

NasoNet, Modeling the Spread of Nasopharyngeal Cancer with Networks of Probabilistic Events in Discrete Time

S. F. Galán¹, F. Aguado², F. J. Díez¹, and J. Mira¹

¹Dpto. de Inteligencia Artificial
Facultad de Ciencias de la UNED
Paseo senda del rey 9
28040 Madrid (SPAIN)

{seve, fjdiez, jmira@dia.uned.es}

²Servicio de Oncología Radioterápica
Hospital Clínico Universitario San Carlos
28040 Madrid (SPAIN)
faguado@cgcom.org

February 22, 2002

Abstract

The spread of cancer is a non-deterministic dynamic process. As a consequence, the design of an assistant system for the diagnosis and prognosis of the extent of a cancer should be based on a representation method that deals with both uncertainty and time. The ultimate goal is to know the stage of development of a cancer in a patient before selecting the appropriate treatment. A *network of probabilistic events in discrete time* (NPEDT) is a type of Bayesian network for temporal reasoning that models the causal mechanisms associated with the time evolution of a process. This paper describes NasoNet, a system that applies NPEDTs to the diagnosis and prognosis of nasopharyngeal cancer. We have made use of *temporal noisy gates* to model the dynamic causal interactions that take place in the domain. The methodology we describe is general enough to be applied to any other type of cancer.

Key words: cancer diagnosis and prognosis, Bayesian networks, causality, probabilistic temporal reasoning, temporal noisy gates

1 Introduction

The diagnosis and prognosis of the extent of a cancer are tasks full of uncertainty. This is due, on the one hand, to the deeply non-deterministic nature

of this disease and, on the other hand, to the incomplete, imprecise, or erroneous information that the oncologist may obtain. This situation is even more complicated in the case of nasopharyngeal cancer, since the nasopharynx is a hidden and difficult to enter cavity located in the highest part of the pharynx. Therefore, early detection of a malignant nasopharyngeal tumor is not common. Generally, patients only seek medical attention at advanced stages, when symptoms become evident.

1.1 Bayesian networks for cancer diagnosis

Bayesian networks [21, 24] are a probability-based knowledge representation method, appropriate for the modeling of causal processes with uncertainty, such as those determining the evolution of a cancerous disease. A Bayesian network is an acyclic directed graph whose nodes represent random variables and whose links define probabilistic dependences between variables. These relations are quantified by associating a conditional probability table with each node. Each conditional probability table contains the probability of a node, given any possible configuration of values for its parents. For root nodes, only their a priori probabilities are needed. Bayesian networks allow probabilistic dependence and independence relations to be specified in a natural way through the network topology. Diagnosis or prediction with Bayesian networks consists of fixing the values of the observed variables and computing the posterior probabilities of some of the unobserved variables. Some applications of Bayesian networks to oncological domains are:

- PATHFINDER [12, 13], an expert system for the diagnosis of lymph node diseases. PATHFINDER makes use of the so-called *probabilistic similarity network*, which represents the possible diagnoses in only one node. Moreover, the patient is supposed to suffer from a unique disease, which is a reasonable hypothesis in this domain.
- MammoNet [16], a Bayesian network to assist in the detection of breast cancer, which integrates mammographic findings, demographic factors, and physical examination to determine the probability of malignancy.
- DynaMoL [19], a general dynamic decision framework based partially on the formalism of *dynamic influence diagrams* [27]. This framework is applied to a case study on deciding the optimal follow-up schedule of colorectal cancer patients who have undergone surgery [5].

Neither PATHFINDER nor MammoNet make use of an explicit representation of time, whereas in DynaMoL the time horizon is defined as a set of discrete time points, each corresponding to a certain decision stage.

1.2 Probabilistic temporal reasoning

Time is a fundamental factor in cancer, since it usually determines the stage of the disease and, consequently, the type of treatment to be applied. Modeling the

process that begins when a malignant tumor arises and ends with the appearance of several typical symptoms, metastasis, or affected lymph nodes, requires representing the causal mechanisms that control this process over time. Some of the most widespread methods for modeling dynamic processes with uncertainty in medical domains [2, 6, 15, 19, 20, 25] are based on the formalism of *dynamic Bayesian networks* [7, 8, 17, 23] or their extension: *dynamic influence diagrams* [27]. These formalisms have the disadvantage of generating highly complex networks, since time is discretized and a node is created for each random variable associated with each time instant. Usually, a copy of a static network is generated for each time point and links are established between nodes in adjacent static networks. In this way, Markovian processes can be modeled so that the future is conditionally independent of the past given the present.

Other extensions of Bayesian networks for temporal reasoning have been proposed over the last few years. Aliferis and Cooper [1] develop the language of *modifiable temporal belief networks* as a structural and temporal extension of Bayesian networks. Ngo *et al.* [22] define a context-sensitive temporal probability logic for representing classes of dynamic Bayesian networks. Arroyo-Figueroa and Sucar [3] propose a model called *temporal nodes bayesian network*, in which each temporal node represents an event or state change of a variable and arcs represent causal-temporal relations between nodes; however, this model lacks a formalization of canonical models (noisy OR-gate, noisy AND-gate, and others) for temporal processes.

A *network of probabilistic events in discrete time* (NPEDT) [11] is a Bayesian network for temporal reasoning that leads to less complex networks than those obtained from the formalism of dynamic Bayesian networks, for domains involving temporal fault diagnosis and prediction. Under the NPEDT approach, time is discretized, nodes are associated with events, and each value of a node represents the occurrence of an event at a particular instant. In our domain, an event is a change of state provoked by an anomaly. The improvement in complexity with respect to dynamic Bayesian networks is a consequence of assuming that each event occurs only once. The value taken on by a variable indicates the time at which the event has occurred. Reversible processes can be represented through multiple events. The links in the network represent temporal causal mechanisms between neighboring nodes. Therefore, each conditional probability table expresses the most probable delays between parent events and the corresponding child event. Two major advantages of an NPEDT are that this model is not restricted to Markovian processes, and that we can make use of different *temporal noisy gates* [11] that facilitate knowledge acquisition and representation. Temporal noisy gates (temporal noisy OR-gate, temporal noisy AND-gate, and others) constitute a generalization for temporal processes of traditional canonical models. The process of cancer spread, previous to the application of therapy, is formed by a set of irreversible events. Each of these events can be represented by a node in an NPEDT. In this work, we show the application of the NPEDT approach to the modeling of nasopharyngeal cancer evolution. The network assists the clinician in the diagnosis and prognosis of the extent of this disease in a patient.

