Diagnostic reasoning based on a genetic algorithm operating in a Bayesian belief network

E.S. Gelsema *

Department of Medical Informatics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Received 15 April 1996; revised 3 June 1996

Abstract

The feasibility of diagnostic reasoning in a Bayesian belief network, based on a genetic algorithm is demonstrated. The reasoning process described here is an example of approximate reasoning. Since exact abduction in a network modelling the "classical diagnostic problem" is NP-hard, inexact or approximate reasoning attracts much attention. The results of the present study indicate that in a given context of observed symptoms, a genetically generated population of possible solutions retains much of the diagnostic power contained in the full model: the disease probabilities as occuring in this population and as calculated from the full model are strongly rank-correlated. Moreover, the disease–symptom correlations are retained in the genetically generated population. This is important, since these probabilities and correlations are dynamic quantities which depend on the context of observed symptoms. The genetic algorithm may be seen as a procedure to dynamically generate diagnostic protocols.

1. Introduction

To determine the diagnosis in a patient based on the presenting symptoms is a daily occuring problem in medical practice. More often than not it is solved in a stepwise fashion. Ideally, an initial differential diagnosis is made, possibly based on a subset of symptoms, and as yet undetermined symptoms are selected which potentially may either confirm or eliminate one or more of the remaining possibilities. In practice, however, such a next set of symptoms is not always so carefully selected; rather, a whole battery of new tests is ordered, even if not all of the outcomes are potential discriminants. For the sake of

comfort to the patient and for economical reasons, a careful determination of a diagnostic protocol has great advantages.

The problem of determining the optimal next set of symptoms is not an easy one. The reason is that the dependencies between diseases and symptoms are not static quantities, but depend on the context of the presence or absence of other symptoms. For a clinician, it is impossible or at least impractical to keep track of this dynamic behaviour of the symptom–disease correlations as the diagnostic workup progresses.

Computer assisted techniques have been proposed in the past to help solving the diagnostic problem. Initially, most of these systems were based on a straightforward application of Bayes' theorem, converting a priori disease probabilities into a posteriori probabilities in the light of additional evidence. Such

---

* E-mail: gelsema@mi.fgg.eur.nl.

early systems commonly had at least three serious limitations. First, the symptoms were treated as independent. Second, it was assumed that only one disease was present; the single disease with the highest a posteriori probability was selected. Finally, the disease–symptom correlations were treated as static quantities. For these reasons the simple Bayesian approach sometimes referred to as crude Bayesian or idiot Bayesian is now widely recognized as sub-optimal.

Bayesian belief networks (BBN's, or causal networks) have been proposed as an alternative. A Bayesian belief network modelling the diagnostic problem is a directed acyclic graph in which diseases and symptoms are graphically represented as nodes. The nodes may or may not be connected by directed arcs, depending on whether or not a causal relationship is believed to exist between them. An arc is directed from cause to effect. Absence of an arc between two nodes implies that no direct dependence between these nodes exists. The graphical part of a Bayesian belief network must be supplemented by tables of probabilities. For each root node, a priori probabilities for each of its possible states must be supplied. For non-root node states, a table specifying the probabilities, conditional on the states of all of this node's parent nodes is required.

A special form of a Bayesian belief network is a network modelling the "classical diagnostic problem" as shown in Fig. 1. It consists of a series of root nodes (d-nodes), modelling a set of diseases and a series of non-root nodes (s-nodes) modelling a set of symptoms. A d-node is connected to an s-node if the disease modelled by the d-node is believed to be a possible cause of the symptom modelled by the

s-node. The absence of such a causal relationship between a disease d_i and a symptom s_j is represented by the absence of an arc between their respective node representations. Disease nodes are not directly interconnected, nor are symptom nodes.

