Artificial Intelligence 61 (1993) 187–208
Elsevier

ARTINT 964

Objective probabilities in expert systems

L.E. Sucar and D.F. Gillies

Department of Computing, Imperial College of Science Technology and Medicine, 180 Queen's Gate, London SW7 2AZ, UK

D.A. Gillies

Centre for Logic and Probability in Information Technology, King's College, Manresa Road, London SW3 6LX, UK

Received December 1990
Revised July 1992

Abstract

Sucar, L.E., D.F. Gillies and D.A. Gillies, Objective probabilities in expert systems, Artificial Intelligence 61 (1993) 187–208.

In this paper we present a general methodology for handling uncertain knowledge in expert systems, which is based upon objective probability theory. The use of objective probabilities helps to overcome some of the difficulties in the subjective Bayesian approach. The basic idea is to refine a qualitative assessment of uncertainty made by a domain expert into a quantitative objective probability by measuring frequencies in data sets. Knowledge is represented as a probabilistic network where the structure is elucidated from the experts, and the probability distributions are estimated from a set of representative samples from the domain. We test the hypothesis of independence between variables using linear regression analysis techniques. Having identified dependencies we modify the structure of the network to account for them. We have tested our methodology by implementing an expert system for providing diagnostic advice during colon endoscopy. Our results show strong empirical evidence supporting our approach.

1. Introduction

The aim of the present paper is to describe a new method of handling uncertain knowledge in expert systems. It is called the QUALQUANT methodology, and is based on the use of objective probability. Some use is made of ideas from the philosophy of science due to Kuhn and Popper. We have tested

Correspondence to: D.F. Gillies, Department of Computing, Imperial College of Science Technology and Medicine, 180 Queen's Gate, London SW7 2AZ, UK.

it experimentally, using the specific application of an expert system to give advice during medical endoscopic examinations.

The general problem of handling uncertainty in artificial intelligence, and particularly in expert systems, is very much to the fore at the moment. Genesereth and Nilsson in their excellent book *Logical Foundations of Artificial Intelligence* [2] devote three chapters to the question, specifically in the areas of induction, reasoning with uncertain beliefs, and knowledge and belief. A very clear account of the different approaches, together with some penetrating observations as to their merits and deficiencies, is given by Ng and Abramson [11]. They highlight the problem that the input variables must be assumed to be independent for most methods to work. The new feature of our method is that the independence of the variables is tested, and dependent variables combined or removed from the system.

Although our method is quite general, we do not claim that it is of universal applicability. Even if we confine ourselves to probability-based approaches, there are many areas, for example investment decisions, where it may not be possible to introduce objective probabilities. Expert systems in these areas would have to use subjective, or perhaps intersubjective, probabilities. There does, however, seem to be a wide range of cases to which it could be applied.

2. The QUALQUANT methodology

In this section, we will outline a methodology for constructing expert systems which we shall call the QUALQUANT methodology. This is short for *qual*itative orientation leading to *quant*itative improvement. It is based on the following three principles:

(1) As far as possible only qualitative suggestions should be sought from the domain expert, and it should be left to the computer scientist to give this a more precise quantitative form.
(2) Objective probabilities should be used wherever possible.
(3) All assumptions should be tested and modified if they fail the test (testing principle).

The testing principle is based on Popper's theory of scientific method [14]. The second and third principles are fundamental to the so-called classical approach to statistics of Fisher and Neyman–Pearson. This approach is in fact Popperian or falsificationist. Its application to expert systems will be contrasted with the subjective Bayesianism which is at present dominant among those who use the probability approach to expert systems.

In building an expert system we start with a skilled task performed by some domain expert. We find out the rules which the domain expert uses, and then attempt to incorporate these into a program in such a way that the task can be performed by a computer. A key difficulty in this procedure is that of obtaining the requisite rules from the domain expert. The domain expert may not be fully aware of the rules which he or she is actually using in practice. The expert may not be able to formulate very precisely even those rules which he or she can remember. Other rules may be temporarily forgotten, while the expert may be using unconsciously some rules of which he or she is not, and never has been, consciously aware.[1] This last possibility is well-illustrated by what may be an apocryphal story told by Michie. Since the story is amusing and illustrates a most important methodological point, it is worth quoting what Michie says in full:

> Unfortunately, human practitioners tend to describe their own rules of operation in terms which do not subsequently stand the test of practice. The story is told of a large cheese factory whose Camemberts were a by-word. Crucial to their renown was the company's procedure for quality control, by which every hundredth cheese was sampled to ensure that the production process was still on the narrow path separating the marginally unripe from the marginally over-ripe. Success rested on the uncanny powers developed by one very old man, whose procedure was to thrust his index finger into the cheese, close his eyes, and utter an opinion.
>
> If only because of the expert's age and frailty, automation seemed to be required, and an ambitious R&D project was launched. After much escalation of cost and elaboration of method, which included lowering into the cheese various steel probes wired to strain gauges and other sensors, no progress had been registered. Substantial inducements were offered to the sage for a precise account of how he did the trick. He could offer little, beyond the advice: "It's got to feel right!" In the end it turned out that feel had nothing to do with it. After breaking the crust with his finger, the expert was interpreting subliminal signals from his sense of smell. [10, p. 217]

The key thing to note here is that the expert thought he was using one characteristic, *feel*, while in fact he was using another, *smell*. As long as a rule in terms of the feel of the cheese was being sought, no progress could be made despite all the complexity of the investigation. As soon as the qualitative orientation was changed from feel to smell, data collection and analysis would, no doubt, have revealed the exact quantitative composition of the smell corresponding to the perfectly ripe Camembert. Kuhn has quite correctly, in

[1] Jeff Paris (personal communication) pointed out that the notion of an expert using a rule unconsciously is somewhat questionable. It may be that the expert does not use a rule at all. This is a fair point, but it may nonetheless be possible to simulate what the expert does using a rule.

our opinion, emphasized the need for a qualitative underpinning of the quantitative in his 1961 paper, where he writes [5, p. 180]: "... large amounts of qualitative work have usually been prerequisite to fruitful quantification in the physical sciences."

