FS V: Fuzzy rule-based systems

Triangular norms

Triangular norms were introduced by Schweizer and Sklar to model the distances in probabilistic metric spaces. [Associative functions and abstract semigroups, *Publ. Math. Debrecen*, 10(1963) 69-81].

In fuzzy sets theory triangular norm are extensively used to model the logical connective **and**.

Definition 1. *(Triangular norm.) A mapping*

$$T\colon [0,1] \times [0,1] \to [0,1]$$

is a triangular norm (t-norm for short) iff it is symmetric, associative, non-decreasing in each argument and $T(a,1) = a$, for all $a \in [0,1]$. In other words, any t-norm T satisfies the properties:

Symmetricity:

$$T(x,y) = T(y,x),\ \forall x,y \in [0,1].$$

Associativity:

$$T(x,T(y,z)) = T(T(x,y),z),\ \forall x,y,z \in [0,1].$$

Monotonicity:

$$T(x,y) \le T(x',y') \text{ if } x \le x' \text{ and } y \le y'.$$

One identy:

$$T(x,1) = x,\ \forall x \in [0,1].$$

These axioms attempt to capture the basic properties of set intersection. The basic t-norms are:

- minimum: $\min(a, b) = \min\{a, b\}$,
- Łukasiewicz: $T_L(a, b) = \max\{a + b - 1, 0\}$
- product: $T_P(a, b) = ab$
- weak:

$$T_W(a, b) = \begin{cases} \min\{a, b\} & \text{if } \max\{a, b\} = 1 \\ 0 & \text{otherwise} \end{cases}$$

- Hamacher:

$$H_\gamma(a, b) = \frac{ab}{\gamma + (1 - \gamma)(a + b - ab)}, \ \gamma \geq 0$$

- Dubois and Prade:

$$D_\alpha(a, b) = \frac{ab}{\max\{a, b, \alpha\}}, \ \alpha \in (0, 1)$$

- Yager:

$$Y_p(a, b) = 1 - \min\{1, \sqrt[p]{[(1 - a)^p + (1 - b)^p]}\}, \ p > 0$$

- Frank:

$$F_\lambda(a, b) = \begin{cases} \min\{a, b\} & \text{if } \lambda = 0 \\ T_P(a, b) & \text{if } \lambda = 1 \\ T_L(a, b) & \text{if } \lambda = \infty \\ 1 - \log_\lambda \left[1 + \dfrac{(\lambda^a - 1)(\lambda^b - 1)}{\lambda - 1}\right] & \text{otherwise} \end{cases}$$

All t-norms may be extended, through associativity, to $n > 2$ arguments. The minimum t-norm is automatically extended and

$$T_P(a_1, \dots, a_n) = a_1 \times a_2 \times \dots \times a_n,$$

$$T_L(a_1, \dots a_n) = \max\{\sum_{i=1}^{n} a_i - n + 1, 0\}$$

A t-norm T is called strict if T is strictly increasing in each argument.

Triangular co-norms are extensively used to model logical connectives **or**.

Definition 2. *(Triangular conorm.) A mapping*

$$S\colon [0,1] \times [0,1] \to [0,1],$$

is a triangular co-norm (t-conorm) if it is symmetric, associative, non-decreasing in each argument and $S(a, 0) = a$, for all $a \in [0, 1]$. In other words, any t-conorm S satisfies the properties:

$S(x, y) = S(y, x)$ *(symmetricity)*

$S(x, S(y, z)) = S(S(x, y), z)$ *(associativity)*

$S(x, y) \leq S(x', y')$ *if* $x \leq x'$ *and* $y \leq y'$ *(monotonicity)*

$S(x, 0) = x$, $\forall x \in [0, 1]$ *(zero identy)*

If T is a t-norm then the equality

$$S(a, b) := 1 - T(1 - a, 1 - b),$$

defines a t-conorm and we say that S is derived from T. The basic t-conorms are:

- maximum: $\max(a,b) = \max\{a,b\}$
- Łukasiewicz: $S_L(a,b) = \min\{a+b, 1\}$
- probabilistic: $S_P(a,b) = a + b - ab$
- strong:

$$STRONG(a,b) = \begin{cases} \max\{a,b\} & \text{if } \min\{a,b\} = 0 \\ 1 & \text{otherwise} \end{cases}$$