In practice, designing the graphical representation of a given diagnostic problem (a set of disease nodes and a set of symptom nodes) is fairly straightforward, using medical knowledge of the domain. However, supplying the numerical part, especially all conditional probabilities of symptoms, given the parent disease states, is more problematic. The concept of causation events as suggested by Peng and Reggia (1987a,b) makes this task much easier. Instead of having to specify a complete table of conditional probabilities for each symptom node (2^n independent numbers if the states are all binary and where n is the number of parent d-nodes), only one "causation strength" has to be specified for each arc in the network. If the network contains an arc from d_i to s_j, the causation strength associated with the arc is the conditional probability that d_i causes s_j, given that d_i is present. Apart from the fact that fewer probabilities have to be specified, the causation strengths are intuitively easier to estimate than the conditional probabilities of the symptoms given the states of the parent diseases. On the other hand, the tables of these latter conditional probabilities can be reconstructed from the causation strengths.

Once the graphical representation has been supplemented by the numerical information, the Bayesian belief network contains a full description of the probability space: for any combination of states (present or absent) of the diseases and symptoms, the probability of occurence of this configura-
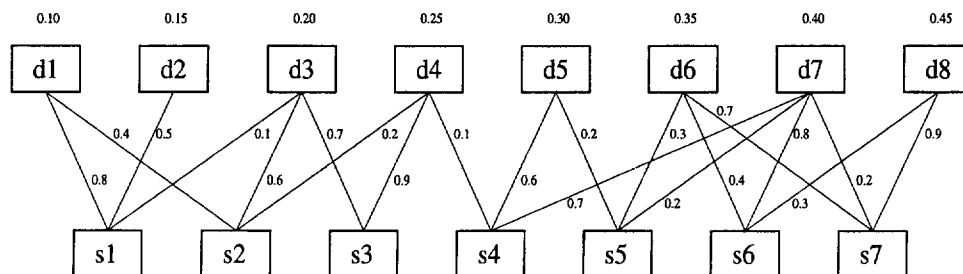

Fig. 1. The Bayesian belief network modelling the "classical diagnostic problem". The d-nodes model diseases, the s-nodes model symptoms. The prior disease probabilities are indicated in the top row. The arcs between nodes d_i and s_j model a causal dependence between disease d_i and symptom s_j with a causation strength (see text) indicated along the arcs.

tion can be computed. If the network is initially uninstantiated (the presence/absence state of all nodes is unknown), this probability is the product of the prior probabilities of all disease states and the appropriate conditional probabilities of the symptom states. In partially instantiated networks, for compatible states this product yields the correct probability up to a scaling factor equal to the probability of that particular network instantiation.

Exact methods of probability propagation in Bayesian belief networks exist. See, e.g., (Pearl, 1988; Neapolitan, 1989) for a treatment of probabilistic reasoning. By probability propagation, the state probabilities of all nodes can be calculated, given the states of any subset of nodes. This does not yet constitute a full answer to the diagnostic problem, if combinations of diseases are considered as valid explanations of the observed symptoms. In that case, abductive reasoning (reasoning from effects to causes) is required. It has been demonstrated that probability propagation (Cooper, 1990) as well as abductive reasoning (Shimony, 1994) in non-singly connected BBN's is NP-hard. In general, a network modelling the classical diagnostic problem is non-singly connected. As an alternative to extract solutions to the problem by exact reasoning, approximate methods have been proposed (Pearl, 1988). It has been shown that even such approximate methods may be NP-hard (Dagun and Luby, 1993).

The operation of a genetic algorithm (Davis, 1991) in Bayesian belief networks, using the steady state variant with an elitist strategy was outlined in (Gelsema, 1995). Summarizing, a possible solution of the network (an individual) is represented as a chromosomal structure. Each gene in the chromosome represents a node in the network and can take (in the network presented) one of two possible values, representing the absence or presence of the disease or symptom modelled by the gene. It is straightforward to calculate the probability (fitness) of a chromosome thus completed as the product of n multipliers, one for each node in the network (see above). After having generated an initial population of n_p solutions, at each step of the algorithm, two chromosomes are selected for reproduction. The two parent individuals undergo cross-over and the offspring is then possibly subjected to mutation. If a resulting individual is not already a member of the

population, its fitness is calculated and if it is higher than the fitness of the least fit individual in the current population, it replaces this latter individual. Otherwise, the next reproduction step is initiated. This process is iterated until some stopping criterion is met. The resulting population of n_p individuals is called the genetically selected population. In (Gelsema, 1995), it was shown that genetic selection of possible solutions, using the calculated probability of such solutions (see above) as the fitness function, results in a subpopulation of solutions in which the MAP (maximum a posteriori) solutions appear with probabilities which may be orders of magnitude higher than when potential solutions were sampled at random.

