General environment for probabilistic predictive monitoring

Silvano Mussi

Abstract**—**The proposal presented in the paper concerns a general environment for probabilistic predictive monitoring. More precisely, the paper is conceptually subdivided in three parts. The first part presents the theoretical model underlying the proposal. In particular, the model is turn presented as a hierarchy of three conceptual levels. The first conceptual level is represented by a set of basic concepts and definitions. This first level is used like a platform on which the second conceptual level, represented by the definition of time-slices based causal network, is built. This second level is, in turn, a platform on which the third conceptual level, represented by the definition of probabilistic network, is built. This last level contains the mathematical foundations of the model and defines a general probabilistic prediction algorithm that can be applied to real world problems in heterogeneous domains. The second part of the paper presents a general predictive monitoring tool in which the predictive algorithm, defined in the first part, is embedded. Since such a general tool needs to be equipped with specific domain knowledge in order to be usefully applied to real world problems, the third part of the paper presents a general environment in which users can easily build, use and administer specific predictive monitoring tools equipped with proper domain knowledge related to specific application fields.

Keywords**—**Computer applications, Knowledge engineering, Decision support systems, Predictive monitoring.

I. Introduction

THE possibility of getting early warnings before an undesired event may occur has always been very appealing. Let us think, for example, of prevention of high risk events for health, or serious faults or anomalies of costly and strategic industrial equipments or plants. Similarly, the possibility of getting predictions about the occurrence of a desired event is useful for taking suitable measures in order to favor the event occurrence. Let us think, for example, of passing an exam or reaching a certain athletic performance in the sport field. The proposal considers predictive monitoring applied to both preventing undesired events and favouring desired events.

The proposal presented in the paper concerns both a general probabilistic model for producing predictions and a general predictive monitoring tool embedded in a general environment for building, using and administering specific probabilistic predictive monitoring tools addressing specific real world problems.

The concept of prediction considered in the paper, expressed in an intuitive and brief way, refers to the following scenario. Let us suppose there is a population of subjects, for example: persons, machines, etc. There is an event E, desired or undesired, that can happen to everyone of the subjects. For example, if subjects are persons an undesired event might be "first cardiac infarct". The probability of E occurrence is, in general, affected by both the mere aging of a subject and the contexts, i.e. the conditions, in which a subject ages (for example, a person who smokes ages in the context: "cigarette smoke"). Let us suppose that a domain expert monitors (considering certain aspects as, for example, state of health, degree of performance, etc.) each subject, not necessarily at constant time intervals. During a monitor session of a subject X the expert hypothesizes that the future time of X elapses in certain contexts and, as a consequence, wants to know the probability that E occurs to X in the future. The idea of the proposal is that such a goal can be reached by exploiting statistical information collected by all the subjects whose history is equal to the history X would have in the simulated future.

A. Paper organization

The paper is conceptually structured in three parts. The first part, sections II, III, IV, presents the theoretical model underlying the proposal. In particular, such presentation is structured in three conceptual levels: basic concepts and definitions (sect. II), time-slices based causal network (sect. III), probabilistic network (sect. IV). The result of this first part is the definition of a general probabilistic predictions algorithm. The second part (sect. V) presents a general predictive monitoring tool that uses the predictive algorithm, defined in the first part. Section VI illustrates a simulated-case study in order to better explain how the general predictive monitoring tool works. The third part (sect. VII) presents a general environment for probabilistic predictive monitoring in which users can easily build, use and administer specific probabilistic predictive monitoring tools oriented to specific application fields. Section VIII concerns related work and discussion and finally section IX draws some conclusions.

Manuscript received January 30, 2013: Silvano Mussi is member of AICA (Associazione Italiana per l'Informatica e il Calcolo Automatico) – Piazza Rodolfo Morandi 2, 20121, Milano, Italy. Author e-mail: mussisilvano@tiscali.it.

II. Basic concepts and definitions

This section presents the set of the basic concepts and definitions that will be used throughout the paper for building the theoretical model that constitutes the proposal.

