T-norms and t-conorms in multilayer perceptrons

C.J. Mantas
Dept. of Computer Science and Artificial Intelligence
University of Granada
cmantas@decsai.ugr.es

Abstract

The definition of t-norms and t-conorms of the class of Hamacher with multilayer feedforward artificial neural networks is achieved in this work. This fact lets to insert fuzzy knowledge into neural network before its training.

Keywords: Neural Networks, Fuzzy Logic, Linguistic Hedges.

1 Introduction

Artificial Neural Networks (ANN's) [5, 9, 11] are computational models that have exhibited excellent behavior in classification and approximation problems. They are interesting due to the training starting from input-output examples and its generalization capacity.

However, ANN's have the shortcoming of being "black boxes". This implies two restrictions when they are used for solving a problem:

- It is not possible to insert knowledge into ANN's.
- It is not possible to understand how a trained ANN solves a problem.

In this paper, we are going to insert fuzzy knowledge into multilayer feedforward ANN's. In order to achieve this aim, we need:

- To build membership function with ANN's.
- To define a t-norm (fuzzy intersection) and a t-conorm (fuzzy union) [7] with neurons.
- To use linguistic hedges in our knowledge.

These tasks will be solved in this paper. Before it, an introduction to multilayer feedforward artificial neural network is presented.

2 Artificial Neural Networks

Multilayer feedforward ANN's are the most common model of neural nets. Let us consider an ANN with n input neurons (x_1, ..., x_n), h hidden neurons (z_1, ..., z_h), and m output neurons (y_1, ..., y_m). The function the net calculates is:

$$F : \Re^n \rightarrow \Re^m; \quad F(x_1, ..., x_n) = (y_1, ..., y_m)$$

$$y_k = g_A\left(\sum_{j=1}^{k}(z_j \beta_{jk}) + \varphi_k\right), \quad z_j = f_A\left(\sum_{i=1}^{n}(x_i w_{ij}) + \tau_j\right)$$

where g_A and f_A are activation functions:

- g_A is usually implemented as $g_A(x) = x$.
- f_A is usually a non-linear, antisymmetric, monotonic, and limited function [5]. Often, it is a *generalized sigmoid* [10]:

$$f_A(x) = \frac{F_{\min} \cdot e^{-x/\rho} + F_{\max} \cdot e^{+x/\rho}}{e^{-x/\rho} + e^{+x/\rho}}$$

which coincides with either the *sigmoid*, for $F_{min}=0$, $F_{max}=1$, $\rho=2$, or with the *hyperbolic tangent*, for $F_{min}=-1$, $F_{max}=1$, $\rho=1$, or with the *logic threshold*, for $\rho \rightarrow 0$. The parameter ρ is inversely proportional to the *steepness* of the activation function in the origin.

3 Construction of membership functions from neurons

In this section, it is presented how to construct membership functions for representing three kinds of linguistic propositions: a) *x is about greater than*

K, b) *x is about lower than K* and c) *x is approximately K.*

The kind a) and b) are directly represented using a neuron that uses the *sigmoid* activation function:

a) $\mu_{\text{is about greater than K}}$ (x) = *sigm*((x-K)·w^+ + 8) (Fig. 1), and

b) $\mu_{\text{is about lower than K}}$ (x) = *sigm*((-x+K)·w^+ + 8) (Fig. 2),

where the parameter w^+>0.0 determines the magnitude of the slope of the membership function (the higher w^+ is, the higher magnitude is). The magnitude of slope of the membership function is practically (w^+/16).

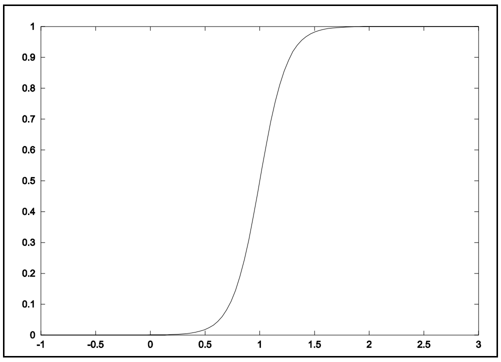

Figure 1: "x is about greater than K" with K=2 and w^+=8.

