Contextual Decision Making With Degrees of Belief

Robert M. Haralick
Intelligent Systems Laboratory
Department of Electrical Engineering
University of Washington
Seattle, WA 98195

Abstract

This paper gives a brief overview of the classical contextual pattern recognition problem. We show that the difficulty of this problem is really associated with the determination and use of the support of the joint prior distribution of the category labels. We indicate how the consistent labeling framework can be used to define the support of the joint prior. Then we show that this formulation of the problem can be generalized and we bring in a general propostional logic framework which not only defines the support of the joint prior but which also permits a calculation to be made evaluating the joint prior for any given set of joint labelings. We show that this formulation is indeed a formulation relating to the degree of belief. We develop a formal system for the degree of belief in terms of an operational probability meaning. The degree of belief in a proposition is exactly the probability with which one can assert the proposition. We then show how the classical contextual problem can be generalized in the belief framework.

1 Statement of Problem

Let the universe be divided up into recognizable and measurable pieces which we call units. Let U be such a set of units. Each unit in U can be characterized by

- its relationships with other units,
- an n-tuple of measurements determined by some local measuring process, and
- the appropriate category interpretation for the unit.

We call each category interpretation a label. We denote by L the set of possible category interpretations for a unit.

The Bayesian framework poses the labeling problem as follows: given the measurement n-tuple made on each unit, and given the prior world knowledge Q which specifies allowable category interpretations for each group of related units, determine the functional assignment, f, $f : U \rightarrow L$ having highest probability of being correct given the sets of measurements and relationships. This is a pattern recognition problem involving context. Further discussion of this problem statement can be found in Haralick (1983).

2 Bayesian Model

We begin our discussion of this problem by describing the general Bayesian model. Corresponding to a unit u_i, there is its assigned interpretation z_i, and its measurement n-tuple x_i. Given the n-tuple for each unit and the world model Q, we would like to assign labels $z_1, \ldots, z_M$ to units $u_1, \ldots, u_M$ respectively, which maximize $P(z_1, \ldots, z_M | x_1, \ldots, x_M, Q)$, the probability that the assigned labels are the true labels given the information we have about the units.

By Bayes theorem, maximizing

$$P(z_1, \ldots, z_M | x_1, \ldots, x_M, Q)$$

is equivalent to maximizing

$$P(z_1, \ldots, z_M, x_1, \ldots, x_M | Q) = P(x_1, \ldots, x_M | z_1, \ldots, z_M, Q) P(z_1, \ldots, z_M | Q).$$

We make the following assumptions about the world and unit measurement process. The first assumption states that the description process is local. When the unit u_i is being examined, no characteristics from any other unit but unit u_i affects the description obtained from unit u_i. Hence:

$$P(x_1, \ldots, x_M | z_1 \ldots, z_M, Q) = \prod_{i=1}^{M} P(x_i | z_1, \ldots, z_M, Q).$$

The second assumption states that the n-tuple measurement of unit u_i depends only on the interpretation associated with unit u_i and does not depend upon any relationships unit u_i may have with other units or upon the interpretations associated with any other unit. Thus,

$$P(x_i | z_1, \ldots, z_M, Q) = P(x_i | z_i).$$

Hence, the optimal decision rule determines interpretations $z_1, \ldots, z_M$ for units $u_1, \ldots, u_M$ which maximize

$$P(z_1, \ldots, z_M, x_1, \ldots, x_M | Q) = P(z_1, \ldots, z_M | Q) \prod_{i=1}^{M} P(x_i | z_i).$$

The assumption often made in pattern recognition without context is that the units themselves are independent. The interpretation given to any one unit does not constrain the interpretation given to any other unit. In this case:

$$P(z_1, \ldots, z_M|Q) = \prod_{i=1}^{M} P(z_i)$$

and the best decision procedure is to give interpretation z_i to unit u_i where label z_i maximizes $p(x_i|z_i)P(z_i)$:

$$P(x_i|z_i)P(z_i) \geq P(x_i|z)P(z) \quad \forall z.$$

This assumption is clearly inappropriate in pattern recognition problems which have a rich context.