Later in this paper, we will describe how this QUALQUANT methodology was used to construct a rule-based expert system for colon endoscopy. At this stage, however, it might be helpful to introduce one simple rule of the system by way of illustration. An endoscope is a flexible tube with viewing capability (fibreoptic or video). In the application under consideration it is inserted into the patient's colon, and one problem is then to steer the endoscope along the colon so that the doctor can observe this organ from the inside. In moving the endoscope, it is very important to direct its tip towards the opening of the next section of the colon—called the *lumen*. If this is not done correctly it can be very painful and dangerous to the patient, and could even cause perforation of the colon wall. One problem then is to pick out from the picture being transmitted by the endoscope's camera the position of the lumen. A domain expert will use his or her judgment, based on experience, to decide where the lumen is located. When our domain expert was consulted about how this was done, his first protocol was that the lumen is a "large, uniform, dark region". This is clearly a qualitative rule, and it was turned into a precise quantitative form by finding suitable mathematical equivalents for "large" and "uniformly dark". Some of the parameters used here were estimated from data in the form of video films of colonoscopy sessions on many different patients. At a later stage, probabilities were incorporated into the mathematics using a method which will be described in the next section.

This simple example shows how the qualitative rules provided by the domain expert can be given a more precise mathematical and quantitative formulation by the computer scientist constructing the expert system. Once these quantitative formulations have been devised, however, the next important step is to test them out against the data of cases which have been handled by a human expert, and whose outcome is known. There are three possible outcomes of such testing which are illustrated in Fig. 1. Outcome 1 is that a quantitative formulation is discovered which performs as well as, if not better than, a

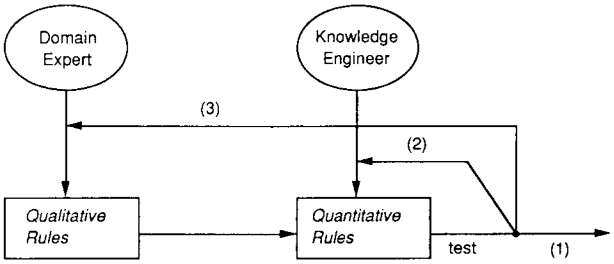

Fig. 1. The QUALQUANT methodology.

human expert. This provides something which can be used in practice, though the possibility of further improvement should not be overlooked. Outcome 2 is that no quantitative formulation works as well as the human expert, but there is some indication that improvements in the quantitative formulation will be successful. In this case the problem returns to the computer scientist who has to try to find a better mathematical model. Outcome 3 is that every existing mathematical model seems to be hopeless. This is illustrated by Michie's Camembert example at the stage when the qualitative orientation is still towards *feel* rather than *smell*. In this case an attempt must be made to elicit new suggestions from the domain expert, always remembering that the domain expert may be using rules of which he or she is not consciously aware.

3. Use of objective probabilities and the testing principle

Those who use the probability approach to expert systems almost always adopt the subjective interpretation of probability. We, however, want to advocate the use of objective probabilities when this is possible, while fully acknowledging that it is not always possible. It would be as well therefore to begin by explaining very briefly the difference between objective and subjective probability.

The subjective approach to probability was introduced by Ramsey [16] in England in 1926 and De Finetti [1] in Italy in 1937. Probability is regarded as the degree of belief of a particular individual, Mr Average say. We measure Mr Average's degree of belief in a particular proposition—say that it will rain in London tomorrow—by forcing him to bet on this proposition under specified conditions, and, in particular, to choose a betting quotient. It can be shown that Mr Average has to choose his betting quotients so that they satisfy the axioms of probability or else a cunning opponent will be able to make a Dutch book against him, i.e. choose the stakes so that Mr Average loses whatever happens. This result, known as the Dutch book argument, or Ramsey–De Finetti theorem, provides the foundation for probability theory on the subjective approach. The subjective approach can be extended to include intersubjective probabilities which represent the degree of belief not of an individual but of a social group which has reached a consensus.

The objective approach to probability, by contrast, denies that probability has anything to do with human beliefs, and regards probabilities as objective features of the external world. Objective probabilities manifest themselves in the observed statistical frequencies with which particular events occur. Take, for example, the probability of a specific biased coin giving heads. This, on the objective approach, has nothing to do with any human beliefs or any human betting, the probability is an objective property of the coin and the way in which it is tossed. The probability is thus analogous to the mass of the coin, to its melting point or its electrical resistance, or to any other of its objective properties. The objective probability of getting heads manifests itself in the frequency with which heads appears in a long sequence of tosses of the coin. In effect the probability is manifested in an experimental outcome whose value is in no way determined by the beliefs of the experimenters.

In recent years the objective approach to probability has been mainly advocated by members of the Popperian school (Popper [13], Gillies [3], and Popper's "Part II: The Propensity Interpretation of Probability" [15, pp. 281–401]). Probability assignments are, on this approach, regarded as conjectures which need to be validated by statistical tests.

Let us now see how these distinctions apply in the field of artificial intelligence, and, more specifically, of expert systems. Lindley is one of the most distinguished advocates of the subjective approach to probability, and recently he has argued strongly for the use of subjective probability to represent uncertainty in information technology, while strongly criticizing alternative approaches, such as fuzzy logic [7, 8]. Ng and Abramson [11] cite Lindley [8], and advocate a similar position—though rather more cautiously. They say:

> In expert systems, a knowledge base stores human knowledge. Thus, in representing an expert's knowledge with probability theory, the only appropriate interpretation of probability is subjective belief. [11, p. 31]

Let us take a medical example to illustrate this. Suppose we ask a doctor (Dr Camphor), who is expert in the field, whether a patient with a particular group of symptoms is likely to have a specified disease. Dr Camphor assigns a probability q to the patient having the disease given those symptoms. Is it not reasonable to regard q as expressing Dr Camphor's degree of belief that the patient has the disease given the symptoms? This is certainly one way to proceed, but another approach is possible.