- Hamacher:

$$HOR_\gamma(a,b) = \frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab},\ \gamma \geq 0$$

- Yager:

$$YOR_p(a,b) = \min\{1, \sqrt[p]{a^p + b^p}\},\ p > 0.$$

Lemma 1. *Let T be a t-norm. Then the following statement holds*

$$T_W(x,y) \leq T(x,y) \leq \min\{x,y\},\ \forall x,y \in [0,1].$$

Proof. From monotonicity, symmetricity and the extremal condition we get

$$T(x,y) \leq T(x,1) \leq x,\ T(x,y) = T(y,x) \leq T(y,1) \leq y.$$

This means that $T(x,y) \leq \min\{x,y\}$. □

Lemma 2. *Let S be a t-conorm. Then the following statement holds*

$$\max\{a,b\} \leq S(a,b) \leq STRONG(a,b),\ \forall a,b \in [0,1]$$

Proof. From monotonicity, symmetricity and the extremal condition we get

$$S(x,y) \geq S(x,0) \geq x,\ S(x,y) = S(y,x) \geq S(y,0) \geq y$$

This means that $S(x,y) \geq \max\{x,y\}$. $\square$

Lemma 3. *$T(a,a) = a$ holds for any $a \in [0,1]$ if and only if T is the minimum norm.*

Proof. If $T(a,b) = \min(a,b)$ then $T(a,a) = a$ holds obviously. Suppose $T(a,a) = a$ for any $a \in [0,1]$, and $a \leq b \leq 1$. We can obtain the following expression using monotonicity of T

$$a = T(a,a) \leq T(a,b) \leq \min\{a,b\}.$$

From commutativity of T it follows that

$$a = T(a,a) \leq T(b,a) \leq \min\{b,a\}.$$

These equations show that $T(a,b) = \min\{a,b\}$ for any $a, b \in [0,1]$. $\square$

Lemma 4. *The distributive law of t-norm T on the $\max$ operator holds for any a, b, $c \in [0,1]$.*

$$T(\max\{a,b\},c) = \max\{T(a,c),T(b,c)\}.$$

Definition 3. *(t-norm-based intersection) Let T be a t-norm. The T-intersection of A and B is defined as*

$$(A \cap B)(t) = T(A(t),B(t)),$$

for all $t \in X$.

Example 1. *Let*

$$T(x, y) = LAND(x, y) = \max\{x + y - 1, 0\}$$

be the Łukasiewicz t-norm. Then we have

$$(A \cap B)(t) = \max\{A(t) + B(t) - 1, 0\},$$

for all $t \in X$.

Let A *and* B *be fuzzy subsets of*

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$$

and be defined by

$$A = 0.0/x_1 + 0.3/x_2 + 0.6/x_3 + 1.0/x_4 + 0.6/x_6 + 0.3/x_6 + 0.0/x_7$$

$$B = 0.1/x_1 + 0.3/x_2 + 0.9/x_3 + 1.0/x_4 + 1.0/x_5 + 0.3/x_6 + 0.2/x_7.$$

Then $A \cap B$ *has the following form*

$$A \cap B = 0.0/x_1 + 0.0/x_2 + 0.5/x_3 + 1.0/x_4 + 0.6/x_5 + 0.0/x_6 + 0.2/x_7.$$

The operation *union* can be defined by the help of triangular conorms.

Definition 4. *(t-conorm-based union) Let* S *be a t-conorm. The* S*-union of* A *and* B *is defined as*

$$(A \cup B)(t) = S(A(t), B(t)),$$

for all $t \in X$.

Example 2. *Let*

$$S(x, y) = LOR(x, y) = \min\{x + y, 1\}$$

be the Łukasiewicz t-conorm. Then we have

$$(A \cup B)(t) = \min\{A(t) + B(t), 1\},$$

for all $t \in X$.

