Journal of Intelligent Manufacturing (1993) **4**, 323–340

Principles and methodology of fuzzy sets

WITOLD PEDRYCZ

Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada R3T 2N2

Received July and accepted September 1992

The principles of fuzzy sets and their role in processing uncertain information will be discussed. The question of knowledge representation that is of significant importance in problems of system modelling will be formulated and considered at the level of fuzzy sets. Modelling and simulation realized with the aid of fuzzy sets are studied in a unified methodological framework. First a notion of the cognitive perspective is applied to articulate the problem in terms of specialized linguistic labels. Fuzzy models are constructed to capture logical relationships between the elements (linguistic labels) of the cognitive perspective. Several different classes of the models distinguished with regard to their structural dependencies will be analysed in depth. Finally a linguistic–numerical transformation constituting a type of model–environment interface will be studied.

Keywords: Fuzzy sets, uncertainty, fuzzy models, linguistic modelling, soft computations

1. Introductory remarks

Uncertainty is an inherent component of real-life problems. Human reasoning involves some general categories within which individual objects are ill-structured with imprecisely defined boundaries. Our abilities to carry out reasoning processes in the presence of incomplete and/or uncertain information are extraordinary. A vast number of tasks ranging from the almost trivial in our human sense (like driving a car, recognizing objects, avoiding obstacles), to the complex (e.g. managing manufacturing processes, scheduling, designing, etc.) still constitute a continuous challenge for computer algorithms.

The reasoning processes of humans and those traditionally implemented at the level of computers are realized at two essentially disjointed conceptual platforms. The former handle all pieces of information at a linguistic level in a symbolic-like fashion. The specific representation can be, and usually is, adjusted according to the problem at hand. The underlying concept on which all computer algorithms are based is that they almost exclusively process rigid and extremely non-modifiable numerical information. In order to progress further one should look carefully at possible ways to narrow the existing gap. A possible alternative lies in the realm of soft computing. In its very essence this computational paradigm encapsulates diverse computational faculties including fuzzy sets, neurocomputations, genetic algorithms and genetic programming, to name just a few. The main objectives one strives to achieve in this area pertain to:

(1) Processing information in linguistic form, coping with the incomplete and heterogeneous character of available knowledge;

(2) Producing a user-friendly computing environment (both in terms of its customization meeting the needs of an individual user and in the sense of its increased interpretation capabilities).

The aim of this paper is to study the principles of fuzzy sets as important concepts for knowledge representation, identify their role in system modelling and present an overview of relevant identification algorithms.

The paper is structured as follows. First in Section 2 we will look at the basic concept of a fuzzy set, analyse two groups of methods useful in membership function elicitation and refine differences between types of uncertainty that are appropriately conveyed by fuzzy sets and probability theory. In Section 3 we will consider basic calculus of fuzzy sets including logical operations realized for fuzzy sets and introduce the representation theorem erecting a bridge between fuzzy sets and set theory. We will then concentrate on transformations of fuzzy sets performed with the use of functions – this gives rise to the extension principle. Section 4 will be devoted to the calculus of fuzzy relations. The methodological aspects of the use of fuzzy sets in modelling and simulation will be covered in Section 5. In this setting we will first address the notion of a cognitive perspective worked out on the basis of fuzzy sets being perceived as basic information granules. Different classes of fuzzy models will then be studied (Section 6) with a strong distinction made with regard to a level of structural relationships conveyed within the domain knowledge. Finally, a method for transforming the results of fuzzy modelling into numerical representatives will be studied in Section 7.

2. An idea of fuzzy sets: origin and basic notions

2.1. *Sets versus fuzzy sets*

In order to introduce the idea of fuzzy sets it is worth starting from the formalism of two-valued logic. In this setting the notion of a set implies that considering any object, no matter how complex it is, we are compelled to assign it to one of the two complementary and exhaustive categories enumerated in advance (for instance, good–bad, normal–abnormal, odd–even, black–white, etc.). Sometimes this discrimination does make sense, in many other situations it could lead to some serious and evident dilemmas. For example, let us consider natural numbers and define two categories of elements (sets) such as odd and even numbers. Within this framework any natural number can be classified without any dawdling. On the other hand, in many tasks in engineering, manufacturing or management we are faced with classes that are ill-defined and do not retain clear and well-defined boundaries. Refer to the following statements (Zimmermann, 1992):

length of the planning horizon is *very long*,
computation time available in the system is *short*,
demand variability is *low*.

They contain terms without well-defined boundaries. Terms of the same nature could be found in rules (conditional statements) describing for example control policies. Two examples follow (Shin *et al.*, 1992):

if waiting time is *long* and slack time is critically *short* then date criteria are *urgent*,
if process variable is *too low* then design variable is *largely increase*.

Even in mathematics we may encounter some broadly accepted and used notions with gradual rather than abrupt boundaries. Refer to such well-known terms as: *sparse* matrix, a linear approximation of a function in a *small* neighbourhood of a point x_0, an *ill-conditioned* matrix. We accept these notions as conveying useful information about the problem to be studied. Furthermore they are not contemplated as essential defects in our everyday language but rather as their remunerative features indicating our ability to generalize and conceptualize knowledge. Nevertheless, we should stress that any of these notions is strongly context-dependent and by no means is its detailed definition universal. These predicates are meaningful and their semantics is obvious within a certain community (such as users, designers, programmers, etc.).

An interesting example appeared in one of Borel's works (Borel, 1950) where he is referring to the ancient Greek sophism of the pile of seeds,

> '. . . one seed does not constitute a pile nor two nor three . . . from the other side everybody will agree that 100 million seeds constitute a pile. What therefore is the appropriate limit? Can we say that 325 647 seeds do not constitute a pile but 325 648 do?'

At a glance, it becomes evident that answers of the 'yes–no' type cannot be viewed as a satisfactory solution to the problem. Clearly, each limit (border) point x_0 used in the definition of the predicate pile (x)

$$\text{pile}\ (x) = \begin{cases} 1, & \text{if } x \geqslant x_0 \\ 0, & \text{if } x < x_0 \end{cases}$$

could contribute to an extremely simplified and unrealistic model of this concept. Any optimization of the limit value x_0 standing in the above definition is therefore futile and does not bring us closer to an acceptable solution – the conceptual shortcoming of this Boolean model remains intact. The key issue of fuzzy sets is that one significantly extends the meaning of a set admitting different grades of belongingness (also known as membership values) of an element in a set. This alleviates the previous problem by embracing all intermediate conceptual situations arising between complete (total) membership and total non-membership.

More formally, a fuzzy set A defined in a universe of discourse X is described by its membership function viewed as a mapping (Zadeh, 1965; 1973)

$$A : X \to [0,1]$$

The degree of membership $A(x)$ expresses the extent to which x fulfils the category described by A. The condition $A(x) = 1$ identifies elements of X which are fully compatible with A. The condition $A(x) = 0$ identifies all the elements which definitely do not belong to A. The higher the membership value at x, the higher the strength of adherence of x to A. Any physical experiment whose realization is a matter of energy or strength (e.g. pulling a rubber band) can serve as a metaphor for the notion of membership function (membership degree).

Usually when discussing a fuzzy set we assume that there exist elements with membership grades equal to 1. Sometimes one may not be able to assign any element with the highest degree of membership. To describe this situation we will introduce a notion of a height of a fuzzy set.

By the height of a fuzzy set A, hgt(A), we mean a maximal value of its membership function,

$$\mathrm{hgt}(A) = \sup_{x \in X} A(x)$$

If hgt(A) = 1 then A is a normal fuzzy set, otherwise we will call it a subnormal fuzzy set.

The support of A, supp(A), describes all elements of X with non-zero grades of membership in A,

$$\mathrm{supp}(A) = \{x \in X \mid A(x) > 0\}$$

Before we proceed with further notions and a formal terminology of fuzzy sets, it is worth while summarizing some basic facts about set theory. In set theory all objects (sets) are fully described by so-called characteristic functions. The characteristic function of set A, χ_A, is defined as a two-valued mapping,

$$\chi_A : X \rightarrow \{0,1\}$$

taking its values in the set {0,1} such that

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

In fuzzy sets, the meaning of the fundamental predicate of set theory '∈' (element of) is significantly expanded by accepting partial membership of an element in a set.

Identifying the universe of discourse X as a set of real numbers R, we can define fuzzy numbers. A fuzzy number A is a fuzzy set defined in R such that:

(1) A is a normal fuzzy set, i.e. there exists at least one element of R for which $A(x) = 1$;

(2) A is convex

$$\forall_{\lambda \in [0,1]} \ \forall_{x,y \in R} \ A[\lambda x + (1-\lambda)y] \geq \min[A(x), A(y)];$$

(3) A is upper semi-continuous;

(4) A has a bounded support.