A. Definition 1 [Basic scenario]

Let us consider a population of subjects (human beings, machines, etc.). For example, let us think of a population of mail persons. The subjects of the population are monitored by a domain expert through monitoring sessions. It is not required that monitoring sessions are separated by constant time intervals. Let us suppose that there is an event E (undesired or desired) that may happen or not to each subject of the population. For example, for a population of mail persons the event E might be represented by “First cardiac infarct”. The event E is represented by a variable with two values (or states): “occurred”, “not-occurred”.

B. Definition 2 [Context]

The probability that E occurs to a subject may be affected by both the mere aging of the subject and the contexts (i.e. conditions) C1, C2, … in which the subject ages. For example, a smoker is a subject that ages in the context “Cigarette smoke”. Ageing in this context increases the probability of having the first cardiac infarct.

C. Definition 3 [Context state]

A context C has a set of possible states s1, s2, … , just like a variable has a set of possible values. The set of possible states of C is denoted by C={s1, s2, ...}, and C=s1 means: “the context C is instantiated to the state s1”. For example, the context “Genetic predisposition” might have two states “yes” (s1), “no” (s2), whereas the context “Cigarette smoke” might have three states: “yes under 10 cigarettes a day” (s1), “yes 10 or more cigarettes a day” (s2), “no” (s3).

D. Definition 4 [History-segment of a subject]

Given a context, say C1, it is reasonably to consider the possibility that a subject elapses a time interval $\Delta t1$ under the state s1, and subsequently a time interval $\Delta t2$ under the state s2, etc. For short, let us adopt the following dot notation: s1.n to represent the fact that a subject has elapsed a time interval of n time units with C1=s1. Let us consider for example the context C1= ”Cigarette smoke”, with time unit = year. Referring to the context C1, let us suppose that a subject X, who is 60, has spent the segment ΔT of his/her life between 20 and 60 years (ΔT = 40 years) in the following way. For 5 years he/she had been smoking less than 10 cigarettes a day, then for 25 years he/she had been smoking more than 10 cigarettes a day, then he/she decided to stop smoking. Given the above definition of the 3 states of the context C1, the formal representation of such a history segment of the life of X with respect to the context “Cigarette smoke” and regarding the time interval (of 40 years) between the age of 20 years and the age of 60 years, is given by: $X_{C1,20,60}$ = (s1.5, s2.25, s3.10). In general, let us define “history segment of a subject X with respect to a context C and regarding the time interval between the age A1 and the age A2” the sequence $X_{C,A1,A2}$ = (st1.n, st2.m, st3.q, ...), where st1, st2, st3 stand for possible states of C, and n + m + q + … = ΔT = A2 – A1. As for A2, it is the age of the subject at the time of the current session. As for A1, it is 0 in case the current session is the first one, the age of the subject at the time of the last session in case the current session is not the first one. Let us notice that there might be some contexts for which the initial state remains constant in time. This is the case, for example, of the context “Genetic predisposition”. If, on the basis of historical family-anamnesis, a subject X is considered to have genetic predisposition to cardiac infarct, then during the first monitoring session of X the context “Genetic predisposition” is instantiated to the state “yes” and such state keeps constant in time (there is no reason to change it in the future).