Let A *and* B *be fuzzy subsets of*

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$$

and be defined by

$$A = 0.0/x_1 + 0.3/x_2 + 0.6/x_3 + 1.0/x_4 + 0.6/x_5 + 0.3/x_6 + 0.0/x_7$$

$$B = 0.1/x_1 + 0.3/x_2 + 0.9/x_3 + 1.0/x_4 + 1.0/x_5 + 0.3/x_6 + 0.2/x_7$$

Then $A \cup B$ *has the following form*

$$A \cup B = 0.1/x_1 + 0.6/x_2 + 1.0/x_3 + 1.0/x_4 + 1.0/x_5 + 0.6/x_6 + 0.2/x_7.$$

If we are given an operator C such that

$$\min\{a, b\} \leq C(a, b) \leq \max\{a, b\}, \ \forall a, b \in [0, 1]$$

then we say that C is a compensatory operator.

A typical compensatory operator is the arithmetical mean defined as

$$MEAN(a, b) = \frac{a + b}{2}$$

Averaging operators

Fuzzy set theory provides a host of attractive aggregation connectives for integrating membership values representing uncertain information. These connectives can be categorized into the following three classes *union, intersection* and *compensation* connectives.

Union produces a high output whenever any one of the input values representing degrees of satisfaction of different features or criteria is high.

Intersection connectives produce a high output only when all of the inputs have high values. Compensative connectives have the property that a higher degree of satisfaction of one of the criteria can compensate for a lower degree of satisfaction of another criteria to a certain extent.

In the sense, union connectives provide full compensation and intersection connectives provide no compensation. In a decision process the idea of *trade-offs* corresponds to viewing the global evaluation of an action as lying between the *worst* and the *best* local ratings. This occurs in the presence of conflicting goals, when a compensation between the corresponding compabilities is allowed. Averaging operators realize trade-offs between objectives, by allowing a positive compensation between ratings.

Definition 5. *An averaging operator M is a function*

$$M\colon [0,1]\times[0,1]\to[0,1],$$

, satisfying the following properties

- *Idempotency*
$$M(x,x)=x,\ \forall x\in[0,1],$$
- *Commutativity*
$$M(x,y)=M(y,x),\ \forall x,y\in[0,1],$$
- *Extremal conditions*
$$M(0,0)=0,\quad M(1,1)=1$$

- *Monotonicity*

$$M(x,y) \leq M(x',y') \text{ if } x \leq x' \text{ and } y \leq y',$$

- M *is continuous.*

Averaging operators represent a wide class of aggregation operators. We prove that whatever is the particular definition of an averaging operator, M, the global evaluation of an action will lie between the *worst* and the *best* local ratings:

Lemma 5. *If M is an averaging operator then*

$$\min\{x,y\} \leq M(x,y) \leq \max\{x,y\},\ \forall x,y \in [0,1]$$

Proof. From idempotency and monotonicity of M it follows that

$$\min\{x,y\} = M(\min\{x,y\}, \min\{x,y\}) \leq M(x,y)$$

and $M(x,y) \leq M(\max\{x,y\}, \max\{x,y\}) = \max\{x,y\}$. Which ends the proof. □

Averaging operators have the following interesting properties:

Property 1. *A strictly increasing averaging operator cannot be associative.*

Property 2. *The only associative averaging operators are defined by*

$$M(x,y,\alpha) = \textit{med}(x,y,\alpha) = \begin{cases} y & \text{if } x \leq y \leq \alpha \\ \alpha & \text{if } x \leq \alpha \leq y \\ x & \text{if } \alpha \leq x \leq y \end{cases}$$

where $\alpha \in (0,1)$.

An important family of averaging operators is formed by quasi-arithmetic means

$$M(a_1, \ldots, a_n) = f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(a_i)\right)$$

This family has been characterized by Kolmogorov as being the class of all decomposable continuous averaging operators. For example, the quasi-arithmetic mean of a_1 and a_2 is defined by

$$M(a_1, a_2) = f^{-1}\left(\frac{f(a_1) + f(a_2)}{2}\right).$$

The next table shows the most often used mean operators.

Name	$M(x, y)$
harmonic mean	$2xy/(x + y)$
geometric mean	$\sqrt{xy}$
arithmetic mean	$(x + y)/2$
dual of geometric mean	$1 - \sqrt{(1 - x)(1 - y)}$
dual of harmonic mean	$(x + y - 2xy)/(2 - x - y)$
median	med$(x, y, \alpha),\ \alpha \in (0, 1)$
generalized p-mean	$((x^p + y^p)/2)^{1/p},\ p \geq 1$

Mean operators.

