Qualitative and interval algebras for robust decision making under uncertainty

Simon Parsons†
Department of Electronic Engineering
Queen Mary and Westfield College
Mile End Road
London E1 4NS

John Fox
Biomedical Computing Unit
Imperial Cancer Research Fund
Lincoln's Inn Fields
London WC2A 3PX

Decision support systems intended for operation under real world conditions require reasoning mechanisms that are robust in the face of degraded data. We present two algebras for reasoning with incomplete and imprecise data that are suitable for such systems. The first is an extended qualitative algebra which includes operations over real numbers. This is appropriate for reasoning with largely symbolic data. The other is an interval algebra, built around the paradigm of interval analysis, which is designed to deal with largely numerical data. We demonstrate how the algebras may be used for robust reasoning, and show how they might be applied to a decision problem in a medical domain.

1 Introduction

Gallie [7] eloquently argues that, in contrast to the cartesian view of the universe, all knowledge is subject to a degree of irreducible uncertainty. Similar arguments may be made for the ubiquity of imprecision and incompleteness; for every piece of information that is not be reliable there is one that is not measured precisely, and one that has not been obtained by the moment it is required. These arguments suggest that any system intended to reason about real world conditions should be able to operate under conditions of uncertainty, imprecision and incompleteness. It is possible to criticise classical decision theory [8] as regards its handling of imprecision and incompleteness. Whilst the decision theoretic mechanism handles uncertainty in a sound manner and degrades gracefully as the quality of information falls, such behaviour is dependent on the availability of a full and unambiguous set of prior and conditional probabilities for all decision options, and expected utility values for all decision outcomes.

Fox [5], for example, contends that such values are frequently difficult to obtain since they rely on controversial objective or subjective judgements. If the values are not available the underlying probability mechanism fractures, and so it is possible to consider decision theory as suspect in the light of everyday imprecision and incompleteness. Such a criticism is particularly valid in systems that are continually encountering new situations and are thus unable to rely on either having the correct probability values or being able to obtain them before a decision is required. In this paper we address the issue of robustness, proposing two algebras that may be used as the basis for extended decision mechanisms. By combining numerical and qualitative [1] data we build in the ability to handle incomplete and imprecise probabilistic information. The algebras will give exact probabilistic assessments if given exact data, but will continue to function if the probabilities are replaced with less precise measures. The use of qualitative techniques also makes it possible to support reasoning that is more intelligible to an untrained human user.

2 The problems of imprecision and incompleteness

To illustrate the problems of propagating imprecise and incomplete information in a probabilistic framework, consider the fragment of medical knowledge [12]: *"Metastatic cancer is a possible cause of a brain tumour, and is also an explanation for increased total serum [calcium] count. In turn, either of these could explain a patient falling into a coma.*

†This research was supported by the project DRUMS, which is funded by a grant from the Commission of the European Communities under the ESPRIT Program, Basic Research Action 3085.

Severe headache is also possibly associated with a brain tumour". This knowledge may be represented as a graph which summarises the underlying causal information (Figure 1):

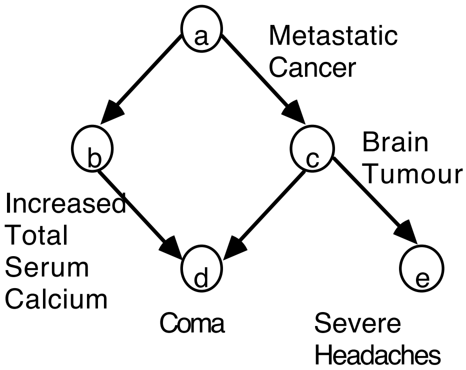

Figure 1 The brain tumour example

A purely probabilistic system requires all the values of Table 1 in order to function, and can deduce [12] the prior probabilities of brain tumour and coma as 0.08 and 0.32 respectively. Given the occurrence of severe headaches the probabilities increase to 0.104 and 0.33. All propagation is halted by a missing value such as that of p(dl b, c). Imprecise data also causes problems since the underlying theory requires point values.

