Note: Fuzzy Sets
(Revised)

A. Inoue

January 6th, 2003

1 Crisp Set

In order to study fuzzy sets, it is necessary to understand (crisp) sets and its operations. In addition, characterization of sets by means of indicator functions is described.

Definition 1 *Let* $\mathbf{X}$ *be a universal set. Then a (crisp) subset* $A \subseteq X$ *is characterized by an indicator function* $\mu_A : \mathbf{X} \mapsto \{0, 1\}$ *such that*

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

Example 1 *Subset of* $\mathcal{Z}$ *(set of integers) include (but are not limited to)*

- $\mathcal{Z}^+$ *(positive integers)*
- $\mathcal{N}$ *(natural numbers)*
- *Odd numbers*
- *Even numbers*
- *Prime numbers*

1.1 crisp partition

Definition 2 *A family of disjoint nonempty subsets of a set* A *denoted by* $\pi(A)$ *such that*

$$\pi(A) = \{A_i | i \in I and A_i \subseteq A\}$$

where $A_i \neq \emptyset$ *is a (crisp) partition on* A *iff*

$$A_i \cap A_j = \emptyset \; \forall \, i, j \in I, \; i \neq j$$

and

$$\cup_{i \in I} A_i = A$$

Example 2 *A partition $\pi(A)$ for set $A = \{a, b, c, d\}$ consists of the following (disjoint nonempty) subsets*

- $A_1 = \{a\}$
- $A_2 = \{b, c\}$
- $A_3 = \{d\}$

1.2 Cardinality

Consider a countable set A: a set whose elements are labeled by positive integers (formally $\exists f : A \mapsto \mathcal{N}$ s.t. f is a bijection). **Cardinality** of set A is the number of elements in this set denoted by $|A|$ and it is computed as follows

$$|A| = \Sigma_{x \in A} \mu_A(x)$$

Example: Let $A = \{a, b, c, d, e\}$. The cardinality of set A is

$$|A| = 5$$

1.3 Crisp Set Operations (Standard-Algebraic)

(Two) Indicator functions of set operations are determined as follows (standard–algebraic)

1. Intersection $A \cap B$

$$\mu_{A \cap B}(x) = min(\mu_A(x), \mu_B(x)) = \mu_A(x) \times \mu_B(x)$$

2. Union $A \cup B$

$$\mu_{A \cup B}(x) = max(\mu_A(x), \mu_B(x)) = \mu_A(x) + \mu_B(x) - \mu_A(x) \times \mu_B(x)$$

3. Complement $\overline{A}$

$$\mu_{\overline{A}}(x) = 1 - \mu_A(x)$$

4. Set difference (relative Complement) $B - A \ = \ B \cap \overline{A}$

$$\mu_{B-A}(x) = min(\mu_B(x), 1 - \mu_A(x))$$

1.4 Properties of Set Operations

Properties of set operations include

1. Commutativity

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

2. Associativity
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

3. Distributivity
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

4. Idempotency
$$A \cap A = A, \; A \cup A = A$$

5. Identity
$$A \cup \emptyset = A, \; A \cap X = A$$

6. Involution
$$\overline{\overline{A}} = A$$

7. *Law of Contradiction*
$$A \cap \overline{A} = \emptyset$$

8. *Law of Excluded Middle*
$$A \cup \overline{A} = X$$

9. De Morgan's Law
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

2 Fuzzy Set

2.1 Basics

Introduced in 1965 by Prof. L. A. Zadeh, as a modest extension of the classical notion of set, the notion of fuzzy set proved to have far reaching, unexpected impact. The idea is that unlike for a crisp set, which is completely determined by an indicator function taking values in $\{0, 1\}$, a fuzzy set is characterized by a membership function taking values in $[0, 1]$. A fuzzy set f is said to be *normal* if there exists $x \in X$ such that $\mu_f(x) = 1$. (it is said to be *subnormal*, otherwise.)