E. Definition 5 [History of a subject]

Given a subject X and a context C, let us call “history of a subject X with respect to a context C and regarding the age A”, and let us denote with $X_{C,A}$, the whole chain of all the history segments of X with respect to C, from the birth to the age A. For example, given the history segments $X_{C1,A1,A2}$ = (s1.n1, s2.m1), $X_{C1,A2,A3}$ = (s2.n2, s3.m2), where (A2 – A1) + (A3 – A2) = A, let us define “history of X with respect to C1 and regarding the age A” the union (in temporal sequence) of all the history segments of X with respect to C1, that is $X_{C1,A}$ = (s1.n1, s2.m1, s2.n2, s3.m2). Let us suppose, for example, that the subject X is 70 and, regarding the context “Cigarette smoke” (C1), he/she has spent his/her life in the following way. The first time interval $\Delta T1$ of 20 years (from the birth to the age of 20) has elapsed without smoking, i.e. $X_{C1,0,20}$ = (s3). The second time interval $\Delta T2$ of 40 years (between the age of 20 and the age of 60) has elapsed according to the history segment above defined, i.e. $X_{C1,20,60}$ = (s1.5, s2.25, s3.10). The third time interval $\Delta T3$ of 10 years (between the age of 60 and the age of 70) has elapsed in state “no” for the first 5 years and in state “yes under 10 cigarettes a day” for the remaining 5 years, so that $X_{C1,60,70}$ = (s3.5, s1.5). On the basis of these 3 history segments let us build the history of the subject X with respect to the context C1 = ”Cigarette smoke” and regarding the age A = 70 years: $X_{C1,70}$ = ($X_{C1,0,20}$, $X_{C1,20,60}$, $X_{C1,60,70}$) = (s3.20, s1.5, s2.25, s3.10, s3.5, s1.5). Let us notice that inside a history there might be some sub-sequences that might be compacted. For example, the sub-sequence “s3.10, s3.5” indicates that after a period of 10 years in state s3, there is a period of 5 years in the same state s3. This is equivalent to consider a single period of 15 years in state s3. We can therefore say that $X_{C1,70}$ = (s3.20, s1.5, s2.25, s3.15, s1.5). Obviously another subject, say Y, that is 70, might have a different history with respect to C1. For example, $Y_{C1,70}$ =

(s3.25, s1.15, s3.15, s1.15).

F. Definition 6 [History variable]

Given a context C, let us define the history variable HC_A that represents "History with respect to the context C and regarding the age A". The set of values of HC_A is defined by the histories of the single subjects with respect to C and regarding the age A. More formally, $HC_A = \{X_{C,A}, Y_{C,A}, Z_{C,A}, \dots\}$, where X, Y, Z, … are the subjects of the population. For example, let us consider again $X_{C1,70}$ = (s3.20, s1.5, s2.25, s3.15, s1.5). Another subject Y might add the history $Y_{C1,70}$ = (s3.25, s1.15, s3.15, s1.15) so that $HC1_{70} = \{X_{C1,70}, Y_{C1,70}\}$ = {(s3.20, s1.5, s2.25, s3.15, s1.5) , (s3.25, s1.15, s3.15, s1.15)}. The concept of history variable allows to consider histories regarding a context C and an age A, independently from the subject the history belongs to.

G. Definition 7 [Profile]

Let SC = {C1, C2, …, CN} be the set of contexts that are considered. Let $HC1_A$, $HC2_A$, …, HCN_A be the set of the related history variables regarding age A. A set of values of these history variables, i.e. {$HC1_A$=(…), $HC2_A$(…), …, HCN_A=(…)}, is said to be a profile with respect to SC and regarding the age A and is denoted by $prof_{SC,A}$, for short: $prof_A$

In general, for an age A we have a set of profiles. The symbol $prof_A$ denotes a single profile among the set of profiles related to age A. For example, beside "Cigarette smoke" (C1) let us consider two other contexts: "Genetic predisposition" (C2) and "Obesity" (C3). Let {"yes" (s1) , "no" (s2)} be the set of states for both C2 and C3. Given A = 70 years, we can write $prof_{70}$ = {$HC1_{70}$ =(s3.20, s1.5, s2.25, s3.15, s1.5), $HC2_{70}$ =(s2.70), $HC3_{70}$ =(s2.50, s1.20)} to denote a profile in the set of profiles regarding age 70. Such a profile represents the case of a subject that, with respect to "Cigarette smoke", has a history given by $X_{C1,70}$ (see above in def. 5), with respect to "Genetic predisposition", indicates that the subject does not have any genetic predisposition, with respect to "Obesity", shows that the subject is obese since he/she is 50.

H. Definition 8 [Counter variable]

At each subject age the model collects profiles regarding that age. Subjects give their contributions by adding their profiles. Let us distinguish between the case in which the profile is added under the condition E = not-occurred (for short E_n) from the case in which the profile is added under the condition E = occurred (for short E_y). For example, it might happen that when a subject adds the related profile the event E has not occurred yet. In this case the subject adds the profile under the condition E_n . Vice versa, it might happen that the profile is added under the condition E_y . Given these considerations let us associate to a profile $prof_A$ two counter variables: E_nprof_A and E_yprof_A . The variable E_nprof_A contains the number of subjects that have added their profiles, related to age A, under the condition E_n . Similarly, the variable E_yprof_A contains the number of subjects that have added their profiles, related to age A, under the condition E_y . For example, if the subjects Z and X, at the age of 70 years, have the same profile $prof_{70}$ under the condition E_n , then E_nprof_{70} = 2, and E_yprof_{70} = 0.