A weaker assumption is the Markov assumption used when the units are linearly ordered and the interpretation of any unit depends only on the interpretation of the previous unit in the order. That is:

$$P(z_i|z_1, \ldots, z_{i-1}) = P(z_i|z_{i-1}) \quad \forall i.$$

Then using the identity

$$\begin{aligned} P(z_1, \ldots, z_M) &= P(z_M, |z_1, \ldots, z_{M-1}) \times \\ &\quad P(z_{M-1}|z_1, \ldots, z_{M-2}) \times \\ &\quad \ldots \times P(z_2|z_1)P(z_1) \end{aligned}$$

we obtain

$$P(z_1, \ldots, z_M) = \prod_{i=1}^{M} P(z_i).$$

In this case, the best decision proceedure chooses interpretations $z_1, \ldots, z_M$ which satisfy the maximality condition:

$$\prod_{i=1}^{M} P(x_i|z_i)P(z_i|z_{i-1}) > \prod_{i=1}^{M} P(x_i|z'_i)P(z'_{i-1}) \quad \forall (z'_1, \ldots, z'_M).$$

The choice of $z_1, \ldots, z_M$ satisfying the maximality condition is a dynamic programming problem (Bellman and Dreyfus, 1963, Forney, 1973). For most pattern recognition problems with context, this too is an inappropriate assumption.

Problems with context require a more sensitive way of handling the prior probability $P(z_1, \ldots, z_M|Q)$. To this end, Duda and Hart (1963), implicitly assumed that the global interpretation $(z_1, \ldots, z_M)$ for units $(u_1, \ldots, u_M)$ is either allowable or not allowable and all allowable global interpretations have equal probability. This is an equal probability of ignorance assumption. But it is one that applies to the entire context and not to each unit individually. This kind of methodology is often used (it is the lexicon or the dictionary) to improve the performance of OCR. Thus, if $A \subset L^M$ is the set of allowable global interpretations, we have:

$$P(z_1, \ldots, z_M|Q) = \begin{cases} \frac{1}{\#A}, & \text{if } (z_1, \ldots, z_M) \in A \\ 0, & \text{if } (z_1, \ldots, z_M) \notin A. \end{cases}$$

Under the context equal probability of ignorance assumption, the optimal decision rule determines interpretations $z_1, \ldots, z_M$ for units $u_1, \ldots, u_M$ which maximize

$$\begin{aligned} &P(z_1, \ldots, z_M, x_1, \ldots, x_M|Q) = \\ &\qquad P(z_1, ..., z_M|Q) \prod_{i=1}^{M} P(x_i|z_i) \\ &\quad = \begin{cases} \frac{1}{\#A} \prod_{i=1}^{M} P(x_i|z_i), & \text{if } (z_1, \ldots, z_M) \in A \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The brute force algorithm to solve this optimization problem must then go sequentially through all consistent interpretations M-tuples $(z_1, \ldots, z_M)$ of A and for each one evaluate $\prod_{i=1}^{M} P(x_i|z_i)$ to find that consistent labeling which maximizes $\prod_{i=1}^{M} P(x_i|z_i)$.

In order to use such a proceedure, we must have an easy way to determine whether or not a labeling $(z_1, \ldots, z_M)$ belongs to A. For this purpose we employ a model involving unit to unit relationships. Two units may be related or not related. If they are related then the possible interpretations for one may constrain or restrict the possible label interpretations for the other. We denote by T the set of all pairs of units which constrain one another; $T \subset U \times U$. We call T the unit constraint relation.

We must next represent the label constraints for those units which do constrain one another; We do this by the relation $R \subset (U \times L)^2$. Category interpretations m and n for units u and v are legal if and only if $(u, m, v, n,) \in R$. We call R the unit-label constraint relation.

If it is appropriate for the unit constraint to be higher order than binary, we can let $T \subset U^N$ and $R \subset (U \times L)^N$. In either the binary or the general higher order case, we can represent the set A as consisting of all M-tuples $(f(u_1), \ldots, f(u_M))$ of labels such that if $(v_1, \ldots, v_N) \in T$, then $(v_1, f(v_1), \ldots, v_N, f(v_N)) \in R$. We call A the set of all (T, R) - consistent labelings. See Haralick and Shapiro (1979, 1980) for more details on the posing of consistent labeling problems and their solution.

The brute force algorithm for optimizing $P(z_1, \ldots, z_M, x_1, ..., x_M|Q)$ is surely computationally expensive because of the exponentially large number of consistent labelings we expect to find in A. Modifying the algorithm to a branch and bound search is helpful, but perhaps not enough.

