Electronic version of an article published in the *International Journal of Information Technology & Decision Making,* volume 3, issue 1, 2004, pages 101-108
DOI: 10.1142/S0219622004000994

Decision Analysis using IC-Bags

Kankana Chakrabarty
School of Mathematics, Statistics & Computer Science
The University of New England
Armidale-2351
New South Wales
Australia.

E-mail: kankanac@turing.une.edu.au

Abstract

In this paper, we discuss the notion of IC-Bags as introduced by the author, some issues related to IC-Bag based systems and their application in decision analysis. Sometimes it is observed that the semi-structured or unstructured nature of the problems addressed by expert systems can be perceived as the cause for the inability to develop precise requirement specifications. IC-Bags can serve as a tool for building rule-based decision analysis systems which can deal with situations where the count of the objects are not necessarily fixed. We discuss the semistructured specifications involving maximum and minimum numbers of applications related to the problem-oriented methodology for the integrated problems involving IC-Bags. In this context, the total bias factor and the confidence level of the knowledge base are considered.

1. Introduction

Bag structure, as introduced by Yager[7], points to a multitude of application areas which deals with object descriptions for collections where the redundancy of the objects play the key role in describing the attribute level specifications for some rule-based systems. It has been noticed that in some uncertain situations where the count of elements are not found to be fixed, but rather represented in the form of intervals of positive integers, then the notion of IC-Bags as introduced by the author[1], are found to be quite useful. While developing the alternatives and grading the feasibility of the possible alternatives in the design phase of decision making, we come across some complex structures, which can be studied and modelled by using the concept of IC-Bags. Consequently, some databases may require the IC-Bag representation.

The validation of expert system techniques is an important issue related to present quality assurance standards and specifications. Sometimes it is observed that expert systems are required to have more more problem-orientated behaviour than process-orientated behaviour and hence they can be easily influenced by the changing requirement specifications. Often, the semi-structured or unstructured nature of the problems addressed by the expert systems can be perceived as the cause for the inability to develop precise requirement specifications and hence IC-Bags can serve as a tool for building rule-based decision analysis systems. These systems can deal with situations where the count of the objects are not necessarily fixed and they can vary over an interval of non-negative integers. In this paper, we discuss the issues related to the verification and validation techniques associated with different modules of this type of systems. This will include the situations where the hesitation factor associated with the count function can play a major role in defining the knowledge-base and the requirement specifications. The analysis of the personal bias factor associated with human judgement patterns are also included and discussed.

2. The Notion of IC-Bags

This section presents a brief overview of the notion of IC-Bags as introduced by the author in [1]. Some IC-Bag operations [1] are also discussed.

Definition of IC-Bag

For any non empty set Ω, an IC bag β drawn from Ω can be characterized by a pair of functions C_l^β and C_u^β such that

$$C_l^\beta : \Omega \longrightarrow N \quad and \quad C_u^\beta : \Omega \longrightarrow N$$

and $C_l^\beta(x) \leq C_u^\beta(x) \quad \forall x \in \Omega$, where N represents the set of non-negative integers.

If $\Omega = (x_1, x_2, ..., x_n)$, then an IC-Bag β drawn from Ω is represented as

$$\beta = \{x_i/(C_l^\beta(x_i), C_u^\beta(x_i))\}$$

where $i = 1, 2, ...n$.

We call $C_l^\beta(x_i)$ the minimum count for x_i in β and $C_u^\beta(x_i)$ the maximum count for x_i in β.

If for each $x \in \Omega, C_l^\beta(x) = C_u^\beta(x)$, then β represents a crisp bag. Clearly the notion of IC-Bag is a generalization of the notion of Bag [7]. This generalization is extremely useful when object counts in a collection are not fixed and can vary over an interval.

Support Sets of IC-Bags

For any IC-Bag β drawn from Ω, a subset $\sigma(\beta)$ of Ω is called the support set of the IC-Bag β if $\forall x \in \Omega$,

$$C_l^\beta(x) > 0 \Longrightarrow x \in \sigma(\beta),$$
$$C_l^\beta(x) = 0 \Longrightarrow x \notin \sigma(\beta).$$

There exists a possibility that different IC bags can have the same set as their support set.

