FUZZ-IEEE 2007

IEEE International Conference on Fuzzy Systems

Tutorial 5: Uncertainty and Information

IEEE

IEEE Computational Intelligence Society

BT

UNCERTAINTY AND INFORMATION

Tutorial at FUZZ-IEEE 2007 in London

George J. Klir
Department of Systems Science & Industrial Engineering
Binghamton University (State University of New York)
Binghamton, New York 13902, USA
gklir@binghamton.edu

OUTLINE

Part 1: Overall Characterization of Generalized Information Theory

1. Generalized Information Theory (GIT): An Overview
2. Two-Dimensional Expansion of Classical Information Theory
3. Uncertainty Theories Viewed as Theories of Imprecise Probabilities
4. Canonical Representation of Imprecise Probabilities

Part 2: Measuring Uncertainty and Information

5. Required Properties of Measures of Uncertainty and Information
6. An Overview of Measures of Uncertainty and Information Developed Within GIT
7. Some Methodological Issues Within GIT
8. Current and Future Research in GIT

GENERALIZED INFORMATION THEORY (GIT)

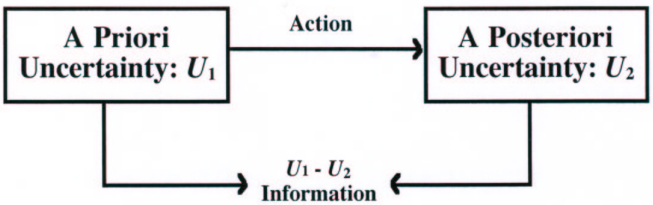
- GIT is a research program whose objective is to develop a formal treatment of the interrelated concepts of uncertainty and information in all their varieties; it is a generalization of two distinct branches of classical information theory, which are based, respectively, on the notions of possibility (crisp) and probability.

- In GIT, as in classical information theory, uncertainty (predictive, retrodictive, diagnostic, prescriptive, etc.) is viewed as a manifestation of some information deficiency, while information is viewed as the capacity to reduce uncertainty. That is, GIT deals with information-based uncertainty and uncertainty-based information.

- The aims of GIT were introduced in 1991 in my paper "Generalized Information Theory" [*Fuzzy Sets and Systems*, 40(1), pp. 127-142].

- Comprehensive and up-to date coverage of results obtained by research within GIT is contained in the text Uncertainty and Information [John Wiley, Hoboken, NJ, 2006].

Uncertainty-based Information



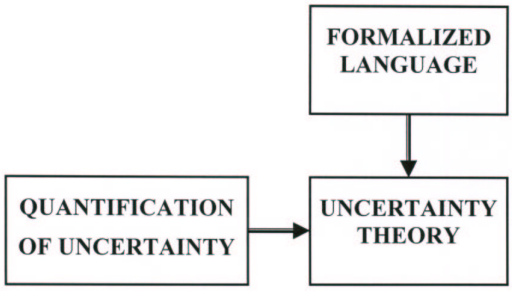
The amount of information obtained by an action

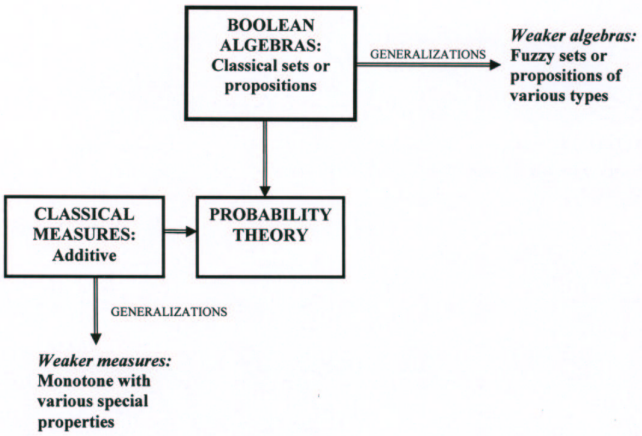
=

The amount of uncertainty reduced by the action

The mathematical theory of information had come into being when it was realized that the flow of information can be represented numerically in the same way as distance, mass, temperature, etc.

(Alfréd Rényi)





MONOTONE MEASURES

Given a universal set X and a non-empty family $\mathcal{C}$ of subsets of X (usually with an appropriate algebraic structure), a *monotone measure* (also called a fuzzy measure), μ, on $\langle X, \mathcal{C} \rangle$ is a function

$$\mu: \mathcal{C} \rightarrow [0, \infty]$$

that satisfies the following requirements:

(1) $\mu(\varnothing) = 0$ (*vanishing at the empty set*);

(2) for all $A, B \in \mathcal{C}$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$ (*monotonicity*);

(3) for any increasing sequence $A_1 \subseteq A_2 \subseteq \ldots$ of sets in $\mathcal{C}$,

$$\text{if } \bigcup_{i=1}^{\infty} A_i \in \mathcal{C}, \text{ then } \lim_{i \to \infty} \mu(A_i) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right)$$

(*continuity from below*);

(4) for any decreasing sequence $A_1 \supseteq A_2 \supseteq \ldots$ of sets in $\mathcal{C}$,

$$\text{if } \bigcap_{i=1}^{\infty} A_i \in \mathcal{C}, \text{then } \lim_{i \to \infty} \mu(A_i) = \mu\left(\bigcap_{i=1}^{\infty} A_i\right)$$

(*continuity from above*).

FUZZY SETS: Basic Characteristics

- Sets whose boundaries are not sharp.
- Sets that allow to distinguish degrees (or grades) of membership.
- Sets that are fully characterized by membership functions of the form $F: D \rightarrow R$.
- Distinct categories of fuzzy sets are distinguished by distinct types of sets D (domains) and R (ranges) that are employed in defining their membership functions.
- For each $x \in D$, $F(x)$ is viewed as the degree of membership of object x in fuzzy set F.
- $F(x)$ may also be interpreted as the degree of compatibility of object x with a given concept represented by fuzzy set F.
- Membership functions of standard fuzzy sets have the form $F: X \rightarrow [0, 1]$, where X is a classical (crisp) set (universal set) whose elements are not fuzzy sets.

α-Cuts of Standard Fuzzy Sets

- For each $\alpha \in [0,1]$, the set ${}^{\alpha}A = \{x \in X \mid A(x) \geq \alpha\}$ is called an α-cut of standard fuzzy set A whose membership function has the form $A: X \rightarrow [0,1]$.
- Any standard fuzzy set is uniquely represented by its α-cuts for all $\alpha \in [0,1]$.
- Properties of classical sets can be extended to fuzzy sets (fuzzified) by requiring that they be preserved in all α-cuts. This kind of fuzzification is called a cutworthy fuzzification.

UNCERTAINTY THEORIES			FORMALIZED LANGUAGES						
			CLASSICAL SETS	NONCLASSICAL SETS					
				STANDARD FUZZY SETS	NONSTANDARD FUZZY SETS				
					INTERVAL VALUED	TYPE 2	LEVEL 2	LATTICE BASED	•••
MONOTONE MEASURES	ADDITIVE	CLASSICAL NUMERICAL PROBABILITY							
	NON ADDITIVE	POSSIBILITY/ NECESSITY							
		SUGENO λ-MEASURES							
		BELIEF/ PLAUSIBILITY (CAPACITIES OF ORDER ∞)							
		CAPACITIES OF VARIOUS FINITE ORDERS							
		INTERVAL-VALUED PROBABILITY DISTRIBUTIONS							
		• • •							
		GENERAL LOWER AND UPPER PROBABILITIES							

DIVERSITY AND UNITY OF UNCERTAINTY THEORIES

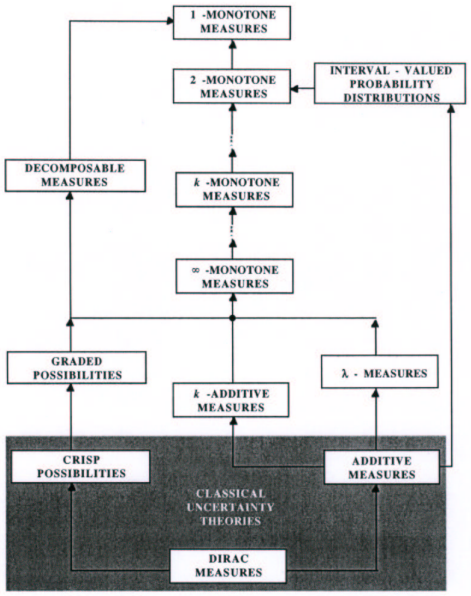
- It is significant that the enormous and ever increasing diversity of uncertainty theories subsumed under GIT is balanced by some common features they share.
- The diversity of uncertainty theories is in some sense desirable. It enables us to focus on the development of those theories that are promising from the standpoint of various applications. However, it makes it increasingly more difficult to get oriented within the many existing or emerging theories and the many prospective theories.
- It turns out that all recognizable uncertainty theories in GIT can be classified in a useful way so that some properties of the theories are invariant within each class. This unity of uncertainty within each class allows us to work within the class as a whole.
- A significant class of uncertainty theory with common properties consists of the various theories of imprecise probabilities defined on finite classical sets.