The rest of this paper is organized as follows. Section 2 describes the domain of nasopharyngeal cancer. Section 3 deals with the characteristics of NasoNet. Specifically, Section 3.1 is devoted to the details of representation of NasoNet as an NPEDT, Section 3.2 considers the process of data acquisition carried out to complete NasoNet, Section 3.3 presents an example of application of NasoNet, and Section 3.4 describes the evaluation of the model. Section 4 discusses the advantages and disadvantages of an NPEDT with respect to other types of Bayesian networks for temporal reasoning. Finally, we conclude with some additional remarks.

2 Nasopharyngeal cancer

The nasopharynx is the highest part of the pharynx, which receives the air breathed through the nose. The nasopharyngeal cavity has a cuboidal shape: the lateral walls are formed by the Eustachian tube and the fossa of Rosenmuller, at the front are the posterior choanae and the nasal cavity, the roof has the base of the skull above, the posterior boundary is formed by the muscles of the posterior pharyngeal wall, and below are the upper surface of the soft palate and the posterior pharyngeal wall

The patient profile that we are interested in corresponds to those patients coming from the Department of Otorhinolaryngology of a hospital, who are admitted to the Department of Radiation Oncology; specifically, our work has been carried out in collaboration with oncologists from San Carlos University Clinical Hospital in Madrid.

A cancer of the nasopharynx [18, 26] appears as a malignant primary tumor localized on one of the nasopharyngeal walls (see Fig. 1). Primary tumors on lateral walls are the most frequent, whereas those on anterior and posterior walls are less likely. As time goes by, an initial primary tumor may either infiltrate the adjacent tissue (infiltrating tumor) or grow in volume inside the nasopharynx (vegetating tumor). Accordingly, any part surrounding the nasopharynx, or even any nasopharyngeal wall, may be affected by the tumor growth. Generally, vegetating tumors may obstruct the ducts connecting the nasopharynx to some of its surrounding parts: nasal cavity, ear, or soft palate. Infiltrating tumors may reach parts of vital importance, like the base of the skull and the cranial nerves. Infiltrating tumors in the nasopharynx are more invasive than vegetating ones, although the latter require a shorter period of time to spread. The usual symptoms of nasopharyngeal cancer are dysfunctions associated with breathing, speech, vision, hearing, and sense of smell, among others. It is therefore crucial to detect the disease at early stages; otherwise, the consequences could be irreversible for the patient. As any other kind of cancer, there is the possibility of regional (lymph node involvement) or distant metastases. The appearance of nasopharyngeal hemorrhage or infection is also evidence of cancer.

The diagnosis of nasopharyngeal cancer consists of three phases:

- Registration of the patient's medical history.

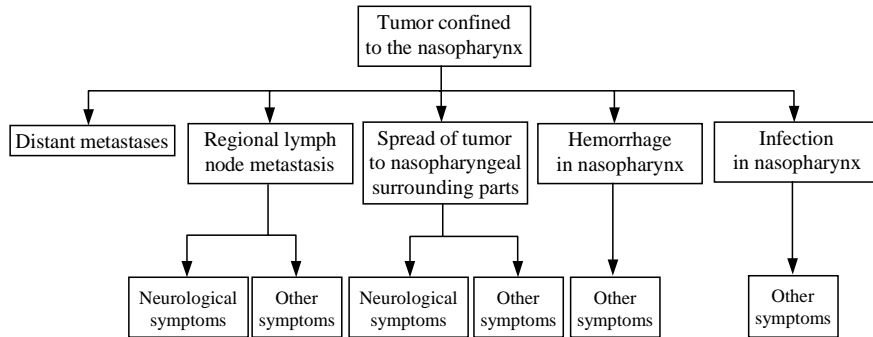


Figure 1: Overview of the evolution of a nasopharyngeal cancer

- Visual examination of the nasopharynx (by mirror or endoscopy) and documentation of the size and location of neck nodes.
- Complementary tests, such as evaluation of hearing and cranial nerve function, biopsy, hemogram, complete computer tomographic (CT) scan or magnetic resonance imaging (MRI) with views delineating the upper and lower extent of the lesion.

Each of the previous phases produces new evidence to assist the oncologist in determining the extent and malignancy of the disease.

Once the diagnosis has been completed, the stage of the cancer can be defined by means of the TNM codification, where T stands for primary tumor, N for regional lymph nodes, and M for distant metastasis. For example, T1N0M1 means “tumor confined to the nasopharynx, no regional lymph node metastasis, and distant metastasis present”. The appropriate treatment (radiation therapy, chemotherapy, surgery, and others) depends on the stage of the cancer.

3 Overview of NasoNet

3.1 Description of the model

NasoNet is an NPEDT that models the process of progression of a nasopharyngeal cancer. The final model assists oncologists in the diagnosis and prognosis of the extent of this type of cancer in a patient.

A primary tumor on any of the nasopharyngeal walls may spread and invade adjacent parts. It may also provoke distant metastases, and hemorrhage or infection in the nasopharynx. The previous processes are characterized by the occurrence of a series of events. These events —before treatment is applied— are generally irreversible, and causally interrelated. For example, a primary vegetating tumor on the anterior wall of the nasopharynx may occupy the nasal fossae and produce anosmia (loss of the sense of smell); a primary infiltrating

tumor on the superior wall of the nasopharynx may spread to the right lateral wall, then to the cavernous sinus, later invade the right inner ear, and finally produce symptoms like tinnitus (ringing in the ears), autophony (resonance of one’s voice), and hypoacusis (diminished acuteness of hearing), associated with abnormalities in the ear. Some examples of relevant events in our domain are: “appearance of a primary vegetating tumor on the posterior wall of the nasopharynx”, “spread of an infiltrating tumor to the left cavernous sinus”, “appearance of rhinolalia”, “appearance of Gradenigo syndrome on the right side”, “appearance of abnormal cervical lymph nodes on the left side”, etc. In NasoNet, these events and all of those causally related to the spread of a nasopharyngeal cancer are represented as nodes in a Bayesian network.