The collection of these conditions is modified quite often (usually some of the listed properties are dropped). In many situations it is worth restricting to piecewise linear membership functions. They give rise to a class of triangular and trapezoidal fuzzy numbers, see Fig. 1. This characterization of a fuzzy number is sufficient to capture uncertainty associated with the studied linguistic term. The triangular fuzzy number, denoted by $A(x;\alpha,m,\beta)$ is uniquely characterized by its three parameters, say m, α and β, where $\alpha < m < \beta$, see Fig. 1a. The first parameter embodies a so-called modal (typical) value of the term. The lower and the upper bounds are denoted by α and β, respectively. For instance, a waiting time W in a queue where it typically takes 15 min to provide service while the bounds are 5 and 29 min, respectively, can be described as a triangular fuzzy number $W(t;5,15,29)$. Since no additional information about the waiting time is available, the choice of the linear relationship is fully legitimate. If there is no uncertainty (fuzziness) then $\alpha = n = \beta$ and the fuzzy number reduces to a single pointwise quantity (real number). The trapezoidal fuzzy number admits an additional degree of freedom that enables us to model a range of equally acceptable typical values. In this class of membership functions the modal value 'm' spreads into a closed interval $[n,m]$, see Fig. 1b.

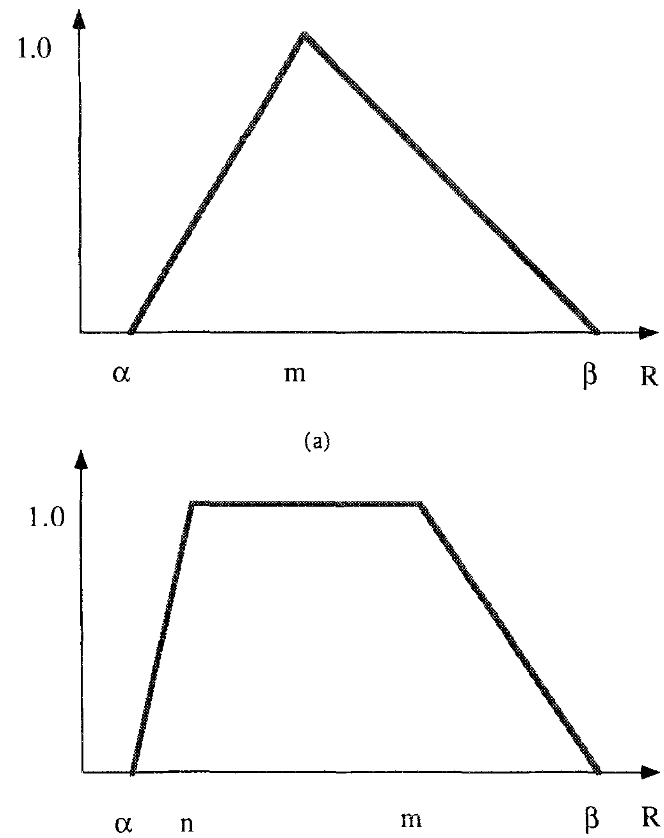

Fig. 1. Examples of (a) triangular and (b) trapezoidal membership functions.

2.2. *Membership function elicitation*

Two essential classes of methods of membership function elicitation can be distinguished:

(1) Horizontal approach. The underlying idea of this method is to gather information about grades of mem-

bership of some elements of a universe of discourse in which a fuzzy set is to be defined. The process of elicitation of these membership functions can be concisely stated as follows,

> Consider a group of 'N' experts. Each of them is asked to answer the following question:
> can x_0 be viewed as compatible with the concept represented by the fuzzy set A?

where x_0 is a fixed element of this universe of discourse and A is a fuzzy set to be determined. The answers are restricted to 'yes' or 'no' statements only. Then, counting the fraction of positive ('yes') responses $n(x_0)$ found in the experiment, the value of the membership function at this element of the universe of discourse is estimated as

$$A(x_0) = \frac{n(x_0)}{N}$$

Thus the method is based on a straightforward counting of the responses of the experts and in this sense it reminds us of the procedure considered in the example cited by Borel (cf. Section 2.1). The evident advantage of this method lies in its simplicity. The experiment can be easily completed and new points of the universe of discourse added, if required. One can determine a standard deviation of the obtained estimates of the membership degrees. Its calculations follow an elementary statistical analysis by noting that the 'yes–no' responses constitute realizations of a certain binomially distributed random variable. The standard deviation of $A(x_0)$ denoted by st_dev(A) (x_0) is given by

$$\text{st\_dev}(A)(x_0) = [A(x_0)/N]^{1/2}$$

The derived values of the standard deviation of A can be utilized towards a simple acceptance criterion: accept $A(x_0)$ as a sound estimate of the grade of membership if its standard deviation [or a ratio of this deviation to the value of $A(x_0)$] does not exceed a threshold level λ.

(2) Vertical approach. The main concept implemented in this approach is to fix a certain level of the membership level α and ask a group of experts to identify a collection of elements in X satisfying the concept carried by A to a degree not lower than α. Denote a collection of the elements of X derived in this way by A_α. We will be referring to this as an α-cut of the fuzzy set A, see also Section 3.2. The experiment is repeated for several levels of α. By virtue of the representation theorem (refer again to Section 3.2) the fuzzy set is 'reconstructed' by aggregating the obtained α-cuts.

Comparing these two approaches we can conclude that they are conceptually simple. The factor of uncertainty reflected by the fuzzy boundaries of A is distributed either vertically (in the sense of the grades of membership) or horizontally (absorbed by the limit points of the elicited α-cuts). The values of α or different elements of the universe of discourse should be selected randomly to avoid any potential bias furnished by the testees (experts).

The evident shortcoming of these two methods resides within the 'local' nature of the experiments. This means that each grade of membership is estimated independently from the others. The results may not then fully comply with a general tendency to maintain the form of a smooth transition from the full membership to the absolute exclusion that is so preponderant in fuzzy sets. In this situation, a pairwise comparison method as introduced by Saaty (1980) can be used to alleviate the inadequacy existent in the above methods.

2.3. *Fuzziness and randomness*

When referring to fuzziness and randomness, one sometimes comes across statements in which fuzziness is viewed as a certain form of randomness. Without moving into mathematical formulae that are evidently distinct for fuzzy sets and probability theory (specifically, set theory and logics on one hand, and measure theory on the other) we can saddle on the following rationalization: 'Randomness has to do with uncertainty concerning membership or non-membership of an object in a non-fuzzy set, while fuzziness has to do with classes in which there may be grades of membership intermediate between full membership and non-membership.'

Yet another argument may help to elucidate a philosophical dissimilitude between probability and fuzziness. Take a finite universe of discourse X and connote

$A(x) = a$ – the value of the membership function of A for a certain element of X is equal to a

$P\{x \in A\} = a$ – probability of x belonging to A is equal to a

Perform now an experiment in which we pick up this element x and observe the outcome, i.e. analyse a position of x with respect to A. The results obtained in this way can be interpreted in the two different ways as summarized in Table 1. Upon observation of x, the *a priori* probability $P\{X \in A\} = a$ becomes *a posteriori* and

Table 1.

	Before experiment	*After experiment*
Fuzziness	$A(x) = a$	$A(x) = a$
Randomness	$P\{x \in A\} = a$	1, if $x \in A$ 0, otherwise

is equal to either 1 if $x \in A$ or 0 otherwise. At the same time, $A(x)$ being treated as a measure of the extent to which x belongs to A remains unchanged.

In general, we can conclude that randomness deals with models of statistical inexactness emerging from the occurrence of random events, while fuzziness deals with models of inexactness arising from the perception processes of the human being.

3. Fuzzy sets calculus

In this section we will depart from basic operations on fuzzy sets (such as union, intersection, negation), study their implementations with the aid of triangular norms and afterwards discuss transformations of fuzzy sets and fuzzy numbers in particular. We will also briefly look at the representation of fuzzy sets resolved in the language of set theory.

3.1. *Logical operations on fuzzy sets*

The basic operations (logical connectives) can be formally defined by replacing the characteristic functions of sets by the membership functions of the fuzzy sets. This gives rise to the following expressions:

$$(A \cup B)(x) = \max[A(x), B(x)]$$
$$(A \cap B)(x) = \max[A(x), B(x)]$$
$$\bar{A}(x) = 1 - A(x)$$
$$x \in X$$

Since the grades of membership extend the two-element set of truth values $\{0,1\}$ into the unit interval, it is worth recalling the collection of properties essential for set theory and investigating whether they are satisfied for fuzzy sets.

The De Morgan law, which is valid for set theory, is preserved for fuzzy sets as well, namely

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B},\ \overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

Distributivity laws are fulfilled and the properties of absorption and idempotency hold as well. However, the exclusion conditions are not satisfied, i.e.

$$A \cup \bar{A} \neq X \qquad \text{(underlap property)}$$
$$A \cap \bar{A} \neq \varnothing \qquad \text{(overlap property)}$$

At a relatively early stage of the development of fuzzy sets it was recognized that the semantics of the logical connectives can be expressed in many ways while this choice is usually driven by a particular application. One of the examples used was the product operation, $A(x)B(x)$, studied as a model used for logic intersection and the probabilistic sum, $A(x)+B(x)-A(x)B(x)$, considered for the union operation. In comparison to the lattice (max and min) operations, the computed degree of membership reflects both values of the membership functions $A(x)$ and $B(x)$. When pursuing these general models of logical connectives, it is justifiable to restrict ourselves to a class of binary operations satisfying a collection of the following sound assumptions:

(1) Boundary conditions

$$A \cup X = X,\ A \cap X = A$$
$$A \cup \varnothing = A,\ A \cap \varnothing = \varnothing$$

(2) Commutativity

$$A \cap B = B \cap A,\ A \cup B = B \cup A$$

(3) Associativity

$$(A \cap B) \cap C = A \cap (B \cap C),$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$

Observe also that interpreting the grades of membership as truth values of the corresponding propositions, all the above conditions take on an intuitively clear interpretation: for instance, the boundary conditions indicate that the logical connectives for fuzzy sets coincide with those applied in the two-valued logic. The property of commutativity states that the truth value of a composite expression does not depend on the order in which the predicates have been placed.