Figure 1 shows how profiles are stored in model memory.

Age	**HC1**	**HC2**	**HC3**	**$E_y$prof**	**$E_n$prof**
….					
61	(s3.40, s1.21)	(s1.61)	(s2.61)	2	113
61	(s3.20, s1.5, s2.25, s3.10, s1.1)	(s2.61)	(s2.40, s1.21)	3	124
…					
62	(s3.20, s1.5, s2.25, s3.10, s1.2)	(s2.62)	(s2.40, s1.22)	4	117
62	(s3.30, s1.10, s2.22)	(s1.62)	(s2.50, s1.12)	5	87
…					

Fig. 1 An example of how profiles are stored in model memory. In the example only three contexts are considered. For each age there is a sub-set of profiles. Given a row, the set of values of HC1, HC2, HC3 related to that row constitutes a profile.

III. Time-slices based causal network

Calculating probabilistic predictions requires time modelling. In fact in real life, time elapses in a continuous way, but in the model we get time to elapse in a discrete way, that is as a sequence of time-slices: time-slice 1, time-slice 2, etc.

A. *Time-slices*

Let us consider different possible time units used to express the age of a subjects: “year”, “month”, “week”, “day”. The choice of the right time unit depends on the type of application. Among the four time units, let *tu* be the one suitable for a given application. The time-slice concept used in the model has the following features.

Feature 1.

The time interval of one *tu* in the real world, gets concentrated in a single point in the model: the related time-slice. For example, the time interval of the first *tu* of life gets concentrated in time-slice 1. For short, let time-slice i be denoted by tsi (e.g. time-slice 1 is denoted by ts1, time-slice 2 by ts2, etc.)

Feature 2.

Each subject age has associated the related time-slice: (age= 1 *tu*) ⇔ ts1, (age= 2 *tu*) ⇔ ts2, etc.

Feature 3.

Each time slice collects the event variable E and the history variables HC1, HC2, … So in ts1 there are E_1, $HC1_1$, $HC2_1$, … in ts2 there are E_2, $HC1_2$, $HC2_2$, … and so forth.

Feature 4.

A value of a history variable in a time-slice tsi, i.e. a value of HC_i , represents the history of a subject, with respect to the context C, from the birth to the time-slice tsi. For example, if a subject X has elapsed the first *tu* of life in state C1=s1, and then the second *tu* of life in state C1=s2, then the values of $HC1_1$ and $HC1_2$ are (s1.1) and (s1.1, s2.1) respectively.

Feature 5.

If in the real world, E occurs to X in the course of the *tu* of life i, in the model E occurs to X in the time-slice tsi, and such occurrence is carried out by assigning E_i=occurred.

Feature 6.

Let us suppose that E occurs to X in the course of the *tu* of life n. If in the first part of the *tu* of life n (i.e. the part of the *tu* of life n before E occurrence) a context C is in state st, then let us assume that X has spent the whole *tu* of life n with C=st (it is an approximation). For example, let us suppose that E occurs to X in the course of the year of life 1. If during the part of the first year of life before E occurrence, C1 is in state s1, then let us perform, in ts1, the assignment: $HC1_1$ = (s1.1).