Consider the set of patients of the appropriate type who have the symptoms in question, and suppose further that the frequency of such patients who have the disease is p. We can then suppose that there is an objective probability p of having the disease given the symptoms. Dr Camphor's stated value q can then be regarded simply as his estimate of the objective probability p. Now if Dr Camphor is a noted expert in the field, we should obviously take his estimate seriously, but it should not be regarded as sacrosanct. After all, even the most notable experts do often make mistakes. The important thing, in accordance with the Popperian ideas developed in the previous section, is to test out the estimate to see if it accords with the objective facts, and to replace it if it does not.

So far we have emphasized the differences between our approach based on objective probabilities and the subjective probability approach. However there are many points in common as well. To begin with both approaches base themselves on the standard mathematical theory of probability, and this contrasts with other ways of handling uncertainty such as fuzzy logic (Zadeh [20]) or Dempster–Shafer theory [17]. There is moreover another point in common which concerns the use of Bayes' theorem. Ng and Abramson write [11, p. 31]: "The term 'Bayesian' is often used as a synonym for subjective probability." It is, however, perfectly possible to use Bayes' theorem in cases where all the probabilities involved are objective. This is indeed the approach which we adopt, and it can be illustrated by pursuing a little further our example of medical diagnosis.

Let D be a disease, and $(S_1, S_2, \ldots, S_n)$ be a group of symptoms. We shall call such a group a syndrome. Now we want to calculate the probability of a patient having the disease given that he or she has the syndrome of symptons, i.e. we want to calculate $P(D \mid S_1 \ \& \ S_2 \ \& \cdots \& \ S_n)$. Applying Bayes' theorem, we get

$$P(D \mid S_1 \ \& \ S_2 \ \& \cdots \& \ S_n) = \frac{P(S_1 \ \& \ S_2 \ \& \cdots \& \ S_n \mid D)P(D)}{P(S_1 \ \& \ S_2 \ \& \cdots \& \ S_n)} . \tag{1}$$

This in effect decomposes the probability we wish to calculate on the left-hand side into three probabilities on the right-hand side. The question now arises as to whether we can regard these three probabilities as objective probabilities, and, if so, whether we can use data to estimate their value. Let us consider the three probabilities in turn.

(1) $P(D)$. This is the probability that an arbitrary member of the underlying population has the disease. This is certainly an objective probability which manifests itself in the frequency of the disease in the population. Since the value of this statistical frequency is known for most major diseases, we should have an estimate of the objective probability based on data without recourse to expert opinion.

(2) $P(S_1 \ \& \ S_2 \ \& \cdots \& \ S_n \mid D)$. This is the probability that an arbitrary patient who has the disease has the syndrome of symptoms. Once again this is a fully objective probability. We are unlikely to have data to estimate its value initially, and may therefore have to rely at first on an expert's estimate. However it is in fact particularly easy to collect data to estimate this probability. The population is those with the disease. Most of them will seek treatment from their family doctor or will be treated in hospital. Of course the patients who report for treatment are a self-selected rather than random sample, and some corrections may be needed here, especially for diseases such as food allergies which would not necessarily be recognized by the patient. Thus, as soon as those concerned in the treatment are alerted to the need for checking whether

such patients have the syndrome in question, they can easily do so. It is interesting to note, however, that without the background of mathematical probability, no one would think of ascertaining this particular statistic, despite the ease with which its value can be found.

(3) $P(S_1 \& S_2 \& \cdots \& S_n)$. This is the probability that an arbitrary member of the underlying population has the syndrome of symptoms whether or not he or she has the disease. Once again this is certainly an objective probability, but, in contrast to the two previous cases, it might well be difficult to estimate this probability from data. Of course there is no difficulty in principle. We need only take a random sample of the population, and check how many have the syndrome. The trouble is that the probability concerned is likely to be small—indeed, as we shall see, we want it to be small—but this means that a very large sample would be needed to get an accurate estimate of the probability. Such a sample might not be feasible on the grounds of expense, etc. This problem could be overcome if it were possible to assume that the various symptoms are independent, for then we have:

$$P(S_1 \& S_2 \& \cdots \& S_n) = P(S_1)P(S_2)\cdots P(S_n) . \quad (2)$$

Now each of the $P(S_i)$ on the right-hand side of this equation is the probability of the particular symptom S_i occurring in the underlying population, and a small sample would in general suffice to obtain an estimate of such a probability. This, of course, is only one simple example of the way in which independence assumptions (or more generally conditional independence assumptions) can simplify calculations and make them more tractable. We will give another example in due course.

The fact that objective probabilities can be used in Bayes' theorem by no means shows that our QUALQUANT methodology is identical to the subjective Bayesian approach. In the rest of this section we will compare the two approaches, and in particular argue that the use of objective probabilities helps to overcome some of the difficulties which Ng and Abramson note in the subjective Bayesian approach. Subjective Bayesians have to elicit subjective probabilities from the domain experts, and, as Ng and Abramson point out, this can be a very problematic process:

> Finding an expert able to accurately quantify personal, subjective, and qualitative information, however, is no mean feat. It has been observed that humans are easily biased, and thus the quality of the knowledge extracted from experts depends greatly on the method used for assessment. Nevertheless, expert system researchers have expended surprisingly minimal effort on studying and deriving appropriate assessment techniques. [11, p. 44]

The QUALQUANT methodology cuts through this problem by trying, wherever possible, to extract only qualitative information from the domain expert, i.e. only the kind of information which the domain expert will find it relatively easy to provide, while leaving the task of turning this into a quantitative mathematical model firmly in the hands of the computer scientist. As far as probabilities are concerned, the attempt is made to estimate these from data wherever possible. If an estimate provided by an expert has to be used, this is regarded as a temporary expedient to be replaced by an estimate from data as soon as this can be obtained. Use was made of Bayes' theorem in our expert system for colonoscopy similar to that just described in the medical diagnosis example, and in this case we were able to estimate all the probabilities involved from the data without asking the domain expert to give any quantitative subjective probabilities. We believe that this would be the case in many practical applications.