THEORIES OF UNCERTAINTY

In order to develop a fully operational theory, T, for dealing with uncertainty of some conceived type requires that a host of issues be addressed at the following four levels:

- LEVEL 1 --- we need to find an appropriate mathematical representation of the conceived type of uncertainty, which is achieved by characterizing, via appropriate axioms, a class of uncertainty functions, say functions u, that represent uncertainty in theory T.
- LEVEL 2 --- we need to develop operating rules (calculus) for manipulating the uncertainty functions u in theory T.
- LEVEL 3 --- we need to find a meaningful way of measuring the amount of relevant uncertainty in any situation formalizable in theory T, which is achieved by finding a justifiable functional, U, which for each uncertainty function u in theory T measures the amount of uncertainty associated with it.
- LEVEL 4 --- we need to develop methodological aspects of theory T by utilizing functional U as an abstract measuring instrument.

Uncertainty Theories



CHOQUET CAPACITIES OF ORDER k

(k=2,3,...,∞)

- **Alternative name: k-monotone measures (k ≥ 2).**
- **2-monotone measures are defined for all pairs A, B of subsets of X by the inequality:**

$$\mu(A\cup B) \geq \mu(A) + \mu(B) - \mu(A\cap B).$$

- **3-monotone measures are defined for all triples A, B, C of subsets of X by the inequality**

$$\mu(A\cup B\cup C) \geq \mu(A) + \mu(B) + \mu(C) - \mu(A\cap B) - \mu(A\cap C) - \mu(B\cap C) + \mu(A\cap B\cap C)$$

- **k-monotone measures are defined for all families of k subsets of X by the inequality**

$$\mu\left(\bigcup_{j=1}^{k} A_j\right) \geq \sum_{\substack{K\subseteq N_k \\ K\neq\emptyset}} (-1)^{|K|+1} \mu\left(\bigcap_{j\in K} A_j\right)$$

- **1-monotone measures is a convenient name for superadditive measures that satisfy for all disjoint pairs of subsets A and B of X the inequality**

$$\mu(A \cup B) \geq \mu(A) + \mu(B).$$

IMPRECISE PROBABILITIES: Canonical Representations

1. **Lower probability function: μ_***
2. **Upper probability function: μ^***
3. **Möbius functions: m**
4. **Convex set of probability distributions: D**

CONVERSIONS

- **1 ⇔ 2: duality equation**
- **1 ⇔ 3: Möbius transform**
- **1 ⇒ 4: constructing extreme points of D**
- **4 ⇒ 1: $\mu_*(A) = \inf_{p\in D}\{\sum_{x\in A} p(x)\}$, ∀A**

LOWER AND UPPER PROBABILITY MEASURES ASSOCIATED WITH A CONVEX SET OF PROBABILITY DISTRIBUTIONS D on X

Notation: $^D\mu_*$ denotes the lower probability measure associated with D

$^D\mu^*$ denotes the upper probability measure associated with D

Basic formulas:

$$^D\mu_*(A) = \inf_{p\in D}\{\sum_{x\in A} p(x)\}, \forall A$$

$$^D\mu^*(A) = \sup_{p\in D}\{\sum_{x\in A} p(x)\}, \forall A$$

Duality of $^D\mu_*$ and $^D\mu^*$:

$$^D\mu^*(A) = 1 - {}^D\mu_*(\bar{A}), \forall A$$

CONVEX SET OF PROBABILITY DISTRIBUTIONS ASSOCIATED WITH A GIVEN LOWER PROBABILITY MEASURE

- Let $X = \{x_1, x_2, ..., x_n\}$ and let $\sigma = (\sigma(x_1), \sigma(x_2), ..., \sigma(x_n))$ denote a permutation by which elements of X are reordered.
- Given any lower probability measure μ_* on the power set of X that is 2-monotone, the convex set of all probability distributions that dominate this measure, $D(\mu_*)$, is determined by its extreme points, which are probability distributions p_σ computed as follows:

$p_\sigma(\sigma(x_1)) = \mu_*(\{\sigma(x_1)\})$,

$p_\sigma(\sigma(x_2)) = \mu_*(\{\sigma(x_1), \sigma(x_2)\}) - \mu_*(\{\sigma(x_1)\})$,

.....

$p_\sigma(\sigma(x_{n-1})) = \mu_*(\{\sigma(x_1), ... ,\sigma(x_{n-1})\}) - \mu_*(\{\sigma(x_1), ... ,\sigma(x_{n-2})\})$

$p_\sigma(\sigma(x_n)) = \mu_*(\{\sigma(x_1), ... ,\sigma(x_n)\}) - \mu_*(\{\sigma(x_1), ... ,\sigma(x_{n-1})\})$

- Each permutation defines an extreme point of $D(\mu_*)$, but different permutations can give rise to the same point.
- $D(\mu_*)$ is the convex hull of the extreme points.

MÖBIUS REPRESENTATION OF LOWER PROBABILITIES

- Given any lower probability $\underline{\mu}$ on P(X) its Möbius representation, m, is obtained for all $A \in P(X)$ via the formula (Möbius transform)

$$m(A) = \sum_{B \mid B \subseteq A} (-1)^{|A-B|} \underline{\mu}(B).$$

- It is guaranteed that $m(\varnothing) = 0$ and

$$\sum_{A \in P(X)} m(A) = 1.$$

- The inverse transform is given for all $A \in P(X)$ by the formula

$$\underline{\mu}(A) = \sum_{B \mid B \subseteq A} m(B).$$

UNCERTAINTY MEASURES: Key Requirements

1. Subadditivity: The amount of uncertainty in a joint representation of evidence cannot be greater than the sum of the amounts of uncertainty in the associated marginal representations of uncertainty.

2. Additivity: The two amounts of uncertainty considered under subadditivity become equal when the marginal representations of evidence are noninteractive according to the rules of the uncertainty calculus involved.

3. Range: The range of uncertainty is [0, M], where 0 must be assigned to the unique uncertainty function that describes full certainty and M depends on the cardinality of the universal set involved and on the chosen unit of measurement.

4. Continuity: Any measure of uncertainty must be continuous.

5. Expansibility: Expanding the universal set by alternatives that are not supported by evidence must not affect the amount of uncertainty.

6. Branching/Consistency: When uncertainty can be computed in several distinct ways, each conforming to the calculus of the theory, the results must be the same (consistent).

7. Monotonicity: When evidence can be ordered in the theory, the measure of uncertainty must preserve this ordering.

8. Coordinate invariance: When evidence is expressed within some Euclidean space, uncertainty must not change under isometric transformation of coordinates.

HARTLEY MEASURE OF UNCERTAINTY

$$H(A) = \log_2 |A|$$

- Among a given universal set X of all considered alternatives (predictions, retrodictions, diagnoses, etc.), only alternatives in set $A \subseteq X$ are possible according to given evidence.
- That is, alternatives in the complement of A are not possible according to given evidence.
- $0 \le H(A) \le \log_2 |X|$
- H(A) measures the degree of nonspecificity (or imprecision).
- H(A) has been extended to fuzzy sets A via their α-cut representations by the formula

$$H(A) = \int_{\alpha \in [0,1]} \log_2 |{}^{\alpha}A| \, d\alpha.$$

- α-cut of fuzzy set A: ${}^{\alpha}A = \{x \in X \mid A(x) \ge \alpha\}$.

HARTLEY MEASURES ON X×Y

Basic Types

- $H(X \times Y)$ — joint
- $H(X)$, $H(Y)$ — marginal (or simple)
- $H(X|Y)$, $H(Y|X)$ — conditional
- $T_H(X,Y)$ — information transmission

Basic Equations and Inequalities

- $H(X|Y) = H(X \times Y) - H(Y)$
- $H(Y|X) = H(X \times Y) - H(X)$
- $T_H(X,Y) = H(X) + H(Y) - H(X \times Y)$
- $H(X \times Y) \le H(X) + H(Y)$
- $H(X|Y) \le H(X)$ and $H(Y|X) \le H(Y)$

Additivity under Independence

- $H(X \times Y) = H(X) + H(Y)$
- $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$
- $T_H(X,Y) = 0$

SHANNON ENTROPY

$$S(p(x)|x \in X) = -\sum_{x \in X} p(x) \log_2 p(x)$$
$$= -\sum_{x \in X} p(x) \log_2 [1 - \sum_{y \ne x} p(y)]$$

- $Con(x) = \sum_{y \ne x} p(y) \in [0, 1]$ for each $x \in X$ expresses the total conflict (aggregated) between the evidential claim focusing on x and all the other evidential claims expressed by the probability distribution $\langle p(x)|x \in X \rangle$.
- Function $-\log_2[1 - Con(x)]$ is monotone increasing with Con(x) and extend the range of Con(x) from [0, 1] to $[0, \infty)$. Hence, it also expresses the total conflict within any given probability distribution $\langle p(x)|x \in X \rangle$, but in a different scale from Con(x).
- This alternative representation of the total conflict is needed to satisfy the additivity requirement of uncertainty measures.
- Shannon entropy can thus be viewed as a measure of the total conflict among evidential claims associated with a probability distribution.

