University of Ostrava

Institute for Research and Applications of Fuzzy Modeling

Fuzzy logical bidirectional associative memory

Radim Bělohlávek

Research report No. 21

April 2, 1999

Submitted/to appear:

Supported by:

grant no. 201/99/P060 of the Grant Agency of the Czech Republic and partly by grant no. VS 96037 of the Ministry of Education of the Czech Republic

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
Bráfova 7, 701 03 Ostrava 1, Czech Republic

tel.: +420-69-6160 233 fax: +420-69-22 28 28
e-mail: radim.belohlavek@osu.cz

Fuzzy logical bidirectional associative memory *

Radim Bělohlávek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Bráfova 7, 701 03 Ostrava, Czech Republic
E-mail: belohlav@osu.cz

We introduce a bidirectional associative memory. The stable points of the memory are naturally interpreted as (non-sharp) concepts—the memory performs association of extents and intents of concepts. We show that this memory is stable and that the set of all stable points forms a complete lattice. We propose a learning algorithm and prove that it enables perfect learning provided the training set forms a consistent conceptual structure. Examples demonstrating the results are presented. Unlike in the case of other associative memories [1], the formal apparatus, architecture, dynamics and convergence proof etc. are based on algebraic structures of fuzzy logic in narrow sense.

Keywords: fuzzy logic, bidirectional associative memory, stability, learning, concepts

AMS Subject classification: 68T05, 68T30, 03B52, 03B80

1. Introduction

Associative memory models are of great interest esp. for neural network community [1]. The reason is that associative memories model a general and powerful phenomenon—the phenomenon of association. Neural network-like associative memory models fit well into the *two-level methodics* [1,3]: studying human behaviour we distinguish two levels—the mental level (macrolevel) and the brain level (microlevel). This idea leads to the endeavour to design systems which mimic humans in that they also exhibit two levels: on the appropriate level of abstraction (macrolevel) we observe phenomena observable on the mental level (e.g. processing uncertainty, performing association) while the processing is "implemented" by a neural network-like architecture (microlevel) inspired by human brain.

Among the many phenomena observable on the mental level one can distinguish the *association phenomenon* (people associate names with faces, causal associations etc.). The models performing associations has been given the name *associative memories* (they are in general discussed and analyzed e.g. in [3]). In [12,13], Kosko proposed so called *bidirectional associative memories* (BAM's). They are to associate two types of phenomena, e.g. names with faces. The memory consists of two layers of neurons, each of the layers representing a phenomenon (e.g. the first layer represents names, the second layer represents faces). The dynamics is as follows: after the neurons are initialized (e.g. from the outside of the net) the net starts sending signals between the layers (from the first to the second, from the second to the first, from the first..., and so on). During the process neurons change their states. Kosko proved [12] that each BAM is stable, i.e. starting from any state the network eventually reaches the state which does not change any more. However, no results concerning the structural description of the set of stable points, characterization of learnable training sets, or learning algorithm with guaranteed properties are available. For further features of the BAM see [12,13] or [1] for other considerations on associative memories.

The aim of this paper is to introduce a new type of bidirectional associative memory based

* This work was supported by grant no. 201/99/P060 of the Grant Agency of the Czech Republic.

strictly on fuzzy logic. The memory is called *fuzzy logical bidirectional associative memory* (FLBAM). The FLBAM network works in discrete time. Unlike other discrete time associative memories (where the excitation values are elements of the set $\{-1,1\}$ by bipolar coding or $\{0,1\}$ by binary coding), the excitation values of neurons are truth values (values of a residuated lattice in general). Section 2 contains the preliminaries. In Section 3, we introduce the architecture and dynamics and show their natural interpretation. In Section 4 we prove stability theorem, describe the structure of stable points, propose a learning algorithm and show its properties. Section 5 contains remarks and examples.