The process of information aggregation appears in many applications related to the development of intelligent systems. One sees aggregation in neural networks, fuzzy logic controllers, vision systems, expert systems and multi-criteria decision aids. In 1988 Yager introduced a new aggregation technique based on the ordered weighted averaging (OWA) operators.

Definition 6. *An OWA operator of dimension* n *is a mapping* $F\colon \mathbb{R}^n \to \mathbb{R}$, *that has an associated weighting vector* $W = (w_1, w_2, \dots, w_n)^T$ *such as* $w_i \in [0,1],\ 1 \leq i \leq n$, *and*

$$w_1 + \cdots + w_n = 1.$$

Furthermore

$$F(a_1, \dots, a_n) = w_1 b_1 + \cdots + w_n b_n = \sum_{j=1}^{n} w_j b_j$$

where b_j *is the* j*-th largest element of the bag* $\langle a_1, \dots, a_n \rangle$.

Example 3. *Assume* $W = (0.4, 0.3, 0.2, 0.1)^T$ *then*

$$F(0.7, 1, 0.2, 0.6) = 0.4 \times 1 + 0.3 \times 0.7 + 0.2 \times 0.6 + 0.1 \times 0.2 = 0.75.$$

A fundamental aspect of this operator is the re-ordering step, in particular an aggregate a_i is not associated with a particular weight w_i but rather a weight is associated with a particular ordered position of aggregate. When we view the OWA weights as a column vector we shall find it convenient to refer to the weights with the low indices as weights at the top and those with the higher indices with weights at the bottom.

It is noted that different OWA operators are distinguished by their weighting function. In 1988 Yager pointed out three important special cases of OWA aggregations:

- F^*: In this case $W = W^* = (1, 0 \dots, 0)^T$ and

$$F^*(a_1, \dots, a_n) = \max\{a_1, \dots, a_n\},$$

- F_*: In this case $W = W_* = (0, 0 \ldots, 1)^T$ and

$$F_*(a_1, \ldots, a_n) = \min\{a_1, \ldots, a_n\},$$

- F_A: In this case $W = W_A = (1/n, \ldots, 1/n)^T$ and

$$F_A(a_1, \ldots, a_n) = \frac{a_1 + \cdots + a_n}{n}.$$

A number of important properties can be associated with the OWA operators. We shall now discuss some of these. For any OWA operator F holds

$$F_*(a_1, \ldots, a_n) \leq F(a_1, \ldots, a_n) \leq F^*(a_1, \ldots, a_n).$$

Thus the upper an lower star OWA operator are its boundaries. From the above it becomes clear that for any F

$$\min\{a_1, \ldots, a_n\} \leq F(a_1, \ldots, a_n) \leq \max\{a_1, \ldots, a_n\}.$$

The OWA operator can be seen to be *commutative*. Let $\langle a_1, \ldots, a_n \rangle$ be a bag of aggregates and let $\{d_1, \ldots, d_n\}$ be any *permutation* of the a_i. Then for any OWA operator

$$F(a_1, \ldots, a_n) = F(d_1, \ldots, d_n).$$

A third characteristic associated with these operators is *monotonicity*. Assume a_i and c_i are a collection of aggregates, $i = 1, \ldots, n$ such that for each i, $a_i \geq c_i$. Then

$$F(a_1, \ldots, a_n) \geq F(c_1, c_2, \ldots, c_n)$$

where F is some fixed weight OWA operator.

Another characteristic associated with these operators is *idempotency*. If $a_i = a$ for all i then for any OWA operator

$$F(a_1, \ldots, a_n) = a.$$

From the above we can see the OWA operators have the basic properties associated with an *averaging operator*.