Returning to our example of medical diagnosis, however, we can observe that although we do not want to elicit from experts quantitative probabilities, which, in general, they would find it difficult to provide, we do want to elicit from them a great deal of very important qualitative information, which, in general, they would find it easy to provide. Let us next examine briefly the nature of this information. Above all the expert has to describe a syndrome (S say) of symptoms appropriate to the disease in question. Our earlier use of Bayes' theorem shows the various criteria which S must satisfy. If S is to be a good indicator of the disease, the $P(D \mid S)$ must be high, but, in general, $P(D)$ will be low, and so, by Bayes' theorem, we require that the ratio $P(S \mid D)/P(S)$ should be high. This means that $P(S \mid D)$ should be large, i.e. there should be a high probability of someone who has the disease exhibiting the syndrome of symptoms, while $P(S)$ should be low, i.e. there should be a low probability of an arbitrary member of the underlying population having that syndrome of symptoms. The computer scientist must obviously consult the domain expert to find a group of symptoms having these characteristics.

This observation helps to overcome another difficulty which Ng and Abramson note in the probability approach. They write:

> The main difficulty in implementing subjective probability theory is the huge number of probabilities that must be obtained to construct a functioning knowledge base. If, for example, some medical diagnosis domain has 100 diseases and 700 relevant, observable symptoms, then at least 70,100 probability values (70,000 conditional probabilities and 100 prior probabilities) must be obtained . . . . [11, p. 34]

This would be true if we have to provide the conditional probability of any disease given any symptom. Indeed if we had to calculate the conditional probability of any disease given any finite group of symptoms, the probabilities required would be still more numerous. If, however, guided by the domain expert, we confine our attention to just a few relevant syndromes, then we should be able to cut down drastically the number of probabilities which need to be estimated. This example illustrates an important general principle which can be stated as follows:

> Qualitative information obtained from the domain expert should be used wherever possible to simplify the corresponding mathematical model.

We come lastly to the difficulty about independence. As we have seen it is often necessary, when using a probability-based approach to artificial intelligence, to make some assumptions about independence in order to render the problem tractable. Unfortunately these assumptions about independence may not be valid. What can be done about this situation?

The first move in the QUALQUANT methodology is familiar enough. We seek advice from the domain expert as to which characteristics might be independent. It is, however, the second move which differentiates our approach from the subjective one. This second move employs the Popperian principle that it is possible to make any assumption we like provided the assumption in question is tested. In this case we begin by assuming that our group of characteristics is independent, but we then test out this assumption against the data. One way of testing for independence is by calculating the correlation of pairs of the characteristics. Although a low correlation does not necessarily imply independence, it provides evidence that independence is a reasonable assumption to make; and, on the other hand, a high correlation indicates that the characteristics are *not* independent and our assumptions are invalid. If we find a high correlation between a pair of characteristics, we must modify our probabilistic assumptions, and there are at least three alternatives:

(a) Eliminate one of the characteristics on the grounds that we can get almost as much information from one characteristic as from the pair.
(b) Find a parameter that combines both characteristics.
(c) Consult the expert(s) and modify the qualitative structure of the rule base.

A reasonable strategy, following our general philosophy of incremental refinement of the rule base shown in Fig. 1, would be to try (a) or (b) and test the system; going on to (c) only if the performance is not satisfactory.

The application of this strategy in the case of our expert system for colon endoscopy produced a remarkable result. Our test for independence using the calculation of correlation coefficients did show that one pair of characteristics were correlated. Applying (a) above, we eliminated one of the characteristics, and found to our surprise that the performance of the system actually improved. At first sight this may seem paradoxical, because, if we eliminate one characteristic, we are using less data to make our decision. However it should be remembered that this extra data is only used via a false assumption of independence. It is thus better to neglect this data altogether than to add it in a way which is mathematically incorrect. This interesting result provides a strong endorsement for the principles of the QUALQUANT methodology, and particularly for the idea that all assumptions should be tested against data. In favourable circumstances such tests can lead to a modified system which is simplified as well as improved.

The QUALQUANT methodology involves two fundamental principles (objective probabilities, and testing) of the classical statistics of Fisher and Neyman–Pearson. The main controversy in the foundations of statistics over the last few decades has been between classical statistics and subjective Bayesianism, so that the above discussion is really just a special case of this debate. The subjective Bayesians constitute a significant minority, but most statisticians prefer the classical approach. This is probably for two interconnected reasons. First of all people always have more confidence in a model or procedure which has been thoroughly tested. Secondly most statisticians find the classical approach simpler in practice. This may well be connected with testing, since testing and the elimination of false assumptions can often simplify the model—as in the present case. At the moment most of those who adopt the probability approach to expert systems are subjective Bayesians. Our suggestion is that the use of objective probabilities and the principle of testing which have proved so effective in other areas of statistics may also be useful in expert systems—at least in some cases.

These points can be clarified by a brief comparison between our approach, and a good recent example of work in the paradigm of subjective Bayesianism [18]. Spiegelhalter and Lauritzen [18] point out that simplifying assumptions are necessary to make their procedure tractable, and go on to say: "In this paper we therefore explore in some detail the simplifying assumptions . . . ." [18, p. 582]. As usual these simplifying assumptions turn out to be independence assumptions. The first is that of global independence, and the second of local independence. They summarize their position as follows:

> An assumption of global independence of θ_vs allows global dissemination to be carried out locally. . . . Assumed local independence allows each conditional probability distribution to be individually updated. Each of these a priori assumptions only remains valid under certain sampling schemes. [18, p. 587]

The point we would make is this. Although Spiegelhalter and Lauritzen state that the *a priori* assumptions of independence are not always valid, they do not propose any tests for seeing whether these assumptions are valid, or even suggest that any such tests be performed. This of course is quite in accordance with the subjective Bayesian paradigm in which they are working. In the subjective Bayesian approach, prior probabilities are interpreted as initial degrees of belief. These are not tested, but are changed into posterior probabilities (i.e. degrees of belief revised in the light of evidence) by the process of Bayesian conditionalization. This is in sharp contrast to our own approach in which all assumptions of independence are conjectural, and have to be tested out against data to see if they really hold.