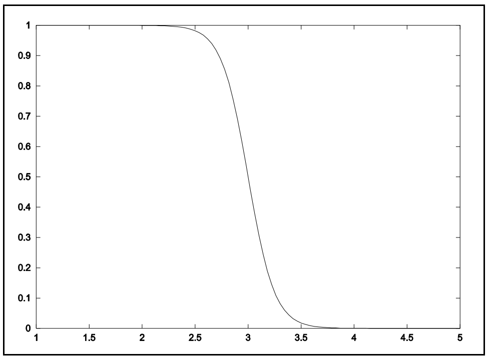

Figure 2: "x is about lower than K" with K=2 and w^+=8.

The basic idea for constructing membership functions that represent the linguistic proposition "*x is approximately K*" starting from sigmoid functions was first suggested by Lapedes and Farber [8]. It consists of taking the difference between two parallel-displaced *sigmoid* functions [3]. That is,

$\mu_{\text{is approximately K}}$ (x) = *sigm*((x-K)·w^++8) - *sigm*((x-K)·w^+-8) (Fig. 3),

where:

- The parameter w^+>0.0 determines the width of the membership function (as the w^+ increases, the width decreases). The width is practically equal to (2· (16/w^+)).

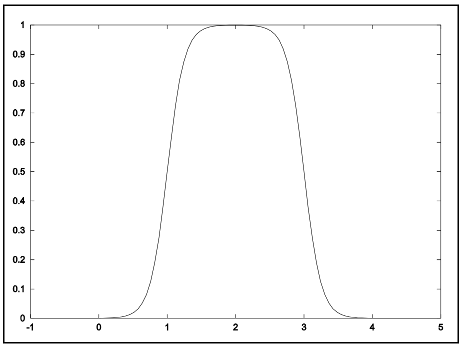

Figure 3: "x is approximately K" with K=2 and w^+=8.

4 Definition of t-norms and t-conorms from neurons

T-norms and t-conorms are binary operators applied on membership degrees belonging to [0,1]. These operators implement the intersection of fuzzy sets (t-norms) and the union of fuzzy sets (t-conorm). In this section, it is presented the construction of t-norms and t-conorms by means of ANN's.

First, we need to define an operator Θ according to the following property:

$$f_A(x_i + x_j) = f_A(x_i)\,\Theta\, f_A(x_j) \qquad (1)$$

where x_i, x_j $\in \Re$. This implies that:

Definition 1

$$a\Theta b = \frac{F_{\min}\cdot(a-F_{\max})\cdot(b-F_{\max}) + F_{\max}\cdot(F_{\min}-a)\cdot(F_{\min}-b)}{(F_{\min}-a)\cdot(F_{\min}-b)+(a-F_{\max})\cdot(b-F_{\max})}$$

with a,b$\in$(F_{min},F_{max}).

It can be noted that the definition of the operator Θ only depends on the parameters F_{min} and F_{max}. This operator is independent of the parameter ρ in function f_A. So, it will be written as $\Theta_{(Fmin,\, Fmax)}$ in the rest of this paper.

T-norms and t-conorms of the class of Hamacher [4, 6] can be defined from this generic operator $\Theta_{(Fmin, Fmax)}$:

a) T-norm of the class of Hamacher

- If F_{min}=0 and F_{max}=2 the operator $\Theta_{(Fmin,Fmax)}$ is equal to:

  $a\,\Theta_{(0,2)}\,b = \frac{a \cdot b}{1+(1-a)\cdot(1-b)}$ with a,b∈(0,2).

  This expression can be generated using the following activation function (*f*) to implement f_A:

  $$f_A(x) = f(x) = 2 \cdot sigm(2 \cdot x).$$

  Besides, if a,b∈(0,1] the operator $\Theta_{(0,2)}$ is a t-norm of the class of Hamacher [4, 6], that is,

  if a,b∈(0,1] ⇒ a $\Theta_{(0,2)}$ b ≡ a AND b

  This property is achieved when the inputs x_i, x_j in equation (1) are **negative**.