One way of reducing the computational problem is to solve a suboptimal problem rather than the optimal problem. For example, we could limit the legal label interpretations for unit u_i to only those interpretations z_i satisfying $P(x_i|z_i) \geq \theta_i$. We can set θ_i to always accept a given number of interpretations. In this case we can define R by:

$$R = \{(u_i, z_i u_j, z_j) \in (U \times L)^2 | \begin{cases} (z_i, z_j) & \text{is legal for } (u_i, u_j) \\ P(x_i|z_i) \geq \theta_i & \\ P(x_j|z_j) \geq \theta_j & \end{cases} \}$$

Because the unit-label constraint relation R defined this way has a smaller number of tuples, the number of (T, R) consistent labelings must be smaller too. Thus, we expect that going through all the labelings in the set A defined with the modified R should be an easier task.

There are other ways in which a relation like R can come about. Each tuple $(v_1, c_1, \ldots, v_N, c_N)$ in a generalized R has the meaning that units $(v_1, \ldots, v_N)$ can have labels $(c_1, \ldots, c_N)$ If we let q_n be the statement label c_n is legal for unit v_n then the tuple $(v_1, c_1, \ldots, v_N, c_N)$ in R just means the assertion q_1 and q_2 and ... and q_N is a possible assertion to use in constructing a labeling. If an assertion is possible to use, then it can be considered as a piece of evidence. The unit-relation constraint R then is just a set of possible assertions which exist in the body of evidence. Assertions in the body of evidence need not be conjunctions. For example the assertion $q_i \rightarrow q_j$ can also be an assertion in a body of evidence. In the next section we develop this idea, contrasting it with the Dempster-Shafer (Shafer,1981) evidential reasoning.

3 Probabilistic Basis for Belief

In this section we describe an operational probabilistic basis for belief: what we believe, we make use of; what we do not believe we make no use of. Hence the degree of belief associated with a proposition in a body of evidence is related to the chance probability that an inference mechanism has access to the proposition for use in making an inference. In this model, propositions in a body of evidence are treated as random propositions. A random proposition has two states: assertable and not assertable. A random proposition whose chance state is assertable is available for making inferences. A random proposition whose chance state is not assertable is not available for making inferences. An assertable proposition is considered to be true. A proposition which is not asserted is neither considered to be true or false. A proposition whose chance state is not assertable is just considered to not exist. The degree of belief in a proposition in a body of evidence is then the chance probability that it is assertable. The degree of belief in a proposition inferable from a body of evidence is the conditional probability that a logical inference mechanism is able to semantically infer the proposition from the conjunction of chance suppositions in the body of evidence, given that the conjunction of assertions is not self-contradictory.

The operational probabilistic model for belief described here is then simultaneously a generalization of Shafer's theory of evidence and a Bayesian approach to belief. To understand this, note that Shafer's frame of discernment contains a set of possibilities, exactly one of which corresponds to the truth. Our body of evidence contains a set of atomic propositions as well as possibly other well-formed formulas of propositions. There is no requirement that the propositions correspond to a set of possibiliites, exactly one of which is true. Shafer uses Dempster's rule of combination which is a generalization of Bernoulli's rule of combination as the mechanism to pool and combine evidence. We will demonstrate how Dempster's rule of combination is a specialization of the natural combination rule when belief in a proposition is considered as the chance probability that the proposition is semantically implied from a set of non–contradictory premises.

We begin our development by considering a model of the legal system, one of whose essential functions is to facilitate the establishment of evidence and the combining of evidence to help a jury decide what to believe. We conclude from this analysis that conditional probability of hypothesis given evidence cannot be used as the basis of a belief system calculus.

3.1 The Legal Court Paradigm

There is an argument or question about h. In order for a court to make a judgment about whether h is true or false, there are two attorneys A and B, a jury to make the judgment, and a judge to administer and make sure that the jury is provided fair and relevant information. Attorney A is paid to uncover, organize, and present all evidence and inferences which are supportive of h. Attorney B is paid to uncover, organize, and present all evidence and inferences which are supportive of $\bar{h}$, the negation of h.

Evidence consists of physically observed facts and statements made by witnesses in response to the questions by the attorneys. The jury listens to all the evidence and all inferences made by the attorneys based on the evidence. In addition, the jury may make inferences from the evidence not stated by the attorneys. Then the jury collectively determines the weight of evidence in favor of h and the weight of evidence in favor of $\bar{h}$ and makes a judgment of h or $\bar{h}$ accordingly.