Equality Types in IC-Bags

Two IC-Bags β_1 and β_2 drawn from Ω are equal if and only if $\forall x \in \Omega$,

$$C_l^{\beta_1}(x) = C_l^{\beta_2}(x), C_u^{\beta_1}(x) = C_u^{\beta_2}(x).$$

If $C_l^{\beta_1}(x) = C_l^{\beta_2}(x)\ \forall x \in \Omega$, but $C_u^{\beta_1}(x) \neq C_u^{\beta_2}(x)\ \forall x \in \Omega$, then β_1 is said to be l-equal to β_2.

If $C_u^{\beta_1}(x) = C_u^{\beta_2}(x)\ \forall x \in \Omega$, but $C_l^{\beta_1}(x) \neq C_l^{\beta_2}(x)\ \forall x \in \Omega$, then β_1 is said to be u-equal to β_2.

Inclusion Types in IC-Bags

The different inclusion types in IC-Bags as shown below are important for data manipulation when object counts are not fixed.

For any two IC-Bags β_1 and β_2 drawn from Ω:

- β_1 is called an IC sub bag of β_2 if $\forall x \in \Omega$, $C_l^{\beta_1}(x) \leq C_l^{\beta_2}(x), C_u^{\beta_1}(x) \leq C_u^{\beta_2}(x)$.
- β_1 is called an l-IC sub bag of β_2 if $\forall x \in \Omega$, $C_l^{\beta_1}(x) \leq C_l^{\beta_2}(x), C_u^{\beta_1}(x) > C_u^{\beta_2}(x)$.
- β_1 is called an u-IC sub bag of $\beta_2 i$ if $\forall x \in \Omega$, $C_l^{\beta_1}(x) > C_l^{\beta_2}(x), C_u^{\beta_1}(x) \leq C_u^{\beta_2}(x)$.

Addition and Removal in IC-Bags

The addition of two IC-Bags β_1 and β_2 drawn from Ω results in the IC-Bag $\beta = \beta_1 \oplus \beta_2$ such that $\forall x \in \Omega$,

$$C_l^{\beta_1 \oplus \beta_2}(x) = C_l^{\beta_1}(x) + C_l^{\beta_2}(x),$$
$$C_u^{\beta_1 \oplus \beta_2}(x) = C_u^{\beta_1}(x) + C_u^{\beta_2}(x).$$

The removal of the IC-Bag β_2 from the IC-Bag β_1 results in the IC-Bag $\beta_1 \ominus \beta_2$ such that $\forall x \in \Omega$,

$$C_l^{\beta_1 \ominus \beta_2}(x) = max[C_l^{\beta_1}(x) - C_l^{\beta_2}(x), 0],$$
$$C_u^{\beta_1 \ominus \beta_2}(x) = max[C_u^{\beta_1}(x) - C_u^{\beta_2}(x), 0].$$

The Null IC-Bag

The IC-Bag β drawn from Ω is called a null IC bag if $\forall x \in \Omega$,

$$C_l^{\beta}(x) = C_u^{\beta}(x) = 0.$$

A null IC bag actually represents the empty bag.

Peak Element Types and Peak Values of IC-Bags

For any IC-Bag β drawn from Ω, $\max_{x\in\Omega} C_l^{\beta}(x)$ and $\max_{x\in\Omega} C_u^{\beta}(x)$ are respectively called the l-peak and the u-peak values of β.

The elements ξ_1 and ξ_2 of Ω satisfying $C_l^{\beta}(\xi_1) = \max_{x\in\Omega} C_l^{\beta}(x)$, and $C_u^{\beta}(\xi_2) = \max_{x\in\Omega} C_u^{\beta}(x)$ are called the l-peak and u-peak elements of β respectively.

If for any $\zeta \in \Omega$, $C_l^{\beta}(\zeta) = \max_{x\in\Omega} C_l^{\beta}(x), C_u^{\beta}(\zeta) = \max_{x\in\Omega} C_u^{\beta}(x)$, then ζ is called the lu-peak element of β.

Selection from IC Bags

Let β be an IC-Bag drawn from Ω and let ω be any subset of Ω. Then a new IC-Bag $\beta \odot \omega$ can be formed with those elements of the IC bag β which belongs to ω such that $\forall x \in \Omega$,

$$C_l^{\beta \odot \omega}(x) = C_l^{\beta}(x) \cdot \mu_\omega(x)$$
$$C_u^{\beta \odot \omega}(x) = C_u^{\beta}(x) \cdot \mu_\omega(x)$$

where μ_ω represents the characteristic function of the set ω and '$\cdot$' indicates the product.