Definition 1 *Let* $\mathbf{X}$ *be a universal set. Then a* <u>*fuzzy (sub)set*</u> $\mathbf{f}$ *is defined by means of its membership function:*

$$\mu_f : \mathbf{X} \mapsto [0, 1] \tag{1}$$

Example 1 *Fuzzy sets tall, average, and short are defined by means of their membership functions as shown in Figure 1:*

$$\mu_{tall}(h) = \begin{cases} 0 & \textit{if } h \leq 170 \\ \frac{h-170}{10} & \textit{if } 170 < h < 180 \\ 1 & \textit{otherwise} \end{cases}$$

$$\mu_{average}(h) = \begin{cases} 0 & \textit{if } h \leq 160, h \geq 180 \\ \frac{h-160}{10} & \textit{if } 160 < h < 170 \\ \frac{180-h}{10} & \textit{if } 170 < h < 180 \\ 1 & \textit{if } h = 170 \end{cases}$$

$$\mu_{short}(h) = \begin{cases} 1 & \textit{if } h \leq 160 \\ \frac{170-h}{10} & \textit{if } 160 < h < 170 \\ 0 & \textit{if } h \geq 170 \end{cases}$$

2.2 Other Types of Fuzzy Sets

There are other types of fuzzy sets as follows

- Interval valued Fuzzy Sets

  $$A : X \mapsto \epsilon[0, 1]$$

  where ϵ is a family of intervals $[0, 1]$

- Type 2 Fuzzy Sets

  $$A : X \mapsto \mathcal{F}([0, 1])$$

  where $\mathcal{F}$ is a set of fuzzy sets defined on the interval $[0, 1]$

- Level 2 Fuzzy Sets

  $$A : \mathcal{F}(X) \mapsto [0, 1]$$

  where $\mathcal{F}$ is a set of fuzzy sets defined on the universal set X

They certainly have more powerful representation of uncertainties, and they are interesting from theoretical aspects. Unfortunately, it will be too computationally expensive from practical aspects. Besides, equivalence models using ordinary fuzzy sets (with certain restrictions) can be drawn from aspects of fuzzy relations.

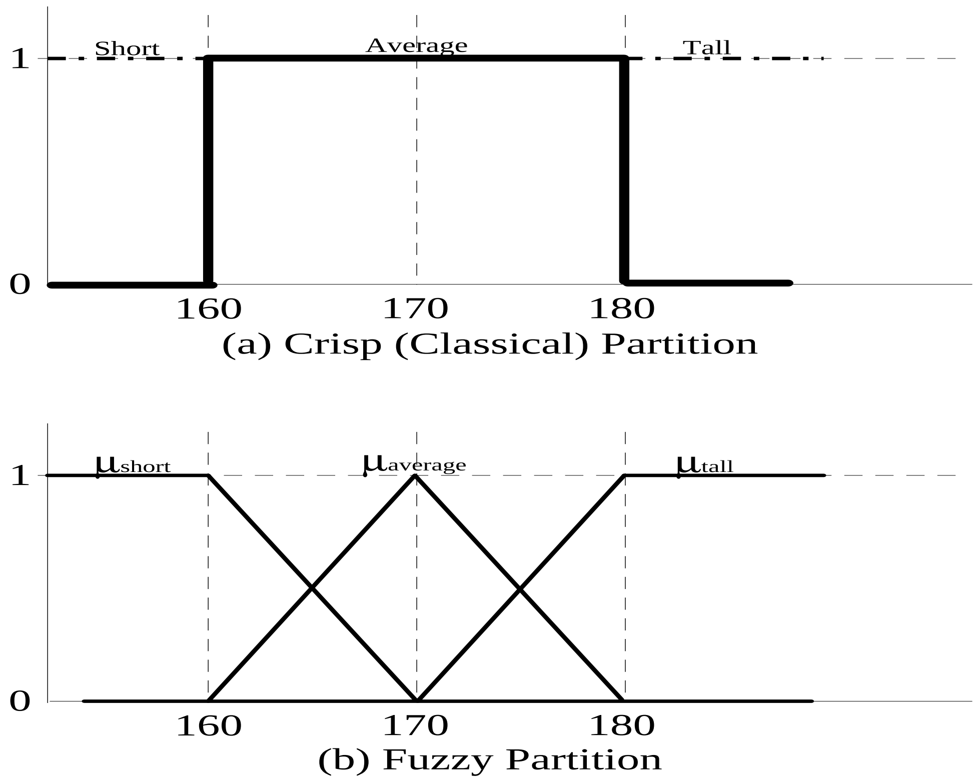

Figure 1: Partitions of Set of Height

2.3 Representation of Fuzzy Set

Definition 2 *Given a fuzzy set f and $0 < \alpha \leq 1$ the α-cut (to be read alpha cut) of f (a.k.a. level set), denoted by f_α is the crisp set defined as*