JOINT, MARGINAL, AND CONDITIONAL UNCERTAINTIES: Basic Formulas

$U(X|Y) = U(X,Y) - U(Y)$
$U(Y|X) = U(X,Y) - U(X)$
$U(X_1, X_2,\cdots, X_n) = U(X_1) + U(X_2|X_1)$
$+ U(X_3|X_1,X_2) + \cdots + U(X_n|X_1, X_2,\cdots, X_{n-1})$

$U(X|Y) \le U(X)$
$U(Y|X) \le U(Y)$
$U(X,Y) \le U(X) + U(Y)$

} The equalities are obtained only in the case of noninteraction.

Information transmission:
$T(X, Y) = U(X) + U(Y) - U(X,Y)$
$T(X, Y) = U(X) - U(X|Y)$
$T(X, Y) = U(Y) - U(Y|X)$

HARTLEY-LIKE MEASURE IN n-DIMENSIONAL EUCLIDEAN SPACE

$$HL(A) = \min_{t\in T}\left\{\log_2\left[\prod_{i=1}^{n}[1+\mu(A_i)] + \mu(A) - \prod_{i=1}^{n}\mu(A_i)\right]\right\}.$$

HARTLEY-LIKE MEASURE

n = 1

$$HL(A) = \log_2[1 + \mu(A)]$$

n = 2

$$HL(A) = \min_{t\in T}\log_2[1 + \mu(A_1) + \mu(A_2) + \mu(A)]$$

GENERALIZED HARTLEY MEASURE IN DST

$$GH(m) = \sum_{A\in F} m(A) \log_2 |A|$$

Conditional forms:

$GH(X|Y) = GH(X,Y) - GH(Y)$
$GH(Y|X) = GH(X,Y) - GH(X)$

Measures of Uncertainty on Finite Sets X

- Hartley measure of nonspecificity in classical possibility theory:

$$H(A) = \log_2|A|\,,\ A \subseteq X.$$

- Shannon measure (entropy) of conflict in classical probability theory:

$$S(p(x) \mid x \in X) = -\sum_{x\in X} p(x) \log_2 p(x).$$

- Generalized Hartley measure in Dempster-Shafer Theory (DST):

$$GH(m) = \sum_{A\subseteq X} m(A) \log_2|A|$$

- All intuitively reasonable candidates for a generalized Shannon measure failed the subadditivity requirement.

- Aggregated total uncertainty in any theory of imprecise probabilities:

$$S^*(D) = \max_{p\in D}\{-\sum_{x\in X} p(x) \log_2 p(x)\}$$

This measure satisfies all the essential requirements for measures of uncertainty, but it is insensitive to changes in evidence.

ENTROPY-LIKE MEASURE IN DST: Attempts

- $C(m) = -\sum_{A\in\mathcal{F}} m(A) \log_2 Bel(A) = -\sum_{A\in\mathcal{F}} m(A) \log_2\left[1 - \sum_{B\neq A} m(B)\right]$

Confusion: Höhle (1982)

- $E(m) = -\sum_{A\in\mathcal{F}} m(A) \log_2 Pl(A) = -\sum_{A\in\mathcal{F}} m(A) \log_2\left[1 - \sum_{A\cap B=\emptyset} m(B)\right]$

Dissonance: Yager (1983)

- $D(m) = -\sum_{A\in\mathcal{F}} m(A) \log_2\left[1 - \sum_{B\in\mathcal{F}} m(B)\frac{|B-A|}{|B|}\right]$

Discord : Klir & Ramer (1990); conjunctive set-valued statements

- $ST(m) = -\sum_{A\in\mathcal{F}} m(A) \log_2\left[1 - \sum_{B\in\mathcal{F}} m(B)\frac{|A-B|}{|A|}\right]$

Strive: Klir & Parviz (1992); disjunctive set-valued statements

DISAGGREGATED UNCERTAINTIES

- Disaggregated total uncertainty with two components:

$$TU = (GH, S^* - GH)$$

- An alternative disaggregated total uncertainty with two components:

$$^aTU = (S^* - S_*, S_*)$$

- Global disaggregated uncertainty that consists of three components:

$$GU = (GH, S^* - S_*, S_*)$$

➢ $S_*(D) = \min_{p\in D}\{-\sum_{x\in X} p(x) \log_2 p(x)\}$

$S^*(D) = \max_{p\in D}\{-\sum_{x\in X} p(x) \log_2 p(x)\}$

$GH(m) = \sum_{A\in F} m(A)\log_2|A|$

INDEPENDENCE OF MARGINAL UNCERTAINTIES VIA THE CLASSICAL PROBABILISTIC INDEPENDENCE APPLIED TO CONVEX SETS OF PROBABILITY DISTRIBUTIONS (STRONG OR CLASSICAL INDEPENDENCE)

- Marginal sets: $X = \{x_i \mid i \in N_n\}$ and $Y = \{y_j \mid j \in N_m\}$
- Joint set: $Z = X \times Y$
- Given: convex sets of marginal probability distributions D_X and D_Y
- Under the assumption of strong independence, the joint set of probability distributions on Z, denoted by D, is defined by applying classical probabilistic independence to D_X and D_Y:

$$D = \{p \mid p(z_{ij}) = p_X(x_i).p_Y(y_j),\ p_X \in D_X,\ p_Y \in D_Y,\ i \in N_n,\ j \in N_m\}$$

- $\underline{\mu}(A) = \inf_{p\in D}\{\sum_{i,j} p(z_{ij}) \mid z_{ij} \in A\},\ \forall A$

Möbius independence (also called mass independence):

$$m(C) = \begin{cases} m_X(A) \cdot m_Y(B) & \text{when } C = A \times B \\ 0 & \text{otherwise.} \end{cases}$$

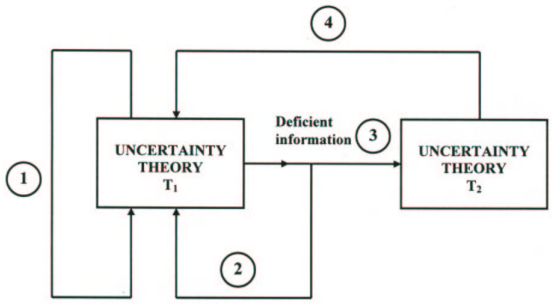
Possibilistic independence:

$$r(x,y) = \min\{r_X(x), r(y)\}, x \in X, y \in Y$$

	GH	S_*	S^*	$S^* - GH$	$S^* - S_*$
Strong independence	No	Yes	Yes	No	Yes
Mass independence	Yes	No	Yes	Yes	No

Principles of Uncertainty

- **Principle of Minimum Uncertainty**
- **Principle of Maximum Uncertainty**
- **Principle of Uncertainty Invariance**
- **Principle of Requisite Generalization**

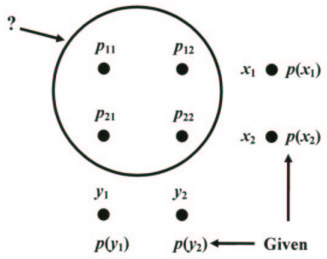


Assumption: Theory T_2 is more general than theory T

1. **Minimum – uncertainty principle**
2. **Maximum – uncertainty principle**
3. **Requisite generalization**
4. **Uncertainty – invariance principle**

THE MARGINAL PROBLEM

Given marginal probabilities (or marginal lower probabilities) what are the joint probabilities (or the joint lower probabilities)?



Example: $p(x_1) = 0.8$, $p(x_2) = 0.2$
$p(y_1) = 0.6$, $p(y_2) = 0.4$

- **Remaining in probability theory: maximum entropy principle**

  $p_{11} = p(x_1) \cdot p(y_1) = 0.48$, $p_{12} = p(x_1) \cdot p(y_2) = 0.32$

  $p_{21} = p(x_2) \cdot p(y_1) = 0.12$, $p_{22} = p(x_2) \cdot p(y_2) = 0.08$.