Spiegelhalter and Lauritzen remark that they: "... have not addressed the crucial area of criticism of the qualitative structure of the model..." [18, p. 601]. This is indeed a problem for the subjective Bayesian approach, since the process of Bayesian conditionalization does not in general alter the qualitative structure of the model, unless alternatives are built in at the beginning (*a priori*). To consider a whole range of alternatives at the beginning, however, is scarcely feasible, and may indeed be unnecessary if the initial assumption proves to be correct. This difficulty does not arise within the QUALQUANT methodology, since, as we stressed in Section 2, one response to the failure of a quantitative model to pass a test is to alter the qualitative structure of the model—perhaps through discussions with the domain expert.

4. Application of the QUALQUANT methodology

We will illustrate the application of our methodology via an expert system for colon endoscopy. This provides a good test case because of the high degree of uncertainty in the knowledge and data, and the availability of real data from many different cases.

Endoscopy is one of the tools available for diagnosis and treatment of gastrointestinal diseases. It allows a physician to obtain direct colour information of the human digestive system. The endoscope is a flexible tube with viewing capability (fiberoptic or video). It consists of a flexible shaft which has a manoeuvrable tip. The orientation of the tip can be controlled by pull wires that bend the tip in two orthogonal directions (left/right, up/down). It is connected to a cold light source for illumination of the internal organs and has an optical system for viewing directly through an eye piece or on a TV monitor. The instrument has more channels for transmitting air to distend the organ, for a water jet to clean the lens and for sucking air or fluid. The consultant controls the instrument by steering the tip with two mechanical wheels, and by pushing or pulling the shaft. The shaft is relatively torque-stable so that he can also apply rotatory movements to the tip. Also, he should control the air supply (inflate or aspirate) for good vision but without excessive air pressure, use the water jet for cleaning the lens when it is dirty, aspirate excess fluid, and realize the diagnostic or therapeutic objective of each particular case. The doctor inserts the instrument estimating the position of the colon centre (lumen) using several visual clues such as the darkest region, the colon muscular curves, the longitudinal muscle, and others. If the tip is not controlled correctly it can be very painful and dangerous to the patient, and could even cause perforations of the colon wall. This is further complicated by the presence of many difficult situations such as the contraction and movement of the colon, fluid and bubbles that obstruct the view, pockets (diverticula) that can be confused with the lumen and the paradoxical behaviour produced by the endoscope looping inside the colon. This requires a high degree of skill and experience that only an "expert" endoscopist will have.

A computer system is being developed to aid a physician in colonoscopy. The primary objective of the system is to help the doctor with the navigation of the endoscope inside the colon by controlling the orientation of the tip via the right/left and up/down controls. As well as a navigation system, it will also serve as an advisory system for learning endoscopists suggesting correct actions. It seems impossible to construct a general and complete model of the human colon, due to its complexity and wide range of variations in different persons. This difficulty points towards the use of expert system techniques to solve the endoscope navigation problem. The main sources of information for endoscope navigation are knowledge about the human colon (for interpretation) and expertise from the expert endoscopist (for planning and control).

The information provided by the visual input and the knowledge compiled from the expert are both incomplete and uncertain. Features obtained from the low-intermediate vision levels are uncertain due to several factors:

- There is not an exact model of the illumination provided by the endoscope inside the colon and we are restricted to a single camera.
- There is noise due to specularities and uneven texture.
- Image acquisition (camera and A/D convertion) distorts the images.
- The edge detection and segmentation techniques used at the lower levels are imprecise due to loss of information caused by poorly selected thresholds.

The expert knowledge is also heuristic, with "fuzzy" concepts (e.g. "*large dark* region") and imprecise rules (e.g. "there is evidence of *possible* diverticula").

The first step for endoscope navigation and advice is to recognize the important features in the images. These are the ones that the expert consultant uses to guide the endoscope inside the human colon. Two main characteristic objects have been found to be very useful, one due to the type of illumination of the endoscope and the second one to the anatomy of the colon. The darkest region generally corresponds to the lumen because there is a single light source close to the camera. The colon has a series of transverse folds (rings) which appear as circular or triangular occluding contours in the image and they also show the correct way to go. Khan and Gillies [4] have developed new methods for extraction of the dark region and the occluding contours (rings). For the dark region extraction they use a Quadtree representation in which the largest quadrant with a certain intensity level and variance is used as a seed region that is extended to "cover" the lumen area.

In the first stage of our advisory system, we are using the data from the dark region detection as main input to the expert system. The features we currently obtain from this process are:

- region size,
- mean intensity level,
- intensity variance in the region,
- location in the image.

This information from the present and past images will be used as a testbed for our knowledge base. Using the experts rules it will decide if the dark region proposed by the low-level vision is the lumen. If it is not (with certain confidence) or if no dark region is detected, it will try to find the position of the lumen by inference from the previous images. From its interpretation of the image it will use the control rules to give advice to the doctor.

A schematic of the system architecture is shown in Fig. 2. It is a modular system with two main components:

(a) *Feature extraction*, which includes the low- and intermediate-level vision modules [4] that obtain the main features from the image (regions and contours) which are integrated in the *symbolic image*. The objects in the symbolic images constitute the input to the expert system.