$$f_\alpha = \{x | \mu_f(x) \geq \alpha\} \tag{2}$$

Definition 3 *Given a fuzzy set f and $0 < \alpha < 1$ the strong α-cut of f (a.k.a. strong level set), denoted by $f_{\alpha+}$ is the classical (crisp) set defined as*

$$f_{\alpha+} = \{x | \mu_f(x) > \alpha\} \tag{3}$$

Example 2 *In the fuzzy sets defined on height shown as Figure 1, α-cuts of these fuzzy sets are defined respectively as follows:*

$$\begin{aligned} f_{0.5}^{tall} &= \{h | \mu_{tall}(h) \geq 0.5\} = [175, \infty) \\ f_{0.5}^{average} &= \{h | \mu_{average}(h) \geq 0.5\} = [165, 175] \\ f_{0.5}^{short} &= \{h | \mu_{short}(h) \geq 0.5\} = (-\infty, 165] \end{aligned}$$

The α-cuts provide the connection between fuzzy sets and classical sets. This is going to be shown as **representation theorem** shortly. In its most general formulation this theorem states the necessary and sufficient conditions for a family of classical sets to be level sets (α-cuts) of a fuzzy set. For the particular case of finite fuzzy sets, these conditions reduce to only one, namely that the classical sets must be nested. This is the base of **Mass Assignment Theory**.

2.4 Convexity

Fuzzy set f is *convex* if all α-cuts of f are convex $\forall\ \alpha \in (0, 1]$.
Shape of membership functions can be any (convex) functions. For the sake of simplifications, the following shapes are often used for computational models without losing too much consistencies

- Trapezoid: 4 parameters
- Triangular: 3 parameters (can be treated as a special case of trapezoid)

(Note that this representation issue is still being discussed among leading scholars.)

2.5 Decomposition (Representation) Theorem

For every fuzzy set $A \in \mathcal{F}(X)$,

$$A = \cup_{\alpha \in (0,1]} F_\alpha$$

where F_α is a fuzzy set defined by means of the membership function

$$\mu_{F_\alpha}(x) = \alpha \cdot \mu_{A_\alpha}(x)$$

and A_α is an α-cut (i.e. level-set) with the degree of α ($\mu_{A_\alpha}(x)$ is the indicator function).

2.6 Extension Principle

A crisp function

$$f : X \mapsto Y$$

is *fuzzified* if it is extended to act on fuzzy sets defined on X and Y such that

$$f : \mathcal{F}(X) \mapsto \mathcal{F}(Y)$$

The membership function is determined such that

$$\mu_f(y) = \vee_{f(x)=y}\mu(x)$$

Consequently, any functions are able to be adopted within fuzzy models.

2.7 Fuzzy Set Operations

The following standard set operations namely complement (not), intersection (and), and union (or) are introduced by Zadeh:

Standard Complement

$$\overline{A}(x) = 1 - A(x) \tag{4}$$

Standard Intersection

$$(A \cap B)(x) = min[A(x), B(x)] \tag{5}$$

Standard Union

$$(A \cup B)(x) = max[A(x), B(x)] \tag{6}$$

These operations perform precisely as the corresponding operations for the crisp sets.

There are classes of binary operations on the unit interval for complement, intersection (generally referred as t-norm) and union (generally referred as t-conorm) satisfying certain axioms such as boundary, monotonicity, commutativity, and associativity. For *complement* operator, the following properties must be satisfied:

Axiom: Boundary Condition

$$\begin{aligned} c(0) &= 1 \\ c(1) &= 0 \end{aligned} \tag{7}$$

Axiom: Monotonicity

$$\forall a, b \in [0, 1], a \leq b \Rightarrow c(a) \geq c(b) \tag{8}$$

Axiom: Continuity (*) c is a continuous function.

Axiom: Involution (**)

$$c(c(a)) = a \; \forall \; a \in [0,1]$$

Axioms with * and ** are required to support the involution property. The following classes of complement supports the involution.

- Sugeno class
$$c_\lambda(a) = \frac{1-a}{1+\lambda \cdot a} \; where \; \lambda \in (-1, \infty)$$

- Yager class
$$c_\omega(a) = (1-a^\omega)^{\frac{1}{\omega}} \; where \; \omega \in (0, \infty)$$

Both classes include the standard fuzzy complement.

