ANGIE: Adaptive Network for Granular Information and Evidence Processing[1]

Alexander Valishevsky[2]
Arkady Borisov[3]

Abstract: In this paper the possibility of using adaptive networks in fuzzy-evidence-based decision support systems is considered. The architecture and learning algorithm underlying ANGIE (adaptive network for granular information processing) is presented. The proposed learning procedure helps in solving a task that can be related to sensitivity analysis in decision aid models. The proposed adaptive network can be used as a decision support system or as a tool for determining the significance and contribution of fuzzy features to the reaching of the desired value of the fuzzy criterion.

Keywords: adaptive network, information granularity, fuzzy evidences, decision support systems, sensitivity analysis.

1. Introduction

One of the first attempts to combine neural networks and fuzzy logic and to bring adaptability to the fuzzy systems was made in [Jang, 1993]. The idea of that paper was to use adaptive network as a fuzzy inference system. Learning algorithm was also proposed, that allowed tuning membership functions parameters. The need for adaptive fuzzy systems is evident, as fuzzy-rule-based systems can rarely be applied to solving real world problems without prior tuning of rules.

This paper proposes adaptive network ANGIE (Adaptive Network for Granular Information and Evidence processing) that can be used in fuzzy-evidence-based decision support systems. The learning algorithm is presented as well. However, the algorithm proposed is intended for determining the contribution of fuzzy features which form the left-hand part of fuzzy rules to achieving the desired value of the fuzzy criterion which is indicated in the right-hand part of fuzzy rules. It has a strong relation to sensitivity analysis in the theory of decision-making [Clemen, 1996], though it is not quite the same. Given the desired value of the criterion, with the aid of the proposed algorithm it is possible to determine the *importance* of each fuzzy feature for the criterion to take the desired value. The algorithm will be described later in the paper.

2. Information Granularity

The idea of information granularity and its application in the context of fuzzy logic is presented in [Zadeh, 1979]. The idea of information granularity can be summarized as follows.

Information is contained in granules. In this paper conditioned π-granules are considered which are characterized by propositions of the following form:

$$g_i \overset{\Delta}{=} \text{If } X = u_i \text{ then } Y \text{ is } G_i \,,$$

---

[1] This research has been financially supported by A/S "DATI"

[2] University of Latvia, Institute of Mathematics and Computer Science, *valisevskis@inbox.lv*
[3] Technical University of Riga, Decision Support Systems Group, *aborisov@egle.cs.rtu.lv*

where G_i is a fuzzy subset of V which is dependent on u_i. Evidence E can be regarded as a collection of these granules:

$$E = \{g_1, \ldots, g_N\}.$$

Variable X assumes its value with a specified probability. Thus, evidence is probability distribution P_X of conditional π- granules. Moreover, each granule can be regarded as conditional possibility $\Pi_{(Y|X=u_i)} = G_i$. Hence, evidence can be regarded as conditional possibility distribution $\Pi_{(Y|X)}$. Thus, evidence can be considered as the following construction:

$$E = \{P_X, \Pi_{(Y|X)}\}.$$

Given a collection of bodies of evidence $E = \{E_1, \ldots, E_K\}$, one would like to ask some questions about the information contained in these evidences. The main question is: "what is the probability of (Y is Q)?", where Q is a fuzzy subset of V. The probability of such an event is an interval value in which the expected possibility $\mathrm{E}\Pi(Q)$ is regarded as the upper boundary, and the expected certainty $\mathrm{EC}(Q)$ is regarded as the lower boundary of the probability sought. In [Zadeh, 1979] the following formulae are proposed:

$$\mathrm{E}\Pi(Q) = \sum_{i,j} p_{ij} \sup\left(Q \cap G_i \cap H_j\right), \quad (1)$$

$$\mathrm{EC}(Q) = 1 - \mathrm{E}\Pi(Q'), \quad (2)$$

where Q' is a complement of set Q. Formula (1) is for a special case in which there are two evidences in the system but it would not present any difficulty to induce a more general form of the formula.