2. Preliminaries

A *fuzzy set* (or **L**-set) in a universe set X is any mapping $A : X \to L$ where, i.e. $A \in L^X$, where L is a support set of some structure **L** of truth values. The value $A(x)$ is understood as the truth value (degree of truth) of "x belongs to A". Similarly, a binary fuzzy relation between X and Y is a mapping $R : X \times Y \to L$. Unlike in the early time [17], fuzzy set theory and fuzzy logic became substantially algebraized [11,10,14]. It has been argued already in [9] that from the logical point of view, **L** should be at least a residuated lattice. Residuated lattices play the role of the algebraic structures of fuzzy logic in narrow sense [10,14]. Recall that a *residuated lattice* is a structure $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \to, 0, 1\rangle$ where (1) $\langle L, \wedge, \vee, 0, 1\rangle$ is a lattice with the least element 0 and the greatest element 1, (2) $\otimes$ is commutative, associative and $1 \otimes x = x$ for all $x \in L$, and (3) $\otimes, \to$ are binary operations (called t-norm and residuum, respectively) which form an adjoint pair, i.e. $x \otimes y \leq z$ iff $x \leq y \to z$. $\otimes$ and $\to$ are intended for modeling of logical conjunction and implication, respectively. The subsequently used identities which hold in residuated lattices may be found e.g. in [14]. The existential and universal quantifiers are modeled using the suprema and infima (in that case, **L** is usually supposed to be a complete lattice w.r.t. the induced lattice ordering). A semantically complete first-order logic with complete residuated lattices as the sets of truth values is described in [11]. Several other special cases of residuated lattices serve as the sets of truth values of other logics [10,14]. The most studied and applied set of truth values is the real interval $[0,1]$ with $a \wedge b = \min(a,b)$, $a \vee b = \max(a,b)$, and with three important pairs of adjoint operations: the Lukasiewicz one ($a \otimes b = \max(a+b-1,0)$, $a \to b = \min(1-a+b,1)$), Gödel one ($a \otimes b = \min(a,b)$, $a \to b = 1$ if $a \leq b$ and $= b$ else), and product one ($a \otimes b = a \cdot b$, $a \to b = 1$ if $a \leq b$ and $= b/a$ else). For the role of these "building stones" in fuzzy logic see [10]. Another important set of truth values is the set $\{a_0 = 0, a_1, \ldots, a_n = 1\}$ ($a_0 < \cdots < a_n$). Two structures are distinguished: Lukasiewicz one with $\otimes$ given by $a_k \otimes a_l = a_{\max(k+l-n,0)}$ and the corresponding $\to$ given by $a_k \to a_l = a_{\min(n-k+l,n)}$, and Gödel one which is a restriciton of the Gödel structure on $[0,1]$.

For $A_1, A_2 \in L^X$ we put $A_1 \subseteq A_2$ iff $A_1(x) \leq A_2(x)$ holds for all $x \in X$.

3. Architecture, dynamics, interpretation

The FLBAM net consists of two layers on neurons. The first layer contains k neurons which we denote by $g_1, \ldots, g_k$. The second layer contains l neurons denoted by $m_1, \ldots, m_l$. Each neuron g_i of the first layer is connected to each neuron m_j of the second layer by a symmetric connection which is assigned a weight I_{ij}. There are no other connections. The architecture is depicted on Fig. 1. The excitation values of the neurons g_i ($i = 1, \ldots, k$) and m_j ($j = 1, \ldots, l$) will be denoted by $A(g_i)$ and $B(m_j)$, respectively. All the values $A(g_i)$, $B(m_j)$ and I_{ij} are values of some structure L of truth values. Hence the excitation values of the layers determines a fuzzy set $A \in L^G, G = \{g_1, \ldots, g_k\}$ ($B \in L^M, M = \{m_1, \ldots, m_l\}$) where the membership degree of g_i in A (m_j in B) is $A(g_i)$ ($B(m_j)$). Similarly, the weights I_{ij} determine a fuzzy relation between

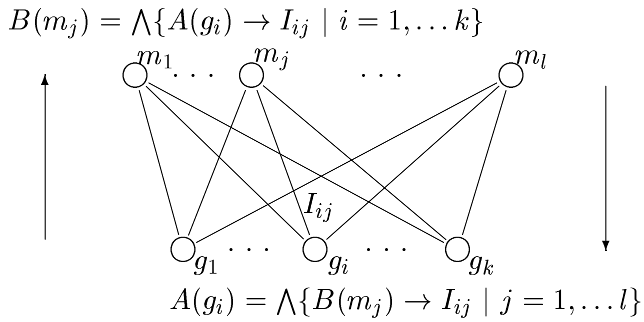

Figure 1. FLBAM neural network scheme.

G and M with the membership degree of $\langle g_i, m_j \rangle$ in I equal to I_{ij}. We choose L to be a support of a residuated lattice $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \to, 0, 1 \rangle$ (cf. Section 2).

Suppose we are given a FLBAM which is described by the weights $I_{ij} \in L$. Suppose the net has been initialized, i.e. all the neurons g_i, m_j have been assigned the truth values $A(g_i)$ and $B(m_j)$, respectively. The pair $\langle A, B \rangle$ of the states of neurons (i.e. fuzzy sets in G and M) is called a *state* of the net. Like in BAM, the signal in FLBAM is fed in two directions, from the first layer to the second layer (up-direction, forward-direction) and from the second layer to the first layer (down-direction, backward-direction). In fact, when initializing, an additional information has to be supplied whether the signal flow has to be started with up- or down-direction.

The up-direction: Suppose the excitation values $A(g_i)$ of all the neurons of the first layer are given. The signal is fed up to the second layer. The excitation values $B(m_j)$ (denoted also $A^{\uparrow}(m_j)$) of the neurons of the second layer are computed by

$$B(m_j) = A^{\uparrow}(m_j) = \bigwedge\{A(g_i) \to I_{ij} \mid i = 1, \dots, k\},$$

where $\bigwedge$ denotes the lattice infimum operation (min in the case of $L = [0, 1]$).