  $TU = \langle 0, 1.693 \rangle$

- **Requisite generalization: the following convex set of joint probability distributions (epistemologically honest representation of evidence – the marginal probability distributions):**

  $p_{11} \in [\max\{0, p(x_1) + p(y_1) - 1\}, \min\{p(x_1), p(y_1)\}] = [0.4, 0.6]$

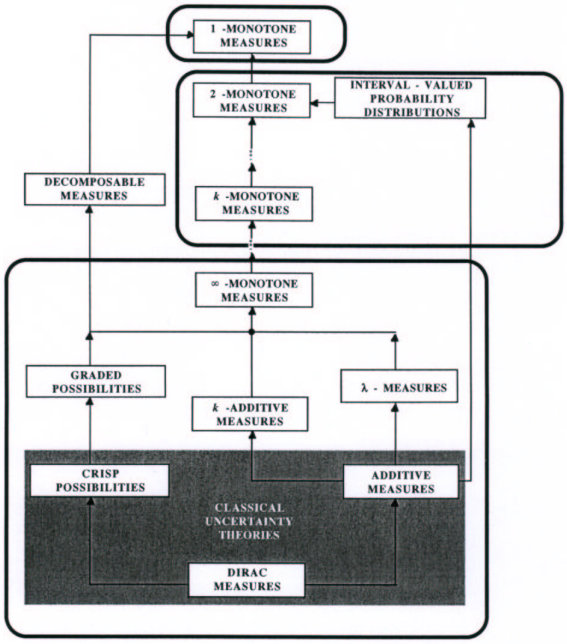
  $p_{12} = p(x_1) - p_{11} = 0.8 - p_{11}$

  $p_{21} = p(y_1) - p_{11} = 0.6 - p_{11}$

  $p_{22} = 1 - p(x_1) - p(y_1) + p_{11} = p_{11} - 0.4$

  $TU = \langle 0.332, 1.683 \rangle$

Uncertainty Theories



Uncertainty and Information: Two Faces of Nondeterministic Systems

George J. Klir

Department of Systems Science & Industrial Engineering
Binghamton University – SUNY
Binghamton, New York 13902, USA
E-mail: gklir@binghamton.edu

Abstract

A research program whose objective is to study uncertainty and information in all their manifestations, primarily within the context of nondeterministic systems, was introduced in the early 1990's under the name "generalized information theory". This research program is based on a two-dimensional expansion of the probabilistic framework of classical information theory. In one dimension, additive probability measures are expanded to various nonadditive measures. In the other dimension, the formalized language of classical set theory is expanded to more expressive formalized languages that are based on fuzzy sets of various types. As in classical information theory, uncertainty is the prinary concept in generalized information theory and information is defined in term
s of uncertainty reduction. The paper introduces principal ideas of generalized information theory and examines the role of generalized information theory in systems science.

Keywords: Nondeterministic systems, information-based uncertainty, uncertainty-based information, generalized information theory, generalized measures, fuzzy sets, imprecise probabilities, measures of uncertainty, principles of uncertainty, systems science.

1. Introduction

One of the insights emerging from systems science is the recognition that knowledge is organized, by and large, in terms of systems of various categories (in the sense of the mathematical theory of categories). Systems in each category are knowledge structures of a particular type. Every particular system is constructed within an *experimental frame*, which consists of a chosen set of variables, their recognized state sets, and a medium within which states of the chosen variables change (such as time, space, or a population). A system based on a chosen experimental frame expresses, in general, some type of *constraint* among the variables, which is characteristic of the category to which the system belongs. This constraint is utilized, in a purposeful way, for restricting states of some variables (output variables) on the basis of known states of other variables (input variables). Some of the most typical purposes for which systems are constructed include prediction, retrodiction, diagnosis, control, decision making, policy making, and planning.

It turns out that the mathematical concept of a *relation* is sufficiently general to express any type of constraint among variables. When for each state of input variables, states of all output variables are uniquely determined from the relation of a given system, the system is called a *deterministic system*; otherwise, it is called a *nondeterministic system*. It is well known that the class of deterministic systems represents only a very small fraction of all conceivable systems. While deterministic systems played an important role in early stages of the history of science, nondeterministic systems are by far more prevalent in contemporary science and technology.

Each nondeterministic system inevitably involves some *uncertainty*, which is associated with the purpose for which the system was constructed. We may thus distinguish predictive uncertainty, retrodictive uncertainty, prescriptive uncertainty, diagnostic uncertainty, decision uncertainty, and the like. The relevant uncertainty (predictive, prescriptive, etc.) must be properly incorporated into the formal description of each nondeterministic system.

There are two classical ways of formalizing uncertainty. In one of them, uncertainty is expressed by a *possibility function* that partitions the set of all relevant states (predictions, diagnoses, etc.) into two subsets. One subset consists of those states that are considered possible; the other one consists of those that, under given evidence, are not possible. In this case, uncertainty is proportional in some way to the size of the set of possible states. Maximum uncertainty is obtained when all relevant states are possible. Minimum uncertainty is obtained when only one of the relevant states is possible. This way of formalizing uncertainty is known in the literature as classical (crisp) *possibility theory*.

The second classical way of formalizing uncertainty is based on *probability function*s, each of which distributes the value 1 to relevant states according to their relative strength of support by given evidence. The maximum uncertainty is obtained when the value is distributed equally to all relevant states. The minimum uncertainty is obtained when the full value of 1 is allocated to one of the relevant states. This way of formalizing uncertainty is known in the literature as classical *probability theory*.

Uncertainty in these two classical theories of uncertainty is viewed as a manifestation of some information deficiency within a given system. It is thus reasonable to refer to it as *information-based uncertainty*. Information, on the other hand, may be viewed in this context as the capacity to reduce relevant uncertainty. It is thus reasonable to refer to it as *uncertainty-based information*. The concepts of uncertainty and information are thus closely connected when dealing with systems. To reduce uncertainty, we need to obtain relevant information. This requires that some action be taken (performing a relevant experiment, conducting a medical test, searching an archive for a relevant historical document, and the like). If an outcome is produced by the action that reduces the uncertainty involved, then the amount of information obtained by the action may be viewed as equivalent to the amount of reduced uncertainty. To measure uncertainty-based information is thus contingent on our ability to measure uncertainty.

Since the mid 20th century, the two classical theories of uncertainty have increasingly been recognized as insufficient for dealing adequately with uncertainty (and uncertainty-based

information) in systems. Various attempts to broaden the classical uncertainty theories were eventually integrated under the name *generalized information theory*.

2. Generalized Information Theory

Generalized Information Theory (GIT) is a research program whose objective is to develop a formal treatment of the interrelated concepts of uncertainty and information in all their varieties; it is a generalization of the two branches of classical information theory, which are based, respectively, on the notion of possibility (crisp) and probability. The aims of GIT were introduced in the early 1990s [Klir, 1991]. A comprehensive and up-to-date coverage of results obtained by research within GIT prior to 2006 is the subject of a recent book [Klir, 2006].

GIT is an outcome of two grand generalizations in mathematics that emerged in the second half of the 20th century. One of them is the generalization of classical measure theory [Halmos. 1950] to the theory of generalized measures by abandoning the requirement of additivity. This generalization was suggested first by Choquet [1953-54]. The second one is the generalization of classical set theory to the theory of fuzzy sets by abandoning the requirement of sharp boundaries between a set and its complement. This very radical generalization in mathematics was suggested first by Zadeh [1965].

A blueprint for the GIT research program is based on a two-dimensional expansion of the classical uncertainty and information theories, which utilizes these two mathematical generalizations. In one dimension, additive probability measures, which are inherent in classical, probability-based information theory, are expanded to various types of nonadditive measures. In the other dimension, the formalized language of classical set theory, within which both of the classical uncertainty theories are formalized, is expanded to more expressive formalized languages that are based on fuzzy sets of various types.

Each possible uncertainty and information theory within the expanded framework of GIT is characterized by choosing a particular formalized language and by expressing relevant uncertainty in this language by a generalized measure of some type. Clearly, the number of distinct theories within the expanded framework grows very rapidly with the number of considered types of formalized languages and the number of considered types of generalized measures. It turns out that this rapidly growing diversity of theories subsumed under the GIT framework is balanced by their unity, which is manifested by their many common properties. The diversity of GIT offers an extensive inventory of distinct theories, each characterized by specific assumptions. This allows us to choose, in any given application context, a theory whose assumptions are in harmony with the application of concern. The unity of GIT, on the other hand, allows us to work within GIT as a whole. That is, it allows us to move from one theory to another as needed.

3. Special Theories Within GIT

All special theories subsumed under GIT are based on various special types of monotone measures. For the sake of simplicity, this paper is restricted to monotone measures defined on the

power set, $P(X)$, of a finite set X. In this restricted case, a *monotone measure* is any set function μ: $P(X) \rightarrow [0, 1]$ that satisfies the following three requirements: (i) $\mu(\varnothing) = 0$; (ii) $\mu(X) = 1$; and (iii) for all $A,B \in P(X)$, if $A \subseteq B$ then $\mu(A) \leq \mu(B)$. Each special theory is then defined by restricting monotone measures via additional requirements. For example, when $\mu(A \cup B) \geq \mu(A) + \mu(B)$ for all $A,B \in P(X)$, measure μ is referred to as a *superadditive measure* and when $\mu(A \cup B) \leq \mu(A) + \mu(B)$ it is called a *subadditive measure*.

Observe that this definition of monotone measures is formulated in terms of classical sets. It can be *fuzzified* – i.e. reformulated in terms of fuzzy sets of some type. Fuzzy sets, in general, are sets whose boundaries are not required to be sharp. That is, they allow us to distinguish degrees (or grades) of membership. Each fuzzy set is fully characterized by a particular membership function $A: D \rightarrow R$. For each $d \in D$, $A(d)$ is viewed as the degree of membership of object d in fuzzy set A or, alternatively, as the degree of compatibility of d with a concept represented by A. Distinct types of fuzzy sets are distinguished by distinct types of classical sets D (domain) and R (range). The most common type of fuzzy sets, referred to in the literature as *standard fuzzy sets*, is based on membership functions $A: X \rightarrow [0, 1]$, where objects of X are not fuzzy sets. When objects of X are fuzzy sets, set A is called a level-2 fuzzy set. When R is the set of subintervals of $[0, 1]$, set A is called an interval-valued fuzzy set; when they are fuzzy subintervals of $[0, 1]$, set A is called a type-2 fuzzy set. Several additional types of fuzzy sets have been recognized in the literature [Klir, 2006], but their coverage is beyond the scope of this paper.