p(elc) p(el~c)	0.80 0.60	Headaches common, but more common if a tumour is present
p(dl b, c) p(dl b, ~c) p(dl~b, c) p(dl~b, ~c)	0.80 0.80 0.80 0.05	Coma rare, but common if brain tumour or increased serum calcium present
p(bl a) p(bl~a)	0.80 0.20	Increased calcium uncommon, but common consequence of metastatic cancer
p(cla) p(cl~a)	0.20 0.05	Brain tumour rare and uncommon consequence of metastatic cancer
p(a)	0.20	Incidence of metastatic cancer in relevant clinic

Table 1. Probabilities for the brain tumour example

3 An extended qualitative algebra

Arithmetic operators over real numbers may be extended to handle qualitative data. Fox and Krause [6], for example, create an extended algebra Q2 that is similar to Williams' Q1 [13]. The crucial difference between these independently developed systems is that Q2 makes no distinction between qualitative expressions and real ones. All types of operands are handled by the same set of operators. Here we augment the original idea, supplying a formal definition of Q2 and a full set of arithmetic operators.

3.1 A formal definition

The qualitative descriptions of Q2 operate on an operand set S' = {-, 0, +, ?} $\cup\, \Re$. S' = S $\cup\, \Re$ where S = {-, 0, +, ?} is Williams' set of qualitative operands, and $\Re$ is the set of real numbers. The operands {-, 0, +} are related to $\Re$ by the mapping [ ] :

$$\text{For any } x \in \Re,\ [x] = \begin{cases} + & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ - & \text{if } x < 0 \end{cases}$$

The qualitative value ? denotes a real value whose sign and value are not known. We also have $\oplus \in OP_{Q2} : \Re \times S \rightarrow S$, where OP_{Q2} is the set of all operators over S'. Since $\Re \not\subset$ S', Q2, unlike Q1, has only one set of arithmetic operators which can operate on all members of S'. We will write the extended operators analogous to the usual arithmetic operators on $\Re$, {+, -, ×, ÷}, as EA_{Q2} = {+', -', ×', ÷'}, and note that along with [ ] they form an algebra that combines arithmetic on real values with the robustness of qualitative algebras. Thus $\otimes \in EA_{Q2}$: S' × S' $\rightarrow$ S' is equivalent to the analogous operation on reals $\otimes \in EA_{Q2} : \Re \times \Re \rightarrow \Re$, qualitative values $\otimes \in EA_{Q2}$: S × S $\rightarrow$ S, and a combined arithmetic and mapping operator converting reals into qualitative values, $\otimes \in EA_{Q2} : \Re \times S \rightarrow S$. To specify the exact results of the operators, we use combinator tables:

+'	R+	+	R-	-	0	?
R+	R+	+	R?	?	R+	?
+	+	+	?	?	+	?
R-	R?	?	R-	-	R-	?
-	?	?	-	-	-	?
0	R+	+	R-	-	0	?
?	?	?	?	?	?	?

(a) Addition

-'	R+	+	R-	-	0	?
R+	R?	?	R+	+	R+	?
+	?	?	+	+	+	?
R-	R-	-	R?	?	R-	?
-	-	-	?	?	-	?
0	R-	-	R+	+	0	?
?	?	?	?	?	?	?

(b) Subtraction

×'	R+	+	R-	-	0	?
R+	R+	+	R-	-	0	?
+	+	+	-	-	0	?
R-	R-	-	R+	+	0	?
-	-	-	+	+	0	?
0	0	0	0	0	0	0
?	?	?	?	?	0	?