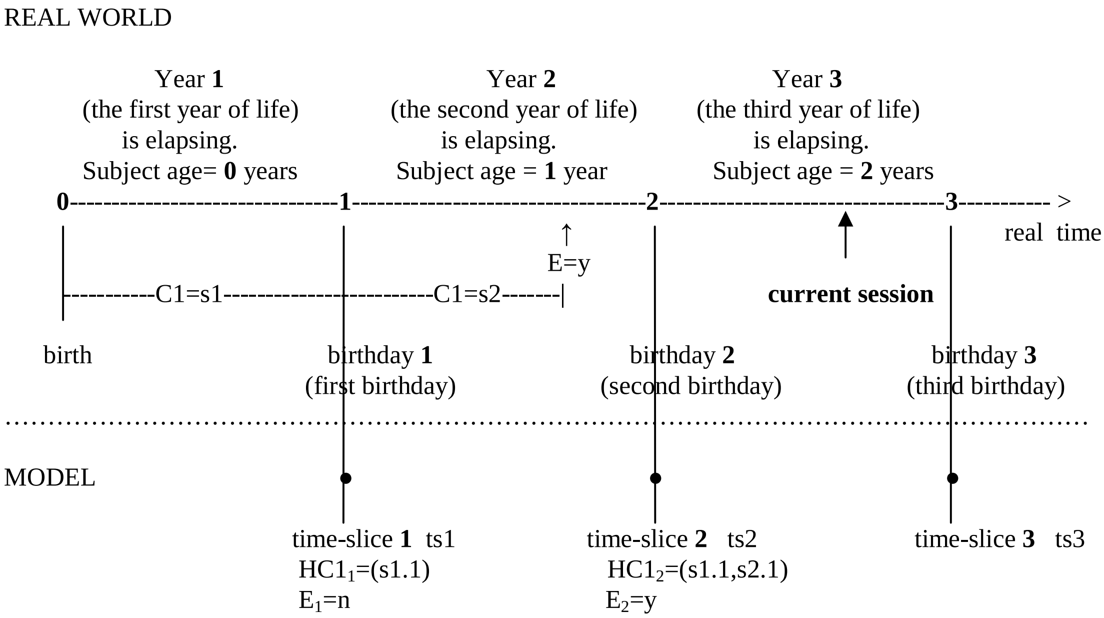


Fig. 2 An application example of the time-slice concept.

Figure 2 illustrates the concept of time-slice. Let us comment it. Let X be a subject and *tu* = year. Let us suppose that C1 is the only context being considered. The interval time defined, in the real world, by the first year of life is concentrated, in the model, in a single point: the time-slice ts1, the second year of life is concentrated in ts2, etc. In the real world we have that the current monitoring session is occurring during the third year of life of X, and since X is 2 years old we have to do with the time-slices ts1 and ts2. In the session the expert acquires the following facts: 1) X spends the first year of life with CT1=s1; 2) in the course of the second year of life the event E occurs to X; 3) the part of the second year of life before E occurrence, elapses with C1=s2. Such a situation is represented in the model by: {HC_1=(s1.1), E_1=n} in ts1; {HC_2=(s1.1,s2.1), E_2=y} in ts2. Let us now suppose that: 1) E occurs to X between the second birthday and the time of the current session; 2) C1 = s2 for all the second year and during the interval between the second birthday and the time of E occurrence. Such a situation is represented in the model by: {HC_1=(s1.1), E_1=n} in ts1; {HC_2=(s1.1,s2.1), E_2=n} in ts2 ; {HC_3=(s1.1,s2.2), E_3=y} in ts3.

When, for a subject, it happens that at a certain time-slice tsi the expert sets E_i = "occurred", then that subject is no longer monitored. For example, if a subject X has his/her first cardiac infarct at the age of 61 (i.e. during his/her sixty second year of life), the expert, during the session related to age 62 (i.e. the session corresponding to the time-slice 62) enters E_{62} = occurred. After this session the subject is no longer monitored. The fact that after E occurrence the subject is no longer monitored, has the following important implication. For each time-slice i (where $i \geq 2$) we are sure that the profiles stored in the model have been entered or updated with the implicit fact E_{i-1}=not-occurred.

By using the concept of time-slice let us define, in the next sub-section, the concept of time-slices based causal network.

the probability distribution on the states of E is constant in time. It may vary due to the only fact that time elapses. For example, the probability of the first cardiac infarct may be affected by the age of the subject. In general, given two time-slices: a present one (tsi) and a future one (tsj), we can state that it might be that $P(E_j=y) \neq P(E_i=y)$ for the only reason that an interval time corresponding to tsj – tsi has elapsed. Let such a situation be represented by a mere causal chain whose definition has the following components.