In order to fully develop a particular uncertainty and information theory T requires that issues at each of the following four levels be adequately addressed:

1. Uncertainty functions u of the theory T are characterized via appropriate axioms. Examples of functions u are probability measures and possibility measures of the classical theories.
2. Calculus is developed for dealing with functions u. An example is the calculus of classical probability theory.
3. A justified functional U in theory T is found, which for each particular function u measures the amount of of uncertainty associated with u. When a particular unit of measurement is chosen, functional U is required to be unique. A visible example of functional U is the well-known Shannon entropy in classical probability theory [Shannon, 1948].
4. Functional U is utilized as an abstract measuring instrument for dealing with various problems in which theory T is involved. An example is the use of the Shannon entropy in the principles of maximum and minimum entropy in classical probability theory [Kapur, 1989, Christensen, 1980-81].

Among the many uncertainty theories that are possible within the expanded conceptual framework of GIT, only a few theories have been sufficiently developed so far. By and large, these are theories based on various types of monotone measures, but they are still expressed in the language of classical set theory. Fuzzification of these theories, which can be done in different ways, has been explored only to some degree [Klir, 2006].

4. Theories of Imprecise Probabilities

One important result of research in the area of GIT is that the tremendous diversity of uncertainty theories that emerge from the expanded framework is made tractable due to some key properties that are invariant across the whole spectrum or, at least, within broad classes of uncertainty theories. One such class consists of theories that can be viewed as theories of imprecise probabilities [Walley, 1991].

All theories of imprecise probabilities that are based on classical set theory share some common characteristics. One of them is that evidence within each theory is fully described by a *lower probability function* P_* or, alternatively, by an *upper probability function* P^*. These functions are always regular monotone measures [Wang and Klir, 1992] that are superadditive and subadditive, respectively, and

$$\sum_{x\in X} P_*(\{x\}) \le 1, \quad \sum_{x\in X} P^*(\{x\}) \ge 1.$$

In the various special theories of uncertainty, they possess additional special properties.

When evidence is expressed (at the most general level) in terms of an arbitrary convex set C of probability distribution functions p (often referred to as a c*redal set)* on a finite set X, functions P_* and P^* associated with C are determined for each $A \in P(X)$ by the formulas

$$P_*(A) = \inf_{p\in C} \sum_{x\in A} p(x), \; P^*(A) = \sup_{p\in C} \sum_{x\in A} p(x).$$

Since

$$\sum_{x\in A} p(x) + \sum_{x\notin A} p(x) = 1,$$

for each $p \in C$ and each $A \in P(X)$, it follows that

$$P^*(A) = 1 - P_*(\overline{A}).$$

Due to this property, functions P_* and P^* are called *dual* (or *conjugate*). One of them is sufficient for capturing given evidence; the other one is uniquely determined by the duality equation. It is common to use the lower probability function to capture the evidence.

It is well known that any given lower probability function P_* is uniquely represented by a set-valued function m for which $m(\emptyset) = 0$ and

$$\sum_{A\in P(X)} m(A) = 1.$$

This function is called a *Möbius representation* of P_* when it is obtained for all $A \in P(X)$ via the *Möbius transform*

$$m(A) = \sum_{B\,|\,B\subseteq A} (-1)^{|A-B|} P_*(B).$$

The inverse transform is defined for all $A \in P(X)$ by the formula

$$P_*(A) = \sum_{B \mid B \subseteq A} m(B).$$

It follows directly from the duality equation that

$$P^*(A) = \sum_{B \mid B \cap A \neq \emptyset} m(B).$$

for all $A \in P(X)$.

Assume now that evidence is expressed in terms of a given lower probability function P_*. Then, the set of probability distribution functions that are consistent with P_*, $C(P_*)$, which is always closed and convex, is defined as follows:

$$C(P_*) = \{p(x) \mid x \in X, p(x) \in [0,1], \sum_{x \in X} p(x) = 1, \text{ and } P_*(A) \leq \sum_{x \in A} p(x) \text{ for all } A \in P(X)\}.$$

A well-defined category of theories of imprecise probabilities is based on Choquet capacities of various orders [Choquet, 1953-54]. The most general theory in this category is the theory based on *capacities of order 2*. Here, the lower and upper probabilities, P_* and P^*, are monotone measures which satisfy the inequalities

$$P_*(A \cup B) \geq P_*(A) + P_*(B) - P_*(A \cap B),$$
$$P^*(A \cap B) \leq P^*(A) + P^*(B) - P^*(A \cup B,$$

for all $A, B \in P(X)$. Less general uncertainty theories are then based on *capacities of higher orders*. For each $k > 2$, the lower and upper probabilities, P_* and P^*, satisfy the inequalities

$$P_*(\bigcup_{j=1}^{k} A_j) \geq \sum_{\substack{K \subseteq N_k \\ K \neq \emptyset}} (-1)^{|K|+1} P(\bigcap_{j \in K} A_j),$$

$$P^*(\bigcap_{j=1}^{k} A_j) \leq \sum_{\substack{K \subseteq N_k \\ K \neq \emptyset}} (-1)^{|K|+1} P^*(\bigcup_{j \in K} A_j)$$

for all families of k subsets of X, where $N_k = \{1, 2, \ldots, k\}$. Clearly, if $k' > k$, then the theory based on capacities of order k' is less general than the one based on capacities of order k. The least general of all these theories is the one in which the inequalities are required to hold for all $k \geq 2$ (the underlying capacity is said to be of order ∞). This theory, which was extensively developed by Shafer [1976], is usually referred to as *evidence theory* or *Dempster-Shafer theory* (DST). In this theory, lower and upper probabilities are called *belief* and *plausibility* measures. An important feature of DST is that the Möbius representation of evidence m (usually called a *basic probability assignment function* in this theory) is a nonnegative function ($m(A) \in [0,1]$). DST is thus closely connected with the theory of *random sets* [Molchanov, 2004]. Any set $A \in P(X)$ for which $m(A) > 0$ is often called in DST a *focal element*, and the set of all focal elements with the values assigned to them by function m is called a *body of evidence*. When we work with nested families of focal elements, we obtain a *theory of graded possibilities*, which is a generalization of classical possibility theory [De Cooman, 1997; Klir, 2006].

In another important theory, which is computationally more efficient than DST, lower and upper probabilities P_* and P^* are determined for all sets $A \in P(X)$ by intervals $[l(x), u(x)]$ of probabilities on singletons ($x \in X$). Clearly, $l(x) = P_*\{(x\}) \in [0, 1]$ and $u(x) = P^*(\{x\}) \in [0, 1]$. Each given tuple of probability intervals, $I = \langle [l(x), u(x)] | x \in X \rangle$, is associated with a closed convex set, $C(I)$, of probability distribution functions, p, defined as follows:

$$C(I) = \{p(x) \mid x \in X, p(x) \in [l(x), u(x)], \sum_{x \in X} p(x) = 1\}.$$

Sets defined in this way are clearly special credal sets. Their special feature is that they always form an $(n-1)$–dimensional polyhedron, where $n = |X|$. In general, the polyhedron may have c extreme points (vertices), where

$$n \le c \le n(n-1),$$

and each probability distribution function contained in the set can be expressed as a linear combination of these extreme points [De Campos at al., 1994].

A given tuple I of probability intervals may be such that some combinations of values taken from the intervals do not correspond to any probability distribution function. This indicates that the intervals are unnecessarily broad. To avoid this deficiency, the concept of reachability was introduced in the theory.

A given tuple I is called *reachable* (or *feasible*) if and only if for each $x \in X$ and every value $v(x) \in [l(x), u(y)]$ there exists a probability distribution function p for which $p(x) = v(x)$. The reachability of any given tuple I can be easily checked: the tuple is reachable if and only if it passes the following tests:

(a) $\sum_{x \in X} l(x) + u(y) - l(y) \le 1$ for all $y \in X$;

(b) $\sum_{x \in X} u(l) - l(y) + u(y) \ge 1$ for all $y \in X$.

If I is not reachable, it can be easily converted to the tuple $I' = \langle [l'(x), u'(y)] \mid x \in X \rangle$ of reachable intervals by the formulas

$$l'(x) = \max\{l(x), 1 - \sum_{y \ne x} u(y)\},$$

$$u'(\{x\}) = \min\{u(x), 1 - \sum_{y \ne x} l(y)\}$$

for all $x \in X$.

Given a reachable tuple I of probability intervals, the lower and upper probabilities are determined for each $A \in P(X)$ by the formulas

$$P_*(A) = \max\{\sum_{x\in A} l(x),\ 1 - \sum_{x\notin A} u(x)\},$$

$$P^*(A) = \min\{\sum_{x\in A} u(x),\ 1 - \sum_{x\notin A} l(x)\}.$$

It is known that the theory based on reachable probability intervals and Dempster-Shafer theory are not comparable in terms of their generalities. However, they both are subsumed under a theory based on Choquet capacities of order 2 (see Figure 1).