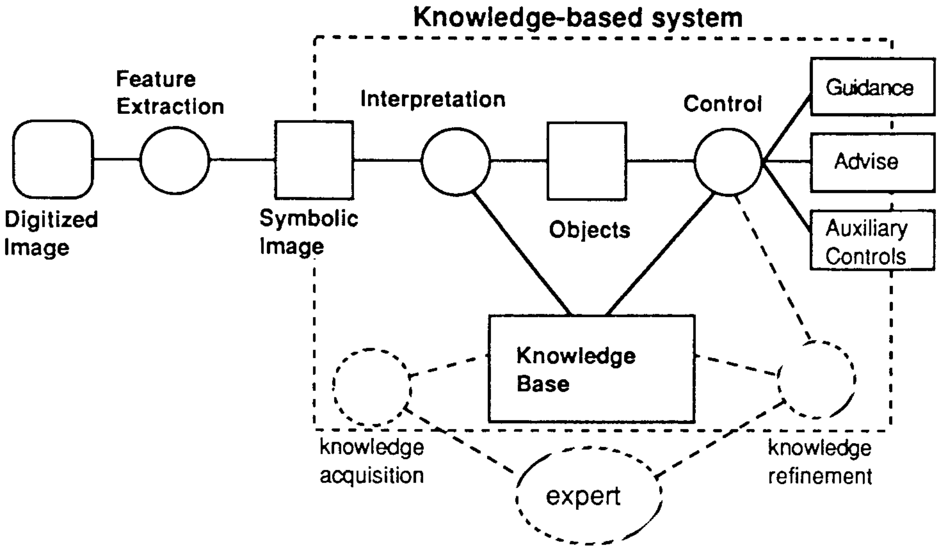

Fig. 2. System architecture.

(b) *The knowledge-based system*, which contains the expert system for advice and control. It is subdivided into:
   - interpretation: which labels the object in the *symbolic image*;
   - control: which is based on the information from the interpretation and the KB, and suggests actions for endoscope guidance, advice on the image and actions for the auxiliary controls (inflation/water wash);
   - knowledge base (KB): initially this contains the colonoscopy rules obtained from the expert. This is refined with data from the endoscopy images and aid from the expert.

The interpretation module has been implemented as a probabilistic network, while the control module is a standard logical inference engine. In the first stage, we have only dark region information from the feature extraction processes and have implemented the lumen recognition part of the KB. The development of the system has been based on incremental refinement. It consisted of three phases, namely:

- *Qualitative knowledge acquisition*. Extraction of the heuristics from expert colonoscopists and their implementation in first-order logic.
- *Estimation of probability distributions*. Statistical analysis of colon images and development of the probabilistic network.
- *Validation of independence assumptions*. Linear regression analysis of the variables and modification of the KB structure.

We start by getting the qualitative knowledge for endoscope navigation from an expert colonoscopist. We are mainly interested in the visual information he uses to guide the endoscope inside the colon. Then we transform this knowledge to probabilistic rules using data from videotapes of endoscopy sessions.

The colonoscopy knowledge-base (KB) consists of heuristics represented as IF–THEN rules mainly because they are closer to the endoscopist's conceptualization. Currently we have extracted about sixty rules. These can be classified in two types:

- *Interpretation rules*: these relate visual features to objects which are relevant to the physician.
- *Control rules*: the advice rules indicate the correct action to follow given an interpretation of the image.

Some examples of typical rules are shown in Table 1.

These sets of rules constitute our starting qualitative KB. A first prototype was implemented by transforming the rules into Prolog Horn clauses. This was tested with colon images but the performance was not satisfactory (Sucar and Gillies [19]). The main problem was that the knowledge was codified as logic rules and did not take into account the uncertain nature of the problem. So as a next stage we incorporated the handling of uncertainty by using objective

Table 1
Examples of colonoscopy rules.

Rules	Interpretation	Control
"normal"	IF: large, uniform, and dark region THEN: lumen	IF: lumen THEN: advance to centre of lumen
"special"	IF: brown or yellow region with straight interface THEN: fluid in colon	IF: fluid and tip below fluid level THEN: aspirate
"special"	IF: bright region surrounding dark region THEN: diverticula	IF: diverticula THEN: pull-back

probabilities. For this, the diagnostic part of the KB was transformed into a *probabilistic network*. This has some features in common with the *belief networks* considered by Pearl [12], but there are some differences as we will point out below. A part of the network, i.e. for lumen and diverticula recognition, is shown in Fig. 3. It basically represents three of the expert's original qualitative rules, namely:

(1) large, uniform dark region→lumen;
(2) several concentric rings→lumen;
(3) small dark region→diverticula.

The circles represent the objects and features, and the arrows the dependencies. Associated with each arrow there is a probability distribution for the conditional probability of the object at the *end* given the object at the *start*. For example the arrow from *Lumen* (L) to *Large Dark Region* (LDR) represents the probability of observing a large dark region given a lumen in the image ($P(LDR \mid L)$).

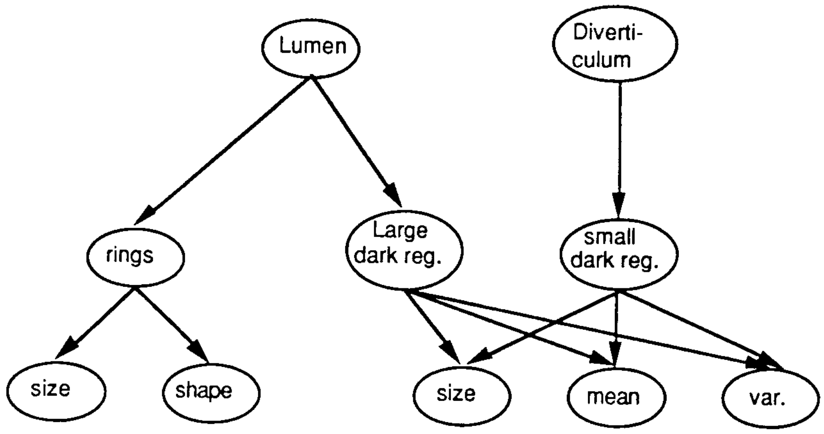

Fig. 3. A section of the probabilistic network for recognition.