(c) Multiplication

÷'	R+	+	R-	-	0	?
R+	R+	+	R-	-	U	U
+	+	+	-	-	U	U
R-	R-	-	R+	+	U	U
-	-	-	+	+	U	U
0	0	0	0	0	U	U
?	?	?	?	?	U	U

(d) Division

Figure 2. Combinator Tables for Q2

R+ and R- designate the occurrence of particular positive and negative numbers respectively. R? represents a real number somewhere between +∞ and -∞ which may be uniquely determined from the operator and operands. This contrasts with ? which represents an indeterminable number between +∞

and $-\infty$. U represents an undefined value, in particular the result of dividing by 0 or an unknown value that may be 0.

3.2 Reasoning with imprecise and incomplete information

Consider using Q2 in the brain tumour example with p(d|b,c) missing. The probability table may be completed by taking p(d| b, c) = + on the grounds that *"either brain tumour or increased total serum calcium could result in a patient falling into a coma"* indicates a positive association between brain tumour, increased calcium and the occurrence of a coma. All inferred values that are unaffected by p(d| b, c) are still precisely available; we can, for instance, still deduce that the prior probability of brain tumour is 0.08. Values affected by p(d|b,c), may not be calculated exactly, but we can deduce something about them. For instance, from p(d) = p(d| b, c). p(b, c), p(d| b, c) = +, and the combinator table for extended multiplication, we can determine that the prior probability of coma is +. Knowing that p(d| b, c) = +, we can also conclude that since headaches make a brain tumour more likely, they also make a coma more likely. A probability value that is not precisely defined may similarly be assumed to be + and the extended operators used to propagate its consequences.

3.3 Choosing between alternatives

Q2 is essentially symbolic, and handles numbers by reducing them to qualitative relations. Thus any mechanism intended to allow the system to choose between competing solutions must diverge from classical methods of decision making. It is possible to use an improper linear model [2] to weigh up arguments for (+) and against (-) particular solutions. Alternatively decisions may be made by the use of some symbolic decision model which includes the recognition of the need for a decision, the classification of the decision options, and the consideration of the pros and cons of each possible decision result. A more detailed description of such a model may be found in [4].

4 An interval algebra

Since Q2 is essentially qualitative it has a tendency to over abstract, converting point quantifications of uncertainty into unbounded +'s. This presents no problems in situations in which few accurate numerical values are available, but may be less attractive when the bulk of the values are known either precisely, or within certain limits. Q3 replaces qualitative values with the [0, 1] interval used by most uncertainty calculi, and operates over exact values x,...,y and bounded intervals [m, n] ,..., [p, q] to provide a unique ability to combine qualitative data with point and interval quantifications of uncertainty.

4.1 A formal definition

We have a set of operands S'' = $\{0, [0, 1], S, \Re_{[0, 1]}\}$ where $S = \{x : x = [y, z] \text{ and } 0 \leq y \leq 1 \text{ and } 0 \leq z \leq 1\}$ and $\Re_{[0, 1]} = \{x : x \in \Re \text{ and } 0 < x \leq 1\}$. There is a mapping [[ ]] that relates $\Re_{[0, 1]}$ to S, $[[\Re_{[0, 1]}]] \rightarrow S$, such that if $x \in \Re_{[0, 1]}$ then [[x]] = [x, x]. Every real number in the zero-one interval is equivalent to a point interval. The distinguished member, 0, is included as a separate value. As above, we have operators that act as mappings; $\otimes \in OP_{Q3} : \Re_{[0, 1]} \times S \rightarrow S$, $\otimes \in OP_{Q3} : S \times S \rightarrow S$, $\otimes \in OP_{Q3} : \Re_{[0, 1]} \times \Re_{[0, 1]} \rightarrow \Re_{[0, 1]}$, where OP_{Q3} is the set of all operators over S''. Members of S may also be created by the quantification of vague information. We have a function {[ ]} which maps from imprecise data to S, so that $\{[\{x : y \leq x \leq z \text{ and } 0 \leq y \leq 1 \text{ and } 0 \leq z \leq 1\}]\} \rightarrow [y, z] \in S$. We distinguish the special case where y = 0 and z = 1 as the vacuous value [0, 1] which expresses complete ignorance about the value of x. Arithmetic operations are defined by Moore's [9] interval analysis, and for any arithmetic combinator $\otimes \in OP_{Q3}$ we have:

$$[a, b] \otimes [c, d] \quad = \quad [\min (a \otimes c, a \otimes d, b \otimes c, b \otimes d), \max(a \otimes c, a \otimes d, b \otimes c, b \otimes d)]$$

where max and min have their usual meanings.