Although Choquet capacities of order 2 do not capture all credal sets, they subsume all the other special uncertainty theories that are examined in this paper (Figure 1). They are thus quite general. Their significance is that they are computationally easier to handle than arbitrary credal sets. In particular, it is easier to compute $C(P_*)$ when P_* is a Choquet capacity of order 2.

Let $X = \{x_1, x_2, \ldots, x_n\}$ and let $\sigma = (\sigma(x_1), \sigma(x_2), \ldots, \sigma(x_n))$ denote a permutation by which elements of X are reordered. Then, it was established [De Campos and Bolaños, 1989] that for any given Choquet capacity of order 2, $C(P_*)$ is determined by its extreme points, which are probability distributions p_σ computed as follows:

$p_\sigma(\sigma(x_1)) = P_*(\{\sigma(x_1)\})$,

$p_\sigma(\sigma(x_2)) = P_*(\{\sigma(x_1), \sigma(x_1)\} - P_*(\{\sigma(x_1)\})$

.....

$p_\sigma(\sigma(x_{n-1})) = P_*(\{\sigma(x_1),\ldots, \sigma(x_{n-1})\} - P_*(\{\sigma(x_1), \ldots., (\{\sigma(x_{n-2})\})$,

$p_\sigma(\sigma(x_{n-1})) = P_*(\{\sigma(x_1),\ldots, \sigma(x_n)\} - P_*(\{\sigma(x_1), \ldots., (\{\sigma(x_{n-1})\})$.

Each permutation defines an extreme point of $C(P_*)$, but different permutations can give rise to the same point. The set of distinct probability distributions p_σ is often called an *interaction representation of* P_* [Grabisch, 2000].

5. Measures of Uncertainty: An Overview

In each given uncertainty theory, a considered functional U for measuring uncertainty must satisfy several intuitively essential axiomatic requirements to be acceptable as an uncertainty measure in the theory. Specific mathematical formulation of each of these requirements depends on the uncertainty theory involved. However, the requirements can be described informally, independent of the various uncertainty calculi, in the following generic form:

1. *Subadditivity* – the amount of uncertainty embedded in a joint uncertainty function (defined on a Cartesian product) must be smaller than or equal to the sum of the amounts of uncertainty embedded in the associated marginal uncertainty functions.
2. *Additivity* – given two universal sets and marginal uncertainty functions defined on them that are independent, the sum of the amounts of uncertainty embedded in these functions must be equal to the amount of uncertainty embedded in the associated joint uncertainty function (defined on the Cartesian product of the two universal sets).

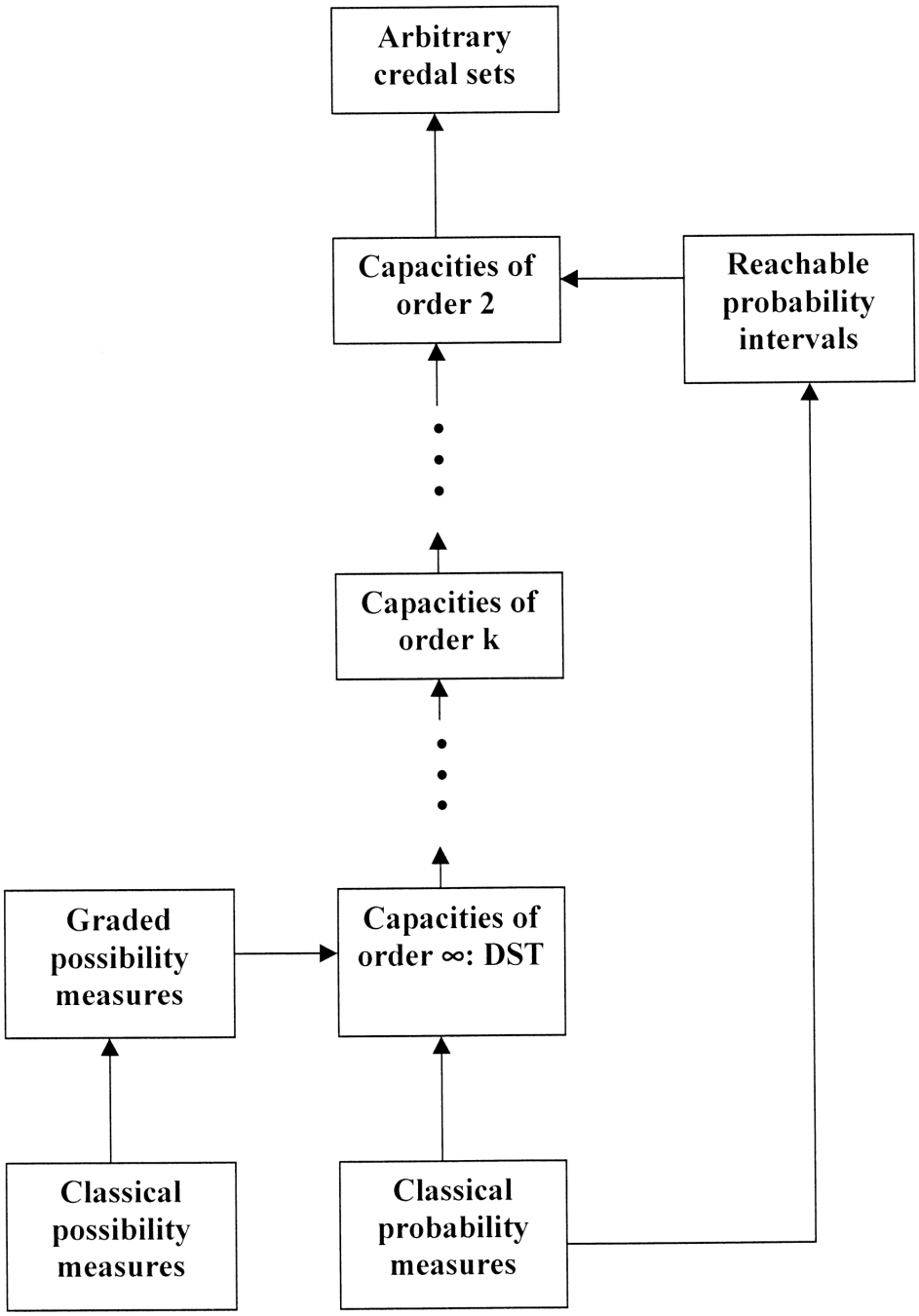


Figure 1. Ordering of principal monotone measures used for representing imprecise probabilities by their levels of generality.

3. *Continuity* – an uncertainty measure must be a continuous functional.

4. *Range* – the range of the amounts of uncertainty embedded in uncertainty functions of a given uncertainty theory must be a closed interval $[0, r]$ of real numbers, where 0

corresponds to full certainty and the value r depends on the chosen measurement unit and the size of the universal set involved.

5. *Expansibility* – expanding the universal set by alternatives that are not supported by given evidence must not affect the amount of uncertainty.
6. *Branching/Consistency* – when the amount of uncertainty embedded in uncertainty functions of a given uncertainty theory can be computed in several distinct ways, all of which conform to the calculus of the theory, the results must be the same (consistent).
7. *Monotonicity* – when uncertainty functions in a given uncertainty theory can be ordered in the same way as we can intuitively order the amount of evidence, the uncertainty measure must preserve this ordering.
8. *Measurement unit* – a suitable measurement unit is defined by specifying what the amount of uncertainty should be for a particular (and usually very simple) uncertainty function.

When distinct types of uncertainty coexist in a given uncertainty theory, as is common in the various non-classical uncertainty theories, it is not necessary that these requirements be satisfied by each uncertainty type. However, they must be satisfied by a measure that appropriately aggregates measures of the individual uncertainty types.

The strongest justification of a functional as a meaningful uncertainty measure of a considered type of uncertainty in a given uncertainty theory is obtained when we can prove that it is the only functional that satisfies all the relevant axiomatic requirements formulated in the calculus of the theory.

It is well established that uncertainty in classical possibility theory is quantified by the *Hartley measure* [Hartley, 1928]. For each nonempty and finite set $A \subseteq X$ of possible alternatives, the Hartley measure $H(A)$, is defined by the formula

$$H(A) = \log_2 |A|,$$

where $|A|$ denotes the cardinality of A. Since $H(A) = 1$ when $|A| = 2$, H measures uncertainty in bits. The type of uncertainty measured by H is usually called *nonspecificity*.

In classical probability theory, a justifiable measure of uncertainty was derived by Shannon [1948]. This measure, which is usually referred to as *Shannon entropy* and denoted by S, is defined for each given probability distribution function p on a finite set X by the formula

$$S(p) = -\sum_{x \in X} p(x) \log_2 p(x).$$

Since $S(p) = 1$ when $|X| = 2$ and $p(x) = 1\text{–}p(x) = 0.5$, S measures uncertainty in bits. However, the type of uncertainty measured by the Shannon entropy is different from the uncertainty type quantified by the Hartley measure; it is well captured by the term *conflict*.

When the classical uncertainty theories are generalized, both types of uncertainty coexist. This required that the Hartley measure and Shannon entropy be properly generalized in the various theories.

