Probability-Possibility Transformation:
Application to Bayesian and Possibilistic Networks

Yosra Ben Slimen, Raouia Ayachi, and Nahla Ben Amor

LARODEC, Institut Supérieur de Gestion Tunis 41 Avenue de la liberté, 2000 Le Bardo, Tunisie

Abstract. Probability-possibility transformation is a purely mechanical transformation of probabilistic support to possibilistic support and vice versa. In this paper, we apply the most common transformations to graphical models, i.e., Bayesian into possibilistic networks. We show that existing transformations are not appropriate to transform Bayesian networks to possibilistic ones since they cannot preserve the information incorporated in joint distributions. Therefore, we propose new consitency properties, exclusively useful for graphical models transformations.

Keywords: Probability-Possibility transformation, Bayesian networks, Possibilistic networks.

1 Introduction

Probability and possibility theories are two ways to express uncertainty. Several bridges between these two frameworks were established. Especially, several researches addressed the problem of transformation of possibilistic distributions into probabilistic ones and vice versa. The first interest underlying these transformations is to study the coherence between these frameworks and, more precisely, the consistency of derived distributions. Another interest is to make a benefit advantage of each framework. Following this idea, we are interested by transformations between Bayesian networks [13] and their adaptation in the possibilistic framework i.e. possibilistic networks [13]. In fact, these graphical models, which share the same graphical component i.e. Directed Acyclic Graph (DAG), are quantified using different distributions (i.e., probability distributions in the case of Bayesian networks and possibility ones in the case of possibilistic networks). Recently, the inference topic in possibilistic networks has been explored using compilation techniques [1]. It has been shown that the qualitative setting of possibility theory goes beyond the probabilistic framework and the quantitative possibilistic framework since it takes advantage of specific properties of the minimum operator. So, our objective in this paper is to study the possibility of switching from one model to another in order to reason in an efficient way.

This paper is organized as follows: Section 2 presents most common transformations. Section 3 presents some basics of Bayesian and possibilistic networks. Section 4 studies the particular case of transforming Bayesian networks into possibilistic ones.

F. Masulli, G. Pasi, and R. Yager (Eds.): WILF 2013, LNAI 8256, pp. 122–130, 2013.

2 Probability-Possibility Transformation

Possibility theory introduced by Zadeh [14] and developed by Dubois and Prade [6] lies at the crossroads between fuzzy sets, probability and non-monotonic reasoning. The basic building block in possibility theory is the notion of *possibility distribution* [6]: let $V = \{X_1, ..., X_N\}$ be a set of state variables whose values are ill-known such that $D_1 \ldots D_n$ are their respective domains. $\Omega = D_1 \times \ldots \times D_N$ denotes the universe of discourse, which is the cartesian product of all variable domains in V. Vectors $\omega \in \Omega$ are often called *realizations* or simply "states" (of the world). In what follows, we use x_i to denote possible instances of X_i. The agent's knowledge about the value of the x_i's can be encoded by a possibility distribution $\pi : \Omega \to [0, 1]$ where $\pi(\omega) = 1$ means that ω is totally possible and $\pi(\omega) = 0$ means that ω is an impossible state. It is generally assumed that there exist at least one state ω which is totally possible, π is then said to be *normalized*. We denote by $\top(\pi)$ the set of totally possible states in π. From π, one can compute, for any event $A \subseteq \Omega$, the possibility measure $\Pi(A) = \sup_{\omega \in A} \pi(\omega)$ that evaluates to which extend A is *consistent* with the knowledge represented by π. The particularity of the possibilistic scale is that it can be interpreted twofold: either in an *ordinal* manner, when the possibility degree reflects only an ordering between the possible values, so the *minimum* operator is used to combine different distributions, or, in a *numerical* manner, so possibility distributions are combined using the *product* operator.

Several researchers tackle different bridges between probability and possibility theory. When we deal with those transformations, two cases can be distinguished, those relative to subjective probabilities [8] and those relative to objective ones. In this paper, we focus on these latters which were used in several practical problems such as: *constructing a fuzzy membership function from statistical data* [11], *combining probabilities and possibilities information in expert systems* [9] and *reducing the computational complexity* [7]. Roughly speaking, transforming probabilistic distributions to possibilistic ones, denoted by $p \to \pi$, is useful when weak source of information makes probabilistic data unrealistic or to reduce the complexity of the solution or to combine different types of data. However, transformation, denoted by $\pi \to p$, is useful in the case of decision making. Interestingly enough, when transforming $p \to \pi$, some information is lost because we transform point value probabilities to interval values ones. In contrast, $\pi \to p$ adds information to some possibilistic incomplete knowledge.