By accepting the above conditions, a broad class of models for logical connectives (union and intersection) is formed by triangular norms, cf. Dubois and Prade, 1988. The triangular norms (Menger, 1942) (t- and s-norms) originated in the theory of probabilistic metric spaces can be introduced consequently.

By a t-norm we mean a function of two arguments

$$t: [0,1] \times [0,1] \rightarrow [0,1]$$

such that it is

(1) Non-decreasing in each argument

$$\text{for } x \leq y,\ w \leq z,\ xtw \leq ytz$$

(2) Commutative

$$xty = ytx$$

(3) Associative

$$(xty)tz = xt(ytz)$$

(4) Satisfies the set of boundary conditions

$$xt0 = 0,\ xt1 = x$$
$$x,y,z,w \in [0,1]$$

By virtue of the introduced definition all the properties of the t-norm can be easily identified with the relevant characteristics of the intersection operation (logic AND).

As n-norm is defined as a function of two arguments

$$s:[0,1]\times[0,1]\rightarrow[0,1]$$

satisfying the following properties:

(1) It is a non-decreasing function in each argument.
(2) It is commutative.
(3) It is associative.
(4) It satisfies the boundary conditions.

$$x\,s\,0 = x,\; x\,s\,1 = 1$$

From this definition, we can deduce that (1)–(4) express the properties of the union operation. An interesting fact is that for each t-norm one can define an associated s-norm such that

$$x\,s\,y = 1-(1-x)\,t\,(1-y)$$

The above relation is nothing but De Morgan law existing in set theory.

Having an infinite family of triangular norms, one has a broad repertoire of formal models of logical connectives. The choice of a certain AND or OR operator can be influenced by the specificity of the problem itself.

Some other operations on fuzzy sets encountered in various applications are briefly summarized below:

(1) Normalization

$$\mathrm{Norm}(A)(x) = A(x)/\mathrm{hgt}(A)$$

(we assume that A is a non-empty fuzzy set, namely at least a single element of X belongs to A with a non-zero degree of membership, $\mathrm{hgt}(A) \neq 0$). This operation converts a subnormal fuzzy set A into its normal counterpart;

(2) The two successive operations called concentration and dilution

$$\mathrm{CON}(A)(x) = A^2(x)$$

$$\mathrm{DIL}(A)(x) = A^{1/2}(x)$$

yield new fuzzy sets with suppressed or elevated grades of membership. These operations are frequently utilized in linguistic approximation for modelling linguistic hedges (e.g. *more or less*, *very*, etc.). In general, one can introduce a pth power of a fuzzy set A^p. Values of the parameter p less than 1 produce a dilution effect, powers greater than 1 cause concentration. An operation called contrast intensification, $\mathrm{INT}(A)$, affects an original fuzzy set more selectively than the two others. It suppresses the grades of membership lower than $\frac{1}{2}$ and elevates those values that are above this threshold.

$$\mathrm{INT}(A)(x) = \begin{cases} 2A^2(x), \text{ if } A(x)<\frac{1}{2} \\ 1-2[1-A(x)]^2, \text{ otherwise} \end{cases}$$

Parameterization of this definition by adding an extra parameter p, $p>1$, makes the intensification more radical,

$$\mathrm{INT}(A,p)(x) = \begin{cases} 2^{p-1}A^p(x), \text{ if } A(x)<\frac{1}{2} \\ 1-2^{p-1}[1-A(x)]^{2p}, \text{ otherwise} \end{cases}$$

3.2. *Representation theorem*

A direct and essential relationship forming a bridge between fuzzy sets and sets is formulated by the representation theorem (Zadeh, 1973; Zimmermann, 1987). Firstly, we have to introduce the notion of an α-cut. By an α-cut, A_α, we mean a set of elements of A belonging to it with degrees of membership not less than α.

$$A_\alpha = \{x\in X\,|\,A(x)\geq\alpha\},\ \alpha\in[0,1]$$

Sometimes it is convenient to distinguish a strong α-cut formed by a collection of elements

$$\{x\in X\,|\,A(x)>\alpha\}$$

The α-cut operation selectively converts a fuzzy set into its Boolean version. The elements with the grades of membership above or equal to the threshold α are elevated to 1, whereas the remainder are eliminated from the set.

The representation theorem states that any fuzzy set A can be represented by a union of its α-cuts, namely,

$$A = \bigcup_{\alpha\in[0,1]} \alpha A_\alpha$$

This relationship is also referred to as a resolution identity. It is used quite frequently in situations where a fuzzy set needs to be translated into a collection of sets. These sets in turn facilitate further use of some standard optimization methods in processing fuzzy sets (like linear programming).

3.3. *Transformation of fuzzy sets and fuzzy numbers*

An important issue arises when one transforms fuzzy sets defined in a certain space X through a mapping f acting from X into Y. The resulting fuzzy set $Y = f(X)$ takes on a membership function computed as

$$[f(X)](y) = \sup_{x\in X:y=f(x)} [X(x)] \tag{1}$$

where the supremum is taken over all xs for which $y = f(x)$ holds. By definition, we accept that the supremum over an empty set gives a zero membership value. We can refer to Equation 1 as the extension principle (Zadeh, 1973).

This basic definition can be naturally expanded to cope with functions of many variables, say $Y = f(X, Z, W)$ etc. The supremum standing in Equation 1 is now taken with respect to all the elements for which the constraint $f(x, z, w) = y$ holds for y fixed. More precisely, one gets

$$[f(X,Z,W)](y) = \sup_{x,y,w: f(x,z,w)=y} \{\min[X(x), Y(y), W(w)]\} \quad (2)$$

One can look at Equation 2 as an optimization task involving constraints formed by 'f'.

For transformations applied to fuzzy numbers we will follow the same route. When fuzzy numbers A and B are given and transformed by the function f we derive

$$C = F(A, B)$$

such that

$$C(z) = \sup_{x,y \in \boldsymbol{R}: z=f(x,y)} \{\min[A(x), B(y)]\}$$

where $z \in \boldsymbol{R}$ and f: $\boldsymbol{R} \times \boldsymbol{R} \rightarrow \boldsymbol{R}$ is a function induced by F such that

$$F(\{x\}, \{y\}) = f(x, y)$$

The arithmetic operations (addition, subtraction, multiplication and division) will be of special interest as they are frequently employed in constructing relationships between a system's variables. The formal notation is the same as before; however for these specific situations one can derive more simplified expressions. Take, for instance, addition

$$C = A + B$$

The membership function of C is computed as

$$C(z) = \sup_{x,y \in \boldsymbol{R}: z=x+y} \{\min[A(x), B(y)]\}$$

The above constraint can be incorporated into the membership function of A or B so that computations of $C(z)$ are converted into a problem of unconstraint optimization

$$C(z) = \sup_{y \in \boldsymbol{R}} \{\min[A(z-y), B(y)]\}$$

The determination of the membership function of the fuzzy number could be a rather tedious task requiring a significant number of computations. To lower the computational burden we will introduce a parametric representation of fuzzy numbers contemplating L (left side) and R (right side) fuzzy numbers. By an L-fuzzy number, we mean a fuzzy number with the following membership function (Dubois and Prade, 1988):

(1) $L(-x) = L(x)$.
(2) $L(0) = 1$.
(3) L is increasing in $[0, +\infty]$.

Some frequently used forms of the L-fuzzy numbers include

$$L(x) = \max[0, (1 - |x|^p)]$$

$$L(x) = \frac{1}{1 + |x|^p}$$

$$p > 0$$

The same definition holds for R-fuzzy numbers.

An L–R fuzzy number is defined by means of the following membership function:

$$A(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right), & \text{if } x \leqslant m \\ R\left(\dfrac{x-m}{\beta}\right), & \text{if } x \geqslant m \end{cases}$$

where α, $\beta > 0$ are parameters controlling 'fuzziness' of the fuzzy number. For a degenerated case, $\alpha = \beta = 0$, one gets a genuine real number. For such an L–R representation, all the basic algebraic operations can be performed making use of the respective parameters of the representation. For instance, for addition and subtraction, we arrive at the following results for the fuzzy numbers $A(m, \alpha, \beta)$ and $B(n, \delta, \gamma)$

addition: $A + B = (m + n, \alpha + \delta, \beta + \gamma)$
subtraction: $A - B = (m - n, \alpha + \delta, \beta + \gamma)$

The relevant expressions for multiplication and division are more complicated and the results do not preserve the piecewise linear character of the original membership functions even though the arguments might be expressed in this way.

3.4. *Linguistic variable and linguistic approximation*

The term 'linguistic variable' coined by Zadeh (1978a) denotes a variable defined in a given universe of discourse taking on some linguistic values such as *small*, *medium*, *large*, etc. These values are modelled by fuzzy sets. Each linguistic variable involves a finite, usually very small, collection of generic linguistic terms (they are sometimes called primary terms). Syntactic and semantic rules describe how to process linguistic variables. With the use of syntactic rules one builds well-formed sentences and constructs non-primary terms. The semantic rules specify the way in which the meaning (membership functions) of the term can be encompassed.