The down-direction: Is dual. Given $B(m_j)$, the signal is fed down by

$$A(g_i) = B^{\downarrow}(g_i) = \bigwedge\{B(m_j) \to I_{ij} \mid j = 1, \dots, l\}.$$

Given an initial state $\langle A(0), B(0) \rangle$ and the starting direction, e.g. "up", the net starts to evolve feeding the signal up, then down, up, down, etc. This leads to a sequence of net states $\{\langle A(t), B(t) \rangle \mid t = 0, 1, 2, \dots\}$. A state $\langle A, B \rangle$ is called *stable* if $A^{\uparrow} = B$ and $B^{\downarrow} = A$, i.e. the net does not change its state when further evolving from this state.

It can be directly seen that FLBAM is derived from the same principles as BAM (in that there are two layers of neurons, and that the dynamics is based on a bidirectional pass of signal). However, there is an important interpretation of FLBAM. Namely, FLBAM can be understood as a memory for storing a collection of concepts together with a hierarchy of these concepts. According to the extent-intent understanding of concepts (which is due to the logical school of Port-Royal [2]), a *concept* (e.g. a DOG) is determined by its *extent*, i.e. a collection of all objects covered by the concept (the collection of all dogs), and its *intent*, i.e. a collection of all attributes covered by the concept (the collection of all properties common to all dogs). By definition, a pair of extent and intent forms a concept iff the intent is the collection of all the attributes common to all the objects of the extent and, conversely, the extent is the collection of all the objects sharing all the attributes of the intent. Note that empirical concepts (like BIG DOG) are non-sharp, i.e. their extents and intents are non-sharp collections. From the point of view of fuzzy approach it is natural to represent the non-sharp extents and intents by fuzzy sets (e.g. the extent of BIG DOG by a fuzzy sets of big dogs, various dogs belong to this extent in various degrees).

Let the neurons g_i represent objects and the neurons m_j represent attributes (cf. the grandmother cell idea [1]). Consider a state $\langle A, B\rangle$. Let A and B be interpreted as the fuzzy sets (non-sharp collections) of objects and attributes, respectively, i.e. $A(g_i)$ and $B(m_j)$ represent the truth value of "the object represented by g_i belongs to A" and "the attribute represented by m_j belongs to B". Let further the weight I_{ij} represents the truth value of "the object represented by neuron g_i has the attribute represented by neuron m_j". Starting by A on the first layer, the network feeds the logical signals up and obtains $A^{\uparrow}$ on the second layer. Conversely, from B on the second layer the net obtains $B^{\downarrow}$ on the first layer. It is immediately (cf. e.g. the definition of truth values of formulas in [11]) seen that $A^{\uparrow}(m_j)$ and $B^{\downarrow}(g_i)$ are the truth values of "the attribute represented by m_j is shared by all the objects which belong to A" and "the object represented by g_i shares all the attributes which belong to B", respectively. We conclude that the state $\langle A, B\rangle$ is stable (i.e. $A^{\uparrow} = B$ and $B^{\downarrow} = A$) iff the pair $\langle A, B\rangle$ represents a (in general non-sharp) concept in the sense of Port-Royal logic. Note that the formalization of the Port-Royal notion of concept for the case of $\mathbf{L}$ being the two-element Boolean algebra of classical logic is due to Wille [16,7] (the developed theory is known as concept lattices or formal concept analysis). The theory has been then generalized for the case of fuzzy approach (for $\mathbf{L}$ being a complete residuated lattice) in [4–6].

4. Stability, sensitivity and learning

First, we analyze the stability of FLBAM. We define: A FLBAM is stable if for each initial state $\langle A(0), B(0)\rangle$ it holds that starting in any of the directions up or down, the FLBAM eventually reaches a stable state $\langle A(t^*), B(t^*)\rangle$. Note that each BAM is stable [12] and that the proof is based on finding an appropriate energy function. The following theorem asserts the stability of FLBAM, however the proof is completely algebraic:

Theorem 1. Each FLBAM is stable. Moreover, the stable point is reached after two (discrete time) steps.

Proof. Let a FLBAM be given by the collection I_{ij} of weights of connections between the sets G and M of neurons. As shown in [5], the mappings $^{\uparrow}$ (assigning to each fuzzy set $A \in L^G$ a fuzzy set $A^{\uparrow} \in L^M$) and $^{\downarrow}$ (assigning $B^{\downarrow} \in L^G$ to each $B \in L^M$) which define the dynamics form a fuzzy Galois connection between G and M (i.e. for each $A_1, A_2 \in L^G$, it holds (i) $A_1 \subseteq (A_1^{\uparrow})^{\downarrow}$, $B_1 \subseteq (B_1^{\downarrow})^{\uparrow}$, (ii) $\mathit{Subs}(A_1, A_2) \leq \mathit{Subs}(A_2^{\uparrow}, A_1^{\uparrow})$, and dually for $B_1, B_2 \in L^M$; here, $\mathit{Subs}(A_1, A_2) = \bigwedge_{g \in G}(A_1(g) \to A_2(g))$ is the subsethood degree of A_1 in A_2 [14]). From the properties of fuzzy Galois connections it follows [5] that $A^{\uparrow} = A^{\uparrow\downarrow\uparrow}$, $B^{\downarrow} = B^{\downarrow\uparrow\downarrow}$ which implies the assertion. □