The Hartley measure was first generalized for graded possibilities by Higashi and Klir [1983] and later to Dempster-Shafer theory by Dubois and Prade [1985]. Its generalized form, *GH*, is defined in terms of the Möbius representation, *m*, by the formula

$$GH(m) = \sum_{A \subseteq X} m(A) \log_2 |A|$$

The uniqueness of this generalized Hartley measure *GH* was proven for graded possibilities by Klir and Mariano [1987 and for DST by Ramer [1987].

Efforts to generalize the Shannon entropy to DST were less successful. Although several intuitively promising candidates for the generalized Shannon measure, *GS*, were proposed in the literature in the 1980s and early 1990, each of them was found to violate the essential property of subadditivity. This would have been acceptable if subadditivity were satisfied for the sum *GH* + *GS*. Unfortunately, this was not the case for any of the proposed measures. A digest of these frustrating efforts is given in Klir [2006].

In the early 1990s, the unsuccessful attempts to generalize the Shannon entropy in DST were replaced with attempts to find an aggregated measure of both types of uncertainty. An aggregate measure that satisfies all the required properties (additivity, subadditivity, monotonicity, proper range, etc.) was eventually found around the mid 1990s by several authors (see Klir [2006] for details). This aggregate uncertainty measure is a functional S^* that for each belief function *Bel* in DST is defined as follows:

$$S^*(\text{Bel}) = \max_{P_{Bel}} \{-\textstyle\sum_{x \in X} p(x) \log_2 p(x)\},$$

where the maximum is taken over the set P_{Bel} of all probability distribution functions *p* that dominate the given function *Bel* (i.e., $Bel(A) \leq \sum_{x \in A} p(x)$ for all $A \subseteq X$). This functional can be readily generalized to any given convex set of probability distributions, as was shown by Abellán and Moral [(2003]. Useful algorithms for computing S^* were developed by various authors for DST, reachable interval-valued probability distributions, and the theory based on Choquet capacities of order 2 [Klir. 2006].

Although the functional S^* is acceptable on mathematical grounds as an aggregate measure of uncertainty in any uncertainty theory where evidence can be represented in terms of arbitrary convex sets of probability distributions, it is highly insensitive to changes in evidence due to its aggregated nature and, moreover, it does not show explicitly measures of the two coexisting types of uncertainty — nonspecificity and conflict. It is thus desirable to disaggregate it. Clearly, $S^* = GH + GS$, where *GH* and *GS* denote, respectively, a generalized Hartley measure (measuring nonspecificity) and a generalized Shannon entropy (measuring conflict). Since S^* and *GH* are well established, at least in DST, it is suggestive to define GS indirectly as the difference $S^* - GH$, providing that it is nonnegative. It was proven by Smith [2000] that $S^* - GH \geq 0$ and, hence, it is meaningful to take $GS = S^* - GH$ as the generalized Shannon entropy. Then, the disaggregated total uncertainty measure, *TU*, is defined as the pair

$$TU = \langle GH, GS \rangle,$$

where $GS = S^* - GH$. Now, it is guaranteed that $GH + GS$ satisfies all the required mathematical properties (since $GH + GS = S^*$) and it does not matter whether each of the two components of TU satisfies them as well. This is important since subadditivity of GH is not guaranteed beyond DST.

The idea of disaggregating S^* into two components, measures of nonspecificity and conflict, opened new possibilities. One of them is based on the recognition that the following two functionals can be defined for each credal set C:

$$S^*(C) = \max_{p \in C} \{-\sum_{x \in X} p(x) \log_2 p(x)\},$$

$$S_*(C) = \min_{p \in C} \{-\sum_{x \in X} p(x) \log_2 p(x)\}.$$

The significance of these functionals and their difference, $S^* - S_*$, for capturing uncertainty associated with convex sets of probability distributions was first discussed by Kapur [1994] and Kapur et al. [1995]. Their role for disaggregating S^* was suggested by Smith [2000] and Klir and Smith [2001]. More recently, Abellán and Moral [2005] further investigated properties of the difference $S^* - S_*$ and described an algorithm for calculating the value of S_*, which is applicable to any convex set of probability distributions whose lower probability function is a Choquet capacity of order 2. They suggested that it is reasonable to view this difference as an alternative measure of nonspecificity. That is, they suggested defining a measure of nonspecificity, N, for each credal set C of probability distributions by the formula

$$N(C) = S^*(C) - S_*(C).$$

They also showed that functional N possesses the following properties:

1. $N(C) \in [0, \log_2 |X|]$, where X denotes the set of all alternatives (elementary events) on which the probability distributions in C are defined: $N(C) = 0$ when C consists of a single probability distribution; $N(C) = \log_2 |X|$ when C consists of all probability distributions that can be defined on X (total ignorance expressed by vacuous probabilities).
2. N is continuous.
3. N is additive.

These properties, which every measure of nonspecificity must possess, motivated the suggestion that this functional may be viewed as a measure of nonspecificity. Unfortunately, contrary to the generalized Hartley measure, functional N violates the essential requirement of subadditivity in virtually any uncertainty theory, including DST. This means that N is not acceptable *alone* as a measure of nonspecificity. However, when considered as one component of a disaggregated total uncertainty measure, then the lack of subadditivity of the individual components is of no consequence. It only matters that the aggregated uncertainty S^* satisfies all the essential requirements, including subadditivity. This suggests defining an alternative disaggregated total uncertainty, ${}^aTU$, as the pair

$${}^aTU(C) = \langle S^*(C) - S_*(C), S_*(C) \rangle.$$

Observe that the first component of ${}^{a}TU$ is the alternative nonspecificity measure N, while the second component, S_*, is a generalized Shannon measure (a general measure of conflict). When the two components are aggregated, we obtain S^* and, clearly, this functional satisfies all the essential mathematical requirements. Hence, even though neither of the components of ${}^{a}TU$ is subadditive, this does not matter since the aggregated uncertainty S^* is subadditive.

It is interesting to observe that the functional S_* has often been considered as one of the candidates for the generalized Shannon entropy. It was dismissed since it is not subadditive, and neither it is subadditive when aggregated with the generalized Hartley measure GH. However, it is perfectly justifiable when aggregated with the alternative measure of nonspecificity N. In fact, some of the other candidates considered for the generalized Shannon entropy could now be considered on similar grounds, although the functional S_* seems to be better justified than its competitors not only by its properties, but also by its behavior and its applicability to all credal sets. Nevertheless, viewing the measure of nonspecificity, in general, as the difference of the aggregate uncertainty S^* and the generalized Shannon entropy GS, opens a new area of research, whose purpose is to compare the various candidates for GS with the functional S_*.

6. The Role of GIT in Systems Science

Uncertainty-based information, which is the subject of investigation in GIT, does not capture the rich notion of information in human communication and cognition, but it is essential for dealing with systems. It allows us, for example, to examine informativeness of comparable systems (constructed within the same experimental frame) with respect to answers given by the systems to various questions regarding predictions, retrodictions, diagnoses, etc. When asked a relevant question, every system gives some answer. If the answer is not unique, it involves some uncertainty. This uncertainty is generally smaller than its counterpart based solely on the experimental frame of the system. The difference between these uncertainties expresses then the degree of informativeness of the system with respect to the given question.

In systems science, the various functionals U for measuring uncertainty, which are surveyed in Sec. 5, are particularly useful in their conditional forms. While functionals U are applied to different types of uncertainty functions u in different uncertainty theories, it turns out that the relationship between their conditional, joint and marginal forms is invariant with respect to the type of uncertainty function u involved. To describe this relationship generically, it is useful (and a common practice in the literature) to identify only sets on which functions u are defined rather than the functions themselves. For example, if a joint uncertainty function u is defined on $X\times Y$, we write $U(X\times Y)$ instead of $U(u)$. Similarly, we write $U(X)$ *and* $U(Y)$ for marginal uncertainties instead of $U(u_X)$ and $U(u_Y)$. Clearly, $X\times Y$, X, and Y are not arguments for computing the respective type of uncertainty U; they are just indicators of the sets on which some type of uncertainty function u is defined. Using this convenient generic notation, conditional uncertainty, denoted as $U(X|Y)$ and $U(Y|X)$, are then determined in all theories of uncertainty by the following equations:

$$U(X|Y) = U(X\times Y) - U(Y),$$

$$U(Y|X) = U(X\times Y) - U(X).$$

In the context of systems science, sets X and Y in these equations may be Cartesian products of several sets, each representing overall states of some chosen subsets of variables of a given system. If $U(X|Y) = U(X)$ and $U(Y|X) = U(Y)$, then the chosen subsets of variables do not interact. In all other cases, $U(X|Y) < U(X)$ and $U(Y|X) < U(Y)$, which means that the two subsets of variables do interact. The strength of this interaction is measured by another functional, T, which is defined by the equation

$$T(X,Y) = U(X) + U(Y) - U(X \times Y).$$

This functional, which is usually referred to as *information transmission*, can also be expressed in terms of differences between the marginal uncertainties and their conditional counterparts:

$$T(X,Y) = U(X) - U(X|Y),$$

$$T(X,Y) = U(Y) - U(Y|X).$$

The capabilities of measuring conditional uncertainties and information transmissions for any groups of variables of a given systems are essential for dealing with nondeterministic systems. These capabilities have not been broadly utilized as yet, even though their significance was recognized a long time ago in terms of the classical uncertainty and information theories by Ashby [1965, 1970, 1972], Conant [1976] and others. It is fair to say that the classical theories are rather restrictive. Results emerging from GIT open many new possibilities and allow us to deal with nondeterministic systems in a broader and more natural way.