More formally, let us consider a linguistic variable of pressure defined over the range of pressure values of interest. We will first recognize generic terms such as *small*, *medium*, *large* and admit a collection of hedges (modifiers) such as *slightly*, *very*, *more or less*, etc.

Well-formed sentences (wfs) will be constructed using these terms being combined according to the syntax rules. For example, a compound term *very high* consisting of the hedge *very* and the generic linguistic value *high* is a wfs.

The semantics describes how the membership function of this wfs is computed. In particular, we have already defined semantics of the three basic logic operators (AND, OR, NOT). Quite frequently the semantics of the hedges admits a somewhat powering effect on the membership function. Selected cases are summarized below:

$$\textit{very } A = A^2,\ \textit{plus } A = A^{1.25},\ \textit{more or less } A = A^{1/2},\ \textit{minus } A = A^{3/4}$$

or in general

$$h(A) = A^p,\ [h(A)](x) = A^p(x)$$

where h denotes a hedge. Hedges with $p > 1$ imply a concentration type of operation (*plus*, *very*, etc.). For $p < 1$ the corresponding hedges 'dilute' the fuzzy set on which they operate (*more or less*, *minus*, etc.). The semantics of the hedges has been extensively studied in Lakoff (1973). It has been noted that this model can be generalized by incorporating some translation of the generic term along the universe of discourse, say

$$[h(A)](x) = A^p(x - \tau)$$

where τ is used to denote a shift of the original membership function. The need for this shift has been discussed in Martin-Clouaire (1987).

The notion of linguistic approximation is used to express the process of matching (approximation) of a given fuzzy set by a collection of fuzzy sets (linguistic terms) $A_1, A_2, \ldots, A_c$ and a group of hedges $h_1, h_2, \ldots, h_p$. The approximation procedure for the fuzzy set B leads to its expression in terms of the most suitable A_1, s and h_js. The straightforward approach one can propose here consists of two steps:

(1) We approximate B by one of the A_is. This selection is guided by the obtained results of matching achieved for B and A_i. B is then approximated by the A_{i0} for which

$$\max_{i=1,2,\ldots,n} [\mathrm{Match}(A_i, B)] = \mathrm{Match}(A_{i0}, B)$$

where Match(.,.) denotes the matching operation performed for these two fuzzy sets;

(2) We determine the best hedge which, applied to A_{i0}, enhances this approximation:

$$\max_{j=1,2,\ldots,p} \{\mathrm{Match}[h_j(A_{i0}), B]\} = \mathrm{Match}[h_{j0}(A_{i0}), B]$$

As a result of this two-step approximation the fuzzy set B is represented as a single properly modified generic fuzzy set.

4. Calculus of fuzzy relations

As fuzzy sets constitute a conceptual extension of set theory so do fuzzy relations. Let us remind ourselves that relations in set theory are treated as elements of the Cartesian products of the spaces in which they are defined. By a fuzzy relation R defined in the Cartesian product $\boldsymbol{X} \times \boldsymbol{Y}$ we mean a mapping

$$R\colon \boldsymbol{X} \times \boldsymbol{Y} \to [0,1]$$

The fuzzy relation assigns a grade of membership $R(x, y)$ to each pair (x, y) of the elements of the above Cartesian product of the universes. The interpretation of the fuzzy relation is analogous to that provided for fuzzy sets. One can think about those membership values as representing strengths of connections (ties) occurring between elements x and y. The closer the value of membership $R(x, y)$ to 1, the stronger the link between x and y.

The relation 'x is similar to y' where x and y are two real numbers, $x, y \in \boldsymbol{R}$, is an example of a fuzzy relation. Its membership function can be of the following form,

$$R(x,y) = \begin{cases} \dfrac{1}{1 + (x-y)^4}, & \text{if } |y - x| \leq 5 \\ 0, & \text{otherwise} \end{cases}$$

Fuzzy relations can be composed. Two main types of the composition operations are defined below:

(1) Sup–min composition of two fuzzy relations R: $\boldsymbol{X} \times \boldsymbol{Z} \to [0,1]$ and S: $\boldsymbol{Z} \times \boldsymbol{Y} \to [0,1]$ generates another fuzzy relation denoted by $R \times S$ such that

$$(R \times S)(x, y) = \sup_{z \in \boldsymbol{Z}} \{\min[R(x,z), S(z,y)]\}$$

(2) Inf–max composition of R and S, $R \otimes S$

$$(R \otimes S)(x, y) = \inf_{z \in \boldsymbol{Z}} \{\max[R(x,z), S(z,y)]\}$$

These operations are dual, that is,

$$\overline{R \times S} = \bar{R} \otimes \bar{S}$$

The sup–min composition has the following properties that result from the relevant operations (max and min) used there:

(1) Distribution with respect to union

$$(R \cup T) \times S = R \times S \cup T \times S$$

(2) Preservation of the inclusion property, i.e.

$$\text{if } R_1 \subset R_2, \text{ then } R_1 \times S \subset R_2 \times S$$

$$R_1, R_2, R, T: \boldsymbol{X} \times \boldsymbol{Z} \to [0,1],\ S: \boldsymbol{Z} \times \boldsymbol{Y} \to [0,1]$$

However, no distributivity with respect to intersection holds,

$$(R \cap T) \times S \subset (R \times S) \cap (T \times S)$$

$$R, T: \boldsymbol{X} \times \boldsymbol{Z} \to [0,1],\ S: \boldsymbol{Z} \times \boldsymbol{Y} \to [0,1]$$

A generalization of the above compositions can be obtained by taking t- and s-norms. This yields:

(1) Sup-t composition

$$(R \circ S)(x, y) = \sup_{z \in \boldsymbol{Z}} [R(x,z) t S(z,y)]$$

(2) Inf-s composition

$$(R \oplus S)(x,y) = \inf_{z \in \boldsymbol{Z}} [R(x,z) s S(z,y)]$$

Let us discuss the sup–min composition performed for fuzzy set X and a particular fuzzy relation R of the following form

$$R(x, y) = \begin{cases} 1, & \text{if } f(x) = y \\ 0, & \text{otherwise} \end{cases}$$

(viz. R is a function). We immediately derive

$$(X \circ R)(y) = \sup_{x \in \boldsymbol{Z}} \{\min[X(x), R(x,y)]\} =$$

$$\sup_{x \in \boldsymbol{X}: f(x) = y} [X(x) \wedge 1] \vee \sup_{x \in \boldsymbol{X}: f(x) = y} [X(x) \wedge 0]$$

$$= \sup_{x \in \boldsymbol{X}: f(x) = y} X(x)$$

This gives another viewpoint of the derivation of the extension principle.

5. Methodological aspects of the use of fuzzy sets in modelling and simulation

As underlined in the previous sections, fuzzy sets deal with collections of objects with varying degrees of membership. As such they allow us to handle more complex concepts effectively. First we will concentrate on the main aspects of knowledge representation that are provided by fuzzy sets. A notion of a cognitive perspective emerging in this context will be utilized to select the most suitable granularity of information to be captured in the modelling process. The general three-phase scheme consists of the level of knowledge representation and knowledge processing carried out at the linguistic level. The

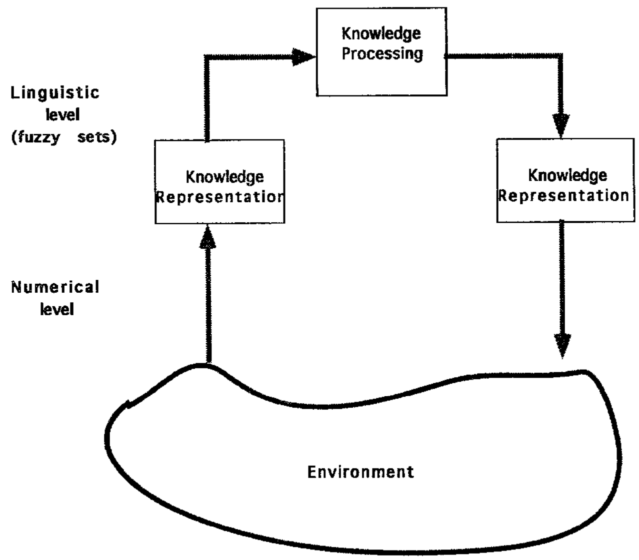

Fig. 2. Linguistic and numerical levels in fuzzy modelling: knowledge representation and knowledge processing.

models built with the aid of fuzzy sets (specifically fuzzy models) will be organized according to the level of their structural relationships and correspond directly with the amount of knowledge existing about the system. Finally, we will concentrate on the transformation of the results of simulation from the elements of the cognitive perspective to the numerical level. The entire scheme viewed as essential to the methodology of fuzzy modelling and simulation is visualized in Fig. 2.

The functional blocks of the scheme relate phenomena occurring at a numerical level with their models (as they are perceived by the user or model developer) formed at the level of the linguistic labels (fuzzy sets). The input interface realizes processes of knowledge representation by transforming all available data into coherent pieces of information that are afterwards used for building detailed model relationships. At this stage a cognitive perspective (called also a frame of cognition) plays a primordial role. The processing activity concentrates on the conceptual level within which all the essential relationships between the variables of the fuzzy models are constructed. The fuzzy models can take on various forms. Their detailed estimation algorithms can also be very distinct. Finally, the third conceptual block of the discussed scheme converts the outcomes of the processing realized at the conceptual level into the detailed numerical results.