What is happening here? The cases made by the two attorneys correspond to the establishment of a set of elementary propositions which constitute the evidence. Associated with each elementary proposition in the evidence, the jury collectively associates an elementary or initial belief. In our model, this initial degree of belief is the probability with which the elementary proposition is asserted and, therefore, available for use in an inference process.

Attorney A organizes an argument from a selected consistent subset of the body of evidence to show that h can be semantically implied from the body of evidence. Attorney B organizes an argument from a different selected subset of the body of evidence to show that $\bar{h}$ can be semantically implied from the body of evidence.

The attorneys' job is not just to present and select evidence, but to present and select jury–believable evidence. Evidence, for example, which is second–hand is considered hearsay. Juries should not find it believable. Therefore, it is legally inadmissable. Statements which cannot be inferred from the evidence is jury–unbelievable. Therefore, the attorneys must present all the evidence relative to the inferences which they would like to make.

Since h and $\bar{h}$ are contradictory, there must obviously be inconsistencies between the selected subsets of propositions in support of h and $\bar{h}$. Both subsets cannot be simultaneously believed. Indeed, the dependencies of each subset on the other influence the degree of belief in each subset and the inferences which might be drawn from each subset.

A theory for evidential reasoning, if it follows the pattern of the legal paradigm, must then have the following five features:

- Each piece of evidence is a proposition.
- The question to be decided is a proposition.
- Each proposition in the body of evidence has associated with it a measure of its assertability.
- There must be a logic calculus for inferring propositions from conjunctions of other propositions.
- There must be a belief calculus for computing a degree of belief

for each proposition the inference calculus is able to infer. The degree of belief in an inferred proposition will depend on the measure of assertability for the proposition in the body of evidence.

It is natural that the logic calculus be based on symbolic logic. A proposition r can be inferred from proposition $q_1, \ldots, q_N$ if and only if r is semantically implied by the proposition $q_1, \ldots, q_N$. It is also natural that the degree of belief in a proposition q should be equal to the degree of belief in a proposition r if q and r are logically equivalent.

The commitment to these two constraints has an important consequence: the mechanism which computes a measure of belief for any inferred proposition cannot use conditional probability as its degree of belief in the proposition $q \rightarrow r$. For belief in $q \rightarrow r$, must be equal to belief in $\overline{r} \rightarrow \overline{q}$, since the two are logically equivalent. Yet $P(r|q)$ is independent of $P(\overline{q}|\overline{r})$.

To insure consistency in belief values, belief in $q \rightarrow r$ must be taken to be belief in $\overline{q} \vee r$ which is the logical definition of $q \rightarrow r$. How to compute belief in $\overline{q} \vee r$ is discussed in the next sections on the belief calculus.

3.2 The Belief Calculus

The belief calculus is a set of rules from which the degree of belief in any inferred proposition can be computed from the measure of assertability of propositions in the body of evidence. Our definition for degree of belief is that the degree of belief in our inferred proposition is the conditional probability of the chance inference of the proposition given a chance state of non–contradictory propositional assertions. The key to understanding the belief calculus then revolves around the meaning of chance state of propositional assertions and the way to compute its probability.

We found the belief calculus on the following stochastic model. Each proposition in the body of evidence can either be in the state of asserted or the state of non–asserted. The probability with which a proposition in the body of evidence can be asserted is what we have earlier called its measure of assertability. The body of evidence itself consists of the sets of its propositions each associated with its probability of being asserted. A state of assertion for E is given by a subset S of E. The propositions in the subset S are considered to be asserted and the propositions in the subset $E - S$ are considered to be non–asserted. We consider that the propositions in E are mutually independent insofar as their assertability is concerned. The probability for the chance state of assertion S can then be computed by

$$\prod_{q \in S} m(q) \prod_{r \in E-S} [1 - m(r)].$$

As there are $2^{\#E}$ subsets of E, there are $2^{\#E}$ chance states of assertions. Each chance state of assertion S contains propositions which are considered to be asserted. Let C be the collection of those states of assertions from which the contradiction cannot be inferred. Let H be the collection of those states of assertions from which the proposition h can be inferred. Then the degree of belief in h is defined by

$$Bel\ (h) = \frac{m(H \cap C)}{m(C)}$$

$$= \frac{\sum_{S \in H \cap C} \prod_{q \in S} m(q) \prod_{r \in E-S} [1-m(r)]}{\sum_{S \in C} \prod_{s \in S} m(s) \prod_{r \in E-S} [1-m(r)]}$$

We now go through a variety of examples to concretely illustrate this calculation so that we can fully appreciate its meaning and consequences.