4. Semistructured Specifications and IC-Bags

We consider a rule base and the concerning criteria for ensuring the validity range regarding the derived and/or restricted interval frequencies of application for each particular rule that is applied in case of solving the different modules of an integrated semistructured decision analysis problem. Each of these problems includes a finite set of parallel modules.

We assume the fact that each of the concerned rules are required by the parallel modules M_1, M_2, and M_3 of a semistructured problem specification. Associated with each module M_i, for each rule $R_i(i = 1, 2, ...n)$ in the rule base two non-negative integers R_i^l and R_i^u ($R_i^l \leq R_i^u$, for each $i = 1, 2, ...n$) indicating the minimum and maximum numbers of possible applications of R_i, representing the values of count functions, are obtained from the process model via the I/P of human experts. These numbers may depend upon the analysis of the interactive communication patterns between the agents, the controllable variables and the environmental variables concerning the specific problem which can be highly contextual in nature and can have functional dependancies between their concerned attributes. Also for R_i in the rule base, there exists an associated rule importance grade concerning a particular module of the original semistructured problem which is also obtained from the process model via the I/P of human experts.

The model relational representations of the IC-Bags β_{M_1}, β_{M_2} and β_{M_3} specifying the requirements regarding the allowable values for maximum and minimum counts of application for each $R_i(i = 1, ...10)$ in the parallel modules M_1, M_2 and M_3 are furnished in the following tables 1, 2, and 3:

ID	Type	$C_l^{M_1}(R_i)$	$C_u^{M_1}(R_i)$
I01	R1	5	8
I02	R2	3	5
I03	R3	5	12
I04	R4	7	8
I05	R5	3	8
I06	R6	8	15
I07	R7	7	7
I08	R8	1	9
I09	R9	2	4
I10	R10	0	5

Table 1

ID	Type	$C_l^{M_2}(R_i)$	$C_u^{M_2}(R_i)$
I01	R1	4	7
I02	R2	6	10
I03	R3	4	5
I04	R4	14	15
I05	R5	6	6
I06	R6	12	13
I07	R7	0	2
I08	R8	0	0
I09	R9	1	3
I10	R10	2	5

Table 2

ID	Type	$C_l^{M_3}(R_i)$	$C_u^{M_3}(R_i)$
I01	R1	6	7
I02	R2	4	7
I03	R3	4	9
I04	R4	3	5
I05	R5	6	6
I06	R6	1	9
I07	R7	2	7
I08	R8	8	8
I09	R9	0	2
I10	R10	5	5

Table 3

It is evident that $S = \beta_{M_1} \oplus \beta_{M_2} \oplus \beta_{M_3}$ (shown in Table 4) indicates the maximum and minimum allowable counts of application for each concerned rule in the problem-oriented methodology for the integrated problem. If necessary, the system can also generate an IC-Bag specification representing the maximum and minimum number of application counts reserved for each of these rules for process-oriented application only. In this case, that generated IC-Bag will be further added to the IC-Bag S resulting from the addition of β_{M_1}, β_{M_2} and β_{M_3}.

ID	Type	$C_l^{S}(R_i)$	$C_u^{S}(R_i)$
I01	R1	15	22
I02	R2	13	22
I03	R3	13	26
I04	R4	24	28
I05	R5	15	20
I06	R6	21	37
I07	R7	9	16
I08	R8	9	17
I09	R9	3	9
I10	R10	7	15

Table 4

For each rule $R_i (i = 1, ...n)$,

$$C_l^S(R_i) = C_l^{M_1}(R_i) + C_l^{M_2}(R_i) + C_l^{M_3}(R_i)$$

$$C_u^S(B_i) = C_u^{M_1}(R_i) + C_u^{M_2}(R_i) + C_u^{M_3}(R_i)$$

and thus it is evident that β_{M_1},β_{M_2} and β_{M_3} are IC subbags of S.