4.2 Robustness under imprecision and incompleteness

When we apply Q3 to the medical example all calculations involving exact probabilities produce degenerate intervals [n, n] equivalent to the exact results. When the missing value is encountered the system can substitute p(d| b, c) = [0, 1] applying the bounds on its possible value. Evaluating as before, we get p(d) = [0.288, 0.688], a far more precise estimate than +. Knowing that severe headaches occur allows us to recalculate the probability of coma using Spiegelhalter's method [12] with interval arithmetic to arrive at $p^*(d) = [0.366, 0.766]$. The system can also accept probability estimates within given bounds. If we know that $p(d|b, c) = [0.6, 0.9]$, we can compute the prior $p(d) = [0.312, 0.324]$ and, in the case of known headaches, the updated value $p^*(d) = [0.039, 0.402]$.

4.3 Ordering solutions in Q3

Interval arithmetic can only compare non-overlapping intervals, giving $[a, b] < [c, d]$ iff $b < c$ [9]. This is not suitable for comparing all interval valued probabilities. Instead we define a flat second order probability distribution over each interval, indicating that the value lying within a given range is no more likely to take any one value than it is to take any other. This allows us to order intervals [a, b] and [c, d], with $[a, b] >_{Q3} [c, d]$ if the value in [a, b] is more likely to be greater than the value in [c, d] than the value in [c, d] is to be greater than that in [a, b]. Thus $[a, b] >_{Q3} [c, d]$ iff $p(V_R^{[a, b]} > V_R^{[c, d]}) > p(V_R^{[c, d]} > V_R^{[a, b]})$ where $V_R^{[a, b]}$ is the actual value taken by the probability which is known to fall in the interval [a, b]. The comparison may also be affected between intervals and point values and it is possible to take all decisions on a well defined basis. The following results have been established:

> **Theorem 4.1**: For a comparison between an interval with lower bound 0 and a point value, $[0, b] >_{Q3} [c, c]$ iff $b > 2c$.
>
> **Theorem 4.2**: Comparing two overlapping intervals we have $[a, b] >_{Q3} [c, d]$ iff $(b - d) > (c - a)$.

Proofs may be found in [10]. Of course, if we wish to carry out some kind of argumentation procedure, it is possible to abstract intervals back into qualitative values. The only difference between propagating interval values that are then abstracted to qualitative values and propagating qualitative values is the added computational expense of the interval calculation which will take longer than its qualitative counterpart; the end result of the inference will be the same.

5 Decision making in gastroenterology

The following example illustrates the performance of the algebras in more detail. We will consider a clinic specialising in gastroenterological complaints. These complaints have a number of possible origins which may be classified as gastric cancer, peptic ulcers (both gastric and duodenal ulcers), gallstones, and functional disorders. The latter are conditions with no identifiable organic cause, and are often stress related. Over many years, a number of symptoms and signs which provide useful information for discriminating between complaints have been recorded from many patients. These are signs of jaundice, pain after meals, weight loss and the age of the patient.

The clinic has a research interest in gastric disorders and has established estimates of the number of patients with confirmed diagnoses of peptic ulcers and gallstones that exhibited the above symptoms on arrival. Figures are also available for the relation of the symptoms to gastric cancer, though the low incidence of the disease casts doubt on the accuracy of the figures, and to functional disorders. A conditional probability table (Table 2) may be constructed from the frequency data.