Example 4. *A window type OWA operator takes the average of the* m *arguments around the center. For this class of operators we have*

$$w_i = \begin{cases} 0 & \textit{if}\ i < k \\ \dfrac{1}{m} & \textit{if}\ k \leq i < k+m \\ 0 & \textit{if}\ i \geq k+m \end{cases} \tag{1}$$

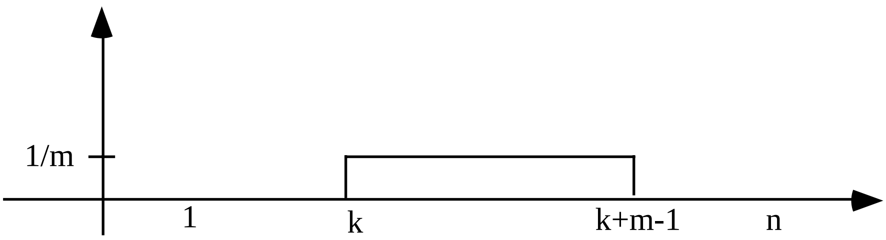


Figure 1: Window type OWA operator.

In order to classify OWA operators in regard to their location between *and* and *or*, a measure of *orness*, associated with any vector W is introduce by Yager as follows

$$\text{orness}(W) = \frac{1}{n-1} \sum_{i=1}^{n} (n-i) w_i.$$

It is easy to see that for any W the $\text{orness}(W)$ is always in the unit interval. Furthermore, note that the nearer W is to an *or*, the closer its measure is to one; while the nearer it is to an *and*, the closer is to zero.

Lemma 6. *Let us consider the the vectors*

$$W^* = (1, 0 \ldots, 0)^T, \ W_* = (0, 0 \ldots, 1)^T,$$

$$W_A = (1/n, \ldots, 1/n)^T.$$

Then it can easily be shown that

$$\text{orness}(W^*) = 1, \ \text{orness}(W_*) = 0$$

and $\text{orness}(W_A) = 0.5$.

A measure of *andness* is defined as

$$\text{andness}(W) = 1 - \text{orness}(W).$$

Generally, an OWA opeartor with much of nonzero weights near the top will be an *orlike* operator, that is,

$$\text{orness}(W) \geq 0.5$$

and when much of the weights are nonzero near the bottom, the OWA operator will be *andlike*, that is,

$$\text{andness}(W) \geq 0.5.$$

Example 5. *Let* $W = (0.8, 0.2, 0.0)^T$. *Then*

$$\text{orness}(W) = \frac{1}{3}(2 \times 0.8 + 0.2) = 0.6,$$

and

$$\text{andness}(W) = 1 - \text{orness}(W) = 1 - 0.6 = 0.4.$$

This means that the OWA operator, defined by

$$F(a_1, a_2, a_3) = 0.8b_1 + 0.2b_2 + 0.0b_3 = 0.8b_1 + 0.2b_2,$$

where b_j *is the* j*-th largest element of the bag* $\langle a_1, a_2, a_3 \rangle$, *is an* orlike *aggregation.*

The following theorem shows that as we move weight up the vector we increase the orness, while moving weight down causes us to decrease $orness(W)$.

Theorem 1. *(Yager, 1993) Assume* W *and* W' *are two* n*-dimensional OWA vectors such that*

$$W = (w_1, \dots, w_n)^T,$$

and

$$W' = (w_1, \dots, w_j + \epsilon, \dots, w_k - \epsilon, \dots, w_n)^T$$

where $\epsilon > 0$*,* $j < k$*. Then* $orness(W') > orness(W)$*.*

Proof. From the definition of the measure of *orness* we get

$$orness(W') = \frac{1}{n-1} \sum_i (n-i) w'_i =$$

$$\frac{1}{n-1} \sum_i (n-1) w_i + (n-j)\epsilon - (n-k)\epsilon,$$

$$orness(W') = orness(W) + \frac{1}{n-1} \epsilon (k-j).$$

Since $k > j$, $orness(W') > orness(W)$. □

In 1988 Yager defined the measure of dispersion (or entropy) of an OWA vector by

$$disp(W) = -\sum_i w_i \ln w_i.$$

We can see when using the OWA operator as an averaging operator $Disp(W)$ measures the degree to which we use all the aggregates equally.