If we had decided not to represent time explicitly in the Bayesian network, each random event variable could take on the values *present* or *absent*; accordingly, we would have only needed binary random variables. On the contrary, as the causal processes we are modeling are not instantaneous and there is uncertainty as to their duration, we need to represent time explicitly. To this end, we consider the instant at which the primary tumor may appear as the initial or reference instant, and we define the occurrence time of any other event with respect to the mentioned initial instant. We suppose there is only one primary tumor, which is a reasonable hypothesis. According to expert opinion, the temporal range of interest in our domain are the three years following the appearance of the primary tumor. We divide this period into trimesters, according with the expert. The final time range and the final time unit were selected as a result of taking into account both computational tractability and temporal expressiveness of the model. Each event represented in the network has its own typical period of occurrence. All these different periods are within the three-year term we have selected as time horizon. As an example, lung metastasis may arise during the second or third year, and abnormal cervical lymph nodes may appear during the first semester.

The approach of NPEDTs allows absolute time to be used as an alternative to time instants relative to the occurrence of a determined initial event (appearance of the primary tumor, in the case of NasoNet). With this new option, each value taken on by a variable represents an absolute time instant at which its associated event may occur. An advantage of using absolute time is that scenarios (see Section 3.3) are not required. However, when using absolute time, each event in the network can take place at any of the instants belonging to the temporal range of interest (3 years for NasoNet). Therefore, all of the variables can take on the same number of values (13 in our domain). Consequently, a more complex network than in the case of relative time would be obtained. This is the reason why we decided to use relative time in NasoNet.

Given an event node E with its occurrence period divided into trimesters, the random variable associated with E can take on the values $\{e[a], \dots, e[b], e[never]\}$ where $a, b \in \{1, \dots, 12\}$ and $a \leq b$. $E = e[j]$, with $j \in \{a, \dots, b\}$, means that event E takes place in the j -th trimester after the appearance of the primary tumor, and $E = e[never]$ means that E does not take place. For example, $Anosmia = anosmia[3]$ expresses the appearance of anosmia during the third

trimester. As each random variable can take on a set of exclusive values, each event associated with a variable can occur only once over time. This condition is satisfied in our domain, since the processes involved, prior to treatment, are irreversible. (Reversible processes could be represented by multiple events but we do not have them in NasoNet.) The current version of NasoNet contains 15 nodes associated with tumors confined to the nasopharynx, 23 nodes representing the spread of tumors to nasopharyngeal surrounding sites, 4 nodes symbolizing distant metastases, 4 nodes related to abnormal lymph nodes, 11 nodes expressing nasopharyngeal hemorrhages or infections, and 50 nodes referring to symptoms or syndromes (combinations of symptoms). The root nodes in the network correspond to events related to the appearance of primary infiltrating or vegetating tumors on each wall of the nasopharynx. As we assume that there is one primary tumor at most, the probabilities of the root nodes should add up to 1. To this end, we introduce an additional parent node for the previous root nodes. The leaf nodes in the network represent the appearance of different symptoms or syndromes. Finally, the intermediate nodes are events related to the spread of the tumor to parts adjacent to the nasopharynx, infections, and metastases.

NasoNet models the evolution of a nasopharyngeal cancer so that each arc represents a causal relation between one parent event and one child event. For instance, in Fig. 2, the appearance of infection in the nasopharynx may produce rhinorrhea (excessive mucous secretion from the nose). If these causal relations were static, we could apply the noisy OR-gate [24] to model the interactions between an effect and its causes. In the noisy OR model, each cause acts independently of the remaining causes to produce a determined effect. This independence of causal interactions is satisfied in our domain, according to expert opinion. For example, the appearance of anosmia (see Fig. 2) may be provoked by either a vegetating tumor occupying the right nasal fossa or by the spread of an infiltrating tumor to that fossa; both processes act independently of each other. In a family of nodes with N parents interacting through the noisy OR model, we only need to specify N independent parameters. This number rises to 2^N in the case of a family of nodes with N parents interacting through the general case.

As the causal relations in our domain are not instantaneous and, furthermore, the nodes in the network correspond to temporal events, we use the temporal noisy OR-gate [11] as a model of causal interaction in the network. The temporal noisy OR-gate represents the case in which the effect is present as soon as any of its causes provokes it to be present. According to Fig. 2, if a primary vegetating tumor on the right lateral wall provoked the appearance of nasopharyngeal infection at trimester i , and a primary infiltrating tumor provoked the same infection at trimester j , with $i \neq j$, then the event “appearance of infection in the nasopharynx” would be considered to occur at trimester $\min(i, j)$. For this reason, a temporal noisy OR-gate becomes a noisy MIN-gate [10].

Let us consider a family of nodes with n causes X_1, \dots, X_n and one effect Y . In principle, each of these event nodes may take place at any of the trimesters

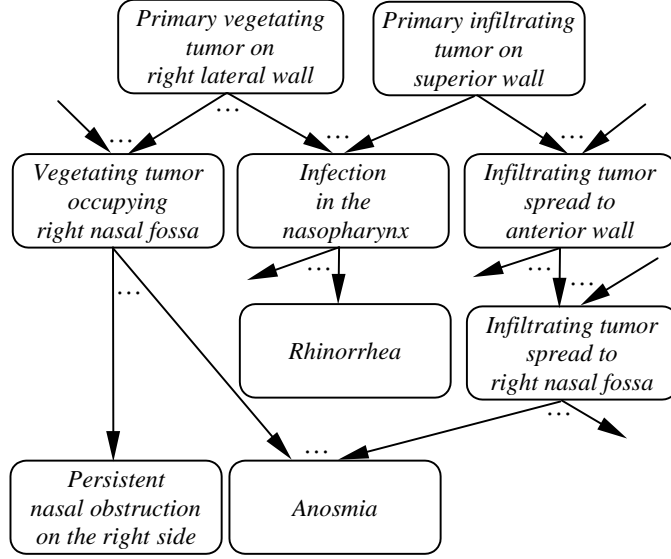


Figure 2: Part of the Bayesian network modeling the evolution of a cancer of the nasopharynx