- In order to achieve a fuzzy intersection with neurons, we must define the function

  $$f_{norm}(\sum_{i=1}^{n} u_i) = f(4 \cdot \sum_{i=1}^{n}(u_i - 1)),$$

  where:

  - u_i∈[0,1], i=1, …,n.
  - f_{norm} can be implemented with the *sigmoid* function, so:

  $$f_{norm}(\sum_{i=1}^{n} u_i) = f(4 \cdot \sum_{i=1}^{n}(u_i - 1)) =$$

  $$2 \cdot sigm(2 \cdot 4 \cdot \sum_{i=1}^{n}(u_i - 1)) = 2 \cdot sigm(\sum_{i=1}^{n}(8 \cdot u_i) - 8 \cdot n) \cdot$$

  - $f_{norm}(0) = f(4\cdot(-1)) = 2\cdot sigm(-8) \approx 2\cdot 0.0 = 0.0$, and

    $f_{norm}(1) = f(4\cdot 0) = 2\cdot sigm(0) = 1.0$

    (these two conditions will allow to consider this function like a linguistic hedge).

  Starting from this function f_{norm} and the membership functions of the fuzzy sets A and B (μ_A(x) and μ_B(x)), we can obtain:

  $f_{norm}(\mu_A(x) + \mu_B(x)) =$

  $f(4\cdot((\mu_A(x)-1) + (\mu_B(x)-1))) =$

  $f(4\cdot(\mu_A(x)-1) + 4\cdot(\mu_B(x)-1))$.

  As $4\cdot(\mu_A(x)-1)$ and $4\cdot(\mu_B(x)-1)$ are negative, we have

  $f(4\cdot(\mu_A(x)-1) + 4\cdot(\mu_B(x)-1)) =$

  $f(4\cdot(\mu_A(x)-1))$ AND $f(4\cdot(\mu_B(x)-1)) =$

  $f_{norm}(\mu_A(x))$ AND $f_{norm}(\mu_B(x))$.

  Hence, if the fuzzy sets A and B are modified by the function f_{norm}, a fuzzy intersection of these fuzzy sets is achieved. In the next section, it is explained as the function f_{norm} can be considered a linguistic hedge.

b) T-conorm of the class of Hamacher

- If F_{min}= −1 and F_{max}=1 the operator $\Theta_{(Fmin,Fmax)}$ is equal to:

  $a\,\Theta_{(-1,1)}\,b = \frac{a+b}{1+a \cdot b}$ with a,b∈(-1,1).

  This expression can be generated using the *hyperbolic tangent* function (*tanh*) to implement f_A (f_A(x)=*tanh*(x)).

  Besides, if a,b∈[0,1) the operator $\Theta_{(-1,1)}$ is a t-conorm of the class of Hamacher [4, 6], that is,

  if a,b∈[0,1) ⇒ a $\Theta_{(-1,1)}$ b ≡ a OR b.

  This property is achieved when the inputs x_i, x_j in equation (1) are **positive**.

- In order to achieve a fuzzy union with neurons, we must define the function

  $$f_{conorm}(u) = tanh(4 \cdot u),$$

  where:

  - u∈[0,1],
  - f_{conorm} can be implemented with the *sigmoid* function so:

    $f_{conorm}(u) = tanh(4\cdot u) = 2\cdot sigm(8\cdot u) - 1$,
  - $f_{conorm}(0) = tanh(0) = 0$ and

    $f_{conorm}(1) = tanh(4) \approx 1.0$

(these two conditions will allow to consider this function like a linguistic hedge).