Component 1.

The E variables E_i , E_{i+1} ,…, E_j, that are present in the respective time-slices tsi, tsi+1, …, tsj, become nodes of the chain (E nodes).

Component 2.

E nodes are connected by causal links: $E_i \rightarrow E_{i+1} \rightarrow \ldots \rightarrow E_j$

Component 3.

Links connecting E nodes represent time elapsing. For this reason they are called temporal links.

Let us recall what has been stated above (sect. 2, definition 2): "the occurrence probability of E for a subject may be affected by both the mere aging of the subject and the contexts C1, C2, … in which the subject ages". For example, the value of $P(E_j=y)$ may be affected by both the fact that the interval

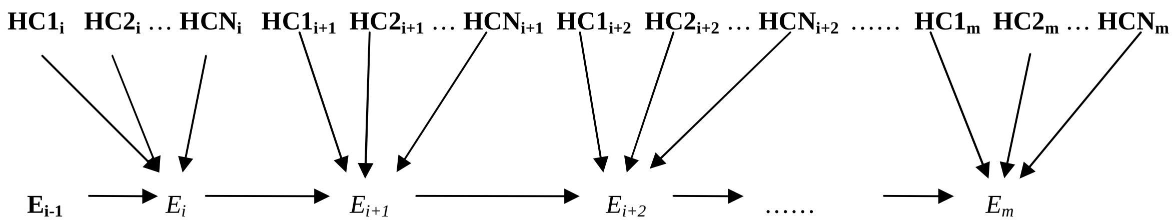


Fig. 3. The structure of the causal network used to produce predictions. The time-slice i-1 (i.e. age i-1) relates to the age of the subject at the time of the present session. The time-slice i (i.e. age i) ($i \geq 2$) relates to the first future age of the subject. The time-slice m (i.e. age m) relates to the last future age. The arrows connecting E nodes represent temporal links. The nodes in bold are instantiated: the set of nodes $HC1_i$, …, HCN_i are instantiated by a profile $prof_i$, etc. The probability values of the E nodes in italic are to be calculated.

B. *Time-slices based causal network*

Let us call "probability of E occurrence at time-slice j" the probability that the expert has to assign E=occurred in time-slice j. Such a probability is denoted by $P(E_j$=occurred). For short let us adopt the following abbreviations: E=n stands for E= not-occurred, whereas E=y stands for E= occurred. If at present we are in time-slice i with E_i=n, the ultimate purpose of the model is to calculate the values of $P(E_j=y)$ for each j > i. In general, for many real world domains, we are not sure that

time between the birth and the age related to tsj has elapsed, and the fact that such time has elapsed in the context-states sequence defined by the values of $HC1_j$, $HC2_j$, … . We are so prompted to enrich the causal chain by adding the following components.

Component 4.

The history variables that are present in the time-slices become nodes of the network (HC nodes).

Component 5.

In each time-slice, HC nodes are connected to the E node of the same time-slice by causal links. For example, in tsj we have: $HC1_j \rightarrow E_j$; $HC2_j \rightarrow E_j$; …

Putting all the components together we get a causal network. The causal network that is used by the model for producing predictions is illustrated in Figure 3.

IV. Probabilistic network

Given two events: A, B, where B causes A, the probability of A occurrence conditioned to B occurrence, for short the probability of A given B, denoted by P(A|B), is defined as $P(A \mid B) = \dfrac{P(A,B)}{P(B)}$ From this definition we get the so-called Product Rule $P(A \mid B) \cdot P(B) = P(A,B)$.

The concept of conditioned probability needs to be embedded in a suitable conceptual structure in order to be actually useful when facing real world problems. This section illustrates both how the concept of conditioned probability is embedded in the conceptual structure of the time-slices based causal network, and the mathematical foundation of the model.