Let i be an operator in t-norm (intersection) $\forall a, b, c \in [0,1]$. Then we have the following properties:

Axiom: Boundary Condition

$$i(a,1) = a \tag{9}$$

Axiom: Monotonicity

$$b \le d \Rightarrow i(a,b) \le i(a,d) \tag{10}$$

Axiom: Commutativity

$$i(a,b) = i(b,a) \tag{11}$$

Axiom: Associativity

$$i(a, i(b,d)) = i(i(a,b), d) \tag{12}$$

The following operators are examples of t-norm (intersection) other than the standard intersection operator:

Algebraic Product

$$(A \cap B)(x) = A(x) \cdot b(x) \tag{13}$$

Bounded Difference

$$(A \cap B)(x) = max[0, A(x) + B(x) - 1] \tag{14}$$

Drastic Intersection

$$(A \cap B)(x) = \begin{cases} A(x) & if \quad B(x) = 1 \\ B(x) & if \quad A(x) = 1 \\ 0 & \quad\; otherwise \end{cases} \tag{15}$$

Similarly for t-conorm (unions) u, the following properties must be satisfied: Then we have the following for t-norm:

Axiom: Boundary Condition

$$u(a, 0) = a \tag{16}$$

Axiom: Monotonicity

$$b \leq d \Rightarrow u(a, b) \leq u(a, d) \tag{17}$$

Axiom: Commutativity

$$u(a, b) = u(b, a) \tag{18}$$

Axiom: Associativity

$$u(a, u(b, d)) = u(u(a, b), d) \tag{19}$$

Examples for t-conorm operators are as follows:

Algebraic Sum

$$(A \cup B)(x) = A(x) + B(x) - A(x) \cdot B(x) \tag{20}$$

Bounded Sum

$$(A \cup B)(x) = min[1, A(x) + B(x)] \tag{21}$$

Drastic Union

$$(A \cap B)(x) = \begin{cases} A(x) & if \quad B(x) = 0 \\ B(x) & if \quad A(x) = 0 \\ \quad 1 & \quad otherwise \end{cases} \tag{22}$$

There are classes of t-norm and t-conorm. As of now, only the standard operations are consistent on the operations of α-cuts (level sets). In addition, it is important to maintain compatibilities of fuzzy set operations and operations on mass assignments described later.

2.8 Aggregation

Aggregation operations on fuzzy sets combine several fuzzy sets in order to generate a single fuzzy set.

$$\mathcal{H} : [0, 1]^n \mapsto [0, 1] \; where \; n \geq 2$$

We denote aggregation function $\mathcal{H}$

$$A(x) = \mathcal{H}(A_1(x), \ldots, A_n(x)) \; \forall \; x \in X$$

There are three required axioms and two additional essential axioms

- **Axiom: Boundary conditions**

$$\mathcal{H}(0,\ldots,0) = 0$$
$$\mathcal{H}(1,\ldots,1) = 1$$

- **Axiom: Monotonicity**

$$\mathcal{H}(a_1,\ldots,a_n) \leq \mathcal{H}(b_1,\ldots,b_n)\ if\ a_i \leq b_i\ \forall\ i \in [1,n]$$

- **Axiom: Continuity** $\mathcal{H}$ is a continuous function.
- **Axiom: Symmetric (*)**

$$\mathcal{H}(a_1,\ldots,a_n) = \mathcal{H}(a_{p(1)},\ldots,a_{p(n)})$$

  for any permutation p on $[1,n]$.

- **Axiom: Idempotency**

$$\mathcal{H}(a,\ldots,a) = a$$

Examples

- Generalized mean

$$\mathcal{H}(a_1,\ldots,a_n) = \left(\frac{\sum_{i=1}^{n} a_i^{\alpha}}{n}\right)^{\frac{1}{\alpha}}$$

- Harmonic mean

$$\mathcal{H}(a_1,\ldots,a_n) = \frac{n}{\sum_{i=1}^{n}\frac{1}{a_i}}$$

- Arithmetic mean

$$\mathcal{H}(a_1,\ldots,a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i$$

- Weighted average

$$\mathcal{H}(a_1,\ldots,a_n) = \sum_{i=1}^{n} w_i \cdot a_i$$

  where $\sum w_i = 1$

- Ordered weighted average (OWA)

$$\mathcal{H}(a_1,\ldots,a_n) = \sum w_i \cdot b_i$$

  where b_i is the i-th largest element $\in a_1,\ldots,a_n$

Aggregations can be viewed as a generalization of fuzzy set operations. This view offers interfaces in order for fuzzy sets to adopt or to be adopted by other computational frameworks such as neural networks and genetic algorithms. (Note that such consortium of computational methods is the principle of so-called **Soft Computing**.)