3. Architecture of the ANGIE Adaptive Network

Let us consider a set that consists of two evidences:

I. Evidence 1:
   IF (X_1 is B_{11}) with p_{11} THEN (Y is G_{11})
   IF (X_1 is B_{12}) with p_{12} THEN (Y is G_{12})

II. Evidence 2:
   IF (X_2 is B_{21}) with p_{21} THEN (Y is G_{21})
   IF (X_2 is B_{22}) with p_{22} THEN (Y is G_{22})

The adaptive network shown in Figure 1 can be used to determine the upper boundary of the probability (1).

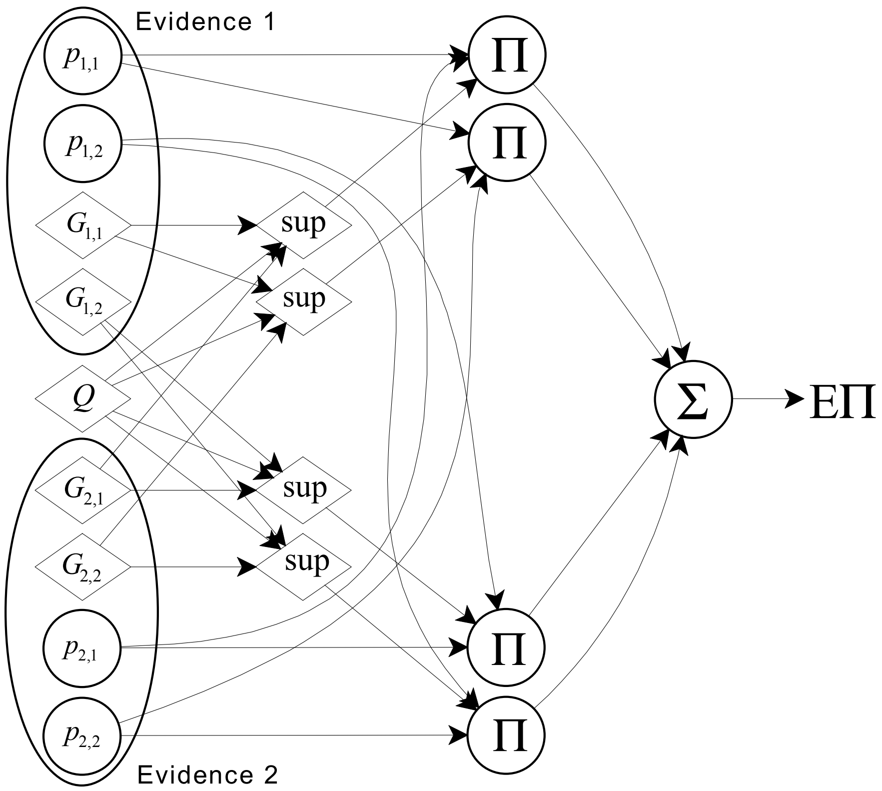

Figure 1. Adaptive network for the calculation of the upper boundary of the probability

Explanation of the meaning of the network processing elements follows.

Oval elements are meant for crisp information processing and diamond elements are meant for fuzzy information processing.

Layer 1:

The output of element $\boldsymbol{G_{ij}}$ is a fuzzy value which corresponds to the right-hand side of the j-th granule in the i-th evidence.

The output of element $\boldsymbol{p_{ij}}$ is the value of the probability associated with the left-hand side of the j-th granule in the i-th evidence.

The output of element $\boldsymbol{Q}$ is the desired value of the criterion.

Layer 2:

The element denoted ***sup*** is used for determining the supremum of the intersection of membership functions fed to the element. Fuzzy values received through dendrites are aggregated by the function of intersection. The transfer function of this element is the supremum function. Thus, the output of this element is the supremum of the intersection of the membership functions received through dendrites.

Layer 3:

The output of element Π is the product of its input values.

Layer 4:

The output of the element Σ is the sum of its input values.

The value of expected certainty EC can be determined in a similar manner. The only difference is that instead of element $\boldsymbol{Q}$ there should be element $\boldsymbol{Q'}$, which denotes a complement of fuzzy set Q, and the output of the network should be defined as $(1-\Sigma)$, instead of Σ.