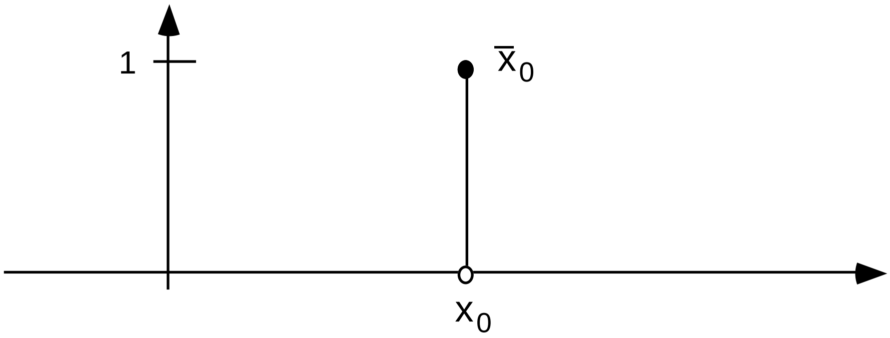


Figure 2: Fuzzy singleton.

Suppose now that the fact of the GMP is given by a fuzzy singleton. Then the process of computation of the membership function of the consequence becomes very simple.

For example, if we use Mamdani's implication operator in the GMP then

rule 1:	if x is A_1 then	z is C_1
fact:	x is $\bar{x}_0$	
consequence:		z is C

where the membership function of the consequence C is computed as

$$C(w) = \sup_u \min\{\bar{x}_0(u), (A_1 \to C_1)(u, w)\} =$$

$$\sup_u \min\{\bar{x}_0(u), \min\{A_1(u), C_1(w)\}\},$$

for all w. Observing that $\bar{x}_0(u) = 0$, $\forall u \neq x_0$, the supremum turns into a simple minimum

$$C(w) = \min\{\bar{x}_0(x_0) \wedge A_1(x_0) \wedge C_1(w)\} =$$

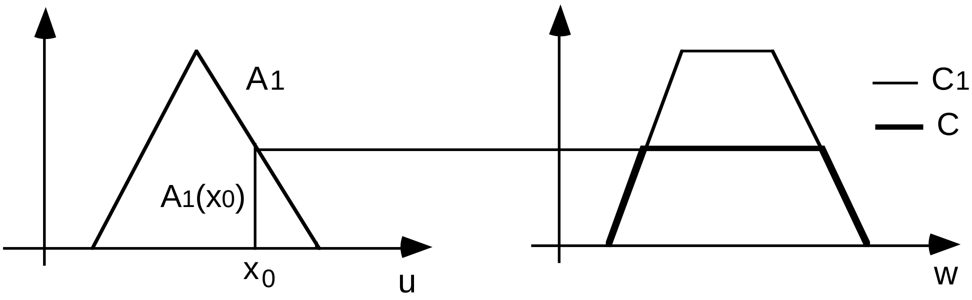


Figure 3: Inference with Mamdani's implication operator.

$$\min\{1 \wedge A_1(x_0) \wedge C_1(w)\} = \min\{A_1(x_0), C_1(w)\}$$

for all w.

and if we use Gödel implication operator in the GMP then

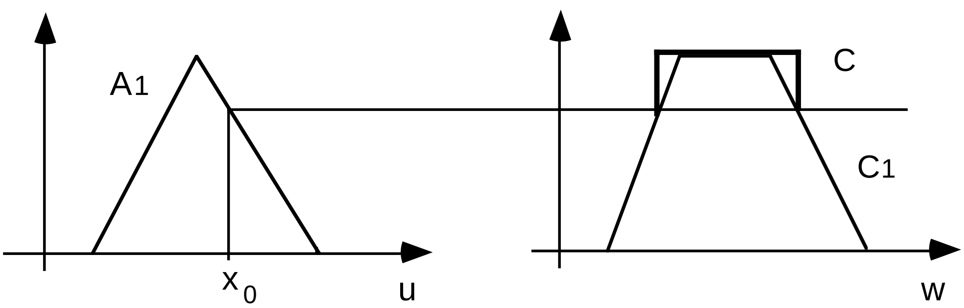
$$C(w) = \sup_u \min\{\bar{x}_0(u), (A_1 \to C_1)(u,w)\} =$$

$$A_1(x_0) \to C_1(w)$$

for all w.