$\{1, \dots, 12\}$. For each cause, in the temporal noisy OR model it is necessary to specify the parameters:

$$c_{y[k_i]}^{x_i[j_i]} \quad i \in \{1, \dots, n\}, j_i \in \{1, \dots, 12, never\}, k_i \in \{1, \dots, 12\} \quad (1)$$

Each parameter is defined as the probability of Y taking place at k_i , given that X_i takes place at j_i , and the rest of the causes are absent. The conditional probability table for Y can be computed as follows (see Fig. 3 for a family with two causes):

$$P(y[k]|x_1[j_1], \dots, x_n[j_n]) = \sum_{(k_1, \dots, k_n) | \min(k_1, \dots, k_n) = k} \prod_i c_{y[k_i]}^{x_i[j_i]} \quad k, j_1, \dots, j_n \in \{1, \dots, 12, never\} \quad (2)$$

By ordering temporal indices from future to past ($never, 12, \dots, 2, 1$), just as illustrated in Fig. 3, a noisy MAX-gate [9, 14] leads to a temporal noisy OR-gate.

If there are non-explicit causes of Y in the model, they can be grouped together and represented through a vector of leaky parameters:

$$c_{y[k]}^* \quad k \in \{1, \dots, 12\} \quad (3)$$

Each leaky parameter is the probability of the effect Y occurring at k , given that all the explicit causes are absent.

$P(y[k] x_1[j_1], x_2[j_2])$	$c_{y[never]}^{x_1[j_1]}$	$c_{y[12]}^{x_1[j_1]}$...	$c_{y[2]}^{x_1[j_1]}$	$c_{y[1]}^{x_1[j_1]}$
$c_{y[never]}^{x_2[j_2]}$	y[never]	y[12]	...	y[2]	y[1]
$c_{y[12]}^{x_2[j_2]}$	y[12]	y[12]	...	y[2]	y[1]
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$c_{y[2]}^{x_2[j_2]}$	y[2]	y[2]	...	y[2]	y[1]
$c_{y[1]}^{x_2[j_2]}$	y[1]	y[1]	...	y[1]	y[1]

Figure 3: Temporal noisy OR-gate for two causes ($n = 2$) and one effect

3.2 Data acquisition

In principle, each arc forming part of a temporal noisy OR-gate in NasoNet requires $(12 + 1) \cdot 12 = 156$ independent parameters. Among these parameters, those satisfying

$$c_{y[k]}^{x_i[j_i=never]} \quad k \in \{1, \dots, 12\} \quad (4)$$

are zero because the effect cannot take place if none of its causes are present. Furthermore, among the remaining 144 parameters, if $j_i > k$ then the corresponding parameter is zero because the effect cannot precede the cause. Finally, the remaining 78 independent parameters can be reduced to 12, since in our domain, according to expert opinion, it is reasonable to assume the property of *time invariance*:

$$c_{y[k+\Delta t]}^{x_i[j_i+\Delta t]} = c_{y[k]}^{x_i[j_i]} \quad \forall j_i, k, j_i + \Delta t, k + \Delta t, \in \{1, \dots, 12\} \quad (5)$$

This property expresses that if we consider a constant delay, $k - j_i$, between cause X_i and effect Y , the parameters defining the arc $X_i \rightarrow Y$ are invariant, independently of the times X_i and Y take place. To summarize, computing the conditional probability table associated with a family of nodes in NasoNet just requires specifying one parameter for each possible delay between cause and effect. Therefore, for a family of nodes with N parents, $12 \cdot N$ independent parameters are needed at most in NasoNet. (Note that in the case of general interaction among the parent nodes this number would rise to $12 \cdot (12 + 1)^N$). The questions that the knowledge engineer has to ask the oncologists are:

Given that X_i takes place in a certain *trimester*, what is the probability of its effect Y occurring in the same trimester, if the rest of its causes are absent? And what is the probability of Y occurring in the next trimester? And so on.

It was difficult for the oncologists to answer these questions. They argued that the answers depend on the cause event occurring in the early or

late trimester. However, they felt more confident when answering the following questions:

Given that X_i takes place in a certain *instant*, what is the probability of its effect Y occurring in the next trimester, if the rest of its causes are absent? And what is the probability of Y occurring in the trimester after the next trimester? And so on.

Let the parameters for these latter questions be:

$$\tilde{c}_Y^{X_i}(\Delta t) \quad \Delta t \in \{1, \dots, 12\} \quad (6)$$

Using a continuous representation of time, if $X_i = x_i[1]$ then we can associate to X_i the probability density function f_1 , depicted in Fig. 4. The integral

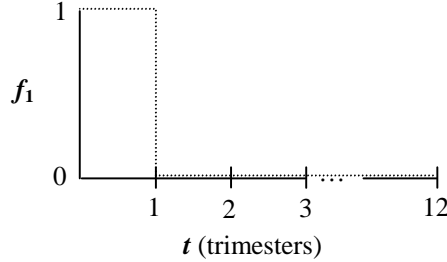


Figure 4: Probability density function for X_i when $X_i = x_i[1]$

of the probability density function between two instants is the probability of occurrence of the corresponding event between those instants. Suppose

$$\tilde{c}_Y^{X_i}(\Delta t) = \begin{cases} k & \Delta t = 1 \\ 0 & \Delta t = \{2, \dots, 12\} \end{cases} \quad (7)$$

where $0 \leq k \leq 1$. These parameters given by the medical experts define a probability transfer function f_2 , between X_i and Y (see Fig. 5), which represents the effect of X_i on Y . The probability density function f for Y (see Fig. 6) is obtained by calculating the convolution of f_1 and f_2 (cf. [28], Section 4.2):

$$f(t) = \int_{-\infty}^{\infty} f_1(\tau) \cdot f_2(t - \tau) d\tau \quad (8)$$

If from Fig. 6 we return to our division of time into trimesters, we obtain:

$$c_{y[1]}^{x_i[1]} = \int_0^1 f dt = \frac{k}{2} \quad (9)$$

$$c_{y[2]}^{x_i[1]} = \int_1^2 f dt = \frac{k}{2} \quad (10)$$

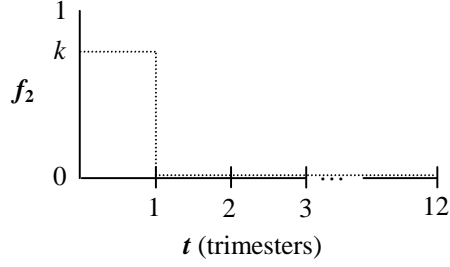


Figure 5: Probability transfer function between X_i and Y

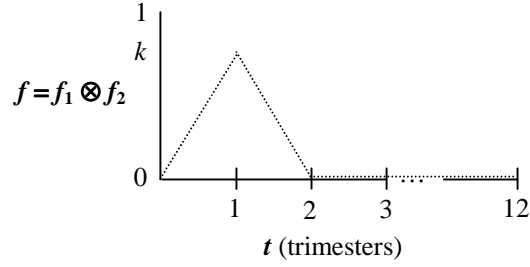


Figure 6: Probability density function for Y

We can conclude that, once we know that X_i has occurred at j_i , each delay Δt contributes the same probability to trimesters $j_i + \Delta t - 1$ and $j_i + \Delta t$. Therefore,

$$c_{y[j_i + \Delta t]}^{x_i[j_i]} = \frac{\tilde{c}_{Y^i}^{X_i}(\Delta t)}{2} + \frac{\tilde{c}_{Y^i}^{X_i}(\Delta t + 1)}{2} \quad (11)$$

In this way, we can compute the conditional probability tables in the network from the knowledge provided by the medical expert.

To cite an example, in the connection *Infiltrating tumor spread to anterior wall* \rightarrow *Infiltrating tumor spread to right nasal fossa* (see Fig. 2), the information provided by the expert is

$$\tilde{c}_{Y^i}^{X_i}(\Delta t) = \begin{cases} 0.48 & \Delta t = 1 \\ 0.24 & \Delta t = 2 \\ 0.12 & \Delta t = 3 \\ 0.06 & \Delta t = 4 \\ 0 & \Delta t = \{5, \dots, 12\} \end{cases} \quad (12)$$

The final parameters for this arc are shown in Table 1.

$c_{y[k]}^{x_i[j_i]}$	$x_i[1]$	$x_i[2]$	$x_i[3]$	$x_i[4]$	$x_i[never]$
$y[1]$	0.48/2	0	0	0	0
$y[2]$	(0.48 + 0.24)/2	0.24	0	0	0
$y[3]$	(0.24 + 0.12)/2	0.36	0.24	0	0
$y[4]$	(0.12 + 0.06)/2	0.18	0.36	0.24	0
$y[5]$	0.06/2	0.09	0.18	0.36	0
$y[6]$	0	0.03	0.09	0.18	0
$y[7]$	0	0	0.03	0.09	0
$y[8]$	0	0	0	0.03	0
$y[never]$	0.1	0.1	0.1	0.1	1

Table 1: Example of parameters for an arc $X_i \rightarrow Y$ in NasoNet

3.3 Example

Oncologists make use of the following information sources: the patient’s medical history, visual examination of the nasopharynx, and the result of different complementary tests. A finding involves determining the occurrence of an event represented by means of a node in NasoNet, and establishing the time it occurred. NasoNet determines, from the available findings, both posterior probabilities and occurrence times for the rest of the events in the network. In order to simplify the example, we will suppose that our aim is twofold: firstly, we want to know whether the primary tumor is vegetating or infiltrating and, secondly, we are interested in finding out on which wall the primary tumor is located. Note that we are assuming that there is one primary tumor at most in the patient.

Consider the portion of NasoNet shown in Fig. 7. Any primary vegetating tumor may grow in volume inside the nasopharyngeal cavity and occupy the nasal fossae. This may produce rhinolalia (nasal voice produced by an alteration in the nasal fossae resonance). The appearance of a primary vegetating tumor may also provoke abnormal cervical lymph nodes on the right side. The parameters of the network in Fig. 7 are shown in Table 2.

Physicians know from the patient’s medical history that on 9/20/99 the patient began suffering from rhinolalia. With this unique finding, there are as many possible scenarios as possible delays between the appearance of the primary tumor and the appearance of rhinolalia, 12 in this case. For each scenario, the posterior probabilities that appear in Table 3 can be obtained in NasoNet. The complexity of the network prevented us from performing evidence propagation through exact algorithms. This fact became evident once we had introduced an explicit representation of time equivalent to nearly a third of the nodes in NasoNet’s graph. The previous result is a consequence of the way development environments for building Bayesian networks deal with evidence propagation through exact algorithms in networks with noisy gates: first, the noisy gate is transformed into a family interacting through the general model, and then an exact algorithm is applied. Note that in the general model both

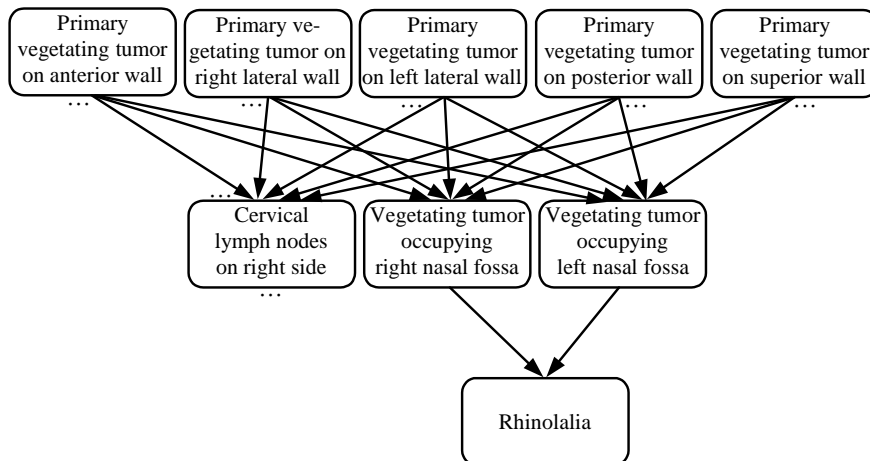


Figure 7: NasoNet subnetwork

the number of conditional probabilities and the time for evidence propagation are exponential with the number of parents. A solution to this problem would be the use of specific exact algorithms for evidence propagation in networks with noisy gates. Anyhow, stochastic simulation algorithms allow acceptable approximate results to be obtained in a few minutes in NasoNet.