3.2.1 Examples

First, we consider the case where the body of evidence consists only of the propositions q and $q \rightarrow r$. There are four possible states of assertion:

q	$q \rightarrow r$	
A	A	$\Rightarrow r$
A	NA	
NA	A	
NA	NA	

where we use A to designate the state "asserted" and NA to designate the state "not asserted." Note again that a proposition whose state is "not asserted" is not a proposition which is denied. For a proposition which is denied is one whose negation is asserted. To be not asserted just means that it is unavaliable to the inference mechanism.

The only subset of $E = \{q,\ q \rightarrow r\}$ from which the inference r can be made is $S = \{q,\ q \rightarrow r\}$. Since no subset of E is contradictory, we can compute

$$Bel\ (r) = m(q)\ m(q \rightarrow r).$$

Next suppose that the question is the proposition q and the entire body of evidence consists of the conjunction qr; $E = \{qr\}$.

qr	
A	$\Rightarrow q$
NA	

Again, the only subset of $E = \{qr\}$ from which the inference r can be made is $S = \{qr\}$ and since S is non–contradictory, we have

$$Bel\ (r) = Bel\ (qr).$$

Now suppose that the question is the proposition $q \vee r$ and the entire body of evidence consists of the propositions q, r; $E = \{q, r\}$.

q	r	
A	A	$\Rightarrow q \vee r$
A	NA	$\Rightarrow q \vee r$
NA	A	$\Rightarrow q \vee r$
NA	NA	

The subsets from which the inference $q \vee r$ can be made are $S_1 = \{q\}$, $S_2 = \{r\}$, and $S_3 = \{q, r\}$. Since each of these is non-contradictory, the chance probability that $q \vee r$ can be inferred is

$$Bel\ (q \vee r) = m(q) + m(r) - m(q)\ m(r).$$

This is not unlike Bernoulli's rule of combination. Some evidence can be supportive of other evidence. Consider the case where the evidence consists of $q, r, q \vee r$. The question is the proposition $q \vee r$. Again, there are no states of assertion which are contradictory and the only state of assertion from which the inference $q \vee r$ cannot be drawn is $\emptyset$. Hence,

$$\begin{aligned} Bel\ (q \vee r) &= 1 - [1 - m(q)][1 - m(r)][1 - m(q \vee r)] \\ &= m(q) + m(r) - m(q)\ m(r) + \\ &\quad \{1 - [m(q) + m(r) - m(q)\ m(r)]\} \times m(q \vee r). \end{aligned}$$

This is clearly greater than what the degree of belief in $q \vee r$ would be if the supporting evidence $q \vee r$ were not present.

The simplest case where the body of evidence contains two conflicting propositions is where $E = \{q, \bar{q}\}$. Consider the case where the question is q. The subsets from which the inference q can be made is only the subset $\{q\}$. The non-contradictory subsets are $\{q\}, \{\bar{q}\}$ and $\emptyset$. Letting $a = m(q)[1 - m(\bar{q})] + m(\bar{q})[1 - m(q)] + [1 - m(q)][1 - m(\bar{q})]$ we obtain

$$\begin{aligned} Bel\ (q) &= \frac{m(q)[1 - m(\bar{q})]}{a} \\ &= \frac{m(q)[1 - m(\bar{q})]}{1 - m(q)m(\bar{q})} \end{aligned}$$

This form is like Demptster's rule of combination. The numerator is the sum of products of the chance probability of non-contradictory assertions which semantically imply the proposition q and the denominator is one minus the product of the chance probability of assertions which are contradictory. By rearranging there results

$$Bel\ (q) = m(q) - \frac{m(q)m(\bar{q})[1 - m(q)]}{1 - m(q)\ m(\bar{q})}.$$

Thus conflicting evidence can make a belief of an inferred proposition have a lower value than the probability of its assertability.