So,

$$C(w) = \begin{cases} 1 & \text{if } A_1(x_0) \le C_1(w) \\ C_1(w) & \text{otherwise} \end{cases}$$



Inference with Gödel implication operator.

rule 1:	if x is A_1 then	z is C_1
fact:	x is $\bar{x}_0$	
consequence:		z is C

where the membership function of the consequence C is computed as

$$C(w) = \sup_u \min\{\bar{x}_0(u), (A_1 \to C_1)(u, w)\} =$$

$$A_1(x_0) \to C_1(w)$$

for all w.

Consider a block of fuzzy IF-THEN rules

$\Re_1$:	if x is A_1 then	z is C_1
also		
$\Re_2$:	if x is A_2 then	z is C_2
also		
	$\dots\dots\dots\dots$	
also		
$\Re_n$:	if x is A_n then	z is C_n
fact:	x is $\bar{x}_0$	
consequence:		z is C

The i-th fuzzy rule from this rule-base

$$\Re_i : \text{if } x \text{ is } A_i \text{ then } z \text{ is } C_i$$

is implemented by a *fuzzy implication* R_i and is defined as

$$R_i(u, w) = (A_i \to C_i)(u, w) = A_i(u) \to C_i(w)$$

for $i = 1, \dots, n$.

Find C from the input x_0 and from the rule base

$$\Re = \{\Re_1, \ldots, \Re_n\}.$$

Interpretation of

- sentence connective ”also”
- implication operator ”then”
- compositional operator ”$\circ$”

We first compose $\bar{x}_0$ with each R_i producing intermediate result

$$C'_i = \bar{x}_0 \circ R_i$$

for $i = 1, \ldots, n$.

C'_i is called the output of the i-th rule

$$C'_i(w) = A_i(x_0) \to C_i(w),$$

for each w.

Then combine the C'_i component wise into C' by some aggregation operator:

$$C = \bigcup_{i=1}^{n} C'_i = \bar{x}_0 \circ R_1 \cup \cdots \cup \bar{x}_0 \circ R_n$$

$$C(w) = A_1(x_0) \to C_1(w) \vee \cdots \vee$$

$$A_n(x_0) \to C_n(w).$$

So, the inference process is the following

- input to the system is x_0
- fuzzified input is $\bar{x}_0$
- firing strength of the i-th rule is $A_i(x_0)$
- the i-th individual rule output is