The DFD for any rule $R_i(i = 1, 2, ...n)$ in the rule base associated with the decision analysis pattern can be represented as in Figure-1:

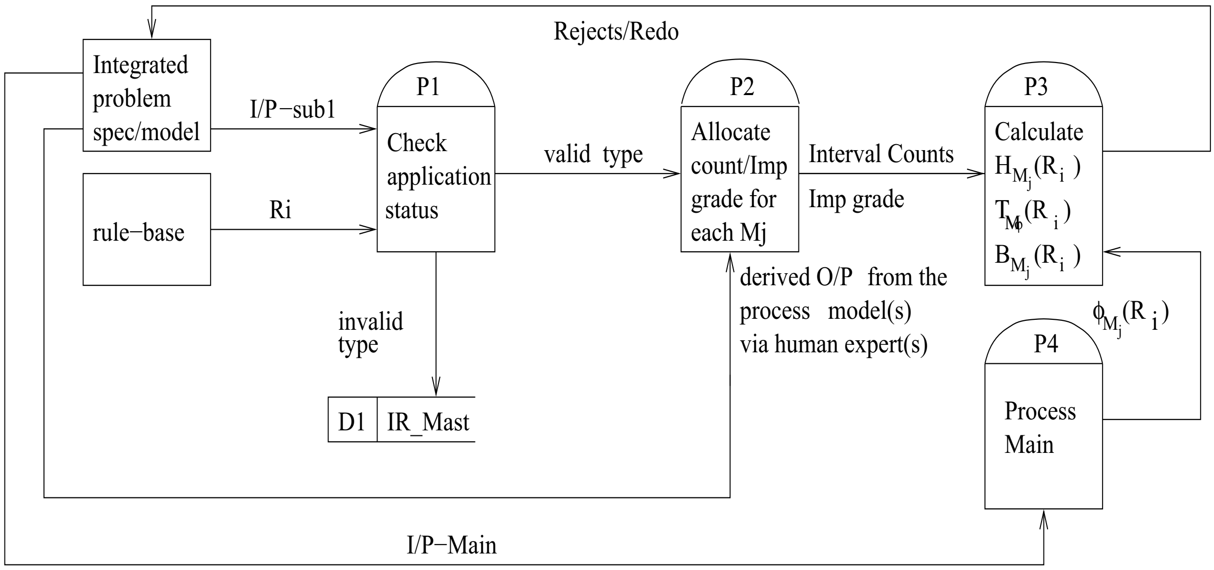

Figure 1

The rules R4 and R6 represent the concerned l-peak and u-peak elements of S. Here S does not include an lu-peak element but however β_{M_1} and β_{M_2} both have lu-peak elements which are R6 and R4 respectively.

These peak elements play significant roles while classifying the rules according to their assigned degree of importance related to variable requirement specification. If we consider that the rule base has a list of rules which are labelled as usable for the concerned problem, then the list constitutes a set V with the characteristic function V_μ, such that:

$$\begin{aligned} V_\mu(R_i) &= 1, \text{where Case: usable} \\ &= 0, \text{where Case: not not usable} \end{aligned}$$

The selection of IC-Bags allows the formation new IC-Bag indicating the minimum and maximum counts of each rule that can actually be applied under the imposed restrictions of usability under the specification set for the semi-structured problem.

If $I_{M_j}(R_i)$ represents the associated rule importance grade for the rule R_i concerning the module M_j, then the hesitation factor $H_{M_j}(R_i)$ associated with the count functions can be obtained by:

$$H_{M_j}(R_i) = I_{M_j}(R_i) - \frac{\sum_{j=1}^{m} I_{M_j}(R_i)}{m}$$

The central tendency of the frequency of application $\tau_{M_j}(R_i)$ for each rule in module M_j can be determined by rounding the value of

$$\frac{C_l^{M_j}(R_i) + C_u^{M_j}(R_i)}{2} - H_{M_j}(R_i)$$

to the nearest integer value.

The rule importance grades are highly process oriented and are subject to bias. The difference between the actual count value $\phi_{M_j}(R_i)$ obtained after the run and the value of $\tau_{M_j}(R_i)$ is called the bias factor for the module M_j and it is denoted by $B_{M_j}(R_i)$. If $B_{M_j}(R_i) > C_u^{M_j}(R_i)$ or $B_{M_j}(R_i) < C_l^{M_j}(R_i)$, then we reject the application and redo it for M_j.