2.1 Consistency Principle

In order to describe different transformations, several properties, called *consistency principle*, were proposed in literature. We retain, in particular, three of them:

Zadeh Consistency Principle: Zadeh [14] defined the probability-possibility consistency principle such as *"a high degree of possibility does not imply a high degree of probability, and a low degree of probability does not imply a low degree of possibility"*. The degree of consistency between p and π is defined by: $C(\pi, p) = \sum_{i=1 \ldots n} \pi_i * p_i$. Zadeh [14] pointed out that $C(\pi, p)$ is not a precise law or a relationship between possibility and probability distributions. It is an approximate formalization of the heuristic connection stating that lessening the possibility of an event tends to lessen its probability but not vice-versa.

Klir Consistency Principle: The concept of consistency condition was redefined by Klir [10]. Assume that the elements of Ω are ordered in such a way that $p_i > 0$ and $p_i \geq p_{i+1}, \ \forall \ i = \{1..n\}$. Any transformation should be based on these assumptions:
– *A scaling assumption* that forces each value π_i to be a function of p_i/p_1 (where $p_1 \geq \ldots \geq p_n$).
– *An uncertainty invariance assumption* according to which p and π must have the same amount of uncertainty.
– *Consistency condition:* $\pi_i \geq p_i$ stating that what is probable must be possible, so π can be seen as an upper-bound of p.
Dubois and Prade [5] gave an example to show that the scaling assumption of Klir may sometimes lead to violation of the consistency principle. The second assumption is also debatable because it assumes that possibilistic and probabilistic information measures are commensurate.

Dubois and Prade Consistency Principle: Dubois and Prade defined the consistency principle, differently, using these assumptions [4]:
– *Consistency condition:* $P_i < \Pi_i, \ \forall \ i = \{1..n\}$.
– *Preference preservation:* Assuming that π has the same form as p, then $\forall(\omega_1, \omega_2) \in \Omega^2, \ p(\omega_1) > p(\omega_2) \ \Rightarrow \ \pi(\omega_1) > \pi(\omega_2)$ and $p(\omega_1) = p(\omega_2) \ \Rightarrow \ \pi(\omega_1) = \pi(\omega_2)$.
– *Maximum specificity:* Let π_1 and π_2 be two possibility distributions, then π_2 is more specific than π_1 iff: $\forall\omega \in \Omega, \ \pi_2(\omega) \leq \pi_1(\omega)$.

2.2 Probability-Possibility Transformation Rules

Several transformation rules are proposed in literature. We present the most common ones, namely: *Klir transformation* (KT), *Optimal transformation* (OT), *Symmetric transformation* (ST) and *Variable transformation* (VT).

Klir Transformation (KT): Assume that the elements of Ω are ordered in such a way that: $\forall \ i = \{1..n\}, \ p_i > 0, \ p_i \geq p_{i+1}$ and $\pi_i > 0, \ \pi_i \geq \pi_{i+1}$ with $p_{n+1} = 0$ and $\pi_{n+1} = 0$. Klir has considered the principle of uncertainty preservation under two scales [10]:

- *The ratio scale:* $p \to \pi$ and $\pi \to p$, named the normalized transformations, are defined by:
$$\pi_i = \frac{p_i}{p_1} \ , \ p_i = \frac{\pi_i}{n \sum_{i=1}^{n} \pi_i} \tag{1}$$
- *The log-interval scale:* $p \to \pi$ and $\pi \to p$ are defined by:
$$\pi_i = (\frac{p_i}{p_1})^\alpha \ , \ p_i = \frac{\pi_i^{\frac{1}{\alpha}}}{\sum_{i=1}^{n} (\pi_i)^{\frac{1}{\alpha}}} \tag{2}$$

where α is a parameter that belongs to the open interval $]0, 1[$.