$$C'_i(w) := A_i(x_0) \to C_i(w)$$

- overall system output (action) is

$$C = C'_1 \cup \cdots \cup C'_n.$$

overall system output = union of the individual rule outputs

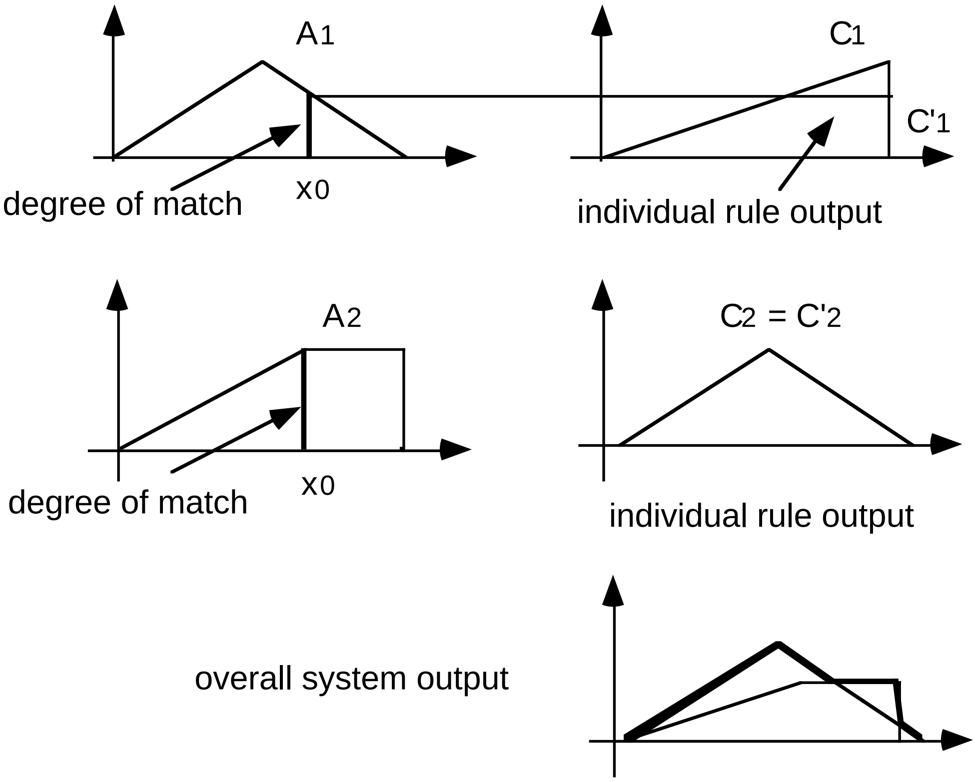
Mamdani $(a \to b = a \wedge b)$

- input to the system is x_0
- fuzzified input is $\bar{x}_0$
- firing strength of the i-th rule is $A_i(x_0)$
- the i-th individual rule output is

$$C'_i(w) = A_i(x_0) \wedge C_i(w)$$

- overall system output (action) is

$$C(w) = \bigvee_{i=1}^{n} A_i(x_0) \wedge C_i(w)$$



Larsen $(a \to b = ab)$

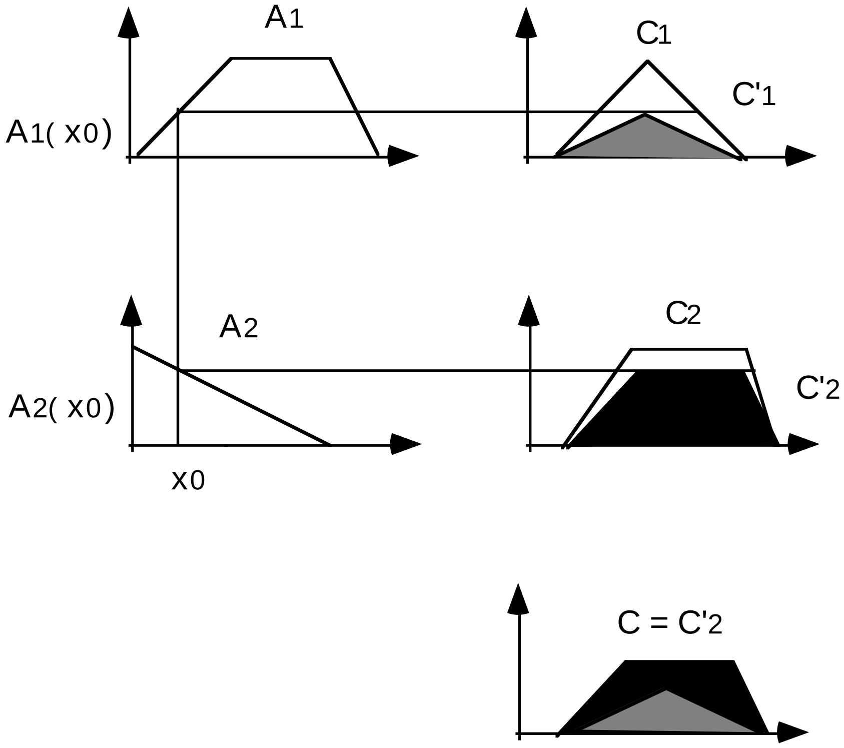
- input to the system is x_0
- fuzzified input is $\bar{x}_0$
- firing strength of the i-th rule is $A_i(x_0)$

- the i-th individual rule output is

$$C'_i(w) = A_i(x_0)C_i(w)$$

- overall system output (action) is

$$C(w) = \bigvee_{i=1}^{n} A_i(x_0)C_i(w)$$



The output of the inference process so far is a fuzzy set, specifying a possibility distribution of the (control) action. In the on-line control, a nonfuzzy (crisp) control action is usually required. Consequently, one must defuzzify the fuzzy control action (output) inferred from the fuzzy reasoning algorithm, namely:

$$z_0 = defuzzifier(C),$$

where z_0 is the crisp action and *defuzzifier* is the defuzzification operator.

Definition 7. *(defuzzification) Defuzzification is a process to select a representative element from the fuzzy output C inferred from the fuzzy control algorithm.*