Whereas in the previous paper (Gelsema, 1995) the properties of individual solutions were studied, we will here concentrate on the diagnostic properties of the genetically generated populations. The question te be addressed in the present paper is: given that a genetic algorithm results in a subpopulation of high probability individuals, to what extent are the diagnostic clues reflected in this subpopulation? The answer to this question will be quantified on the basis of two sets of diagnostically important indicators.

– In different contexts of observed symptoms, the probabilities of diseases (single or in combination with other diseases) are different. Such probabilities are called current disease probabilities, i.e. they are valid in the given context. The rank-orders of the current probabilities of the diseases may be calculated from the genetically selected subpopulation.
– The disease–symptom correlations also depend on the given context. These correlations can also be calculated from the genetically generated subpopulation.

Since any diagnostic process is a stepwise procedure in which after the determination of the presence or absence of a subset of symptoms, a differential diagnosis guides the further clinical workup, it is important to ascertain that at each step in this procedure, the current disease probabilities are at least ordered in the proper way. It is equally important to establish that at each step, high (positive or negative) disease–symptom correlations are properly reflected.

These correlations point the way to promising, discriminating symptoms of which the state (absent or present) remains to be established. As was mentioned earlier, these current probabilities and the disease–symptom correlations are dynamic quantities: they depend on the state of instantiation of the network, or in diagnostic terms, on the states of the symptoms thusfar observed to be present or absent.

2. Methods

2.1. The experiments

The graphical part of the network used in the present study is the same as the one used in the third example in (Gelsema, 1995). It consists of eight disease nodes (d_1 through d_8) and seven symptom nodes (s_1 through s_7), interconnected as shown in Fig. 1. The prior probabilities of the d-nodes (the disease prevalences) were slightly modified so as to avoid equal probabilities, making the rank correlation coefficient between the disease probabilities in the full model and in the selected subpopulations a more powerful indicator. The prior probabilities of the disease nodes and the causation strengths associated with the disease–symptom dependencies are also given in Fig. 1.

The total solution space of the uninstantiated network (case 1) consists of $2^{15} = 32,768$ states. The network was also studied in its state where the symptom corresponding to node s_1 was observed to be present (case 2). In the last case, the cardinality of the search space reduces to 16,384.

In both cases, computational experiments consisting of the following steps were performed 100 times:
1. generation of an initial population of n_p solutions compatible with the given case instantiation and computation of the corresponding probabilities of these solutions;
2. application of a genetic algorithm to the initial population until n_c solutions (including the n_p initial ones) had been evaluated;
3. calculation of the current disease probabilities as occuring in the final population of n_p solutions;
4. establishing the rank-order correlation coefficients between the estimated current disease probabilities and the ones as computed from the full model;
5. calculation of the Spearman rank correlation coefficients between all disease–symptom pairs in the final subpopulation.

After 100 runs of the genetic algorithm, the means and standard deviations of the correlation coefficients as calculated in steps 4 and 5 were obtained.

In each of the cases described above, the experiments were performed for various combinations of n_p and n_c. The values of n_p and n_c used are listed in Table 1. It is seen that within a case, going from one experiment to the next, populations of which the size reduces by a factor of 2 are explored. Experiments 1, 2 and 3 are comparable between the cases in the sense that an equal proportion of the total search space is represented in the population (the fractions are 1/128, 1/256 and 1/512 for experiments 1, 2 and 3, respectively). In all experiments, in both cases, the population size n_p is one quarter of the number of explored solutions n_c.