We have structured the KB as a set of probabilistic trees (multitree or forest) so there is one tree that encapsulates the rules for recognition for each object. Each recognition tree consists of a single root that represents the object of interest, intermediate objects/features, and several leaves that constitute the measured parameters. This structure allows us to evaluate easily the posterior probability of each object given its prior probabilities and the conditional probabilities represented by each arrow in the tree. For this we initially assume that the *sons* (the objects at the end of the arrows) are *conditionally independent* with respect to their *parent*. Afterwards we will test this assumption.

Figure 4 shows the recognition tree for lumen restricted to dark region features, as well as the required probability distributions. Note that we only require the prior probability of the root. The measured inputs are S (the region size), M (the region's mean intensity), and V (the region's variance). The posterior probability is obtained by applying Bayes' theorem:

$$P(L \mid S, M, V) = \frac{P(L)P(S, M, V \mid L)}{P(S, M, V)} . \tag{3}$$

Putting this in terms of the intermediate object (LDR) we get:

$$P(L \mid S, M, V) = \frac{P(L)[P(S, M, V \mid L, LDR)P(LDR \mid L) + P(S, M, V \mid L, \neg LDR)P(\neg LDR \mid L)]}{P(S, M, V)} . \tag{4}$$

If we assume that S, M, and V are mutually independent given LDR as well as independent of L, we obtain:

$$P(L \mid S, M, V) = \frac{P(L)\left[\sum_i P(S \mid LDR_i)P(M \mid LDR_i)P(V \mid LDR_i)P(LDR_i \mid L)\right]}{P(S, M, V)} . \tag{5}$$

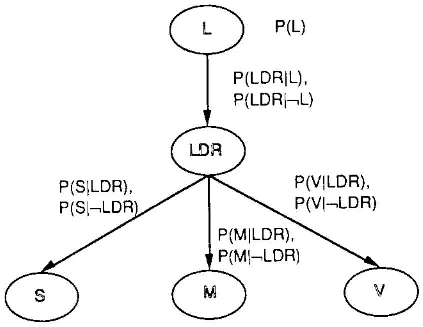

Fig. 4. Lumen recognition and related probabilities.

Similarly:

$$P(\neg L \mid S, M, V) = \frac{P(\neg L)\left[\sum_i P(S \mid LDR_i) P(M \mid LDR_i) P(V \mid LDR_i) P(LDR_i \mid \neg L)\right]}{P(S, M, V)} . \tag{6}$$

Given that (5) and (6) must sum to one, we can eliminate the term $P(S, M, V)$ and finally obtain:

$$P(L \mid S, M, V) = \frac{P(L)\left[\sum_i P(S \mid LDR_i) P(M \mid LDR_i) P(V \mid LDR_i) P(LDR_i \mid L)\right]}{k} . \tag{7}$$

Using (7) we can obtain the posterior probability of the lumen (root) given only its prior probability and the conditional probabilities indicated in Fig. 4, given that our independence assumptions are correct.

This procedure could be extended to a tree of any size and it allows us to *propagate* the probabilities directly from the leaves to the root. To maintain this computational simplicity we treat separately each object of interest and construct a similar tree for each, so at the end the KB is represented as a collection of such trees or a *multitree*.

Our method differs in two important aspects from the other *probabilistic networks* [6, 12, 18]:

(1) The prior and conditional probabilities required for evaluation are objective, that is obtained statistically from test data.
(2) Although we also assume conditional independence according to the network structure, we continuously test for it by using linear regression techniques and if there are dependencies the network structure is modified accordingly.

5. Experimental results

In one study we analyzed a random sample of about 300 colon images and obtained the required probability distributions from the data. We measured:

(1) the number of times a lumen appeared in the image;
(2) the number of times a lumen appeared but was not identified as a large dark region by the low-level vision system;
(3) the number of times a lumen appeared and was identified as a large dark region;

(4) the number of times the low-level vision system identified a large dark region which was not a lumen.

For each large dark region identified we also measured the size, mean and variance.

From these statistics we can easily obtain an objective estimate of the probability distributions required for lumen recognition (Fig. 4). In our example, the variables we are using have discrete values which are further quantised into small ranges and probabilities are computed for each of these. For *mean* and *variance* we divide the range of interest into m such ranges setting m to approximately $\sqrt{n}$, where n is the number of samples. Unlike the continuous case we do not make any assumptions about the initial distribution.

The final stage in our methodology is to validate the structure of our probabilistic network, that is basically checking if the independence assumptions we made are correct. For this we calculate the pair-wise correlation for all the variables, applying the Pearson's product moment (r) and Kendall's tau (τ) correlation coefficients which are defined by McPherson [9].

Both provide different measures of association between random variables. Pearson's correlation coefficient indicates if there is a linear relationship, and Kendall's correlation coefficient detects any increasing or decreasing trend curve present in the data. In the case of our example, we computed the correlation between *size*, *mean*, and *variance* conditioned over *large dark region* and *lumen*. We actually found a high correlation between *mean* and *variance*, so we modify our KB accordingly.

After the second phase, the recognition part had basically the structure shown in Fig. 4. The analysis of the colon images gave us a frequency estimate of the required probability distributions and enabled us to construct the first prototype of the *probabilistic network*. We still had to test its structure and for this we obtained the correlation between the features of the lumen which we were assuming independent. Table 2 summarizes the main results of this analysis.

There is a strong correlation between *mean* and *variance* which questions our independence assumption between these two variables, the other values seem relatively low so we will maintain that *size* is independent from the other two parameters. As a first step we eliminated one of the correlated variables. This resulted in two alternative trees for lumen recognition, one with only *size* and

Table 2
Correlation between lumen parameters.

Correlation coefficients	size and mean	size and variance	mean and variance
Pearson's r	−0.146	0.264	0.482
Kendall's τ	−0.089	0.116	0.342

Table 3
Performance results for different structures.