5.1. *Frame of cognition*

Knowledge about the system as well as the perspective from which one is interested in taking a look at it is articulated with the aid of linguistic labels. These are generic pieces of knowledge which are deemed by the user as being essential in describing and understanding the system. The linguistic labels are represented by fuzzy sets. As demonstrated in Zadeh (1979) they can also be viewed as elastic constraints defined over a universe of discourse and identifying regions with the highest degree of compatibility of elements with the given linguistic term. Sometimes the linguistic labels are also referred to as information granules. All the information granules defined in a certain space constitute a frame of cognition of the variable (Pedrycz, 1990b; 1992a). More formally, the family of fuzzy sets

$$\{A_1, A_2, \ldots, A_c\}$$

(where $A_i: X \rightarrow [0,1]$) constitutes a frame of cognition A if the following two properties are satisfied:

(1) A 'covers' the universe X, namely each element of the universe is assigned to at least one granule with a non-zero degree of membership. It means that

$$\underset{x}{\forall}\ \underset{i}{\exists}\ A_i(x) > 0$$

This property assures that any piece of information defined in X is properly represented (described) by A_is;

(2) The elements of A are unimodal fuzzy sets (unimodal membership functions). By stating that, we identify several regions of X as (one for each A_i) highly compatible with the labels.

The frame of cognition can be developed either on a fully experimental basis or in an algorithmic way. In the first instance the linguistic labels can be specified by studying the problem and recognizing basic relevant information granules as being accepted as necessary in describing and handling it. In this way the subjective evaluation of the membership functions completed by the user of the model becomes a key factor. It is the user who provides relevant membership functions for the variables of the system and therefore creates his own individual cognitive perspective. In this regard the standard methods of membership function estimation discussed in Section 2.2 can be directly utilized.

The second approach which could be helpful when some records of numerical data are available relies on a suitable utilization of fuzzy clustering techniques. Fuzzy clustering (Bezdek, 1981) enables us to discover and conveniently visualize the structure existing in the data set. With the aid of it the numerical data are structured into a number of groups (clusters) according to a predefined proximity measure between data points. The number of clusters is also defined in advance so that they correspond to the linguistic labels constituting the frame of cognition. The algorithm generates grades of membership of the elements of the data set in the given clusters. If necessary, these grades can be also converted into an analytical form of the final membership functions.

The frame of cognition A can be also referred to as a fuzzy partition of X.*

We list three essential features of A:

(1) Specificity of the frame of cognition. The frame of cognition A' is more specific than A if all the elements of A' are more specific (with specificity defined for example in the sense of Yager, 1980) than the elements of A. Usually the number of elements of A' is greater than the number of the labels in A.

For example, the frame

$$A = \{\textit{Negative, Zero, Positive}\}$$

is less specific than the frame

$A' = \{$*Negative Large, Negative Medium, Negative Small, Zero, Positive Small, Positive Medium, Positive Large*$\}$

where now the variable takes on more levels of this linguistic quantification. The partition A' is less general than the previous one. The information granularity of A' is finer than that conveyed by A.

(2) Information hiding of the frame of cognition refers to each element of A. This feature means that some elements of X are made non-distinguishable by associating them with the same level of membership (usually equal to 1.0). For instance, the fuzzy set A_1 with the membership function defined as

$$A_1(x) = \begin{cases} \text{exponentially increasing over } (-\infty, x') \\ \quad \text{such that } A_1(-\infty) = 0, A_1(x') = 1 \\ 1, \text{ if } x \in [x', x''] \\ \text{exponentially decreasing over } (x'', \infty) \\ \quad \text{such that } A_1(x'') = 1, A_1(\infty) = 0 \end{cases}$$

makes all elements from $[x', x'']$ equivalent. By defining this membership function we selectively hide the information about the elements situated within the interval. In other words there is no distinction (at the level of specificity defined by the label A_1) between elements a_1 and a_2, $a_1, a_2 \in X$, in as far as both of them are included in the above interval.

Information hiding is completed on purpose so that all computational processes following this stage will not be carried out below this predefined conceptual level. We have already learned that this is an inherent property of set theory. Fuzzy sets allow us to add an extra flexibility to this term by parameterizing it along allowable grades of membership. In other words, λ-cuts are sets com-

*The fuzzy partition has an additional property: $A_1(x) + A_2(x) + \ldots + A_c(x) = 1$ which holds for every x; this constraint is automatically satisfied by most of the clustering algorithms but is usually not satisfied by the first method.

pleting information hiding at this specified level. Particularly, for the above trapezoidal fuzzy number used to construct the frame of cognition, the λ-cuts with $\lambda = 1$ imply that the information hiding is performed at its highest level.

For a fixed number of labels, the information hiding can be additionally accomplished by enhancing regions of X associated with higher grades of membership. For instance, the operation of contrast intensification applied to A amplifies 'high' membership values (greater than 0.5) and suppresses those which have already been viewed as insignificant.

(3) Robustness. Fuzzy sets constituting the frame A exhibit an interesting property of robustness. Due to smooth transitions in the membership functions of fuzzy sets they allow to a relatively high extent imprecision in the input information. Consider the input numerical datum $x \in \mathbf{R}$ which being exposed to the existing noises is received as x' and as such is mapped onto the frame A. This mapping describes the levels of activation of $A_1, A_2, \ldots, A_c \in A$ which become numbers from the unit interval. The noisy version of x induces

$$A_1(x'), A_2(x'), \ldots, A_c(x')$$

instead of

$$A_1(x), A_2(x), \ldots, A_c(x)$$

Usually $A_i(x)$ [or $A_i(x')$] is used as input information to be processed in the course of fuzzy modelling. The lower the difference between $A_i(x)$ and $A_i(x')$, the higher the robustness of the frame (its ability to tolerate noise). Thus the overall sum of absolute differences

$$r(x) = |A_1(x) - A_1(x')| + |A_2(x) - A_2(x')| + \ldots + |A_c(x) - A_c(x')| \tag{3}$$

can be viewed as a suitable indicator of the robustness property of A.

The overall measure of robustness can be defined by completing standard averaging over the universe of discourse, say

$$\int_X r(x)\,\mathrm{d}x \tag{4}$$

(We assume that this integral does make sense.)

It is also obvious that Equation 3 can attain low values since x and x' can generate relatively similar values of the membership function (for relatively low values of disturbances). In the case of sets, $A_i(x)$ and $A_i(x')$ could have distinct values of membership even for some very close values of x and x', say

$$x \approx x' \quad \text{but} \quad A_i(x) = 0 \quad \text{and} \quad A_i(x') = 1$$

For more details on the robustness problem refer to Pedrycz (1993).

It is worth indicating that the fuzzy partition leads to a homogeneous form of information. To elaborate a bit on this issue let us note that input information can be uniquely expressed in terms of linguistic labels. The converse is not true; knowing the representative of the input information at the level of the information granules one cannot reconstruct it in a unique manner. This phenomenon is due to the level of generality introduced by fuzzy sets. The relevant scheme of transformation can be portrayed qualitatively as follows:

$$\text{input information} \rightarrow \text{frame of cognition} \rightarrow \boldsymbol{x}$$

where the input information could be given in a numerical, interval, or fuzzy set format. The derived output $\boldsymbol{x}$ expresses this input information by providing degrees of matching (activation) of the elements in the fuzzy partition. Thus $\boldsymbol{x}$ becomes associated with the fuzzy partition and its semantics directly reflects the semantics of the linguistic labels. In a simple example where the input information is given precisely as a single numerical quantity x_0, the vector $\boldsymbol{x} \in [0,1]^c$ can be developed by taking the values of the possibility measure (Zadeh, 1978b) of x_0 expressed with respect to $A_1, A_2, \ldots, A_c$, namely

$$\boldsymbol{x} = [A_1(x_0) A_2(x_0) \ldots A_c(x_0)]$$

For interval-valued information $[x_1, x_2]$ the resulting vector $\boldsymbol{x}$ is expressed as

$$\boldsymbol{x} = \left[\sup_{x \in [x_1, x_2]} A_1(x) \quad \sup_{x \in [x_1, x_2]} A_2(x) \quad \cdots \quad \sup_{x \in [x_1, x_2]} A_c(x) \right]$$

The character of this input information results in higher values of most of the entries of $\boldsymbol{x}$. Apparently

$$\sup_{x \in [x_1, x_2]} A_i(x) \geqslant A_i(x_0)$$

for any $x_0 \in [x_1, x_2]$.

6. The paradigm of fuzzy modelling

The basic idea of fuzzy models and fuzzy modelling is to model or represent a problem at the level of linguistic labels. Models of this class are not used to represent relationships between variables at a numerical, pointwise level. The role of fuzzy modelling is to look at the system from a 'distance' by accepting a suitable cognitive perspective. The fuzzy partition constructed for each variable can be adjusted separately to meet the requirements of the modelling and simulation. Some details can be then selectively hidden and will not increase unnecessarily the computational burden of model building.‡ If more details are required one can formulate another more detailed fuzzy partition and re-design the fuzzy model. The fuzzy models developed in this way concur with the principle of incompatibility formulated by Zadeh (1978a); a similar formulation can be found in Puccia and Levins (1985). This principle states that any model building calls for a rational trade-off between significance (relevance) and precision achievable within the model. It should be stressed that one should sacrifice (to a certain degree) precision to reach an acceptable level of generality. Overall, the properties of the frame of cognition applied to modelling are inherited and afterwards become inbuilt into the models. For instance, the increase in robustness of the frame A implies higher robustness of the fuzzy model. Through the frame of cognition one can easily process heterogeneous pieces of data and provide interpretation of the results of modelling in the same linguistic categories. The core part of the identification activity will be therefore concentrated on expressing links between the linguistic labels. In the sequel we will be concerned with deriving these formal relationships. We will review several classes of fuzzy models, highlight their essentials and briefly look at the level of their structural dependencies. The models will be discussed starting from those with the lowest level of imposed structural relationships and proceeding with more structured classes of models such as relational, neural network-oriented and those based on the concepts arising from fuzzy arithmetic.

