}$ is a set of duplicates of A varied over time. A time duplicate of A is termed an *instantiation* and denoted by $\dot{A}_t$, so that $\dot{A}_t(x) = \dot{A}(t, x)$. The perturbation function induces ‘small’ and temporary alterations in $\mu_A(x)$. There are alternative forms of non-stationarity, but the only form used in this paper is *variation in location* in which:

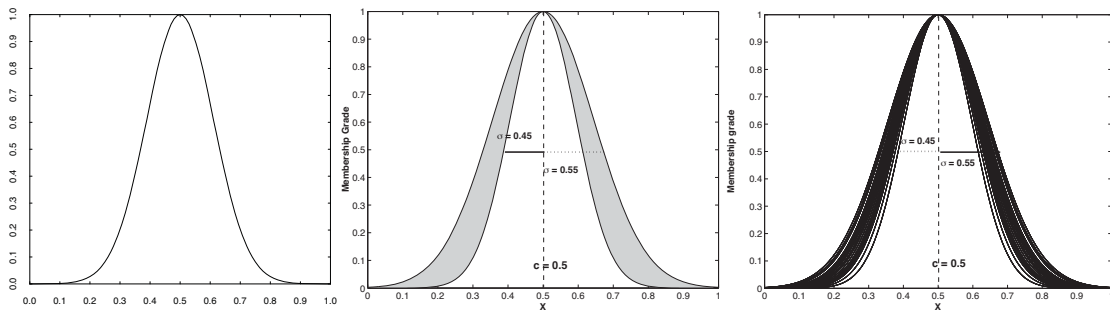
$$\forall_{t \in T} \mu_{\dot{A}}(t, x) = \mu_A(x + a(t)),$$

where $a(t)$ is a constant for any given t . Thus, the membership function is shifted, as a whole, left ($a(t) > 0$) or right ($a(t) < 0$) by small amounts over time.

1.1 Fuzzy Inference Systems

A fuzzy inference system (FIS) is a system that uses fuzzy reasoning to map an input space to an output space, usually consisting of a set of ‘if-then’ rules connecting the inputs to the outputs. Space constraints preclude formal definitions of FISs in this paper, but it suffices to state that a type-1 FIS, type-2 FIS and non-stationary FIS are defined as utilising *only* type-1 fuzzy sets, one or more type-2 fuzzy sets and one or more non-stationary fuzzy, respectively. Type-1, type-2 and non-stationary fuzzy sets are illustrated in Fig. 1.

Figure 1: An illustration of type-1, type-2 and non-stationary fuzzy sets (note that only the footprint of uncertainty of the type-2 set is actually shown)

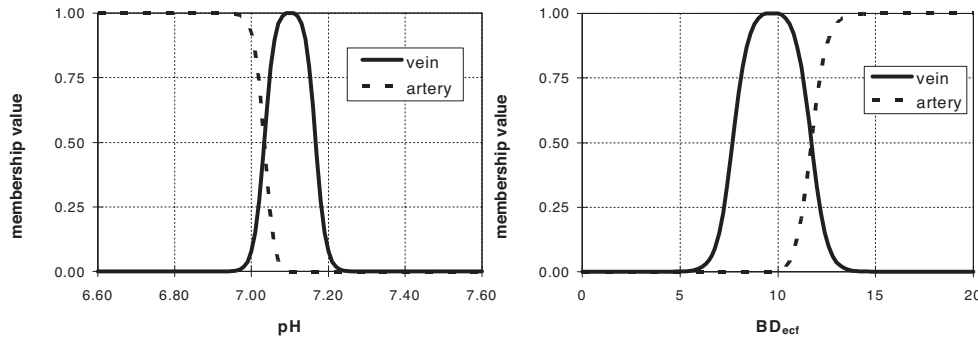


2. Modelling Uncertainty in Umbilical Acid-Base Analysis

In umbilical acid-base (UAB) assessment, the health of a newborn infant in the first few seconds of its life is assessed from physiological measurements of blood taken from the arteries and vein of the umbilical cord. The acidity (pH) and base deficit of extracellular fluid (BD_{ecf}) obtained by a blood gas analysis machine are used for UAB assessment, but this requires considerable expertise and experience. A type-1 FIS was developed encapsulating the knowledge of leading obstetricians, neonatologists and physiologists expert in acid-base interpretation. This FIS combined knowledge of the errors likely to occur in acid-base measurement, physiological knowledge of plausible results and statistical knowledge of a large database of cases.