Although the utility of the various functionals for measuring uncertainty and information is as broad as the utility of any measuring instrument, their role is particularly significant in the following four *principles of uncertainty*:

1. principle of minimum uncertainty;
2. principle of maximum uncertainty;
3. principle of requisite generalization;
4. principle of uncertainty invariance.

In general, these are epistemologically based prescriptive procedures that address such methodological issues involving nondeterministic systems that cannot be resolved by using calculi of the individual uncertainty theories. Due to the connection between information-based uncertainty and uncertainty-based information, these principles also can be interpreted as *principles of information.* Their role is illustrated in Figure 2, where T_2 denotes an uncertainty theory that is assumed to be more general than uncertainty theory T_1.

The *principle of minimum uncertainty* is basically an arbitration principle. It facilitates the selection of the most informative systems from solution sets obtained in problems of systems simplification, conflict-resolution, and the like. The *principle of maximum uncertainty* facilitates ampliative reasoning in dealing with systems problems. This is reasoning in which conclusions are not entailed in the given premises. A typical example is the problem of identifying an overall system from some of its subsystems. Using common sense, the principle may be expressed as follows: in any ampliative inference, use all available information, but make sure that no additional information (unsupported by the given evidence) is unwittingly added. Employing the

connection between information and uncertainty, this definition can be reformulated in terms of uncertainty: any conclusion resulting from ampliative inference should maximize the relevant uncertainty within the constraints representing the premises. This principle guarantees that we fully recognize our ignorance when we attempt to make inferences that are beyond the information contained in the given premises and, at the same time, that we utilize all this information.

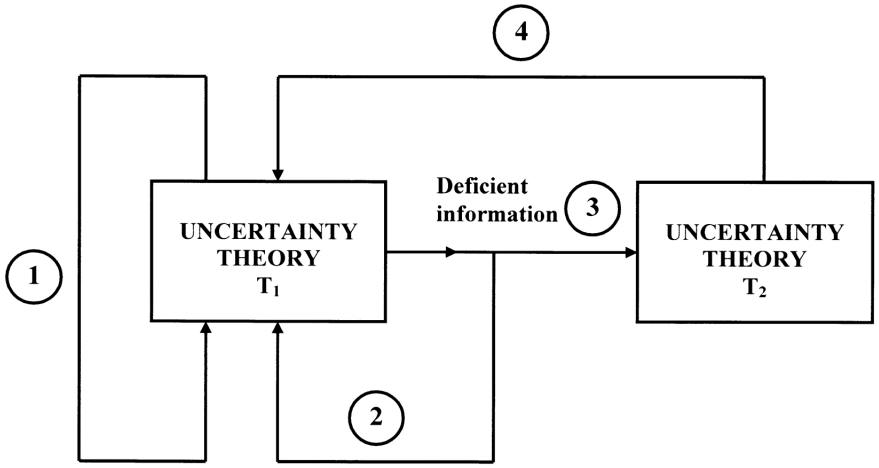


Figure 2. An overview of principles of uncertainty: 1. principle of minimum uncertainty; 2. principle of maximum uncertainty; 3. principle of requisite generalization; 4. principle of uncertainty invariance.

Principles of minimum and maximum uncertainty are fundamentally different from the other principles of uncertainty. While the former are applied within one particular uncertainty theory, the latter involve transitions from one theory to another (Figure 2). Principles of minimum and maximum uncertainty are well developed in classical, probability-based uncertainty theory, where they are called *principles of minimum and maximum entropy* [Christensen, 1980-81, Kapur, 1989].

The *principle of requisite generalization* is based on the assumption that we work within GIT as a whole. According to this principle, we should not a priori commit to any particular uncertainty theory. Our choice should be determined by the nature of the problem we deal with. The chosen theory should be sufficiently general to allow us to capture fully our ignorance. Moreover, when the chosen theory becomes incapable of expressing uncertainty resulting from deficient information at some problem-solving stage, we should move to a more general theory that has the capability of expressing the given uncertainty (Figure 2).

The *principle of uncertainty invariance* (also called the *principle of information preservation*) was introduced in GIT to facilitate meaningful transformations between various uncertainty theories. According to this principle, the amount of uncertainty (and the associated information) should be preserved in each such transformation. The primary use of this principle is to

approximate in a meaningful way a formalization of uncertainty in one theory by its formalization in another, less general theory (Figure 2).

REFERENCES

Abellán, J. and Moral, S. [2003], "Maximum of entropy for credal sets". *Intern. J. of Uncertainty*, *Fuzziness and Knowledge-Based Systems*, **11**(5), pp. 587-597.

Abellán, J. and Moral, S. [2005], "Difference of entropies as a non-specificity function on credal sets". *Intern. J. of General Systems*, **34**(3), pp. 201-214.

Ashby, W. R. [1965], "Measuring the internal informational exchange in a system". *Cybernetica,* **8**(1), pp. 5-22.

Ashby, W. R. [1970], "Information flows within co-ordinated systems". In: Rose, J. (ed.), *Progress in Cybernetics*, Gordon and Breach, London, pp. 57-64.

Ashby, W.R. [1972], "Systems and their information measures". In: Klir, G. J. (ed.), *Trends in General Systems Theory*, Wiley-Interscience, New York, pp. 78-97.

Choquet, G. [1953-54], "Theory of capacities". *Annales de L'Institut Fourier*, **5**, pp. 131-295.

Christensen, R. [1980-81], *Entropy Minimax Sourcebook*. Entropy Ltd., Lincoln, MA.

Conant, R. C. [1976], "Laws of information which govern systems". *IEEE Trans. on Systems, Man, and Cybernetics,* **SMC-6**(4), pp. 240-255.

De Campos, L. M. and Bolaños, M. J. [1989], "Representation of fuzzy measures through probabilities". *Fuzzy Sets and Systems* **31**(1), pp. 23-36.

De Cooman, G. [1997], "Possibility theory". *Inter. J. of General Systems*, **25**(4), pp. 291-371.

Dubois, D. and Prade, H. [1985], "A note on measures of specificity for fuzzy sets". *Intern. J. of General Systems*, **10**(4), pp. 279-283.

Grabisch, M. [2000], "The interaction representations of fuzzy measures on finite spaces, *k*-additive measures: a survey". In: Grabisch, M. et al. (eds.), *Fuzzy Measures and Integrals: Theory and Applications*. Springer-Verlag, New York, pp. 70-93.

Halmos, P. R. [1950], *Measure Theory.* D. Van Nostrand, Princeton, NJ.

Hartley, R. V. L. [1928], "Transmission of information". *The Bell System Technical Journal*, **7**(3), pp. 535-563.

Higashi, M. and Klir, G. J. [1983], "Measures of uncertainty and information based on possibility distributions". *Intern. J. of General Systems*, **9**(1), pp. 43-58.

Kapur, J. N. [1989], *Maximum Entropy Models in Science and Engineering.* John Wiley, New York.

Kapur, J. N. [1994], *Measures of Information and Their Applications.* John Wiley, New York.

Kapur, J. N., Baciu, G., and Kesavan, H. K. [1995], "The MinMax information measure". *Intern. J. of Systems Science,* **26**(1), pp. 1-12.

Klir, G. J. [1991], "Generalized information theory". *Fuzzy Sets and Systems*, **40**(10), pp. 127-142.

Klir, G. J. [2006], *Uncertainty and Information: Foundations of Generalized Information Theory.* Wiley-Interscience, Hoboken, NJ.

Klir, G. J. and Mariano, M. [1987], "On the uniqueness of possibilistic measure of uncertainty and information", *Fuzzy Sets and Systems*, **24**(2), pp. 197-219.

Klir, G. J. and Smith, R. M. [2001], "On measuring uncertainty and uncertainty-based information: recent developments". *Annals of Mathematics and Artificial Intelligence,* **32**(1), pp. 5-33.

Molchanov, I. [2004], *Theory of Random Sets.* Springer, New York.

Ramer, A. [1987], 'Uniqueness of information measure in the theory of evidence". *Fuzzy Sets and Systems*, **24**(2), pp. 183-196.

Shafer, G. [1976], *A Mathematical Theory of Evidence.* Princeton Univ. Press, Princeton, NJ.

Shannon, C. E. [1948], "The mathematical theory of communication". *The Bell System Technical Journal*, **27**(3&4), pp. 379-423, 623-656.

Smith, R. M. [2000], *Generalized Information Theory: Resolving Some Old Questions and Opening Some New Ones.* Ph.D. Dissertation, Binghamton University – SUNY.

Walley, P. [1991], *Statistical Reasoning With Imprecise Probabilities.* Chapman and Hall, London.

Wang, Z. and Klir, G. J. [1992], *Fuzzy Measure Theory.* Plenum Press, New York.

Zadeh, L. [1965], "Fuzzy sets". *Information and Control,* **8**(3), pp. 338-353.

[**Published in**: *IIAS Trans. on Systems Research and Cybernetics,* **VI**(1), 2006, pp. 1-17}