6.1. *Constraint propagation*

The model based on the principle of constraint propagation does not assume any particular structure as this is conveyed by the data set. The frame of cognition is formed for the input variables and all the fuzzy sets involved there are propagated through the data. For some combinations of the labels the data support the generic input fuzzy sets, and generate a corresponding fuzzy set in the output space. In some others, the input combination does not get enough numerical support and is filtered out by the data.

As a concise illustration let us consider a system with two input variables and a single output. For the first input variable we assume the linguistic label given by the fuzzy set A; the label for the second variable we will denote by B. The data set used for identification purposes makes up a collection of N-triples of data: (input variable$_1$, input variable$_2$, output). We accept non-fuzzy data i.e. pointwise values (x_k, z_k, y_k), $k = 1, 2, \ldots, N$.

‡In a so-called qualitative modelling all relationships are defined between symbolic quantities, cf. De Kleer and Brown (1984); the extension including linguistic variables is reported in D'Ambrosio (1989).

Let us construct α-cuts of fuzzy sets A and B, A_α and B_α. They play the role of constraints imposed on the structure of the system. At the same time they direct a focused search through the data. A_α and B_α 'covers' some portion of the data viz. all x_k and y_k such that $x_k \in A_\alpha$, $y_k \in B_\alpha$. The induced subsets of the output y_k denoted symbolically by $(\{y_k\}|A)_\alpha$ will be formally put down as

$$(\{y_k\}|A)_\alpha = \{y_k | x_k \in A_\alpha\}$$

Similarly one gets the induced subsets for the second variable,

$$(\{y_k\}|B)_\alpha = \{y_k | z_k \in B_\alpha\}$$

We jointly propagate the constraints A and B (more exactly their α-cuts) across the data. This is stated as an intersection of the induced subsets,

$$(\{y_k\}|A \;\&\; B)_\alpha = \{y_k | x_k \in A_\alpha\} \cap \{y_k | z_k \in B_\alpha\} \tag{5}$$

The results of this propagation can be reconstructed as a fuzzy set C defined in the output variable. The general idea of constraint propagation is displayed as below:

$$A, B - \underset{\text{the data}}{\text{propagation through}} - C$$

Equation 5 describes α-cuts of the fuzzy set C which, due to the representation theorem, is reconstructed by taking their union,

$$C = \bigcup_{\alpha \in [0,\pi]} \alpha C_\alpha$$

where π stands for the highest value of the membership function of C (height of C)

$$\pi = \mathrm{hgt}(C)$$

and this corresponds to the highest value of α yielding a non-empty intersection,

$$\{y_k | x_k \in A_\alpha\} \cap \{y_k | z_k \in B_\alpha\} \neq \varnothing \quad \text{for } \alpha \leqslant \pi$$

$$\{y_k | x_k \in A_\alpha\} \cap \{y_k | z_k \in B_\alpha\} = \varnothing \quad \text{for } \alpha > \pi$$

The height of the generated fuzzy set π can be used as a performance measure expressing the coincidence of the labels with the data. This type of possibility-based criterion can be accompanied by the probabilistic type of the performance criterion with which one expresses the following probability

$$p(\alpha) = \frac{\text{number of data included in } C_\alpha}{\text{total number of data } (N)}$$

Since $C_{\alpha_1} \subset C_{\alpha_2}$ for $\alpha_1 > \alpha_2$ then $p(\alpha)$ is a non-increasing function of α. Again the constraints A and B with a low value of the probability $p(\alpha)$ should be discarded as being a meaningless combination of the labels.

In the case of 'n' input variables and several linguistic labels, the procedure of constraint propagation should be repeated separately for each of them. The combinations with the highest value of π will be used as allowable components (descriptions) of the model.

For n variables with m linguistic labels (fuzzy sets) we have m^n possible combinations. Usually a significant portion of them are ruled out by the data. The choice of labels is based on analysis of the derived fuzzy sets of the output variable and involves analysis of their overlap as well as their heights or associated probabilities. The trade-off exists between the overlap of these fuzzy sets (which calls for more specific labels for the input variables) and a lack of complete coverage ('explanation') of all the data provided by the constraints. The latter phenomenon suggests the use of coarser constraints (of lower granularity) with a higher overlap.

6.2. *Relational calculus and relational models*

Models based on relational calculus express links between linguistic labels in terms of relations. In other words, the statements about a system's variables read as

'there is a relationship between A and B'
(A and B are related)

or more formally

$$ARB$$

where A and B are the linguistic labels in the two universes of discourse in which these variables are defined. Models of this class capture dependencies between the linguistic labels and express them as fuzzy relations. Fuzzy sets, fuzzy relations, and the calculus of these objects are put together into a form of fuzzy relational equations. These equations, originally developed outside fuzzy sets as a branch of relational calculus and applied vigorously to problems of operations research (cf. Rudeanu, 1974), were reformulated and generalized in Sanchez (1976). Some interesting links with multivalued logic have been underlined in Ledley (1968). Since then many theoretical results have been obtained and the methodology of their use followed by a series of specific applications has been formulated.

We will be concerned with discrete versions of fuzzy relational equations, namely the equations defined in finite universes of discourse. The fuzzy sets $\boldsymbol{x}$ and $\boldsymbol{y}$ include levels of activation of the frames in which input and output variables are expressed, thus $\boldsymbol{x}\in[0,1]^n$, $\boldsymbol{y}\in[0,1]^m$. The general statement

$\boldsymbol{x}$ and $\boldsymbol{y}$ are related (R)

can be translated in many ways. The main classes of relational structures (and subsequently fuzzy relational equations) are summarized below. The fuzzy set $\boldsymbol{y}$ results from $\boldsymbol{x}$ and R as

$$\boldsymbol{y} = \boldsymbol{x} \circ R$$

$R \in [0,1]^{n\times m}$. The composition operator '$\circ$' involves triangular norms (t- and s-norms)

$$y_j = \mathop{\boldsymbol{S}}_{i=1}^{n} [x_i t r_{ij}]$$

$j = 1, 2, \ldots, m$.

Depending on the combination of triangular norms with which one is faced is a well-known max–min composition used in most of the existing constructs in fuzzy sets.

The relational structure can be treated as a relational equation with the two generic problems formulated accordingly:

(1) $\boldsymbol{x}$ and $\boldsymbol{y}$ are given, determine R;
(2) $\boldsymbol{y}$ and R are given, determine $\boldsymbol{x}$.

For the max–min and max-t composition there are analytical solutions available. Furthermore the non-empty family of solutions contains more than a single solution. The extremal solutions (maximal ones) are determined with the use of the residuation operation (φ-operator, Pedrycz, 1985; Di Nola *et al.*, 1989) associated with the t-norm standing in the equation. The fundamental results are then concisely summarized. For problem (1): if the family of solutions $R \neq \phi$ then the maximal element of R, $\hat{R} = \max R$ is determined through applying the φ-operation applied pointwise to $\boldsymbol{x}$ and $\boldsymbol{y}$.

$$\hat{R} = \boldsymbol{x}\varphi\boldsymbol{y}$$

i.e.

$$\hat{r}_{ij} = x_i\varphi y_j$$

where $a\varphi b = \sup\{c \in [0,1]/atc \leq b\}$. For (2): assuming that the family of solution X is non-empty, its maximal element is calculated as

$$\hat{\boldsymbol{x}} = R\varphi\boldsymbol{y}$$

or

$$\hat{x}_i = \min_{j=1,2,\ldots,m} (r_{ij}\varphi y_j)$$

The dual class of the fuzzy relational equations involves the t–s composition. Now

$$\boldsymbol{y} = \boldsymbol{x}\Delta R$$

$$y_j = \mathop{\boldsymbol{T}}_{i=1}^{n} (x_i s r_{ij})$$

$j = 1, 2, \ldots, m$.

Again the two more specialized families of the composition operators include the min–max and min–s aggregation. The analytical solutions are available in these two cases. The character of the results is dual to those reported previously. Briefly, the obtained solutions are minimal in the family of solutions: for (1)

$$\check{R} = x\beta y \quad r_{ij} = x_i \beta y_j$$

for (2)

$$\check{x} = R\beta y \quad \check{x}_i = \max_j [r_{ij} \beta y_j]$$

where the definition of the β-operator uses the s-norm standing in the equation

$$a\beta b = \inf\{c \in [0,1] \mid asc \geq b\}$$

6.3. *Fuzzy neural networks*

In this section we will take another look at a more structured family of fuzzy models by studying fuzzy neural networks. By this class of networks we mean distributed and parallel computing structures employing extensively the logical operations existing in the theory of fuzzy sets. As opposed to standard neural networks, the networks emerging within this framework are usually heterogeneous i.e. they consist of neurons of a different conceptual and numerical character. When put together they exhibit diverse functional characteristics and play quite distinct roles in the network. Fuzzy neural networks originated as a natural extension of the fuzzy relational models, which, in this context, are simple two-layer neural networks. Studies of these models have been initiated in Pedrycz (1990a; 1991); refer also to some application-oriented studies (Pedrycz, 1993).