Next we consider a case where there is a more complex conflict dependency among the propositions of the evidence. Suppose the body of evidence consists of q, r, and $q \rightarrow \bar{r}$. The question is the proposition $q \vee r$.

q	r	$q \rightarrow \bar{r}$	
A	A	A	$\Rightarrow f$
A	A	NA	$\Rightarrow q \vee r$
A	NA	A	$\Rightarrow q \vee r$
A	NA	NA	$\Rightarrow q \vee r$
NA	A	A	$\Rightarrow q \vee r$
NA	A	NA	$\Rightarrow q \vee r$
NA	NA	A	
NA	NA	NA	

There are five non-contradictory states of assertions from which $p \vee r$ can be inferred. They are $\{q, r\}$, $\{q, q \rightarrow \bar{r}\}$, $\{q\}$, $\{r, q \rightarrow \bar{r}\}$, $\{r\}$. The contradictory state is $\{q, r, q \rightarrow \bar{r}\}$. Hence we have

$$\begin{aligned} Bel\ (q \vee r) &= \{m(q)m(r)[1 - m(q \rightarrow \bar{r})] + \\ &\quad m(q)[1 - m(r)]m(q \rightarrow \bar{r}) + \\ &\quad m(q)[1 - m(r)][1 - m(q \rightarrow \bar{r})] + \\ &\quad [1 - m(q)]m(r)m(q \rightarrow \bar{r}) + \\ &\quad \frac{[1 - m(q)]m(r)[1 - m(q \rightarrow \bar{r})]\}}{[1 - m(q)m(r)m(q \rightarrow \bar{r})]} \\ &= \frac{\{m(q) + m(r) - m(q)m(r)[1 + m(q \rightarrow \bar{r})]\}}{[1 - m(q)m(r)m(q \rightarrow \bar{r})]} \\ &= m(q) + m(r) - m(q)m(r) - \\ &\quad \frac{[1 - [m(q) + m(r) - m(q)m(r)]]m(q)m(r)m(q \rightarrow \bar{r})}{1 - m(q)m(r)m(q \rightarrow \bar{r})} \end{aligned}$$

This shows again that when conflicting relevant evidence exists, the degree of belief in a proposition will be less than what it would be if the conflicting evidence did not exist.

For our next example where there is a case of conflict, suppose the evidence

$$E = \{q, q \rightarrow r, q \rightarrow \bar{r}\}$$

and the question is r.

q	$q \rightarrow r$	$q \rightarrow \bar{r}$	
A	A	A	$\Rightarrow f$
A	A	NA	$\Rightarrow r$
A	NA	A	
A	NA	NA	
NA	A	A	
NA	A	NA	
NA	NA	A	
NA	NA	NA	

Here we have

$$Bel\ (r) = \frac{m(q)\ m(q \rightarrow r)[1 - m(q \rightarrow \bar{r})]}{1 - m(q)\ m(q \rightarrow r)\ m(q \rightarrow \bar{r})}.$$

From the form of the above relation, we can see that if the degree of assertablility for $q \rightarrow \bar{r}$ is one, then the belief in r must be zero. And this happens even if the degree of belief in $q \rightarrow r$ is also one. The largest $Bel\ (r)$ can be is $Bel\ (q) Bel\ (q \rightarrow r)$.

For a final example where there is a case of conflict, suppose $E = \{q, q \rightarrow r, s, s \rightarrow \bar{r}\}$ and the question is $q \rightarrow \bar{s}$.

q	$q \rightarrow r$	s	$s \rightarrow \bar{r}$	
A	A	A	A	$\Rightarrow f$
A	A	A	NA	
A	A	NA	A	$\Rightarrow g \rightarrow \bar{s}$
A	A	NA	NA	
A	NA	A	A	
A	NA	A	NA	
A	NA	NA	A	
A	NA	NA	NA	
NA	A	A	A	$\Rightarrow q \rightarrow \bar{s}$
NA	A	A	NA	
NA	A	NA	A	$\Rightarrow q \rightarrow \bar{s}$
NA	A	NA	NA	
NA	NA	A	A	
NA	NA	A	NA	
NA	NA	NA	A	
NA	NA	NA	NA	
A	NA	NA	A	

Here we have

$$Bel\ (q \rightarrow \bar{s}) = \frac{m(q \rightarrow r)m(s \rightarrow \bar{r})[1 - m(q)m(s)]}{1 - m(q \rightarrow r)m(s)m(q)m(s \rightarrow \bar{r})}.$$