Denote $\mathit{Stab}(I) = \{\langle A, B\rangle \mid A^{\uparrow} = B, B^{\downarrow} = A\}$ the set of all stable points (i.e. fuzzy concepts) of a FLBAM with the weights I_{ij}. On $\mathit{Stab}(I)$ define the relation $\leq$ ("to be a subconcept") by $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ iff $A_1 \subseteq A_2$ (iff $B_1 \supseteq B_2$). For instance, a concept BIG DOG is a subconcept of a concept DOG for each element (a big dog) of the extent of BIG DOG is an element of the extent of DOG (a dog). The following theorem describes the hierarchy of concepts.

Theorem 2. The set $\mathit{Stab}(I)$ of all stable points is under $\leq$ a complete lattice.

Proof. The assertion follows from the fact that $\mathit{Stab}(I)$ is the set of fixed points of fuzzy Galois connection, see [5]. □

Theorem 2 states that the stable points (concepts) obey a complete hierarchy: for each set of concepts there is their direct generalization (superconcept, supremum) as well as their specialization (subconcept, infimum).

To describe algebraically the computational power of FLBAM's completely note that the mappings $^{\uparrow}$ and $^{\downarrow}$ which define the dynamics form a fuzzy Galois connection between G and M (cf. proof of Theorem 1). From [5] it follows also the converse: each pair $^{\uparrow}$ and $^{\downarrow}$ which forms a fuzzy Galois connection between G and M can be realized by some FLBAM (which can obtained from $^{\uparrow}$ and $^{\downarrow}$ constructively, see [5]).

The next topic we will be interested in is the sensitivity analysis of FLBAM to input changes. The sensitivity is connected to an important feature of association performed by humans, namely, the preservation of similarity: two phenomena which are associated to two similar phenomena, are similar. This property is often required for a system to be considered as a good model of association. Even Aristotle [15] discussed the role of similarity in association. There is a natural way to measure the similarity among fuzzy sets. For fuzzy sets $A_1, A_2 \in L^X$ put

$$E(A_1, A_2) = \bigwedge_{x \in X} (A_1(x) \leftrightarrow A_2(x)).$$

Here $a \leftrightarrow b = (a \to b) \wedge (b \to a)$ is the biresiduum operation. On the linguistic level, $E(A_1, A_2)$ corresponds to the truth value of the fact that each element belongs to A_1 if and only if it belongs to A_2. It may be shown [3] that E is in fact a $\otimes$-similarity relation on L^X, i.e.

$$\begin{aligned} E(A, A) &= 1 \\ E(A_1, A_2) &= E(A_2, A_1) \\ E(A_1, A_2) \otimes E(A_2, A_3) &\leq E(A_1, A_3) \end{aligned}$$

for every $A, A_1, A_2, A_3 \in L^X$. We take E for measuring similarity both between the states of the first layer (fuzzy sets from L^G) and between the states of the second layer (fuzzy sets from L^M). The question of preserving similarity by the association performed by FLBAM can then be put in a clear way. We say that FLBAM *preserves similarity* if for each $A_1, A_2 \in L^G$, $B_1, B_2 \in L^M$, it holds $E(A_1, A_2) \leq E(A_1^{\uparrow}, A_2^{\uparrow})$ and $E(B_1, B_2) \leq E(B_1^{\downarrow}, B_2^{\downarrow})$. The interpretation is: if the inputs are similar then also the outputs are similar.

Theorem 3. Each FLBAM preserves similarity.

Proof. We prove only $E(A_1, A_2) \leq E(A_1^{\uparrow}, A_2^{\uparrow})$, the second part may be obtained symmetrically.

$$\begin{aligned} &\bigwedge_{g \in G} (A_1(g) \leftrightarrow A_2(g)) = E(A_1, A_2) \leq \\ &\leq E(A_1^{\uparrow}, A_2^{\uparrow}) = \bigwedge_{m \in M} \left(A_1^{\uparrow}(m) \leftrightarrow A_2^{\uparrow}(m)\right) \end{aligned}$$

holds iff for each $m \in M$ it holds

$$\begin{aligned} &\bigwedge_{g \in G} (A_1(g) \leftrightarrow A_2(g)) \leq \\ &\leq A_1^{\uparrow}(m) \leftrightarrow A_2^{\uparrow}(m) = \left(A_1^{\uparrow}(m) \to A_2^{\uparrow}(m)\right) \wedge \left(A_2^{\uparrow}(m) \to A_1^{\uparrow}(m)\right) \end{aligned}$$

which holds iff the left side of the inequality is less or equal than both members of the right side which are connected by $\wedge$. We check only the first one of these inequalities, i.e.