Table 1
Experimental setup and correlations of current disease probabilities

Case.exp	n_p	n_c	card/n_p	P_{np}	$c(P_d)$
1.1 uninstantiated	256	1024	128	0.423 ± 0.038	0.957 ± 0.043
1.2 uninstantiated	128	512	256	0.272 ± 0.053	0.855 ± 0.139
1.3 uninstantiated	64	256	512	0.155 ± 0.050	0.736 ± 0.210
2.1 s_1 present	128	512	128	0.229 ± 0.035	0.844 ± 0.117
2.2 s_1 present	64	256	256	0.128 ± 0.032	0.707 ± 0.215
2.3 s_1 present	32	128	512	0.059 ± 0.029	0.520 ± 0.281

The table contains: the initial state of instantiation of the network (case.exp), the number of individuals in the genetically generated population (n_p), the number of explored individuals (n_c), the reduction factor w.r.t. the cardinality of the search space (card/n_p), the average probability and its standard deviation accumulated in the final population (P_{np}) and the average (± 1 s.d.) rank order correlation coefficient between the current disease probabilities as calculated from the final population and as calculated from the full model ($c(P_d)$).

2.2. The genetic algorithm

Initial populations were generated as follows: for each individual to be generated, a gene corresponding to an instantiated node in the network (if any) was given its corresponding value. Genes corresponding to a disease node were assigned states (present or absent) with equal probability. Genes corresponding to non-instantiated symptom nodes were assigned states (present or absent) with probabilities equal to the relevant conditional probabilities.

In the present case, individuals were selected for reproduction by the roulette wheel technique, using the fitness as a selection criterion (in (Gelsema, 1995), the rank orders of the fitnesses were used; no systematic difference was observed between the two selection criteria). After selection of two parent individuals, cross-over was applied with a probability of 100%. The probability of mutation was set at 0.2%. Genes corresponding to instantiated nodes were excluded from mutation.

As described above, the steady state variant with elitist strategy of generation transition was used. In this variant, the population is updated immediately after the generation of a new accepted offspring individual. This ensures that this individual takes part in the selection process as soon as it is accepted in the population. Using an elitist strategy, a new individual is accepted only if its fitness is better than that of the least fit individual in the current population.

The stopping criterion was implemented as follows: after $n_c - n_p$ reproductions, the genetic search was terminated. Thus, at termination, a maximum of $n_c - n_p$ new individuals were explored.

After termination, the quantities mentioned in points 4 and 5 above were calculated on the basis of the final population of n_p individuals.

2.3. Rank-order correlations of disease probabilities

In order to verify the hypothesis that the population of individuals thus obtained contains sufficient diagnostic information to guide further clinical workup, the current disease probabilities were calculated from the population and compared in rank order to the corresponding probabilities as calculated from the network. In the non-instantiated state of the network (case 1) these latter probabilities are the prior probabilities as shown in Fig. 1; in case 2, they can be calculated by probability propagation or by an exhaustive exploration of all compatible states.

Both series of n ($n = 8$) disease probabilities are assigned rank orders (e.g., p_i and q_i, respectively, $i = 1, \ldots, n$). Spearman's rank-order correlation coefficient between the two series can then be expressed as (Sachs, 1984):

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} (p_i - q_i)^2}{n(n^2 - 1)}.$$

The Spearman rank-order correlation coefficient takes values between -1 and $+1$. After each application of the genetic algorithm, these coefficients were calculated. In each experiment, at the end of 100 runs of the genetic algorithm, the mean and standard deviation were computed.

2.4. Disease–symptom correlations

For each disease and each symptom, the presence or absence in each of the n_p individuals in the final population can be determined. For each disease–symptom combination this leads to two series of n_p numbers:

disease d_i: 0 1 1 0 1 1 1 0 . . . 1 0 0 1,

symptoms s_j: 0 0 1 0 1 1 1 0 . . . 1 0 0 1.

If 0 stands for absent and 1 stands for present, then the sequences above express the fact that in the first individual of the population (first column), d_i and s_j are both absent, in the second d_i is present, while s_j is absent, etc. If the numbers represented by the dots in both sequences were the same, it is clear that disease d_i and symptom s_j are strongly positively correlated (almost perfect co-presence or co-absence). The correlation between sequences in which the same numbers occur frequently can be expressed by a modified Spearman rank-order correlation coefficient:

$$r_{s,c} = 1 - \frac{6 \sum_{k=1}^{n_p} (e_{ik} - e_{jk})^2}{\left(n_p^3 - n_p\right) - \left(S_i + S_j\right)},$$