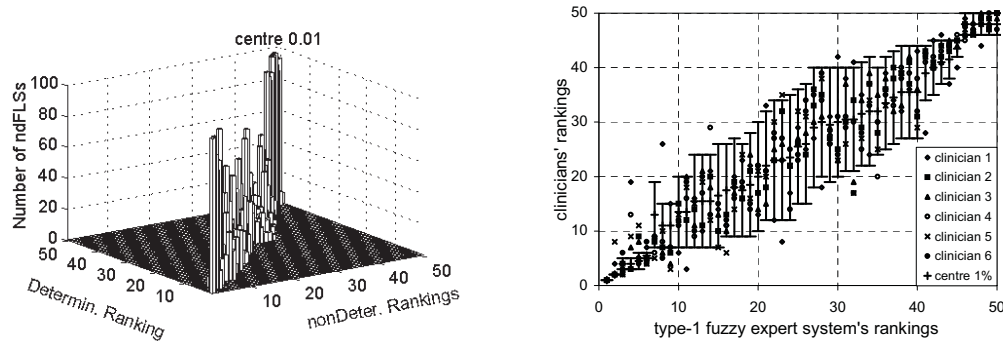
Standard deviations of the various measurements were determined through a range of experiments. Type-1 fuzzy sets were then used to directly represent uncertainty by converting an observed measurement into a fuzzy set located at the observation with a width equal to the standard deviation. Missing measurements were represented by combining a fuzzy 'unknown' value ($\mu = 1$) with the physiologically knowledge that umbilical arterial pH must be lower than venous pH, and arterial BD_{ecf} must be higher than venous BD_{ecf} . An example of the fuzzy sets representing an observation in which the venous pH is 7.10 and venous BD_{ecf} is 9.8, and the corresponding arterial results are missing is shown in Fig. 2. Note that the widths have been exaggerated for illustrative purposes.

Figure 2: *The use of fuzzy sets to represent uncertain data and missing values*



A non-stationary FIS and type-2 FIS were also created in order to explore modelling uncertainty in the clinical knowledge (expressed through the fuzzy sets and rules used within the rule-base of the FIS). Illustrative results are shown in Fig. 3.

Figure 3: *Distribution of outputs obtained from non-stationary fuzzy sets and ranges of outputs from type-2 fuzzy sets*



3. Modelling Uncertain Knowledge in Breast Cancer

The appropriate medical treatment decision for follow-up (adjuvant) therapy can greatly improve the survival rate of patients with breast cancer following diagnosis and primary surgical treatment. In Nottingham City Hospital, the decision is made based on a set of clinical guidelines in the form of rules. An FIS capturing these guidelines was created to provide a recommendation for chemotherapy adjuvant therapy. Five attributes that feature in the medical guidelines were utilised as inputs to the FIS. These are: the Nottingham Prognostic Index (*NPI*), estrogen receptor status (*ER*), degree of vascular invasion (*VI*), patient age at diagnosis (*age*) and number of positive lymph nodes (*LN*). The output of the fuzzy system is the recommendation for chemotherapy treatment, containing three categories: *No*, *Maybe* and *Yes*.

A type-1 FIS was created in which the fuzzy sets for the input and output variables were derived from the clinical guidelines and available clinical knowledge. The performance of the FIS was evaluated by comparing its predictions with the actual clinical data and known clinical recommendations available on a data set of 1310 patients presenting at the Nottingham City Hospital between 2004 and 2008. The type-1 FIS provided the appropriate recommendation in 1108 of 1310 cases (84.6% accuracy). A non-stationary FIS was then created to capture the variability in opinion observed over a diverse range of clinical experts involved in such a decision. The non-stationary fuzzy system was run 30 times with 1% variation in location of the fuzzy sets, and a majority voting mechanism was utilised to reach the overall ensemble decision. It was found that this non-stationary FIS improved performance to result in the appropriate recommendation in 1141 of 1310 cases (87.1% accuracy). The confusion matrices for the original type-1 FIS and the majority vote non-stationary FIS are shown in Fig. 4.

Figure 4: The confusion matrices for the original type-1 FIS and the majority vote non-stationary FIS

		Clinician Decision		
		No	Maybe	Yes
Fuzzy Decision	No	982	35	64
	Maybe	2	2	2
	Yes	83	15	124

		Clinician Decision		
		No	Maybe	Yes
Fuzzy Decision	No	1024	42	74
	Maybe	0	0	0
	Yes	43	10	117

References

- Garibaldi J.M. and Ifeachor E.C. (1999) Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation, *IEEE Transactions on Fuzzy Systems*, 7, 72–84.
- Garibaldi J.M., Jaroszewski M. and Musikasuwan S. (2008) Non-stationary fuzzy sets, *IEEE Transactions on Fuzzy Systems*, 16 (4), 1072–1086.
- Garibaldi J.M. and Ozen T. (2007) Uncertain fuzzy reasoning: A case study in modelling expert decision making, *IEEE Transactions on Fuzzy Systems*, 15, 16–30.
- Wang X.Y., Garibaldi J.M., Zhou S.M. and John R.I. (2009) A novel approach to designing fuzzy inference systems and its application to recommending treatment for breast cancer, *submitted to Artificial Intelligence in Medicine*.