Optimal Transformation (OT): proposed by Dubois and Prade [4] and also called "Asymmetric Transformation", is defined as follows:
$$\pi_i = \sum_{j/p_j \leq p_i} p_j, \ p_i = \sum_{j=1}^{n} \frac{\pi_j - \pi_{j+1}}{j} \tag{3}$$

OT is optimal because it gives the most specific possibility distribution i.e. that loses less information [7], and it's asymmetric since the two formulas in Equation (3) are not converse. Sandri et al. [7] suggested a *Symmetric Transformation* (ST) that needs less computation but it is quite far from the optimum. It is defined by:

$$\pi_i = \sum_{j=1}^{n} min(p_i, p_j) \tag{4}$$

Variable Transformation (VT): It's a $p \to \pi$ transformation proposed by Mouchaweh et al. [12] and expressed as follows: assume that the elements of Ω are ordered in such a way that: $\forall\ i = \{1..n\},\ \ p_i > 0,\ \ p_i \geq p_{i+1}$ and $\pi_i > 0,\ \ \pi_i \geq \pi_{i+1}$ with $p_{n+1} = 0$ and $\pi_{n+1} = 0$, then:

$$\pi_i = (\frac{p_i}{p_1})^{k.(1-p_i)} \tag{5}$$

where k is a constant belonging to the interval: $0 \leq k \leq \frac{logp_n}{(1-p_n).\log(\frac{p_n}{p_1})}$.

Bouguelid [3] proposed VT_i, which is an improvement of VT, to make it as specific as OT. So, a parameter k_i is set for each π_i. Formally: $\forall\ i = \{1..n\}$,

$$\pi_i = (\frac{p_i}{p_1})^{k_i.(1-p_i)} \tag{6}$$

where k_i belongs to the interval: $0 \leq k_i \leq \frac{log(p_i+p_{i+1}+...+p_n)}{(1-p_i).\log(\frac{p_i}{p_1})},\ \ \forall i = \{2..n\}$.

Table 1 summarizes characteristics of KT, OT, ST, VT and VT_i. For each transformation, it is mentioned if it deals with discrete (D) and-or continuous case (C) and if it satisfies consistency principle (Cs), preference preservation (PP) and maximum specificity (MS). Clearly, OT and VT_i are the most interesting rules in the discrete case for $p \to \pi$.

Table 1. Summary of transformations

TR	$p \to \pi$	$\pi \to p$	Properties				
			D	C	Cs	PP	MS
KT	$\times$	$\times$	$\times$	$\times$		$\times$	
OT	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
ST	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	
VT	$\times$		$\times$			$\times$	
VTi	$\times$		$\times$		$\times$	$\times$	$\times$

3 Basics on Bayesian and Possibilistic Networks

Bayesian networks [13] are powerful probabilistic graphical models for representing uncertain knowledge. Studying the possibilistic counterpart of Bayesian networks leads to two variants, namely: min-based possibilistic networks corresponding to the ordinal interpretation of the possibilistic scale and product-based possibilistic networks corresponding to the numerical interpretation [2]. It is well-known that product-based possibilistic networks are close to Bayesian networks since they share the same features (essentially the product operator) with almost the same theoretical and practical results [2]. This is not the case for min-based possibilistic networks due to the particularities of the min operator (e.g. the idempotency). Over a set of N variables $V = \{X_1, .., X_N\}$, Bayesian networks (denoted by BN) and possibilistic networks (denoted by $\Pi G_{\otimes}$ where $\otimes = min$ in the ordinal setting, and $\otimes = *$ in the numerical one) share the same two components:

– A *graphical component* composed of a DAG, $\mathcal{G}= (V, E)$ where V denotes a set of *nodes* representing variables and E a set of *edges* encoding links between nodes.

– A *numerical component* that quantifies different links. Uncertainty of each node X_i is represented by a local normalized conditional probability or possibility distribution in the context of its parents.