Table 2
Efficiency of selection of the currently most probable disease

case.exp	N_{1max}	N_{3max}
1.1 uninstantiated	95	100
1.2 uninstantiated	82	98
1.3 uninstantiated	58	90
2.1 s_1 present	78	100
2.2 s_1 present	58	100
2.3 s_1 present	46	82

The table lists the number of times (out of 100 runs) that the most probable disease, given the context of observed symptoms, was selected (N_{1max}) and the number of times the selected most probable disease was one of the three maximum probability diseases (N_{3max}).

where e_{ik} and e_{jk} ($k = 1, \ldots, n_p$) are the entries in the first and the second sequence, respectively, and where S_i and S_j are given by

$$S_i = \frac{1}{2} \sum_{g=1}^{G} \left(n_{ig}^3 - n_{ig} \right), \qquad S_j = \frac{1}{2} \sum_{g=1}^{G} \left(n_{jg}^3 - n_{jg} \right).$$

In these last expressions, G is the number of groups of different entries in the sequences. In the case above, there are two groups of entries ($G = 2$): one group of 0's and one group of 1's. The numbers n_{ig} are the numbers of entries in group g contained in the first sequence; n_{jg} are the numbers of entries in group g contained in the second sequence. The corrected Spearman rank-order correlations also take values between -1 and $+1$. In the case illustrated above (with all entries represented by the dots assumed to be equal) this yields a corrected correlation coefficient very close to 1.

3. Results

The experiments and the results are summarized in Tables 1 through 4. The first three rows of Table 1 summarize the experiments with the network in its uninstantiated state (case 1). Population sizes of 256, 128 and 64 individuals were used, exploring a total number of individuals of 1024, 512 and 256 in these three experiments, respectively. While the ratio of the population size to the number of explored individuals is kept constant (a factor of 1/4), the ratio of the cardinality of the search space and the population size n_p increases by factors of 2 from 128 to 512. The column headed P_{np} contains the means and standard deviations over 100 runs of the total probability accumulated in the final population of n_p individuals. It is seen that even when the final population has only 64 individuals (1/512 times the cardinality of the search space), this population accumulates 15.5% of the total probability space. The means and standard deviations over 100 runs of the Spearman rank order coefficients $c(P_d)$ between the current disease probabilities as estimated from the final subpopulation and as calculated from the complete model are given in the last column of Table 1. The last three rows of Table 1 summarize the results of the experiments with the network in a state with symptom s_1 instantiated to state present. The ratios card/n_p in these three experiments are the same as those for the experiments with the uninstantiated case.

In each of the two cases, the disease with the highest current probability as calculated from the

Table 3
Average disease-symptom correlations (over 100 runs) for the uninstantiated network

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
d_1	+ + +	+ + +					
d_2	+ + 0						
d_3		+ + +	+ + +		− 0 0		
d_4		+ + +	+ + +	+ 0 0			
d_5				+ + +	+ + +		
d_6					+ + +	+ + +	+ + +
d_7				+ + +	+ + +	+ + +	+ + 0
d_8						+ + +	+ + +

For each (d_i, s_j) combination, the three entries correspond to the experiments 1, 2 and 3, respectively. A correlation is indicated with a + ($-$) if the coefficient is larger (smaller) than 0 by more than one standard deviation. An entry is 0 if the coefficient is compatible with 0 at the 1 s.d. level. No entries for a (d_i, s_j) combination imply that all three entries are compatible with 0.

Table 4
Average disease-symptom correlations (over 100 runs) for the network in the context of the presence of s_1

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
d_1		+ + 0					
d_2		− − −					
d_3		+ + +	+ + +				
d_4		+ + 0	+ + +				
d_5				+ + +	+ + 0		
d_6					+ + +	+ + +	+ + +
d_7				+ + +	+ + +	+ + +	+ + 0
d_8						+ + +	+ + +

The same conventions as in Table 3 apply. The negative (D_2, s_2) correlations are noteworthy.

model can be identified. In case 1 this is obviously d_8 $(P = 0.45)$; in case 2 this is d_2 $(P = 0.50)$, with as close competitors d_1 $(P = 0.49)$ and d_8 $(P = 0.45)$. For each run of the genetic algorithm, the disease with the highest accumulated probability in the final population was recorded and compared to the disease with the highest probability as calculated from the network. The number of runs (out of 100) in which the genetic algorithm resulted in the correct maximum probability disease is indicated in Table 2 (column N_{1max}). The number of runs in which the genetic algorithm selected one of the three most probable diseases (d_8, d_7 or d_6 in case 1; d_2, d_1 or d_8 in case 2) as the maximum probability one is listed in Table 2 (column N_{3max}).