	Gastric Cancer (gc)	Peptic Ulcer (pu)	Gall-stones (gs)	Functional Disease (fd)
Jaundice (j)	$p(j \mid gc) = 0.08$	$p(j \mid pu) = 0.03$	$p(j \mid gs) = 0.30$	$p(j \mid fd) = 0.10$
Pain after meals (m)	$p(m \mid gc) = 0.63$	$p(m \mid pu) = 0.40$	$p(m \mid gs) = 0.18$	$p(m \mid fd) = 0.30$
Weight Loss (w)	$p(w \mid gc) = 0.71$	$p(w \mid pu) = 0.35$	$p(w \mid gs) = 0.03$	$p(w \mid fd) = 0.01$
Elderly Patient (e)	$p(e \mid gc) = 0.68$	$p(e \mid pu) = 0.50$	$p(e \mid gs) = 0.12$	$p(e \mid fd) = 0.20$

Table 2. Conditional probabilities relating symptoms of gastric disease to their causes

We also have a set of prior probabilities for the incidence of each disease among patients referred to the clinic:

$$p(gc) = 0.01 \qquad p(pu) = 0.35$$
$$p(gs) = 0.10 \qquad p(fd) = 0.54$$

We are interested in the case of Fred, an elderly patient referred to the clinic by his doctor. Fred shows no signs of jaundice but has recently lost weight and often has pain after eating. The use of Bayes' rule [8] allows us to aggregate the evidences for each disease to come up with a set of posterior probabilities:

$$p(gc \mid \neg j, m,w,e) = 0.104 \qquad p(pu \mid \neg j, m,w,e) = 0.883$$
$$p(gs \mid \neg j, m,w,e) = 0.002 \qquad p(fd \mid \neg j, m,w,e) = 0.011$$

These figures enable us to conclude that Fred is most likely to have a peptic ulcer, may possibly have gastric cancer, and is extremely unlikely to have either gallstones or some functional disorder.

Given point probability estimates we can carry out an aggregation in Q2 and Q3 as well as the more familiar algebra of real numbers. The robustness of the two new algebras may also be used when the evidence is less precise. As mentioned above the data for gastric cancer and functional disorders is likely to be less accurate than that for ulcers and gallstones in which case assessments made by point values are less acceptable than estimates of the form:

$$p(m \mid fd) = [0.2, 0.6] \qquad p(e \mid fd) = [0.1, 0.42]$$

and: "there is a positive association between the incidence of pain after meals and the occurrence of gastric cancer".

In such a case, conventional probability calculations over real numbers may not be performed. However, Q2 and Q3 may be used to deduce some information about the patient. In Q3 we can use the assessment of a positive association between pain after meals and gastric cancer to give $p(m \mid gc) = [0, 1]$ which may be used with the interval and point estimates in an interval version of Bayes' rule [10] to give:

$$p(gc \mid \neg j, m,w,e) = [0, 0.186]$$
$$p(pu \mid \neg j, m,w,e) = [0.807, 0.996]$$
$$p(gs \mid \neg j, m,w,e) = [0.002, 0.002]$$
$$p(fd \mid \neg j, m,w,e) = [0.012, 0.099]$$

Applying the rules of interval arithmetic and Theorems 4.1 and 4.2 we arrive at the same ordering of hypotheses as in the point probability case, but with an indication of other possibilities. For instance, though Fred is most likely to have an ulcer, if this is eliminated, the overlap between functional disease and cancer indicates that the former should not be discounted.

In Q2, recognition of the imprecision prompts the replacement of the conditional probabilities with a set of qualitative assessments based upon them (Table 3). The + relating weight loss to cancer indicates that the occurence of weight loss increases our belief in the presence of cancer, whilst the - relating jaundice to cancer indicates that should jaundice be present, our belief in the presence of cancer would be reduced. The ? relating pain after meals to gallstones indicates that the occurence of the symptom has little or no effect on the diagnosis.

	Gastric Cancer	Peptic Ulcer	Gall-stones	Functional Disease
Jaundice	-	-	+	?
Pain after meals	+	+	?	+
Weight Loss	+	+	-	-
Elderly Patient	+	+	?	?