$$\bigwedge_{g \in G} (A_1(g) \leftrightarrow A_2(g)) \leq A_1^{\uparrow}(m) \to A_2^{\uparrow}(m).$$

By adjunction, this holds iff

$$A_1^{\uparrow}(m) \otimes \bigwedge_{g\in G} (A_1(g) \leftrightarrow A_2(g)) \leq A_2^{\uparrow}(m)$$

i.e.

$$\left(\bigwedge_{g\in G} A_1(g) \rightarrow I(g,m)\right) \otimes \bigwedge_{g\in G} (A_1(g) \leftrightarrow A_2(g)) \leq \bigwedge_{g\in G} A_2(g) \rightarrow I(g,m)$$

which holds iff for each $g' \in G$ the inequality

$$\bigwedge_{g\in G} (A_1(g) \leftrightarrow A_2(g)) \otimes \left(\bigwedge_{g\in G} A_1(g) \rightarrow I(g,m)\right) \leq A_2(g') \rightarrow I(g',m)$$

holds. The last inequality is equivalent (by adjunction) to

$$A_2(g') \otimes \bigwedge_{g\in G} (A_1(g) \leftrightarrow A_2(g)) \otimes \left(\bigwedge_{g\in G} A_1(g) \rightarrow I(g,m)\right) \leq I(g',m)$$

which holds because

$$\begin{aligned} A_2(g') \otimes \bigwedge_{g\in G} (A_1(g) \leftrightarrow A_2(g)) \otimes \left(\bigwedge_{g\in G} A_1(g) \rightarrow I(g,m)\right) \leq \\ \leq A_2(g') \otimes (A_2(g') \rightarrow A_1(g')) \otimes (A_1(g') \rightarrow I(g',m)) \leq I(g',m) \end{aligned}$$

by applying twice the fact that $\otimes$ and $\rightarrow$ is an adjoint pair. □

The following corollary is immediate.

Corollary 4. Under the conditions of Theorem 3, it holds $E(A_1, A_2) \leq E(A_1^{\uparrow\downarrow}, A_2^{\uparrow\downarrow})$ and $E(B_1, B_2) \leq E(B_1^{\downarrow\uparrow}, B_2^{\downarrow\uparrow})$.

Note that even without Theorem 3 we have due to the properties of E that $E(A_1^{\uparrow\downarrow}, A_1) \otimes E(A_1, A_2) \otimes E(A_2, A_2^{\uparrow\downarrow}) \leq E(A_1^{\uparrow\downarrow}, A_2^{\uparrow\downarrow})$, i.e. if $A_1^{\uparrow\downarrow}$ and A_1, A_1 and A_2, A_2 and $A_2^{\uparrow\downarrow}$ are pairwise similar then also $A_1^{\uparrow\downarrow}$ and $A_2^{\uparrow\downarrow}$ are similar. Corollary 4 asserts a stronger result.

Corollary 5. Under the conditions of Theorem 3, it holds $E(A_1, A_2) \otimes E(A_1, A_1^{\uparrow\downarrow}) \leq E(A_1, A_2^{\uparrow\downarrow})$ and $E(B_1, B_2) \otimes E(B_1, B_1^{\downarrow\uparrow}) \leq E(B_1, B_2^{\downarrow\uparrow})$.

Proof. $E(A_1, A_2) \otimes E(A_1, A_1^{\uparrow\downarrow}) \leq E(A_1^{\uparrow\downarrow}, A_2^{\uparrow\downarrow}) \otimes E(A_1, A_1^{\uparrow\downarrow}) \leq E(A_1, A_2^{\uparrow\downarrow})$. The second part may be proved analogously. □

A further question we will be interested in is that of learning. The transparent interpretation of FLBAM weights (i.e. truth values of "an object has a property") enables us to set the weights by experts. The stable points of the net are then just the concepts determined by the weights. The problem may be inverted: Given a set of net states

$$T = \{\langle A^p, B^p\rangle \mid p \in P, A^p \in L^G, B^p \in L^M\} \tag{1}$$

find a FLBAM (i.e. its weights I_{ij}) so that $Stab(I) = T$ (or some weaker condition, e.g. $Stab(I) \supseteq T$) holds. Thus, T describes the intended behaviour. The training set T should be in some sense consistent. In our case, since the elements are interpreted as concepts and since the properties of $^\uparrow$ and $^\downarrow$ (cf. proof of Theorem 1) we should require e.g.: if we have $A_1 \subseteq A_2$ for $\langle A_1, B_1\rangle$, $\langle A_1, B_1\rangle \in T$ then it must hold $B_1 \supseteq B_2$ (i.e. the more objects the less common properties). We call T *consistent* (*complete*, resp.) if it is learnable in principle (completely, resp.), i.e. if there is a FLBAM given by I such that $T \subseteq Stab(I)$ (each element of T becomes a stable point), ($T = Stab(I)$, resp.). The following theorem states that a complete T can be learned perfectly.