Tables 3 and 4 summarize the results for the disease–symptom correlations. For each of the experiments, the means and standard deviations were obtained over 100 runs of the genetic algorithm. For each disease–symptom combination, the three entries in Tables 3 and 4 relate to the situations with $card/n_p = 128$, 256 and 512, respectively. Table 3 represents the uninstantiated network, Table 4 the network with s_1 instantiated to state present. The correlation coefficients are represented by the symbols +, 0 and −. An entry is + (−) if the mean correlation coefficient is positive (negative) and at least one standard deviation from the value 0. An entry is 0 if the mean value is within one standard deviation from the value 0. When there are no entries for a disease–symptom combination, this implies that all three entries were equal to 0. Table 4 has no entries for (d_i, s_1) combinations, since all individuals in the genetically generated population have s_1 instantiated to present, rendering correlations with s_1

meaningless. It is seen that even in unfavourable situations $(card/n_p = 512)$, Table 3 reproduces most of the causation strengths as implemented in the model: in the final subgroup of genetically selected individuals, the causation strengths are retained as correlation coefficients different from 0 (at the one standard deviation level). It is interesting to observe that in Table 4, the (d_2, s_2) correlations become negative. It will be explained in the next section that this must indeed be the case and that this is a reflection of the dynamic behaviour of the disease–symptom correlations.

4. Discussion

An approximate method for diagnostic reasoning on the basis of a Bayesian belief network, using a genetic algorithm, has been described. In a previous article (Gelsema, 1995), the efficiency of this method to determine maximum a posteriori individual solutions was presented. In order to study the diagnostic properties of the entire genetically selected subpopulation of solutions, two cases were considered. In case 1, the network was in its a priori state, i.e., no nodes were instantiated; in case 2, the node corresponding to symptom s_1 was instantiated to present. Thus, case 2 corresponds to making a diagnosis in a patient in which symptom s_1 is observed to be present. For both cases, the experiments 1, 2 and 3 are comparable in the sense that the ratio of the cardinality of the search space and the population size is the same for corresponding experiments. From Table 1 it may be observed that the results for the case with the higher cardinality (case 1) are more

favourable than those for case 2: the accumulated probability in the final population P_{np} is higher. Also, the rank order correlation coefficients $c(P_d)$ between the current disease probabilities as computed from the final population and as computed from the model are higher. This behaviour of increasing performance with increasing complexity of the search space was also observed in (Gelsema, 1995), where under otherwise comparable conditions, genetic selection was shown to operate better in more complex situations. The results presented in Table 2 show that the currently most probable disease is contained in the genetically selected population with a high probability.

The values of the disease–symptom correlation coefficients cannot be compared between corresponding experiments in the two cases. These correlations are dynamic quantities which depend on the instantiation, i.e., the symptoms for which the presence or absence has been determined so far. However, an interesting phenomenon is observed in Table 4, representing the disease–symptom correlations in the network with s_1 instantiated to state present. In all three experiments, the correlation coefficient between d_2 and s_2 becomes significantly (at the one standard deviation level) negative, implying that the presence of s_2 is an indication for the absence of d_2 and conversely. It can easily be understood that this must indeed be the case. Symptom s_1 being instantiated to present, only solutions with s_1 present are allowed. Possible causes of s_1 are d_1, d_2 or d_3 or combinations thereof. If either d_1 or d_3 is the cause, the presence of s_2 is also fairly probable (causation strengths of 0.4 and 0.6, respectively) but at the same time solutions with d_2 also being present have low probability, due to the multiplicative effect of the prior probabilities. On the other hand, if d_2 is the cause of s_1, the co-presence of d_1 and/or d_3 and/or d_4 is improbable. Since only d_1, d_3 or d_4 (or combinations thereof) can cause s_2, the presence of s_2 is improbable. Indeed, probability propagation in this network with s_1 instantiated to present and s_2 alternatively instantiated to present and absent reveals that the posterior probability of d_2 is 0.30 (making it the least probable disease) and 0.61 (making it the most probable disease), respectively. This is an expression of the dynamic behaviour of disease–symptom correlations discussed in the intro-