2.9 Comparison of Fuzzy Sets

Unlike comparing classical sets, there are multiple ways of determining the comparison for fuzzy sets because of their membership values. Ways of determining a similarity degree between two fuzzy sets, **F** and **C**, include:

- index of intersection/inclusion

$$d = \frac{|F \cap C|}{|F|} \tag{23}$$

- centroid

$$d = \mu_{F\cap C}(x_{\lfloor c \rfloor}) + (c - \lfloor c \rfloor) \cdot (\mu_{F\cap C}(x_{\lfloor c \rfloor + 1}) - \mu_{F\cap C}(x_{\lfloor c \rfloor}))$$

  where

$$c = \frac{\sum_{j=1}^{n} \mu_{F\cap C}(x_j) \cdot j}{\sum_{j=1}^{n} \mu_{F\cap C}(x_j)} \tag{24}$$

  assuming that membership degrees $\mu(x_j)$ are in non-decreasing order.

- truth value interval (support pairs)

$$d = (n, p) \tag{25}$$

  where

$$n = \bigwedge_x (1 - \mu_C(x)) \vee \mu_F(x) \tag{26}$$

  and

$$p = \bigvee_x \mu_F(x) \wedge \mu_C(x) \tag{27}$$

  In equations (26) and (27), $\bigwedge$ and $\bigvee$ denote minimum and maximum operations respectively. In fuzzy set theory, **n** and **p** are referred to as necessity and possibility respectively.

In examples of comparison methods shown above, the cardinality of fuzzy set becomes an issue. Note that several definitions of cardinality of fuzzy sets are available. They include:

- sigma count: fuzzy cardinality of a fuzzy set A with support $\{x_1, \ldots, x_n\}$ is given as a real number such that

$$Card_{\Sigma-count}(A) = \sum_{i=1}^{n} \mu_A(x_i) \tag{28}$$

- fuzzy number: fuzzy cardinality is given as a fuzzy number (introduced by A. L. Ralescu), i.e. a fuzzy set on natural numbers, such that

$$Card_R^X(i) = \begin{pmatrix} 0 & \ldots & i & \ldots & n \\ & & & & \\ \mu_X(x_0) & & \mu_X(x_i) & & \mu_X(x_n) \\ \wedge & \ldots & \wedge & \ldots & \wedge \\ (1-\mu_X(x_1)) & & (1-\mu_X(x_{i+1})) & & (1-\mu_X(x_{n+1})) \end{pmatrix} \tag{29}$$

where

$$1 = \mu_X(x_0) \geq \mu_X(x_1) \geq \ldots \geq \mu_X(x_n) > \mu_X(x_{n+1}) = 0$$

The similarity degree **d** is given by a fuzzy number when the second definition $\mathbf{Card_R}$ is chosen. In such a case, a defuzzification process is necessary to come up to single value in $[0,1]$, which represents the fuzzy set.

2.10 Fuzzy Partitions

Fuzzy partitions are introduced as the counterpart of the crisp partitions. They provide a kind of summary and sufficient description of the universal set X. For comparison, we have the following for classical (crisp) and fuzzy partitions of a universe of discourse X:

Definition 4 *Let* $\mathbf{X}$ *be a universal set,*

Classical (crisp) Partition of X *A collection* $(A_i)_{i=1}^n$, $A_i \subseteq X$ *forms a partition of* X *if*

$$A_i \cap A_j = \emptyset \; \forall \; i \neq j \tag{30}$$

and

$$\bigcup_{i=1}^{n} A_i = X \tag{31}$$

Fuzzy Partition of X *A collection of fuzzy sets* $(\mu_i)_{i=1}^n$ *forms a fuzzy partition of* X *if*

$$\forall i \neq j \;\forall x \in X \;\; \mu_i(x) \leq 1 - \mu_j(x) \tag{32}$$

and

$$\forall x \in X \;\; MAX_{i=1,\ldots,n}\mu_i(x) > 0 \tag{33}$$

With the definition of set operations on fuzzy sets obtained as a straight forward extension of those on classic sets, conditions of fuzzy partition of X are nothing but fuzzy set version of conditions of classical (crisp) partition of X. (30, 31).