Theorem 6. Let T be a complete training set, i.e. such that there is a FLBAM given by I such that $Stab(I) = T$. Let the weights I^T are given by $I^T_{ij} = \bigvee_{\langle A,B\rangle \in T} A(g_i) \otimes B(m_j)$. Then it holds $I^T = I$ and hence $T = Stab(I^T)$.

Proof. We have to prove $I^T = I$. First, we show $I^T \subseteq I$. Let $g \in G$, $m \in M$ be arbitrary elements. For each $\langle A, B\rangle \in T$ we have $A(g) = \bigwedge\{B(m') \to I(g, m') \mid m' \in M\}$. From the well-known lattice properties it follows $A(g) \leq B(m) \to I(g, m)$ from which we get by adjunction property $A(g) \otimes B(m) \leq I(g, m)$. Since g and m are chosen arbitrarily we conclude

$$I^T(g, m) = \bigvee_{\langle A,B\rangle \in T} (A(g) \otimes B(m)) \leq I(g, m),$$

i.e. $I^T \subseteq I$ holds.

Conversely, consider the concept $\langle \{\,{}^1/g\}^{\uparrow\downarrow}, \{\,{}^1/g\}^{\uparrow}\rangle$. We have

$$\begin{aligned}\{\,{}^1/g\}^{\uparrow}(m) &= \bigwedge\{\{\,{}^1/g\}(g') \to I(g', m) \mid g' \in G\} \\ &= 1 \to I(g, m) = I(g, m)\end{aligned}$$

and $\{\,{}^1/g\}^{\uparrow\downarrow}(g) = 1$. Thus for the concept $\langle A, B\rangle = \langle \{\,{}^1/g\}^{\uparrow\downarrow}, \{\,{}^1/g\}^{\uparrow}\rangle \in T$ we have $A(g) \otimes B(m) = 1 \otimes I(g, m) = I(g, m)$ which gives

$$I^T(g, m) = \bigvee_{\langle A,B\rangle \in T} (A(g) \otimes B(m)) \geq I(g, m),$$

i.e. $I^T \supseteq I$. We have $I^T = I$ completing the proof. □

Theorem 6 gives and justifies a learning algorithm for FLBAM's: given T, set weights I^T as in the assertion of Theorem 6. The following theorem shows that each consistent T can be learned by this algorithm.

Theorem 7. Let T be consistent and I^T be as in Theorem 6. Then it holds $T \subseteq Stab(I^T)$.

Proof. Suppose T is consistent, i.e. $T \subseteq Stab(I)$ for some I. Since, by Theorem 6,

$$\begin{aligned}I^T(g, m) &= \bigvee_{\langle A,B\rangle \in T} A(g) \otimes B(m) \leq \\ &\leq \bigvee_{\langle A,B\rangle \in Stab(I)} A(g) \otimes B(m) = I^{Stab(I)}(g, m) = I(g, m)\end{aligned}$$

we have $I^T \subseteq I$. Consider any $\langle A, B\rangle \in T$. We have to prove $\langle A, B\rangle \in Stab(I^T)$, i.e. $A^{\uparrow_{I^T}} = B$ and $B^{\downarrow_{I^T}} = A$ (where $^{\uparrow_{I^T}}$ and $^{\downarrow_{I^T}}$ are induced by I^T). Due to symmetry we prove only $A^{\uparrow_{I^T}} = B$ by checking both inequalities. Take any $m \in M$. By $I^T \subseteq I$,

$$A^{\uparrow_{I^T}}(m) = \bigwedge_{g \in G} \left(A(g) \to I^T(g, m)\right) \leq$$

$$\leq \bigwedge_{g\in G} (A(g) \to I(g,m)) = A^{\uparrow_I}(m) = B(m).$$

On the other hand,

$$B(m) \leq A^{\uparrow_{I^T}}(m) = \bigwedge_{g_i\in G} \left(A(g_i) \to I^T_{ij}\right)$$

holds iff for each $g \in G$ it holds

$$B(m) \leq A(g) \to I^T(g,m) = A(g) \to \bigvee_{\langle A',B'\rangle\in T} A'(g) \otimes B'(m)$$

which holds (applying $b \leq a \to (a \otimes b)$) since

$$B(m) \leq A(g) \to (A(g) \otimes B(m)) \leq A(g) \to \bigvee_{\langle A',B'\rangle\in T} A'(g) \otimes B'(m).$$

□

As a corollary we obtain a criterion for a training set T to be consistent.

Corollary 8. A training set T is consistent iff for each $\langle A,B\rangle \in T$, $g \in G$, $m \in M$ it holds $A(g) = \bigwedge_{m\in M}(B(m) \to \bigvee_{\langle A',B'\rangle\in T}(A'(g) \otimes B'(m)))$ and $B(m) = \bigwedge_{g\in G}(A(g) \to \bigvee_{\langle A',B'\rangle\in T}(A'(g) \otimes B'(m)))$.

The foregoing statements lead to the following learning algorithm for FLBAM.