The aforementioned adaptive network is intended for processing the crisp values of probabilities (in a general case, probabilities can be fuzzy or linguistic).

To summarize the aforesaid we can mention the following facts about the ANGIE topology.

The first layer consists of elements that describe a collection of fuzzy evidences specifying one of the criteria and the desired value of the criterion.

In *the second layer*, the values of the suprema of every possible intersection of the desired value of the criterion and the values of the criteria in the right-hand part of granules are determined.

In *the third layer,* the values of the suprema are multiplied by the corresponding probabilities.

In *the fourth layer*, the values obtained are summed and, as a result, we get the value of $\mathrm{E\Pi}(Q)$.

If there are N evidences in our collection of bodies of evidence and each one consists of Q_i, $i=1,\ldots,N$ granules, then there are $Q_1\ Q_2\ \ldots\ Q_N$ elements in the second layer. The number of elements in the third layer is equal to $Q_1\ Q_2\ \ldots Q_N$ as well.

4. Training algorithm

The gradient descent method is used to tune the parameters of ANGIE. That means that in order to tune the parameters we have to determine the partial derivatives of the error function with respect to the parameters we want to tune.

We will use the same error function as in the backpropagation neural networks:

$$\mathrm{E} = \tfrac{1}{2}(d-o)^2,$$

where d is the desired output value and o is the actual value of the system output.

The output of the system is (1). It is obvious that we cannot tune the parameters of the membership functions since the signal is passed through the ***sup*** element.

Calculation of the partial derivatives follows.

According to the chain rule,

$$\frac{dE}{dp_i} = \frac{dE}{do}\frac{do}{dp_i},$$

$$\frac{dE}{do} = (o-d), \quad \text{and} \quad \frac{do}{dp_i} = \sum_j p_{ij} \sup\left(Q \cap G_i \cap H_j\right),$$

$$\text{thus,} \quad \frac{dE}{dp_i} = (o-d)\sum_j p_{ij} \sup\left(Q \cap G_i \cap H_j\right). \tag{3}$$

Again, (3) is for a special case in which there are two evidences in the system, but it would not present any difficulty to induce a more general type of the formula. It is evident that the value of the partial derivative with respect to the given probability is a sum of terms of $\mathrm{E\Pi}(Q)$ which contain the given probability.

In order to keep the probabilities consistent and to make sure their sum is always equal to 1, the normalization procedure should be applied after every change of the probability value.

Now, what do we mean when we say 'to tune parameters of the decision support model' or 'to tune values of the probabilities'? We will consider these questions in the next section.

5. Semantic view of the decision support model training

Let us consider for what purposes one could tune values of the probabilities with the aid of an adaptive network.

It is possible to tune values of the probabilities in the case when there is a collection of bodies of evidence, specifying a criterion, and when the desired value of the criterion can be determined. In this case tuning values of the probabilities can be interpreted as finding probabilities under which the evidences would maximally satisfy the desired value of the criterion. Furthermore, the values of the probabilities obtained can be regarded as *weight* or *significance* of a given granule in reaching the desired criterion value. In other words, the bigger the value of the probability the bigger the *permissibility* or the *preference* of the left-hand feature value in the corresponding granule.

As an example, let us consider a decision support model which is used to determine the optimal alternative for the construction of a high-voltage electrical network. The alternatives represent different electrical network projects. Each alternative is defined using a set of bodies of evidence. As we are going to solve a sensitivity analysis task, we will consider only one set of bodies of evidence (as we have to determine the *importance* of each parameter value, which is represented by the probability value). We will consider a simplified version of this decision support model.

Let us assume that we have the following features:

- X_1 = density of population
- X_2 = length of the power line

In addition, we have a collection of evidences specifying criterion 'Y = visual effect':

I. IF X_1 = BIG with probability p_{11} THEN Y = BIG
   IF X_1 = MIDDLE with probability p_{12} THEN Y = MIDDLE

II. IF $X_2 \approx 30$ km with probability p_{21} THEN Y = BIG
   IF $X_2 \approx 50$ km with probability p_{22} THEN Y = VERY BIG,

where $p_{11} + p_{12} = p_{21} + p_{22} = 1$.