Using fuzzy sets, transition from membership to non-membership to a class defined as a fuzzy set is gradual (on the other hand, it is abrupt in classical sets). Comparing fuzzy sets to classical (crisp) sets, set elements belong to more than one fuzzy set. For instance, the height of 173.33 can be treated as both average and tall with degree of 0.67. Such a feature represents these concepts more naturally and smoothly comparing to classical sets defined as interval values such as *tall* $= [180, 200]$, *average* $= [160, 180]$, and *short* $= [140, 160]$.

Fuzzy partitions are utilized in many successful applications including (but are not limited to) fuzzy controls and data analysis methods such as fuzzy clustering.

2.11 Fuzzy Sets and Probability

Probability, including Bayesian probability, and fuzzy sets are known to be fundamentally different. From the probability model (axioms) it follows that the probability of an event A completely determine the probability of its complement $\overline{A}$ and vice versa. For example, when the probability of 1 by rolling a dice is determined, the complement i.e. the probability of either 2, 3, 4, 5, or 6, is determined at the same time. In contrast, fuzzy sets are not complementary, i.e. a fuzzy set and its complement do not construct the universe. In Example 1, fuzzy sets *tall* and *short* are complements of each other, i.e. $1 - \mu_{tall} = \mu_{short}$. However, their union, $\mu_{tall} \vee \mu_{short}$, does not cover the entire universe, i.e. membership values of their union are not always 1.0, as it is seen in Figure 1.

Second, fuzzy sets do not have to be uniquely determined like probability. Fuzzy sets for height shown above should be defined based on user's perception because a particular height, say 170, is not considered to be tall for a user but it might be for another. This property enables us to represent user's perception on computers.

According to J. F. Baldwin,

> 'Fuzzy set theory expects one to be able to give a degree of applicability and therefore have some measure of closeness of match. This assumption accepts that there can be degrees of possibility and not simply possible and not possible.'

A voting model model introduced by J. F. Baldwin meets such criteria and gives an interpretation of fuzzy sets. It says that the membership degree of an element in a fuzzy set is determined based on how many people of a representative population accept it. For example, we have 10 representatives and 4 items, namely a,b,c, and d. Let's say that 10 people vote for item a, 7 people vote for item b, 4 people vote for item c, and 2 people vote for item d. Then the membership values of these items are 1.0, 0.7, 0.4, and 0.2 respectively under the constant threshold assumption such that a person who votes for an item is assumed to vote all other items whose membership values are higher than that item.

Finally, there is a relation between fuzzy sets and probabilities. In the example of height, there is a certain threshold such that everybody agrees that it is tall although the fuzzy set of *tall* is different for each user. This means that the order of significance is preserved in both fuzzy sets and probabilities.

3 Computing with Words

Fuzzy sets offer maps between linguistic labels and actual data known as *linguistic values*. For example, height of 173.33 is interpreted as 'tall' with a degree of 0.67. Using linguistic values relieves computation of handling numeric value intervals and generates descriptive results in a natural language.

Moreover, a variable containing a linguistic value is called *linguistic variable*. For example, a linguistic variable 'HEIGHT' contains some linguistic values such as 'tall', 'average', and 'short'. Considering the range of heights shown in Figure 1, i.e. $[140, 200]$, it is now represented as a simple discrete set of linguistic values $\{tall, average, short\}$ instead of the continuous value interval.

The notion of *Computing With Words* is introduced by Zadeh by treating linguistic variables as constraints between detailed numerical data and macro linguistic description. Fuzzy sets play key roles in this computational scheme as mediator between numerical data and linguistic representation.

From a different aspect, fuzzy sets are viewed as a bag of data covered by these sets, and this determines the granulation of data.

The map between numerical values and linguistic labels is a representation of user's concepts biased based on their perception. In this framework it is possible to represent a concept with the same linguistic description but the associated numerical data is different based on user's perception.