Algorithm 9.

```
INPUT(T); training set (1), we suppose P = {1,...,|P|}.
for i:=1 to k do
   for j:=1 to l do
     begin
      I_ij:=0
      for p:=1 to |P| do
        I_ij:=I_ij ∨ (A^p(g_i) ⊗ B^p(m_j))
     end
 for p:=1 to |P| do
   if A^p↑ ≠ B^p or B^p↓ ≠ A^p then
     OUTPUT("T not consistent"); exit
 OUTPUT(I)
```

It is immediate that the time complexity of this algorithm is in $O(k \cdot l \cdot |P|)$.

5. Remarks, examples

First, due to the clear interpretation the net can be used as a concepts-storing module which might be used as a part of more complex network architectures [1]. Second, due to their transparency, the weights I of FLBAM may be set directly by experts. From this expert knowledge of the form "an object has a property", the net can determine all concepts which are hidden in the supplied information. The net may thus be used for conceptual data analysis of non-sharp (fuzzy) data (for more on this subject see [3,7]). Third, by Algorithm 9, the net can extract a knowledge from a given set of concepts (training set). The given set can be then completed into a complete hierarchy of concepts which stays beyond the particular concepts listed in the set.

no.	extent									intent			
	Me	V	E	Ma	J	S	U	N	P	ss	sl	df	dn
T_1	0	0	0	0	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	1	1	$\frac{1}{2}$
T_2	1	1	1	1	0	0	0	0	0	1	0	0	1
T_3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0	$\frac{1}{2}$	0
T_4	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1	$\frac{1}{2}$	0	1	0
T_5	0	0	0	$\frac{1}{2}$	1	1	1	1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	0
T_6	0	0	0	$\frac{1}{2}$	1	1	1	1	1	0	0	1	0
T_7	0	0	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0	1	0
T_8	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
T_9	0	0	0	$\frac{1}{2}$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$
T_{10}	1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0	1
T_{11}	1	1	1	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	0	$\frac{1}{2}$
T_{12}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1
T_{13}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
T_{14}	1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$

Table 1
Training set for FLBAM (Example 1).

I^T	size		from sun	
	small (ss)	large (sl)	far (df)	near (dn)
Mercury (Me)	1	0	0	1
Venus (V)	1	0	0	1
Earth (E)	1	0	0	1
Mars (Ma)	1	0	$\frac{1}{2}$	1
Jupiter (J)	0	1	1	$\frac{1}{2}$
Saturn (S)	0	1	1	$\frac{1}{2}$
Uranus (U)	$\frac{1}{2}$	$\frac{1}{2}$	1	0
Neptune (N)	$\frac{1}{2}$	$\frac{1}{2}$	1	0
Pluto (P)	1	0	1	0

Table 2
Weight matrix learned from T (Example 1).

Example 10. Consider a FLBAM with 9 neurons in the first layer and 4 neurons in the second layer. The neurons represent the planets (Mercury (Me), ..., Pluto (P)) and their attributes ("size small", "size large", "far from sun", "near to sun"). Take $L = \{0, \frac{1}{2}, 1\}$ and the Lukasiewicz operations $\otimes, \rightarrow$ (see Section 2). Consider the training set T given in Table 1. Elements of T can be easily interpreted as concepts. For instance, consider T_6 which can be linguistically described as "a planet which is far from sun" (J,...,P belong to its extent in degree 1, Ma in degree $\frac{1}{2}$). It can be verified by Corollary 8 that T is consistent. Tab. 2 contains the matrix of weights I^T_{ij} obtained by Algorithm 9. By Theorem 7, the set $Stab(I^T)$ of all stable points contains each element of T, hence T has been completely learned. By Theorem 2, the set $Stab(I^T)$ forms a complete lattice which reflects the hierarchy of stable points. The elements of $Stab(I^T)$ can be easily obtained from I^T, we do not list them explicitly. The lattice structure of $Stab(I^T)$ is depicted in Fig. 2. The elements of T correspond to the following elements of $Stab(I^T)$: T_1 is 13, T_2 is 14, T_3 is 24, T_4 is 27, T_5 is 28, T_6 is 34, T_7 is 18, T_8 is 20, T_9 is 19, T_{10} is 21, T_{11} is 22, T_{12} is 15, T_{13} is 26, and T_{14} is 29. Note that it is rather natural that $Stab(I^T)$ contains some additional points (concepts). Namely, T itself does not obey the complete hierarchy. E.g. the points (concepts) T_3 ("small planet which is at least little bit far from sun") and T_5 ("at least little bit large planet which is far from sun") have no generalization in T. On the other