After determining the collection of evidences and building the ANGIE, the training of the adaptive network is carried out. Let us assume that we have chosen a fuzzy value Q as a desired value for criterion Y. After training, we obtain specific values of the probabilities. Let us assume that the values are the following:

$$p_{11} = 0.79,\ p_{12} = 0.21$$
$$p_{21} = 0.47,\ p_{22} = 0.53$$

We can interpret these values in the following way. In order for Y to take value Q it is enough that the probability of the event (X_1 = BIG) is equal to 0.79, 0.21 for the event (X_1=MIDDLE), 0.47 for ($X_2 \approx 30$ km) and 0.53 for ($X_2 \approx 50$ km).

In other words, the *importance* of the event (X_1 = BIG) is 0.79, the importance of (X_1=MIDDLE) is 0.21.

In its turn, the importance of the events ($X_2 \approx 30$ km) and ($X_2 \approx 50$ km) is almost equal, which could mean that these features do not much influence the probability of the event *(Y is Q)*.

It is obvious that the network intended for the calculation of the expected certainty of the event *(Y is Q)* should be used for the above-mentioned procedure, as this value is the lower boundary of the probability.

Before training, it is recommended to initialise probabilities with the value $\frac{1}{N}$, where N is the number of granules in the evidence.

6. Relation to sensitivity analysis

One of the problems that are considered in the context of sensitivity analysis is the study of preference dependence on the probability of different events [Clemen, 1996]. In other words, a change in preference caused by ranging one or several probabilistic values is considered. For example, in one-way sensitivity analysis all the system parameters, except one, are fixed and the change of preference caused by varying one variable parameter is studied. We can mention tornado diagrams as an example. In two-way sensitivity analysis, there are two variable parameters.

In one-way or two-way sensitivity analysis, one can demonstrate the results obtained graphically. What should we do if we have to determine the interaction of three or more probabilistic parameters? It would be extremely difficult to represent the solution of that problem graphically. A subproblem can be formulated as follows: how to find a domain in the space of probabilistic parameters which would satisfy the condition of reaching the desired criterion value. Moreover, we can carry out several experiments to determine the values of probabilities which lead to different alternative decisions.

That problem can be solved with the help of ANGIE. Values of the probabilities obtained after the training of the system can be considered as approximations to the satisfaction of the condition of reaching the desired value of the criterion.

Further analysis can include the determination of the distance between different points in the space of probabilistic parameters which denote different solutions to the subproblem mentioned above, but this exceeds the bounds of this paper.

7. Alternative architecture of ANGIE

In the adaptive network shown in Figure 1, it is assumed that all the weights in the network are equal to 1. In order to simplify the topology of ANGIE, we can assign corresponding values of the probabilities to all the axons connecting the elements of the first layer with those of the second layer.

That means that all the fuzzy criteria values of every granule will be scaled according to the corresponding value of probability, before reaching the element ***sup***.

Numerically, the result of such calculations will not be the same as that of the adaptive network shown in Figure 1. However, this approach does not contradict our intuition: every membership function is scaled according to the corresponding probability, and only after this operation is the supremum of the intersection of these functions calculated. In the network shown in Figure 1, first the supremum is determined and then the value obtained is multiplied by the corresponding values of probabilities, which can be regarded as scaling as well, with the only difference being that we scale the value after determining the supremum.

The network equivalent to the one in Figure 1 in which the weights are equal to the values of probabilities is shown in Figure 2.