Given a Bayesian network BN on N variables, we can compute its joint probability distribution by the following chain rule :

$$p(X_1, \ldots, X_N) = *_{i=1..N} \ P(X_i \mid U_i) \tag{7}$$

In a similar manner, the joint possibility distribution of a possibilistic network $\Pi G_{\otimes}$ is defined by the $\otimes$-based chain rule, where $\otimes = \min$ for the ordinal setting and $\otimes = *$ for the numerical one, expressed by:

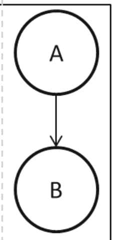
$$\pi_{\otimes}(X_1, \ldots, X_N) = \otimes_{i=1..N} \ \Pi(X_i \mid U_i) \tag{8}$$

One of the most interesting treatments that can be applied for possibilistic networks is to evaluate the impact of a certain event on the remaining variables. Such process, called *inference*, consists on computing a-posteriori possibility distributions of each variable X_i given an evidence e.

Example 1. Let us consider the Bayesian network and the possibilistic network in Fig. 1(a) and Fig. 1 (b), respectively (sharing the same DAG). The joint distributions of BN and $\Pi G_{\otimes}$ using Equations (7) and (8) are presented in Fig. 1 (c).

P(A)	a1	a2
	0.6	0.4

P(B|A)	a1	a2
b1	0.6	0.5
b2	0.3	0.3
b3	0.1	0.2



(a)

Π(A)	a1	a2
	1	0.4

Π(B|A)	a1	a2
b1	1	1
b2	0.4	0.5
b3	0.1	0.2

(b)

A	B	P	π_*	π_{min}
a_1	b_1	0.36	1	1
a_1	b_2	0.18	0.4	0.4
a_1	b_3	0.06	0.1	0.1
a_2	b_1	0.2	0.4	0.4
a_2	b_2	0.12	0.2	0.4
a_2	b_3	0.08	0.08	0.2

(c)

Fig. 1. A Bayesian network (a), a possibilistic network (b) and their joint distributions (c)

4 Transformation from Bayesian to Possibilistic Networks

Probability-possibility transformations can be useful to study the coherence between probabilistic and possibilistic frameworks and, more precisely, the consistency of derived distributions. Our idea consists in applying such transformations from Bayesian networks to possibilistic networks and interpreting their behavior on joint distributions. Formally, using existing transformations, we can define transformation from Bayesian to possibilistic networks in a local manner as follows:

Definition 1. *Let BN be a Bayesian network and p be its joint distribution. Let TR be a transformation rule. Let $BNto\Pi N$ be the function that transforms BN into $\Pi N_{\otimes}^{TR}$ using TR under the setting $\otimes$ s.t. $\otimes = \{*, min\}$. Let $PDto\Pi D$ be the function that transforms a probability distribution into a possibilistic one using TR. Formally, $\Pi N_{\otimes}^{TR}$ is the transformation of BN using TR if, $\forall X_i \in V$,*

$$\Pi(X_i \mid U_i) = PDto\Pi D(P(X_i \mid U_i), TR) \tag{9}$$

$$\Pi N_{\otimes}^{TR} = BNto\Pi N(BN, TR, \otimes) \tag{10}$$

Example 2. Table 2 depicts the transformation of conditional tables of the Bayesian network of Fig. 1 (a) using KT, OT, ST, VT and VT_i.

Table 2. Transformation of conditional distributions

$\Pi(A)$	Π^{KT}	Π^{OT,VT_i}	Π^{ST}	Π^{VT}
a_1	1	1	1	1
a_2	0.66	0.4	0.8	0.4
$\Pi(B \mid A)$	Π^{KT}	Π^{OT,VT_i}	Π^{ST}	Π^{VT}
$b_1 \mid a_1$	1	1	1	1
$b_2 \mid a_1$	0.5	0.4	0.7	0.5
$b_3 \mid a_1$	0.16	0.1	0.3	0.1
$b_1 \mid a_2$	1	1	1	1
$b_2 \mid a_2$	0.6	0.5	0.8	0.27
$b_3 \mid a_2$	0.4	0.2	0.6	0.2

This local transformation does not ensure the same results as a global one. In other words, the transformation of the joint distribution underlying the initial Bayesian network is not equivalent to the transformation of its local conditional distributions, which can affect the inference results. Let π_p^{TR} be the transformation of the joint distribution encoded by a Bayesian network BN using the transformation TR and let $\pi_{\otimes}^{TR}$ be the joint distribution relative to $\Pi N_{\otimes}^{TR}$ obtained using Definition 1. The following example illustrates the problem described above.