On 12/30/99, the patient detects neck nodes on the right side and is received in the Department of Radiation Oncology, where oncologists establish by palpation the presence of abnormal cervical lymph nodes on the right side. Since the presence of abnormal cervical lymph nodes can only occur during the next semester to the appearance of the primary tumor (see Table 2), only one scenario is possible: *rhinolalia*[1], *chnrs*[2] (see Table 4). Therefore, we can suspect that the primary tumor is located on the nasopharyngeal anterior wall. This is a valuable result that permits a better interpretation of the information obtained from subsequent complementary tests.

The process of determining the resulting scenarios from a set of findings requires: establishing for each finding a time period (divided into trimesters) for primary tumor appearance according to the finding, calculating the intersection of the previous periods and, finally, selecting the possible scenarios within the intersection. Fig. 8 illustrates a case with three findings corresponding to events: $A = \{a[1], a[2], a[3], a[4], a[never]\}$, $B = \{b[1], b[2], b[3], b[never]\}$, and $C = \{c[1], c[2], c[never]\}$. The following four scenarios are obtained: $sc1 \equiv \{a[2], b[2], c[2]\}$, $sc2 \equiv \{a[1], b[2], c[2]\}$, $sc3 \equiv \{a[1], b[1], c[2]\}$, and $sc4 \equiv \{a[1], b[1], c[1]\}$. If in our example the patient had begun suffering from rhinolalia after the appearance of abnormal cervical lymph nodes, new scenarios involving primary tumor on other walls than the anterior one could explain the evidence. This fact demonstrates the importance of using an explicit rep-

X_i	Y	$\tilde{c}_Y^{X_i}(\Delta t)$
<i>Pvtaw</i>	<i>Vtornf</i>	0.225 for $\Delta t \in \{1, \dots, 4\}$
<i>Pvtaw</i>	<i>Vtolnf</i>	0.225 for $\Delta t \in \{1, \dots, 4\}$
<i>Pvtaw</i>	<i>Clnrs</i>	0.27 for $\Delta t = 1$, 0.13 for $\Delta t = 2$
<i>Pvtrlw</i>	<i>Vtornf</i>	0.175 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtrlw</i>	<i>Vtolnf</i>	0.05 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtrlw</i>	<i>Clnrs</i>	0.27 for $\Delta t = 1$, 0.13 for $\Delta t = 2$
<i>Pvtllw</i>	<i>Vtornf</i>	0.05 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtllw</i>	<i>Vtolnf</i>	0.175 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtllw</i>	<i>Clnrs</i>	0.14 for $\Delta t = 1$, 0.07 for $\Delta t = 2$
<i>Pvtpw</i>	<i>Vtornf</i>	0.00625 for $\Delta t \in \{5, \dots, 12\}$
<i>Pvtpw</i>	<i>Vtolnf</i>	0.00625 for $\Delta t \in \{5, \dots, 12\}$
<i>Pvtpw</i>	<i>Clnrs</i>	0.27 for $\Delta t = 1$, 0.13 for $\Delta t = 2$
<i>Pvtsw</i>	<i>Vtornf</i>	0.175 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtsw</i>	<i>Vtolnf</i>	0.175 for $\Delta t \in \{3, \dots, 6\}$
<i>Pvtsw</i>	<i>Clnrs</i>	0.27 for $\Delta t = 1$, 0.13 for $\Delta t = 2$
<i>Vtornf</i>	<i>Rhinolalia</i>	0.2, instantaneous
<i>Vtolnf</i>	<i>Rhinolalia</i>	0.2, instantaneous

Table 2: Parameters of the network in Fig. 7

		<i>Pvtaw</i>	<i>Pvtrlw</i>	<i>Pvtllw</i>	<i>Pvtpw</i>	<i>Pvtsw</i>	
sc12:	<i>rh</i> [12]	0	0	0	1	0	$\uparrow_{09/20/96}$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
sc7:	<i>rh</i> [7]	0	0	0	1	0	$\uparrow_{12/20/97}$
sc6:	<i>rh</i> [6]	0	0.4059	0.4072	0.0089	0.1778	$\uparrow_{03/20/98}$
sc5:	<i>rh</i> [5]	0	0.3817	0.386	0.0088	0.2233	$\uparrow_{06/20/98}$
sc4:	<i>rh</i> [4]	0.022	0.364	0.3629	0	0.2509	$\uparrow_{09/20/98}$
sc3:	<i>rh</i> [3]	0.0401	0.3434	0.3477	0	0.2686	$\uparrow_{12/20/98}$
sc2:	<i>rh</i> [2]	1	0	0	0	0	$\uparrow_{03/20/99}$
sc1:	<i>rh</i> [1]	1	0	0	0	0	$\uparrow_{06/20/99}$
							$\uparrow_{09/20/99}$

Table 3: Posterior probabilities for primary vegetating tumors with temporal localization for each scenario

		<i>Pvtaw</i>	<i>Pvtrlw</i>	<i>Pvtllw</i>	<i>Pvtpw</i>	<i>Pvtsw</i>	
sc1':	<i>rh</i> [1], <i>clnrs</i> [2]	1	0	0	0	0	$\uparrow_{06/30/99}$
							$\uparrow_{09/20/99}$

Table 4: New posterior probabilities

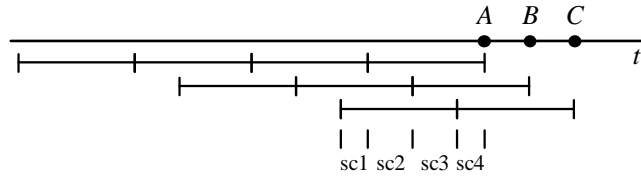


Figure 8: Possible scenarios from three findings

resentation of time in this domain, since the same type of events can produce different diagnoses depending on their occurrence times.