Starting from this function f_{conorm} and the membership functions of the fuzzy sets A and B (μ_A(x) and μ_B(x)), we can obtain:

$$f_{conorm}(\mu_A(x) + \mu_B(x)) = tanh(4\cdot(\mu_A(x) + \mu_B(x))) = tanh(4\cdot\mu_A(x) + 4\cdot\mu_B(x)).$$

As $4\cdot\mu_A(x)$ and $4\cdot\mu_B(x)$ are positive, we have

$$tanh(4\cdot\mu_A(x) + 4\cdot\mu_B(x)) =$$

$$tanh(4\cdot\mu_A(x)) \text{ OR } tanh(4\cdot\mu_B(x)) =$$

$$f_{conorm}(\mu_A(x)) \text{ OR } f_{conorm}(\mu_B(x)).$$

Hence, if the fuzzy sets A and B are modified by the function f_{conorm}, a fuzzy union of these fuzzy sets is achieved. In the next section, it is explained as the function f_{conorm} can be considered a linguistic hedge.

5 Linguistic hedges for inserting knowledge into an ANN

In this section, the functions f_{norm} and f_{conorm} are presented as linguistic hedges.

a) **f_{norm} as linguistic hedge**

In Fig. 4, the linguistic hedge of type concentration $(\mu_A(x))^5$ and the function f_{norm} are illustrated in [0,1]. We can see that the function f_{norm} in [0,1] is very similar to the linguistic hedge $(\mu_A(x))^5$. Therefore, the function f_{norm} applied on the fuzzy set *A* can be considered a linguistic hedge of type *concentration*. It can be denominated as "*x is extremely A*".

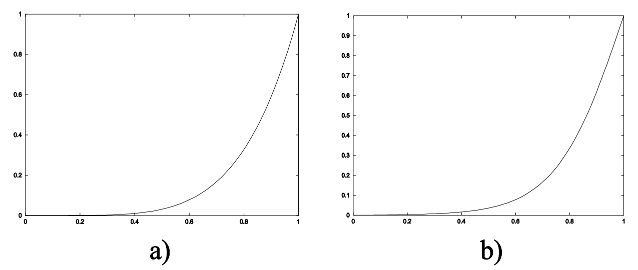

Figure 4: a) Linguistic hedge $(\mu_A(x))^5$. b) Function f_{norm}.

b) **f_{conorm} as linguistic hedge**

On the other hand, in Fig. 5, the linguistic hedge of type dilation $(\mu_A(x))^{0.25}$, the function f_{conorm} and the expression $1-(1-x)^5$ are illustrated in [0,1]. We can see that the function f_{conorm} is a linguistic hedge of type dilation and it is very similar to the expression $1-(1-x)^5$. The main difference with the standard linguistic hedges of type dilation lies in that the function f_{conorm} applied on a fuzzy set *A* increases the magnitude of the grade of membership of x in *A*, which is relatively small for those x with a low grade of membership in *A* (high grade in standard hedges), and relatively large for those x with high membership (low membership in standard hedges).

Therefore, the function f_{conorm} applied on the fuzzy set *A* can be considered a linguistic hedge of type dilation. It can be denominated "*x is somewhat A*".

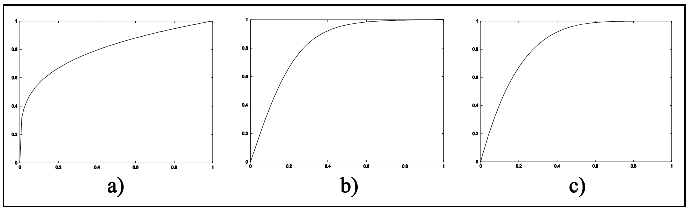

Figure 5: a) Linguistic hedge $(\mu_A(x))^{0.25}$. b) Function f_{conorm}. c) Expression $1-(1-x)^5$.

6 Inserting fuzzy knowledge into ANN's

Multilayer feedforward ANN's are additive fuzzy systems (see [1, 2]). Therefore, if we want to insert fuzzy knowledge into an ANN, this knowledge must be represented with the format of an additive fuzzy system. The previously presented t-norm, t-conorm and membership functions can be used in this additive fuzzy system.

7 Examples

In this section, we present an example where fuzzy knowledge is inserted into a multilayer feedforward ANN. This knowledge is expressed using two input variables (*service* and *food*) and one output variable (*tip*). These variables have the following linguistic values:

a) *Service*: poor, average and good (Fig. 6).

b) *Food*: rancid, normal and delicious (Fig. 7).