duction. This example shows that such behaviour is reflected in the genetically generated populations, even if the population size is small compared to the cardinality of the search space.

It may be argued that in a practical situation, in which the genetic algorithm would be run only once, the standard deviations of the coefficients $c(d_i, s_j)$ will not be available. It is then impossible to select the coefficients which are significantly different from 0. Therefore, for case 1, the average coefficients $c(d_i, s_j)$ corresponding to an arc in the network were compared to those not corresponding to an arc. Of those corresponding to an arc, only two are smaller than 0.10 in absolute value ($c(d_3, s_1)$ and $c(d_4, s_4)$, both corresponding to a causation strength of 0.1). On the other hand, all coefficients for combinations not corresponding to an arc are smaller than 0.10 in average absolute value. These observations apply to all three experiments in case 1. Thus, selection of correlation coefficients with large absolute values seems to result in those disease–symptom combinations which are linked by a high causation strength. The average values of $c(d_i, s_j)$ show very little sensitivity to the degrading conditions going from experiment 1 to experiment 3: between experiments, the average values of corresponding coefficients $c(d_i, s_j)$ differ less than 10%.

In practice, a diagnostic reasoning system might be implemented as follows: given a network describing the medical diagnostic problem at hand, for any patient, the initial subset of symptoms observed to be present or absent defines a partial instantiation of the network. A genetic algorithm may then select a subpopulation consistent with this initial instantiation, giving a ranking of the current disease probabilities and even a list of the most probable individual solutions with the corresponding probabilities. It will also report estimates of the disease–symptom correlations which are valid in that particular context of observed symptoms. The disease probabilities and the list of most probable solutions will guide the definition of further diagnostic goals (which disease to confirm or to exclude), whereas the disease–symptom correlations will help in defining the best set of as yet unobserved symptoms (i.e., the diagnostic workup) to reach that goal. Also, symptom–symptom correlations (not discussed in the present article) may be reported, to avoid the observation of

both members of highly correlated pairs of symptoms. Having established the presence/absence of the next set of symptoms, the non-compatible solutions may be removed from the genetically generated population and replaced by randomly selected compatible ones. A new round of the genetic algorithm may then be initiated. Thus, the solutions of the first round as far as they are compatible are preserved and better compatible ones may be generated in the new context. This process may be repeated as long as the final diagnosis has not yet been reached and symptoms remain to be observed. The advantage of this reasoning process is that no diagnostic protocol needs to be established beforehand for various subgroups of patients with various subsets of observed symptoms. The genetic algorithm generates the protocol dynamically, at each step adapting to the new context.

References

Cooper, G.F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence* 42, 393–405.

Dagun, P. and M. Luby (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence* 60, 141–153.

Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York.

Gelsema, E.S. (1995). Abductive reasoning in Bayesian belief networks using a genetic algorithm. *Pattern Recognition Lett.* 16, 865–871.

Neapolitan, R.E. (1989). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley, New York.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo.

Peng, Y. and J.A. Reggia (1987a). A probabilistic causal model for diagnostic problem solving, Part 1: Integrating symbolic causal inference with numeric probabilistic inference. *IEEE Trans. Syst. Man Cybernet.* 17, 146–162.

Peng, Y. and J.A. Reggia (1987b). A probabilistic causal model for diagnostic problem solving, Part 2: Diagnostic strategy. *IEEE Trans. Syst. Man Cybernet.* 17, 395–406.

Sachs, L. (1984). *Angewandte Statistik*. Springer, Berlin.

Shimony, S.E. (1994). Finding MAP's for belief networks is NP-hard. *Artificial Intelligence* 68, 399–410.