3.4 Evaluation of NasoNet

As a previous step towards the final construction of NasoNet, we developed an atemporal Bayesian network in which each node represents the occurrence or nonoccurrence of a certain event. The graph of this network is the same as that of NasoNet and the type of causal interaction for each family of nodes is modeled through the noisy OR-gate. We checked that the set of nodes in the network described the disease with the appropriate degree of detail. After a number of interviews with the medical experts, we decided to simplify certain parts of the graph and to enlarge others. For example, the initial model detailed the spread of the tumor to each of the skull openings that the cranial nerves pass through. Later, we decided to group together these openings to form more significant parts. Also in the beginning, we only considered the general concept of “primary tumor in the nasopharynx”. Later, a differentiation of primary tumors by walls was necessary. The initial graph was designed following a breadth-first strategy; that is to say, for each cause at a determined depth, all its effects were generated. This way of depicting the graph is analogous to the way cancer spreads over time. The verification of the arcs in the network was carried out following a different strategy. The nodes were ordered alphabetically and each node was associated with its possible causes. The oncologists were thus forced to consider the graph bottom-up. Additionally, the alphabetical order obliged the oncologists to reason locally in the network because each node had nothing to do with its previous one. The atemporal network consists of 276 arcs with multiple loops. Once prior and conditional probabilities were introduced in the network, evidence propagation from clustering algorithms took about one second.

The introduction of an explicit representation of time prompted the use of both multivalued variables and the temporal noisy OR model. The average number of values each variable in the temporal version of NasoNet could take on increased to 9.6. The verification of the parameters in this version of NasoNet was facilitated by the previous construction of the atemporal network, in which a conditional probability was established for each arc. However, the temporal network associated a vector of conditional probabilities with each arc, and each

conditional probability was related to a delay between cause and effect. For each arc, the consistency of the system required the sum of the components of its vector of conditional probabilities in the temporal network to be equal to the conditional probability in the atemporal network.

We have made a preliminary evaluation of NasoNet from eight medical histories, which were contrasted with the results given by the network for each case. The following steps were taken for each particular case:

- Introduction of the available evidence in the network. Generally, between four and eight temporal findings.
- Evidence propagation in the network.
- Comparison of the rest of the information present in the medical history with the posterior probabilities obtained in NasoNet.

For each patient, NasoNet provides us with a set of posterior probabilities $\{p^*(e[t_i^E]), \dots, p^*(e[t_f^E]), p^*([never])\}$, where t_i^E and t_f^E represent the limits of the temporal range for event E . E is any event not included as evidence in the patient's medical history. The information about when event E really happened ($t_{\text{real}}^E \in \{t_i^E, \dots, t_f^E, never\}$) is obtained from the medical history and compared with the posterior probabilities, following three different methods:

1. For each event E , we establish whether

$$p^*(e[t_{\text{real}}^E]) = \max_{j \in \{t_i^E, \dots, t_f^E, never\}} \{p^*(e[j])\} \quad (13)$$

and calculate the percentage of events in the network for which this identity holds.

2. Another option is to consider in the previous method, not just the value with highest posterior probability, but the pair of values with highest posterior probabilities.
3. Finally, it is interesting to study the percentage of events correctly diagnosed or predicted by NasoNet, independently of the time they occurred. In this case, we only consider whether or not events happen. Now, an event E is correctly diagnosed or predicted when

$$\left\{ \begin{array}{l} \left(\sum_{j=t_i^E}^{t_f^E} p^*(e[j]) \right) > p(e[never]) \quad \text{if } t_{\text{real}}^E \neq never \\ \left(\sum_{j=t_i^E}^{t_f^E} p^*(e[j]) \right) < p(e[never]) \quad \text{otherwise} \end{array} \right. \quad (14)$$

Table 5 summarizes the results obtained from each of the three methods.

	Method 1	Method 2	Method 3
<i>patient 1</i>	90.1%	92.8%	93.9%
<i>patient 2</i>	84%	92%	90%
<i>patient 3</i>	82.7%	89.5%	87.5%
<i>patient 4</i>	87%	92.5%	91%
<i>patient 5</i>	83.5%	90%	88.3%
<i>patient 6</i>	90%	92.5%	92%
<i>patient 7</i>	86%	91%	92.7%
<i>patient 8</i>	85.2%	90.1%	89.2%

Table 5: Percentage of events correctly diagnosed or predicted

4 Discussion

Under the NPEDT approach, time is discretized and each value of a variable represents the instant at which a certain event may occur. This is the main difference with respect to dynamic Bayesian networks, in which the value of a variable V_i represents the state of a real-world property at time t_i . Therefore, an NPEDT is more appropriate for temporal fault diagnosis because only one variable is necessary for representing the occurrence of a fault and, as a consequence, the networks involved are much simpler than those obtained by using dynamic Bayesian networks. In contrast, dynamic Bayesian networks are more appropriate for monitoring tasks, since they explicitly represent the state of the system at each moment. An example of this kind of tasks is therapy planning in diabetes.

Dynamic Bayesian networks have the disadvantage of generating highly complex networks [4]. If we had applied this formalism to our domain, it would have been necessary to copy NasoNet’s causal graph —with more than one hundred nodes— twelve times. However, only one copy is needed if an NPEDT is used. Moreover, the application of dynamic Bayesian networks is usually restricted to Markovian processes in which the future is conditionally independent of the past given the present, i.e., only connections between random variables within the same or adjacent time slices are allowed. In our domain, in contrast, a causal mechanism can in general take one trimester, two trimesters, or even years.

NPEDTs are similar to Arroyo-Figueroa and Sucar’s temporal nodes Bayesian networks [3], although the latter lack a formalization of canonical models for temporal processes. In the general case, it is necessary to assign each node of a Bayesian network a set of conditional probabilities that grows exponentially with the number of parents. This complicates the acquisition of the parameters, their storage, and the propagation of evidence. For these reasons, new causal interactions models —called canonical models [24, 10]— were developed in order to simplify both Bayesian network construction and probability computation. In a family of nodes interacting through a canonical model, the number of required parameters grows linearly with the number of parents. The construction of NasoNet led us to developing temporal canonical models [11] within the NPEDT

approach. There are nodes in NasoNet whose number of parents rises to ten. It would have been impossible to use the general model of causal interactions for such families, while the elicitation of the numerical parameters for NasoNet's temporal links was relatively easy.