Example 3. Table 3 presents the transformation of global distributions of the Bayesian network of Fig. 1 (a) and of the resulted possibilistic network $\Pi N_{\otimes}$ using KT, OT, ST, VT and VT_i.

Table 3. Possibility distributions using different transformations

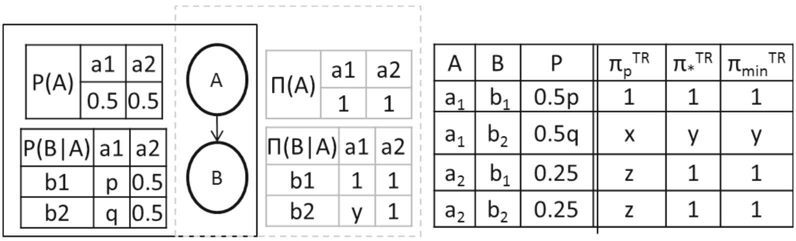
A	B	p	KT			OT, VT_i			ST			VT		
			π_p^{KT}	π_*^{KT}	π_{min}^{KT}	π_p^{OT,VT_i}	π_*^{OT,VT_i}	π_{min}^{OT,VT_i}	π_p^{ST}	π_*^{ST}	π_{min}^{ST}	π_p^{VT}	π_*^{VT}	π_{min}^{VT}
a_1	b_1	0.36	1	1	1	1	1	1	1	1	1	1	1	1
a_1	b_2	0.18	0.5	0.5	0.5	0.44	0.4	0.4	0.8	0.7	0.7	0.38	0.5	0.5
a_1	b_3	0.06	0.16	0.16	0.16	0.06	0.1	0.1	0.36	0.3	0.3	0.06	0.1	0.1
a_2	b_1	0.2	0.55	0.66	0.66	0.64	0.4	0.4	0.84	0.8	0.8	0.45	0.4	0.4
a_2	b_2	0.12	0.33	0.4	0.6	0.26	0.2	0.4	0.62	0.64	0.8	0.19	0.108	0.27
a_2	b_3	0.08	0.22	0.26	0.4	0.14	0.08	0.2	0.46	0.48	0.6	0.09	0.08	0.2

As depicted in Table 3, if we are in a numerical setting, the values of π_p^{TR} are different from those of π_*^{TR} and, if we deal with an ordinal setting, the order between π_p^{TR} and π_{min}^{TR} is not preserved, as well. For instance, for the transformation ST, more precisely for a_1b_2 and a_2b_2, we can see that $0.8 > 0.62$ while $0.7 < 0.8$. It is also the case of VT for a_1b_2 and a_2b_1. Suppose, now, that we have the evidence $B = b_2$, then for π_p^{ST} we have $a_1 > a_2$ while the same evidence implies $a_2 > a_1$ for π_{min}^{ST}. This means that, considering π_{min}^{ST} as the consistent transformation of the initial Bayesian network and using it to infer evidence can lead to erroneous results.

The question that may arise is the following: *Do all transformations suffer from the problem of information loss?* The answer can be found in the following example.

Example 4. Let us consider the BN of Fig. 2 (a) s.t $p > q$. This implies that $p > 0.5$ and $q < 0.5$, which in its turn implies that $0.5p > 0.5q > 0.25$. Fig. 2 (c) shows the joint distributions where $x < 1$, $y < 1$ and $z < 1$ and TR can be any transformation (i.e. KT, OT, ST, VT, VT_i).

We start by interpreting product-based networks which only rely on numerical values. It is obvious, from columns 4 and 5 of Fig. 2 (c), that there is a loss of information since values of π_p^{TR} and π_*^{TR} are different. When we deal with min-based networks, the focus is only on the order induced by values. In fact, the order of π_p^{TR} of the initial network BN is $\{a_1b_1 > a_1b_2 > (a_2b_1 = a_2b_2)\}$, while the order relative to π_{min}^{TR} is $\{(a_1b_1 = a_2b_1 = a_2b_2) > a_1b_2\}$.



(a) (b) (c)

Fig. 2. A BN (a), its transformation into a possibilistic one (b) and their joint distributions (c)

Following this problem, we propose two new properties. The first one (resp. the second one), presented in Definition 2 (resp. Definition 3), is applicable for transforming Bayesian networks into min-based possibilistic networks (resp. product-based possibilistic networks). These properties should be seen as extensions of Dubois and Prade Consistency principle described above.