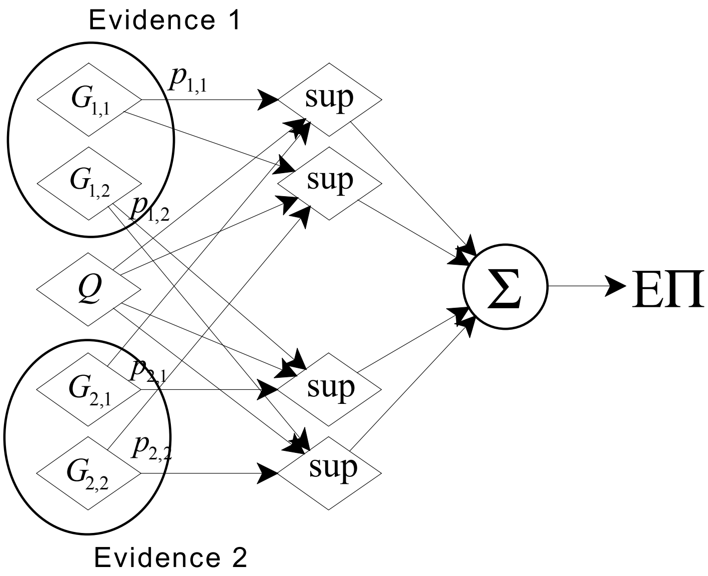

Figure 2. Adaptive network with probabilities as weights

The network in Figure 2 calculates the expected possibility according to the following formula (4):

$$\mathrm{E\Pi}(Q) = \sum_{i,j} \sup\left(Q \cap p_{1,i} G_i \cap p_{2,j} H_j\right). \tag{4}$$

As we can see, the formula (4) is similar to the one shown in (1) and does not contradict our intuition.

8. Alternative way of calculating the EC value

In [Glushkov et al., 1987] an alternative method was proposed for determining the value of the expected certainty. In the new approach, the correcting value is taken into consideration, which guarantees that in case of a total uncertainty the value of *EC* will be equal to zero. Thus, new formula (5) guarantees that the condition $\mathrm{EC} \geq \mathrm{E\Pi}$ is always satisfied:

$$\mathrm{EC}(Q) = \sum_{i,j} p_{ij} \left(\sup\left(G_i \cap H_j\right) - \sup\left(Q' \cap G_i \cap H_j\right)\right). \tag{5}$$

In order to determine the value of (5) with the aid of ANGIE, it is better to represent this formula in the following form (6):

$$\mathrm{EC}(Q) = \sum_{i,j} p_{ij} \sup\left(G_i \cap H_j\right) - \sum_{i,j} p_{ij} \sup\left(Q' \cap G_i \cap H_j\right). \tag{6}$$

In order to determine the first term of (6), we can assume that the membership function $\mu_Q(x)$ is equal to 1 for all values of x and use network shown in Figure 1 for calculations. The second term is quite similar to (2) and can be computed in the same way. Thus, to determine the value of (5), we need to determine the difference between the two values obtained.

9. Conclusion

Adaptive network for a fuzzy-evidence-based decision support model has been proposed in this paper. Topologically, the network represents a calculation network that can be used for determining the upper and the lower bounds of the probability according to (1) and (2).

Moreover, the proposed network-training algorithm enables one to do multi-parameter sensitivity analysis tasks that consist of determining the contribution of fuzzy granules to reaching the desired value of the fuzzy criterion.

A network that can be used to calculate the expected certainty value according to (5) has been described.

Alternative simplified network architecture which yields intuitionally equivalent results has been proposed.

References

[Glushkov et al., 1987] Glushkov V.I., Borisov A.N. *Analysis of Fuzzy Evidence in Decision Making Models*, Optimization Models Using Fuzzy Sets and Possibility Theory, D. Reidel Publishing Company, 1987, pp. 141-153.

[Clemen, 1996] Clemen R.T. *Making Hard Decisions*, Duxbury Press, 1996, pp. 155-178.

[Jang, 1993] Jang J.-S. R. *ANFIS: Adaptive-Network-Based Fuzzy Inference System*, IEEE Transactions on Systems, Man and Cybernetics, Vol. 23, No. 3, May/June 1993, pp. 665-684.

[Zadeh, 1979] Zadeh L.A. *Fuzzy Sets and Information Granularity*, Advances in Fuzzy Set Theory and Applications, Editors: Gupta M.M., Ragade R.K et al., North-Holland Publishing Company, 1979, pp. 3-18.