So if, for example, $m(s) = .9$, $m(q) = .2$, $m(q \rightarrow r) = .8$, $m(s \rightarrow \bar{r}) = .9$, then

$$\begin{aligned} Bel\ (q \rightarrow \bar{s}) &= \frac{.8(.9)[1 - .2(.9)]}{1 - (.8)(.9)(.2)(.9)} \\ &= \frac{.5904}{.8704} = .6783 \end{aligned}$$

while if the probability of the assertability of q is raised to .7 so that $m(2) = .9$, $m(q) = .7$, $m(q \rightarrow r) = .8$, and $m(s \rightarrow r) = .9$, then

$$\begin{aligned} Bel\ (q \rightarrow \bar{s}) &= \frac{.8(.9)[1 - .7(.9)]}{1 - .8(.9)(.7)(.9)} \\ &= \frac{.2664}{.5464} = .4876 \end{aligned}$$

These conflict examples illustrate that the degree of belief in an inferred proposition is not just a function of the degree of the assertability of the proposition from which the inference stems. It is also a function of the context of those propositions in the body of evidence.

4 Putting It All Together

We started with a statement of the pattern recognition contextual decision making problem, showing that its solution depended on how the support of the joint prior $P(z_1, \ldots, z_M)$ was represented. We discussed representing this support in the unit-label constraint relation of the consistent labeling problem. There we saw that tuples in the unit-label constraint relation were just statements which were conjuncions whose terms were of the form "it is legal for unit v to take label c". No probability was associated with this conjunction. Then we generalized the kinds of assertions that might be consistent with the support of $P(z_1, \ldots, z_M)$, allowing any kind of propositional form. In addition we associated with each such proposition a probability. The probability had the meaning that it was the frequency with which the proposition could be used. We then argued that this was indeed the meaning of degree of belief.

This generalization permitted us to actually generalize the contextual decision making problem. Each proposition of the form "label c is legal for unit u_m" constitutes a piece of evidence with an associated belief value $P(x_m|z_m)$ which depends on the locally observed data associated with unit u_m. Contextual information, global world knowledge, consists of a set of propositions each one of which is formed by the rules of logic from the elementary propositions "label c is legal for unit u_m" Each such contextual proposition is also a piece of evidence and has associated with it a degree of belief which does not depend on any locally observed data.

One important component to the generalized contextual decision making problem then is to determine the probability with which any given proposition formed from the elementary propositions and other propositions in the body of evidence can be asserted. We have argued that the probability with which it can be asserted is the degree of belief we have in the proposition. The contextual pattern recognition problem with which we started only considered propositions of a complete conjunctive form of the simple propositions in the body of evidence. Complete here means that each unit u has associated with it some term of the form "label c is legal for unit u" in the conjunction.

So we have been able to layout a mechanism for determining the probability of any complex proposition about the labels for the units given the observed data. And we showed how this probability can be interpreted as a degree of belief arising from the probabilities $P(x_m|c)$ associated with the simple statements "label c is legal for unit u_m" and all the other contextual probabilities of more complex statements about what labels are legal for various subsets of units. If we now want to explain the observed data by making some statements about the labels of the units, we can do that in a form which does not necessarily have to involve all the units. We can explain that part of the context we are interested in and explain it in a more general way than simple conjunctions. Having such great freedom about the kind of contextual statement that can be made, we must either use this mechanism in a mode where a hypothesis is given and the probability that the hypothesis is true must be evaluated, or we must find a way to generate the high probability statements which could then be given to a hypothesis generator.

References

[1] Bellman, R. and S. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton, N.J., 1963.

[2] Duda, R. and P. Hart, "Experiments in the Recognition of Handprinted Text: Part II - Context Analysis,", Proc. AFIPS COnf., Vol. 33, 1968, pp. 1139-1149.

[3] Forney, G., "The Viterbi Algorithm,", Proc. IEEE, Vol. 6, 1973, pp. 268-278.

[4] Haralick, R. M. "Decision Making In Context," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 5, 1983, pp. 417-428.

[5] Haralick, R. M. and L. G. Shapiro, "The Consistent Labeling Problem I" IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 1, 1979, pp. 178-184.

[6] Haralick, R. M. and L. G. Shapiro, "The Consistent Labeling Problem II" IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 2, 1980, pp. 192-203.

[7] Shafer, Glen "Constructive Probability,", Synthese, 1981, pp. 1-60.