5 Conclusions

We have presented NasoNet, a network of probabilistic events in discrete time for the diagnosis and prognosis of the extent of a nasopharyngeal cancer.

The spread of cancer is a process full of uncertainty, where time must be taken into account. NasoNet makes use of the *temporal noisy OR-gate* to model the different dynamic causal interactions that take place during the spread of cancer. NasoNet is the first system to apply the approach of *network of probabilistic events in discrete time* (NPEDT) to a real-world domain. In comparison with other kinds of Bayesian networks that introduce an explicit representation of time, this approach offers important advantages for the representation and management of irreversible processes, like those occurring during the spread of cancer before the appropriate treatment is applied. We have explained how to calculate the numerical parameters needed in NasoNet from the parameters elicited from the oncologists.

Several issues regarding knowledge representation, knowledge acquisition, inference, and verification of the system have been discussed. Testing of clinical cases is the main task to be carried out in the future, in order to perform a thorough evaluation of the system. Another future task is to extend NasoNet to cope with cancer evolution after treatment.

Acknowledgments

This research was supported by the Spanish CICYT, under grants TIC97-0604 and TIC-97-1135-C04-04. We are grateful to the anonymous reviewers for their useful comments and suggestions.

References

- [1] C. F. Aliferis and G. F. Cooper. A structurally and temporally extended Bayesian belief network model: definitions, properties, and modeling techniques. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI'96)*, pages 28–39, Portland, Oregon, 1996. Morgan Kaufmann, San Francisco, CA.
- [2] S. Andreassen, R. Hovorka, J. Benn, K. G. Olesen, and E. R. Carson. A model-based approach to insulin adjustment. In *Proceedings of the Third Conference on Artificial Intelligence in Medicine*, pages 239–248, Maastriech, The Netherlands, 1991. Springer-Verlag.

- [3] G. Arroyo-Figueroa and L. E. Sucar. A temporal Bayesian network for diagnosis and prediction. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 13–20, Stockholm, Sweden, 1999. Morgan Kaufmann, San Francisco, CA.
- [4] J. Binder, K. Murphy, and S. Russell. Space-efficient inference in dynamic probabilistic networks. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 1292–1296, Nagoya, Japan, 1997.
- [5] C. Cao, T. Leong, A. Leong, and F. Seow. Dynamic decision analysis in medicine: a data driven approach. *International Journal of Medical Informatics*, 51:13–28, 1998.
- [6] P. Dagum and A. Galper. Forecasting sleep apnea with dynamic network models. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 64–71, Washington D.C., 1993. Morgan Kaufmann, San Francisco, CA.
- [7] P. Dagum, A. Galper, and E. Horvitz. Dynamic network models for forecasting. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 41–48, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.
- [8] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.
- [9] F. J. Díez. Parameter adjustment in Bayes networks. The generalized noisy OR-gate. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 99–105, Washington D.C., 1993. Morgan Kaufmann, San Francisco, CA.
- [10] F. J. Díez and M. Druzdzel. Canonical interactions for probabilistic expert systems: The OR gate and other models. Technical Report, Decision Systems Laboratory, University of Pittsburgh, 2001. In preparation.
- [11] S. F. Galán and F. J. Díez. Networks of probabilistic events in discrete time. Technical Report UNED-IA-2001-01, Department of Artificial Intelligence, UNED, 2001. Submitted to *International Journal of Approximate Reasoning*.
- [12] D. E. Heckerman, E. J. Horvitz, and B. N. Nathwani. Toward normative expert systems: Part I — The Pathfinder Project. *Methods of Information in Medicine*, 31:90–105, 1992.
- [13] D. E. Heckerman and B. N. Nathwani. Toward normative expert systems: Part II — Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine*, 31:106–116, 1992.

- [14] M. Henrion. Some practical issues in constructing belief networks. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. Elsevier Science Publishers, Amsterdam, 1989.
- [15] M. E. Hernando, E. J. Gómez, F. del Pozo, and R. Corcoy. DIABNET: A qualitative model-based advisory system for therapy planning in gestational diabetes. *Medical Informatics*, 21:359–374, 1996.
- [16] C. E. Kahn Jr., L. M. Roberts, K. A. Shaffer, and P. Haddawy. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*, 27:19–29, 1997.
- [17] U. Kjærulff. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 121–129, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.
- [18] A. Lee, W. Foo, S. Law, Y. F. Poon, S. K. O, S. Y. Tung, W. M. Sze, R. Chappell, W. H. Lau, and J. Ho. Staging of nasopharyngeal carcinoma: From Ho's to the new UICC system. *International Journal of Cancer*, 84(2):179–187, 1999.
- [19] T. Leong. Multiple perspective dynamic decision making. *Artificial Intelligence*, 105(1-2):209–261, 1998.
- [20] P. Magni and R. Bellazzi. DT-Planner: an environment for managing dynamic decision problems. *Computer Methods and Programs in Biomedicine*, 54:183–200, 1997.
- [21] R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley-Interscience, New York, 1990.
- [22] L. Ngo, P. Haddawy, and J. Helwig. A theoretical framework for context-sensitive temporal probability model construction with application to plan projection. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI'95)*, pages 419–426, Montreal, 1995. Morgan Kaufmann, San Francisco, CA.
- [23] A. E. Nicholson and J. M. Brady. Sensor validation using dynamic belief networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 207–214, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.
- [24] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, 1988. Revised second printing, 1991.
- [25] G. M. Provan and J. R. Clarke. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):299–307, 1993.

- [26] S. P. Schantz, L. B. Harrison, and A. A. Forastiere. Tumors of the nasal cavity and paranasal sinuses, nasopharynx, oral cavity, and oropharynx. In V. T. DeVita, S. Hellman, and S. A. Rosenberg, editors, *Cancer: Principles and Practice of Oncology*, pages 741–801. Lippincott-Raven Publishers, Philadelphia, 1997.
- [27] J. A. Tatman and R. D. Shachter. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):365–379, 1990.
- [28] A. Y. Tawfik and E. Neufeld. Temporal reasoning and Bayesian networks. *Computational Intelligence*, 16(3):349–377, 2000.