First we will discuss basic models of neurons (aggregative and reference ones) and afterwards concentrate on logical processors constituting a generic architecture of the fuzzy neural networks. In the sequel, learning algorithms will be studied.

6.3.1. *Aggregative and matching functions of logical neurons*

The logic-based neurons aggregate input signals $x_1, x_2, \ldots, x_n \in [0,1]$ using some basic logic operations. The two basic logical connectives (AND and OR) give rise to so-called AND and OR neurons. The AND neuron ANDs the input signals

$$y = x_1 \text{ AND } x_2 \text{ AND } \ldots \text{ AND } x_n \tag{6}$$

For the OR neuron we obtain

$$y = x_1 \text{ OR } x_2 \text{ OR } \ldots \text{ OR } x_n \tag{7}$$

Both the AND and OR connectives are represented as triangular norms (t- and s-norms). A straightforward extension of the above formulas is to include weights (connections) associated with the inputs. In this way Equations 6 and 7 are translated into the following formulas:

AND neuron

$$y = (x_1 \text{ OR } w_1) \text{ AND } (x_2 \text{ OR } w_2) \text{ AND } \ldots \text{ AND } (x_n \text{ OR } w_n) \qquad *8)$$

OR neuron

$$y = (x_1 \text{ AND } w_1) \text{ OR } (x_2 \text{ AND } w_2) \text{ OR } \ldots \text{ OR } (x_n \text{ AND } w_n) \tag{9}$$

$w_i \in [0,1]$, $i = 1, 2, \ldots, n$.

The weights are used to enhance or eliminate the influence of x_is on the output y:

(1) The lower the value of w_i the more evident the influence of x_i on y (AND neuron);

(2) Higher values of w_i enhance the importance of x_i (OR neuron).

The AND (OR) neuron can be enhanced functionally in two different ways:

(1) The complemented input signals, $\bar{x}_i = 1 - x_i$. This allows us to realize the inhibitory performance of the neuron while still preserving the unit interval as a suitable range of coding for the connections. By choosing appropriate values of the connections in the neuron it can easily exhibit inhibitory and excitatory characteristics.

(2) The second modification involves the inclusion of a non-linear transformation following the AND (OR) neuron. Its role is to modify (calibrate) the obtained grades of membership. This functional block does not affect the logical properties of the neuron. The standard two-parametric sigmoid function

$$z = \frac{1}{1 + \exp(-(y - m)^2/\alpha)}$$

can serve as one of the possible instances. The parameters $m \in [0,1]$ and $\alpha > 0$ are adjusted to arrive at an appropriate calibration.

The standard s–t composition operator applied to the fuzzy set x and the fuzzy relation R has an equivalent representation in terms of m OR neurons where each of them possesses n inputs. Recall that the basic relational equation can be rewritten as a series of m expressions

$$y_j = \underset{i=1}{\overset{n}{S}} (x_i t r_{ij})$$

$j = 1, 2, \ldots, m$.

6.3.2. *Logical processors as basic processing units*

The AND and OR neurons can be put together to form a so-called logic processor (LP). Roughly speaking, the aim of the LPs is to approximate any function with the use of logic-based neurons. This form of approximation reveals logical characteristics of the approximated function (or a collection of experimental data). Essentially there are two structures of the logic processor:

(1) The first one consisting of three layers can be viewed as a sum of products (SOM). The input layer consists of $2n$ nodes and includes both x_is as well as their complements ($\bar{x}_i$). The hidden layer includes p AND nodes. The output layer is built with a single OR node.

(i) the hidden layer forms p minterms z_j

$$z_j = \mathop{\boldsymbol{T}}_{i=1}^{2n} (w_{ij} s x_i') \qquad j = 1, 2, \ldots, p$$

where x' is an extended vector of the $2n$ inputs including complemented values of all x_is,

$$\boldsymbol{x}' = [x_1\ x_2\ \ldots\ x_n\ \bar{x}_1\ \bar{x}_2\ \ldots\ \bar{x}_n]$$

(ii) output layer. The minterms are combined by taking the OR operation on z_js

$$y = \mathop{\boldsymbol{S}}_{j=1}^{p} (v_j t z_j)$$

(2) The dual structure of the logical processor computes y by considering a product of minterms and combining the results of the hidden layer by AND-ing them. We will refer to this structure as a product of maxterms (POM). Its formal model is given accordingly.

(i) hidden layer

$$z_j = \mathop{\boldsymbol{S}}_{i=1}^{2n} (w_{ij} t x_i') \qquad j = 1, 2, \ldots, p$$

(ii) output layer

$$y = \mathop{\boldsymbol{T}}_{j=1}^{p} (v_j s z_j)$$

6.3.3. *Learning*

The discussed neural networks require learning. The basic learning procedures are mainly of a parametric nature and deal with a series of suitable adjustments of the weights (connections) of the network. The supervised learning is carried out on the basis of a so-called learning set of input–output patterns $(\boldsymbol{x}_k, y_k)$, $k = 1, 2, \ldots, N$, and is driven by a specified performance index Q. Usually Q is given as the sum of squared errors measuring distances between y_ks and the values at the output of the network while driven by $\boldsymbol{x}_k$, say $N(\boldsymbol{x}_k)$,

$$Q = \sum_{k=1}^{N} [y_k - N(\boldsymbol{x}_k)]^2$$

where $N(\bullet)$ stands for a general notation of the output of the network obtained for $\boldsymbol{x}_k$. The adjustments of the connections are made according to a standard Newton-like method. The abbreviated form of the scheme looks as follows

$$(\text{connections})_{\text{new}} = (\text{connections}) - \alpha \frac{\partial Q}{\partial(\text{connections})}$$

while α specifies a rate of learning. This rate implies a suitable speed of learning. Too high values of α could result in oscillations in learning, too small values could cause a very slow learning.

The general learning formula can be applied to different networks upon specification of all details (such as e.g. triangular norms and the topology of the network).

In the logic processor the size of the hidden layer determines its representation capabilities, i.e. uniquely specifies a number of the generalized minterms (maxterms) of the hidden layer that are used to approximate the data (function). The determination of the size of p is out of the stream of the parametric learning. Its choice should be directed by the values of Q. Two basic strategies are worth considering in this regard:

(1) Successive expansions: the values of Q are used to guide a growth of the network. Starting from some small values of p we successively increase the size of the hidden layer. This process is terminated once Q viewed as a function of p tends to stabilize;

(2) Successive reductions: starting from a large number of nodes in the hidden layer it is successively reduced up to a point at which the values of the performance index Q incline to increase significantly.

6.4. *Fuzzy regression models*

The essential concept used in fuzzy regression models is to describe relationships between input–output data using parameters viewed as fuzzy numbers (Tanaka *et al.*, 1982; Heshmaty and Kandel, 1985; Tanaka, 1987; Savic and Pedrycz, 1991),

$$Y = A_1 x_1 + A_2 x_2 + \ldots + A_n x_n$$

where $x_1, x_2, \ldots, x_n \in \boldsymbol{R}$ are independent variables and the parameters $A_1, A_2, \ldots, A_n$ are fuzzy numbers. In fact, we usually treat A_j as symmetrical triangular fuzzy numbers such that

$$A_j(a) = \begin{cases} 1 - \dfrac{|\alpha_j - a|}{c_j} & \text{if } \alpha_j - c_j \leq a \leq \alpha_j + c_j \\ 0 & \text{otherwise} \end{cases}$$

$a \in \boldsymbol{R}$. Note that α_j is a modal value of the jth parameter whereas c_j becomes its spread. The spread characterizes the precision of the model parameter. The values of c_js are used in order to absorb deficiencies of the constructed model, as well as to incorporate ('neutralize') noise existing in the data set.

The fuzzy regression models are characterized by a higher level of structural linkage: they utilize extensively the linear type of input–output relationships (even though they are expressed in terms of fuzzy numbers). The identification algorithms are well defined and computationally efficient. It has been shown that they reduce to standard schemes of linear programming.

6.5. *Local regression models*

The idea behind this class of fuzzy models, introduced in Takagi and Sugeno (1985), is to replace a single 'global' model by a series of 'local' models. These are usually easier to construct and verify. Each of these models is valid in a certain local and fairly limited region of the input variables $\boldsymbol{x} \in \boldsymbol{R}^n$. The resulting global model is aggregated with the aid of the conditional statements about the local models constructed so far,

$$\begin{aligned} &\text{if } \boldsymbol{x} \in \Omega_1 \quad \text{then } y = \psi_1(\boldsymbol{x}, \boldsymbol{a}_1) \\ &\text{if } \boldsymbol{x} \in \Omega_2 \quad \text{then } y = \psi_2(\boldsymbol{x}, \boldsymbol{a}_2) \\ &\quad\vdots \\ &\text{if } \boldsymbol{x} \in \Omega_p \quad \text{then } y = \psi_p(\boldsymbol{x}, \boldsymbol{a}_p) \end{aligned}$$

The conditional parts of the above rules specify regions Ω_j of the input variables, namely $\Omega_j \subset \boldsymbol{R} \times \boldsymbol{R} \times \ldots \times \boldsymbol{R}$ specific for the individual local models. The conclusion parts include linear or non-linear relationships $\psi_j(\boldsymbol{x}, \boldsymbol{a}_j)$ pertaining to the local models.