Configuration	Number of samples correct (percentage)			
	Lumen	¬Lumen	Diagnosis	Advice
All parameters (size, mean, variance)	89%	54%	86%	89%
Eliminate *variance* (size, mean)	93%	79%	91%	92%
Eliminate *mean* (size, variance)	97%	50%	92%	92%

mean and the other with *size* and *variance*. We then tested the system's performance for lumen recognition with these three possible configurations. To evaluate it we compared the results of the image interpretation and control to the experts opinion in a random sample of cases. For this we used a different sample of more than 130 colonoscopy images. These results are presented in Table 3.

As will be seen there is an improvement in all cases when the variance is eliminated, and this difference is highly significant in the case of ¬Lumen recognition. The elimination of the mean provides an improvement in lumen recognition which is significant with a confidence interval of just above 95%. In both alternative modifications, the image interpretation and suggestions to the physician were equal to the expert's opinion in a larger number of cases. Additionally, the network size was reduced and consequently object recognition was faster, so at the same time we obtain a more reliable and efficient system.

For a Bayesian approach to be practical, we usually need to assume independence or at least conditional independence. But as the previous example demonstrated, this should be done with caution. We should treat our independence assumption as conjectural, in the same way that we consider the subjective estimates of probabilities, and use real data to corroborate it or change the structure of the network if the results are not satisfactory.

6. Conclusions

We can now briefly recapitulate the main points in favour of our approach which involves the QUALQUANT methodology and the use of objective probabilities.

First of all the approach is easy to use. Experts are only asked to provide a qualitative orientation, and this they will in general find easy to do. In most cases experts will not be required to provide any quantitative estimates, and, even if they have to do so, these estimates are treated as provisional and subject to revision as more data is collected.

Secondly the approach has a sound theoretical foundation in Popper's falsificationist methodology, and the theory of objective probability. The views of experts are treated with respect, but are still subjected to criticism and experimental testing. This cannot but result in improvement, and in more reliable systems.

Thirdly the approach makes some progress in solving what are recognised as the principal problems in the use of probabilities to handle uncertainty in expert systems. The first such problem concerns the large number of probabilities which apparently need to be evaluated. Here our suggestion is to use qualitative suggestions from a domain expert to limit consideration to relatively small groups of characteristics which are known to be relevant to the question in hand. The second main problem concerns the need to make assumptions of independence (or conditional independence) to render the calculations tractable. The difficulty is that these independence assumptions may not be valid. Here our suggestion is that independence assumptions can be made (perhaps in consultation with the domain expert) provided they are regarded as conjectures which must be tested out against the data. This testing methodology was strongly vindicated in the example given, since the discovery of significant correlation between two characteristics enabled us to simplify and improve the system merely by eliminating one of the characteristics. Naturally it is not to be expected that every case of dependence will be so easily dealt with, and in fact we are at present working on the elaboration of further, more complicated, strategies for handling cases of strong correlation. Even as it stands at present, however, we can recommend our approach to workers in the field of expert systems.

Acknowledgements

The authors would like to thank CONACYT, Mexico and the Olympus Optical Company, Tokyo for the financial support that made this research possible.

References

[1] B. De Finetti, Foresight: its logical laws, its subjective sources (1937); reprinted in: H.E. Kyburg and H.E. Smokler, eds., *Studies in Subjective Probability* (Wiley, New York, 1964) 95–158.

[2] M.R. Genesereth and N.J. Nilsson, *Logical Foundations of Artificial Intelligence* (Morgan Kaufmann, Los Altos, CA, 1987).

[3] D.A. Gillies, *An Objective Theory of Probability* (Methuen, London, 1973).

[4] G.N. Kahn and D.F. Gillies, A highly parallel shaded image segmentation method, in: P.M. Dew and R.A. Earnshaw, eds., *Parallel Processing for Vision and Display* (Addison-Wesley, Reading, MA, 1989) pp. 180–189.

[5] T.S. Kuhn, The function of measurement in modern physical science (1961); reprinted in: T.S. Kuhn, *The Essential Tension* (University of Chicago Press, Chicago, IL, 1977) 178–224.

[6] S.L. Lauritzen and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, *J.R. Stat. Soc. B* **50** (2) (1988) 157–224.

[7] D.V. Lindley, Scoring rules and the inevitability of probability, *Int. Stat. Rev.* **50** (1982) 1–26.

[8] D.V. Lindley, The probability approach to the treatment of uncertainty in artificial intelligence and expert systems, *Stat. Sci.* **2** (1) (1987) 17–24.

[9] G. McPherson, *Statistics in Scientific Investigation* (Springer, New York, 1990).

[10] D. Michie, The state of the art in machine learning, in: D. Michie, ed., *Introductory Readings in Expert Systems* (Gordon and Breach, New York, 1982) 208–229.

[11] K. Ng and B. Abramson, Uncertainty management in expert systems, *IEEE Expert* **5** (2) (1990) 29–48.

[12] J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, CA, 1988).

[13] K.R. Popper, The propensity interpretation of probability, *Brit. J. Philos. Sci.* **10** (1959) 25–42.

[14] K.R. Popper, *Conjectures and Refutations* (Routledge and Kegan Paul, London, 1963).

[15] K.R. Popper, *Realism and the Aim of Science* (Hutchinson, London, 1983).

[16] F.P. Ramsey, Truth and probability (1926); reprinted in: H.E Kyburg and H.E. Smokler, eds., *Studies in Subjective Probability* (Wiley, New York, 1964) 63–92.

[17] G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976).

[18] D.J. Spiegelhalter and S.L. Lauritzen, Sequential updating of conditional probabilities on directed graphical structures, *Networks* **20** (5) (1990), 579–605.

[19] L.E. Sucar and D.F. Gillies, Knowledge-based assistant for colonoscopy, in: *Proceedings Third International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems* (1990) 665–672.

[20] L.A. Zadeh, Fuzzy sets, *Infor. Control* **8** (1965) 338–353.