The model relational data entries for contextual rule importance grades, associated hesitation factors are furnished in table 4:

Type	Module	$C_l^{M_j}(R_i)$	$C_u^{M_j}(R_i)$	$I_{M_j}(R_i)$	$H_{M_j}(R_i)$	$\tau_{M_j}(R_i)$	$\phi_{M_j}(R_i)$	$B_{M_j}(R_i)$
R1	M1	5	8	0.6	+0.1	6	8	+2
R1	M2	4	7	0.6	+0.1	5	9	R
R1	M3	6	7	0.3	-0.2	7	7	0
R2	M1	3	5	0.8	+0.4	4	4	0
R2	M2	6	10	0.4	0	8	7	-1
R2	M3	4	7	0.1	-0.3	6	2	R
R3	M1	5	12	0.7	+0.2	8	6	+2
R3	M2	4	5	0.1	-0.4	5	4	-1
R3	M3	4	9	0.6	+0.1	6	8	+2
R4	M1	7	8	0.2	-0.3	8	2	R
R4	M2	14	15	0.9	+0.4	14	12	-2
R4	M3	3	5	0.5	0	4	6	R
R5	M1	3	8	0.5	+0.1	5	5	0
R5	M2	6	6	0.4	0	6	7	R
R5	M3	6	6	0.3	-0.1	6	6	0
R6	M1	8	15	0.1	-0.2	12	10	-2
R6	M2	12	13	0.6	+0.3	12	12	0
R6	M3	1	9	0.2	-0.1	5	8	+3
R7	M1	7	7	0.1	-0.3	7	7	0
R7	M2	0	2	0.7	+0.3	1	1	0
R7	M3	2	7	0.4	0	5	2	-3
R8	M1	1	9	0.6	+0.3	5	2	-3
R8	M2	0	0	0	-0.3	0	2	R
R8	M3	8	8	0.3	0	8	5	R
R9	M1	2	4	0.9	+0.4	3	4	+1
R9	M2	1	3	0.5	0	2	1	-1
R9	M3	0	2	0.1	-0.4	1	5	R
R10	M1	0	5	0.5	0	3	2	-1
R10	M2	2	5	0.7	+0.2	3	5	+2
R10	M3	5	5	0.2	-0.3	5	5	0

Table 4

If the total number of R's equals the total number of rules in the rule-base, then the system is considered to be in inconsistent state and in this case the minimum and maximum numbers of possible applications of R_i, representing the values of count functions, are all set to null and the new knowledge base is obtained from the domain expert. Otherwise each module with one or more reject tags undergoes incremental testing with unit increments in the values of $C_u^A(R_i)$ where $\phi_{M_j}(R_i) > C_u^A(R_i)$, and unit decrement in the values of $C_l^A(R_i)$ where $\phi_{M_j}(R_i) < C_l^A(R_i)$ subject to the constraint $C_l^A(R_i) \geq 0$ for each R_i till all the reject tags are removed. Closer the value of the bias factor to zero, higher is the confidence level of the knowledge base. If the bias factor equals zero, then the knowledge base receives the highest degree of confidence. If the total bias factor is positive, then the knowledge base is called positively biased, otherwise it is called negatively biased.

Acknowledgement

The author is extremely thankful to the referees for their very valuable comments which helped in the preparation of the final version of this paper.

References

[1] K. Chakrabarty, Bags with Interval Counts, *Foundations of Computing and Decision Sciences*, 25, 23-36, 2000.

[2] K. Chakrabarty, On Bags and Fuzzy Bags, *Advances in Soft Computing: Soft Computing Techniques and Applications*, Physica-Verlag, 201-212, 1999.

[3] K. Chakrabarty, IC-Bag Based Systems and Validation Issues, *10th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic, and Validated Numerics*, Paris, 2002.

[4] K. Chakrabarty, R. Biswas, S. Nanda, On Yager's theory of Bags and Fuzzy bags, *Computers and Artificial Intelligence*, 18(1), 1-17, 1999.

[5] K. Chakrabarty, On IC-Bags, *Proc. International Conference on Computational Intelligence for Modelling, Control and Automation*, USA, 2001.

[6] K. Chakrabarty, R. Biswas, S. Nanda, Fuzzy Shadows, *Fuzzy Sets and Systems*, 101/3, 413-421, 1999.

[7] R. R. Yager, On the Theory of Bags, *International Journal of General Systems*, 13, 23-37, 1986.

Electronic version of an article published in the *International Journal of Information Technology & Decision Making,* volume 3, issue 1, 2004, pages 101-108
DOI: 10.1142/S0219622004000994

http://www.informaworld.com/smpp/title~content=t713408216