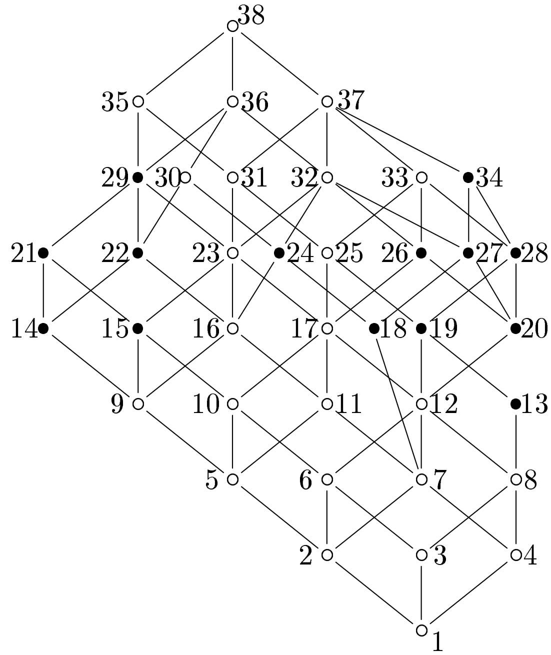

Figure 2. Complete lattice $Stab(I^T)$. The elements of T correspond to filled circles (Example 1).

hand, their generalization in $Stab(I^T)$ is 37 which is a concept which would be represented by $\langle \frac{1}{2}\frac{1}{2}\frac{1}{2}111111|00\frac{1}{2}0\rangle$ in the form used in Tab. 1, i.e. the concept "planet which is at least little bit far from sun".

Example 11. Let $L = \{0, \frac{1}{2}, 1\}$ and consider the Gödel operations $\otimes, \rightarrow$ on L. (see Section 2). Consider further a FLBAM with 9 neurons in the first layer and 25 neurons in the second layer. We will represent each state $\langle A, B\rangle$ of the net by a pair of 3-level grayscale bitmap pictures, a 3×3-bitmap for A and 5×5-bitmap for B. Each neuron corresponds to one pixel. The neuron states is 0, $\frac{1}{2}$, and 1 are represented by white, gray, and black pixels, respectively. Consider the training set T consisting of four pairs $\langle A, B\rangle$ represented by the pairs 13, 14, 16, and 25 of Fig. 3. Algorithm 9 yields I^T such that all the stable states of $Stab(I^T)$ are depicted in Fig. 3. The conceptual hierarchy of the stable points is visualized in Fig. 4.

References

[1] M. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press, London, 1995.
[2] A Arnauld, P. Nicole, *La logique ou l'art de penser*, 1662.
[3] Bělohlávek R.: Networks Processing Indeterminacy, Ph.D. thesis, Technical University of Ostava, viii+198 pp., Ostrava (1998).
[4] R. Bělohlávek, Lattices generated by binary fuzzy relations, Tatra Mount. Math. Publ. (to appear, 1999).
[5] R. Bělohlávek, Fuzzy Galois connections, Math. Logic Quarterly **45**(1999) (to appear).
[6] R. Bělohlávek, L-concept lattices, Research report 15, IRAFM, University of Ostrava (1999).
[7] B. Ganter, R. Wille, Applied lattice theory: formal concept analysis, in: G. Grätzer, *General Lattice Theory*, to appear.
[8] J. A. Goguen, L-fuzzy sets, Journ. Math. Anal. Appl. 18 (1967) 145–174.
[9] J. A. Goguen, The logic of inexact concepts, Synthese 19 (1968–69) 325–373.
[10] P. Hájek, *Metamathematics of Fuzzy Logic*, Kluwer, 1998.
[11] U. Höhle, On the fundamentals of fuzzy set theory, Journ. Math. Anal. Appl., 201 (1996) 786–826.
[12] B. Kosko, Bidirectional associative memories, IEEE Trans. SMC vol.18, no.1 (1988) 49–60.
[13] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice-Hall, 1992.

1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30

Figure 3. Stable points (concepts) $Stab(I^T)$ (Example 2).

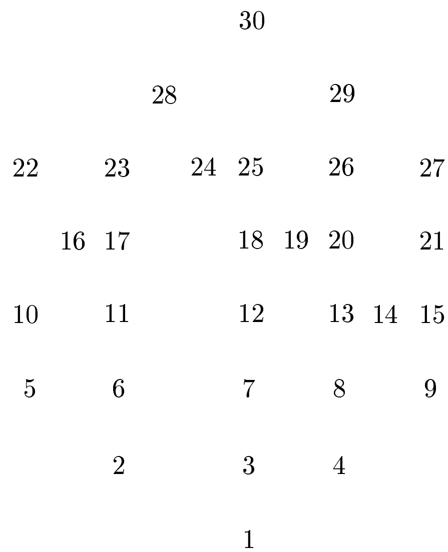

Figure 4. Complete lattice $Stab(I^T)$. The elements of T correspond to filled circles (Example 2).

[14] V. Novák, I. Perfilieva and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, 1999.
[15] R. Sorabji R, *Aristotle on Memory*, Brown University Press, Providence, 1972.
[16] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: Rival I, *Ordered Sets*, Reidel, Dordrecht, Boston, 1982, 445—470.
[17] L. A. Zadeh, Fuzzy sets, Information and Control 8(3) (1965) 338–353.