Definition 2. *Let TR be a transformation rule used in order to transform a Bayesian network BN into a min-based possibilistic network* ΠN_{min}^{TR}*. Let p be the joint distribution relative to BN computed using Equation (7) and* π_p^{TR} *be its transformation by TR. Let* π_{min}^{TR} *be the joint distribution relative to* ΠN_{min}^{TR} *using Equation (8) (s.t* $\otimes = min$*). Let* $\delta(\pi_p^{TR})$ *and* $\delta(\pi_{min}^{TR})$ *be the order underlying* π_p^{TR} *and* π_{min}^{TR}*, respectively. Then TR is said to be consistent iff: (i)* $\delta(\pi_p^{TR}) = \delta(\pi_{min}^{TR})$ *and (ii)* $\top(\pi_p^{TR}) = \top(\pi_{min}^{TR})$

Definition 3. *Let TR be a transformation rule used in order to transform a Bayesian network BN into a product-based possibilistic network* ΠN_*^{TR}*. Let p be the joint distribution relative to BN computed using Equation (7) and* π_p^{TR} *be its transformation by TR. Let* π_*^{TR} *be the joint distribution relative to* ΠN_*^{TR} *using Equation (8) (s.t* $\otimes = *$*). Then TR is said to be consistent iff:* $\pi_p^{TR} = \pi_*^{TR}$

Clearly, the formulas (ii) in Definition 2. guarantees the normalized values in both ordinal and numerical settings. We point out that this property is ensured by existing transformations.

5 Conclusion

Our objective in this paper is to study the transformation of Bayesian networks into possibilistic networks using existing transformations proposed in literature. We found out that switching from one model to another does not preserve the information incorporated in joint distributions (either numerical values for ΠN_* or the order induced by values for ΠN_{min}). Such result allows us to conclude that such transformations are inappropriate in the case of graphical models. Indeed, we have shown that it leads to erroneous inference results. A deep study on this behavior shows that this loss of information is due to the non-compatibility of product and min operators, in the ordinal setting. In our future work, we will deeply explore the impact of this loss of information on inference result for both product-based possibilistic networks and min-based possibilistic networks and propose two new transformations that respect the properties we proposed in order to transform Bayesian networks into possibilistic ones.

References

1. Ayachi, R., Ben Amor, N., Benferhat, S.: Possibilistic local structure for compiling min-based networks. In: SMPS, pp. 479–487 (2012)
2. Borgelt, C., Gebhardt, J., Kruse, R.: Possibilistic graphical models. In: Proceedings of the International School for the Synthesis of Expert Knowledge (1998)
3. Bouguelid, M.S.: Contribution à l'application de la reconnaissance des formes et la théorie des possibilités au diagnostic adaptatif et prédictif des systèmes dynamiques. PhD thesis, Université de Reims Champaghe-Ardenne (December 2007)

4. Dubois, D., Foulloy, L., Mauris, G., Prade, H.: Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. Reliable Computing (2004)
5. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
6. Dubois, D., Prade, H.: Possibility theory: An approach to computerized, Processing of uncertainty. Plenium Press, New York (1988)
7. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. Fuzzy Logic (1993)
8. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. International Journal of Approximate Reasoning 48, 352–364 (2008)
9. Klir, G.J.: Probability-possibility transformations: a comparison. Int. J. General Systems 21, 291–310 (1992)
10. Klir, G.J.: Information-preserving probability-possibility transformations: recent developments. Fuzzy Logic, 417–428 (1993)
11. Mouchaweh, M.S.: Diagnosis in real time for evolutionary processes using pattern recognition and possibility theory. Invited paper in International Journal of Computational Cognition, 79–112 (2004)
12. Mouchaweh, M.S., Bouguelid, M.S., Billaudel, P., Riera, B.: Variable probability-possibility transformation. In: 25th European Annual Conference on Human Decision-Making and Manual Control (EAM 2006), pp. 417–428 (September 2006)
13. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco (1988)
14. Zadeh, L.A.: Fuzzy sets as a biaisis for a therory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)