Table 3. Qualitative assessments relating symptoms of gastric disease to their causes

With the addition of the re-assessed priors:

p(gc)	=	-	p(pu)	=	?
p(gs)	=	?	p(fd)	=	+

these values can be used in some form of improper linear model with uniform weights and no constant term to give an assessment of the likelihood of each disease matching the symptoms that agrees with the probabilistic approaches.

6 Conclusions

The examples demonstrate the strengths of Q2 and Q3 in maintaining performance in the face of moderately and highly degraded data whilst allowing precise calculations when the data permit. Q2 is most effective in domains where the degraded data are best represented by qualitative values since its tendency to abstract will naturally degrade interval data. In contrast Q3, although capable of handling qualitative data to a degree, performs best when dealing with interval estimates, and so will be most effective in domains where imprecise and largely complete data are available. It is, of course, possible to combine the power of both algebras by using them in conjunction. This will require the establishment of a mapping between qualitative and interval probability values which is likely to be domain dependent.

As a final word, it should not be assumed that algebras described in this paper may only be applied to systems that deal with probabilistic information. There is no theoretical reason why, given the necessary extension of the operator sets to include the relevant mathematical operations, the algebras may not be applied to systems reasoning with possibilistic or fuzzy logics [3] or belief functions [11]. It is, however, true that whilst the meaning of a probabilistic interval is reasonably clear, that of a belief function or possibility/necessity interval is less transparent.

References

[1] Bobrow, D. G., Qualitative Reasoning about Physical Systems (North-Holland, Amsterdam, 1984).
[2] Dawes, R. M. The robust beauty of improper linear models in decision making, in: Kahnemann, D., Slovic, P. and Tversky, A. (eds.) Judgement under uncertainty: Heuristics and biases (Cambridge University Press, Cambridge, 1982) pp. 391-407.
[3] Dubois, D. and Prade, H. An introduction to possibilistic and fuzzy logics (with discussions) in: Smets, P., Mamdani, E. H., Dubois, D. and Prade, H. (eds.), Non-Standard Logics for Automated Reasoning (Academic Press, London, 1988) pp. 287-326.
[4] Fox, J. Decision making and autonomous systems, this volume.
[5] Fox, J. Automating assistance for safety critical decisions. Phil. Trans. R. Soc Lond. (1990) 327.
[6] Fox, J. and Krause P. J. Combining symbolic and numerical methods for defeasible reasoning (IEE Colloqium on Reasoning Under Uncertainty, 1990).
[7] Gallie, W., B. Uncertainty as a philosophical problem:1, in: Carter, C. F., Meredith, G. P. and Shackle, G. L. S. (eds.), Uncertainty and business decisions : the logic, philosophy and psychology of business decision-making under uncertainty (Liverpool University Press, 1957).
[8] Lindley, D., V. Making Decisions (John Wiley and Sons, London, 1985).
[9] Moore, R. E. Interval Analysis (Prentice-Hall, Inc. Englewood Cliffs, N.J. 1966).
[10] Parsons, S. On using qualitative algebras in place of metatheories for reasoning under uncertainty (Technical Report, Dept. of Electronic Engineering, Queen Mary and Westfield College. 1990).
[11] Smets, P. Belief Functions (with discussions) in: Smets, P., Mamdani, E. H., Dubois, D. and Prade, H. (eds.), Non-Standard Logics for Automated Reasoning (Academic Press, London, 1988) pp. 253-286.
[12] Spiegelhalter, D. J., Probabilistic Reasoning in predictive Expert Systems. in: Kanal, L. and Lemmer, J. F. (eds.) Uncertainty in Artificial Intelligence (Elsevier Science Publishers, North-Holland, Amsterdam, 1986) pp. 47-67.
[13] Williams, B. C., MINIMA. A Symbolic Approach to Qualitative Algebraic Reasoning. (AAAI 1988).