To make the model complete, the family $\Omega_1, \Omega_2, \ldots, \Omega_p$ should form a Boolean partition of the n-fold Cartesian product of $\boldsymbol{R}$. The replacement of the Boolean regions by their fuzzy-set-based versions enables us to avoid eventual discontinuities occurring when moving from one local model to another – in this way the fuzzy partition preserves a highly desired property of 'continuity' of the global model. The method described in Takagi and Sugeno (1985) allows us to determine the parameters of the linear local models, $\psi_i(\boldsymbol{x}, \boldsymbol{a}_i) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}_i$ through the use of standard regression techniques.

7. Linguistic–numerical conversion in fuzzy models

Fuzzy models, as emphasized at the beginning of the discussion, are constructed at the level of linguistic labels. The output variable $\boldsymbol{y}$ summarizes a distribution of the activation levels of the fuzzy sets forming the frame of cognition. In some applications it might be also desirable to translate this linguistic information into its equivalent numerical form. This process of transformation occurs quite often in fuzzy controllers (or some classes of expert systems) where a final non-fuzzy control or decision value has to be inferred. The standard route (being a relative of the centre-of-gravity method) followed can be described accordingly. Let $\boldsymbol{y}$ be defined over a space of the fuzzy sets (linguistic labels) $A_1, A_2, \ldots, A_n : \boldsymbol{R} \to [0,1]$. Assume that each A_i also has a non-fuzzy (viz. numerical) representative (this could be viewed as e.g. a modal or a mean value of its membership function). Denote these representatives by $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, respectively. The vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$ is converted into a single numerical quantity x_0 by considering a weighted sum of $\bar{x}_i$s

$$x_0 = \frac{\bar{x}_1 y_1 + \bar{x}_2 y_2 + \ldots + \bar{x}_n y_n}{y_1 + y_2 + \ldots + y_n}$$

In limit cases one derives:

(1) If only a single label becomes active i.e. $\boldsymbol{y}$ contains all the entries but one (namely j_0) equal to zero, then the resulting x_0 is equal to the non-fuzzy representative of the j_0th linguistic label, $\boldsymbol{x}_0 = \bar{\boldsymbol{x}}_{j0}$;

(2) If all the entries of y are equal, say $y = \beta$, then x_0 happens to be an average of the numerical representatives of A_is ($\bar{x}_i$);

$$x_0 = \frac{\bar{x}_1\beta + \bar{x}_2\beta + \ldots + \bar{x}_n\beta}{\beta + \beta + \ldots + \beta} = \frac{\bar{x}_2 + \bar{x}_2 + \ldots + \bar{x}_n}{n}$$

This transformation from the level of fuzzy sets to numbers could constitute a potential source of errors which is, in any case, unavoidable and caused by the transition realized between the two distinct conceptual levels of information processing.

8. Conclusions

We have investigated fuzzy sets as a formalism capable of representing and processing the uncertainty visible in many systems, particularly those where a factor of human interaction with the environment plays a central role. The methodology of fuzzy sets in developing fuzzy models has been discussed in depth. A particular emphasis put on distinguishing between the linguistic (conceptual) and numerical (real-world) levels makes it possible to comprehend the role of fuzzy sets in developing fuzzy models. The models discussed in the paper have been arranged according to the levels of their structural dependencies. By considering an amount of initial knowledge available about the system to be modelled one can select the most suitable class of models. Subsequently, the straightforward adjustment of the cognitive perspective (easily accomplished by changing the number as well as the form of the corresponding fuzzy sets) allows the fuzzy models to be customized to meet the requirements of the individual user. The flexibility of the customized models achieved in this way is significant. While studying several types of fuzzy models the discussion has been kept at a general level. The intention was that the level of the structural relationships to be grasped in order to work with a specific class of models should be made clear. Based on this global selection criterion the chosen model can be effectively constructed through the use of one of the discussed methods. Fuzzy sets realize subsymbolic computations; the linguistic terms can be thus either treated as pure symbols or could include membership functions reflecting the semantics residing within the terms. This feature is of special value in constructing links between the architectures of artificial intelligence (e.g. knowledge-based systems) and purely numerical structures such as neural networks.

Acknowledgement

The author acknowledges support from the Natural Sciences and Engineering Research Council of Canada and MICRONET.

References

Bezdek, J. C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.

Borel, E. (1950) *Probabilité et Certitude*, Press Université de France, Paris.

D'Ambrosio, B. (1989) *Qualitative Process Theory Using Linguistic Variables*, Springer-Verlag, New York, Berlin.

De Kleer, J. and Brown, J. S. (1984) A qualitative physics based on confluence. *Artificial Intelligence*, **24**, 7–83.

Di Nola, A., Sessa, S., Pedrycz, W. and Sanchez, E. (1989) *Fuzzy Relational Equations and Their Applications in Knowledge Engineering*, Kluwer Academic Press, Dordrecht.

Dubois, D. and Prade, H. (1988) *Possibility Theory – An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York.

Heshmaty, B. and Kandel, A. (1985) Fuzzy linear repression and its application to forecasting in uncertain environment, *Fuzzy Sets and Systems*, **15**, 159–191.

Lakoff, G. (1973) Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, **2**, 458–508.

Ledley, R. S. (1968) *Digital Computer and Control Engineering*, McGraw-Hill, New York, p. 196.

Martin-Clouaire, R. (1987) Semantics and computation of the generalized modus ponens: efficient deduction in fuzzy logic, in *Uncertainty in Knowledge-Based Systems*, Bouchon, B. and Yager, R. R. (eds), Springer-Verlag, Berlin, pp. 123–136.

Menger, K. (1942) Statistical metric spaces. *Proceedings of the National Academy of Science, USA*, **28**, 535–537.

Pedrycz, W. (1985) On generalized fuzzy relational equations and their applications. *Journal of Mathematical Analysis and Applications*, **107**, 520–536.

Pedrycz, W. (1990a) Processing in relational structures: fuzzy relational equations. *Fuzzy Sets and Systems*, **40**, 77–106.

Pedrycz, W. (1990b) Fuzzy sets framework for development of a perception perspective. *Fuzzy Sets and Systems*, **37**, 123–137.

Pedrycz, W. (1991) Neurocomputations in relational systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 289–296.

Pedrycz, W. (1992a) Selected issues of frames of knowledge representation realized by means of linguistic labels. *International Journal of Intelligent Systems*, **7**, 155–170.

Pedrycz, W. (1992b) Fuzzy neural networks with reference neurons as pattern classifiers. *IEEE Transactions on Neural Networks* (in press).

Pedrycz, W. (1993) *Fuzzy Control and Fuzzy Systems*, Research Studies Press/J. Wiley, Taunton/New York.

Puccia, Ch. J. and Levins, R. (1985) *Qualitative Modeling of Complex Systems*, Harvard University Press, Cambridge MA, London.

Rudeanu, S. (1974) *Boolean Functions and Equations*, North Holland, Amsterdam.

Saaty, T. L. (1980) *The Analytic Hierarchy Processes*, McGraw-Hill, New York.

Sanchez, E. (1976) Resolution of composite fuzzy relation equations. *Information and Control*, **34**, 38–48.

Savic, D. A. and Pedrycz, W. (1991) Evaluation of fuzzy linear regression models. *Fuzzy Sets and Systems*, **39**, 51–63.

Shin, Y. C., Chen, Y.-T. and Kumara, S. (1992) Framework of an intelligent grinding process advisor. *Journal of Intelligent Manufacturing*, **3**, 135–148.

Takagi, T. and Sugeno, M. (1985) Fuzzy identification of systems and the applications to modelling and control. *IEEE Transactions on Systems, Man and Cybernetics*, **15**, 116–132.

Tanaka, H. (1987) Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets and Systems*, **24**, 363–375.

Tanaka, H., Uejima, S. and Asai, K. (1982) Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics*, **6**, 903–907.

Yager, R. R. (1980) On measuring specificity, Technical Report, Iona College, New Rochelle, NY.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control*, **8**, 338–353.

Zadeh, L. A. (1973) Outline of a new approach to analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*, **1**, 28–44.

Zadeh, L. A. (1978a) PRUF-a meaning representation language for natural language. *International Journal of Man–Machine Studies*, **10**, 395–446.

Zadeh, L. A. (1978b) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3–28.

Zadeh, L. A. (1979) Fuzzy sets and information granularity, in *Advances in Fuzzy Set Theory and Applications*, Gupta, M. M., Ragade, R. K. and Yager, R. R. (eds), North Holland, Amsterdam, pp. 3–18.

Zimmermann, H. J. (1987) *Fuzzy Sets, Decision Making and Expert Systems*, Kluwer Academic Publishers, Boston.

Zimmermann, H. J. (1992) Approximate reasoning in manufacturing, in *Intelligent Design and Manufacturing*, Kusiak, A. (ed.), J. Wiley & Sons, New York, pp. 701–722.