Let us consider the causal network of Figure 3 and let us enrich it with the concept of conditioned probability in the following way. Let us establish that a causal relation like B → A (B causes A) has associated a quantitative aspect: the value of P(A|B), value that represents the strength of the causal relation. A causal network enriched by such quantitative aspect is called probabilistic causal network (for short: probabilistic network). The probabilistic network of Figure 3 is automatically created by the model (when the expert requires prediction), and instantiated to the future ages for which it is possible to simulate the future. For each future age the related history variables are automatically instantiated with the simulated profile produced for that age. The instantiated network is so ready for the algorithm of probability calculus, but a question arises: how can such a network be used to produce predictions? If i-1 (where $i \geq 2$) is the time-slice of the present time, probabilistic predictions consist in calculating, for each future time-slice k, where $i \leq k \leq m$, the value of

$$P(E_k = y \mid E_{i-1} = n, HC1_i, ...HCN_i, ..., HC1_k, ..., HCN_k)$$

In order to be able to calculate the value of such a probability we have to accomplish some probabilistic reasoning. The following section illustrates two different but equivalent approaches to such a probabilistic reasoning: the evidence-propagation based approach, and the global joint probability table based approach.

A. Conditioned probability based reasoning: evidence-propagation based approach

Let us define the following theorem characterized by the fact that it uses evidence-propagation rules in a causal network. Let $E_n prof_{A,}$ and $E_y prof_{A,}$ be two counter variables (related to time-slice A) containing the numbers of subjects with profile $prof_{A,}$ under the conditions E=n and E=y respectively.

Theorem 1 [reasoning based on evidence-propagation]

$$P(E_k = y \mid E_{i-1} = n, HC1_i, ...HCN_i, ..., HC1_k, ..., HCN_k)$$

=

L_k if k=i

$L_k \cdot (1 - X_{k-1}) + X_{k-1}$ if $i < k \leq m$

where

1)

$$L_k = \frac{E_y prof_k}{E_y prof_k + E_n prof_k}$$

where $prof_k$ represents the profile instantiating the set of history variables $HC1_k$,…,HCN_k related to time-slice k (i.e. age k)

2)

$$X_{k-1} = P(E_{k-1} = y \mid E_{i-1} = n, HC1_i, ..., HCN_{k-1})$$

that is: X_{k-1} is the prediction value calculated for the time-slice k-1.

Proof

Let us premise the following considerations.

For each profile $prof_k$ ($2 \leq i \leq k \leq m$) it can be stated the implicit fact

$$E_{k-1} = n$$

Given that, by adopting the Frequency Probability definition, it can be stated that

$$P(E_k = y \mid E_{k-1} = n, HC1_k, ...HCN_k) = L_k$$

Obviously it has to be intended as an empirical probability value approximating the theoretical probability value, approximation that is as smaller as greater the sum

$E_y \, prof_k + E_n \, prof_k$ is.

Given this premise, let us enter the theorem proof.

Let us consider first the case of k = i.

If k=i we have

$$P(E_i = y \mid E_{i-1} = n, HC1_i, ...HCN_i)$$

but this is just L_i

Let us now consider the case $i < k \le m$.

For short let us use the symbol A to denote the sequence:

$$E_{i-1} = n, HC1_i, ..., HCN_i, ..., HC1_k, ..., HCN_k$$

It can be stated that

$$\begin{aligned} &P(E_k = y \mid A) = \\ &P(E_k = y \mid E_{k-1} = n, A) \cdot P(E_{k-1} = n \mid A) + \\ &P(E_k = y \mid E_{k-1} = y, A) \cdot P(E_{k-1} = y \mid A) \end{aligned} \quad (1)$$

In fact:

1) by applying the product rule we have

$$\begin{aligned} &P(E_k = y \mid E_{k-1} = n, A) \cdot P(E_{k-1} = n \mid A) = \\ &P(E_k = y, E_{k-1} = n \mid A) \end{aligned}$$

and similarly

$$\begin{aligned} &P(E_k = y \mid E_{k-1} = y, A) \cdot P(E_{k-1} = y \mid A) = \\ &P(E_k = y, E_{k-1} = y \mid A) \end{aligned}$$

2) since the two joint events (E_k=y, E_{k-1}=n) and (E_k=y, E_{k-1}=y) are mutually exclusive, on the basis of the addition axiom we have:

$$\begin{aligned} &P(E_k = y, E_{k-1} = n \mid A) + P(E_k = y, E_{k-1} = y \mid A) = \\ &P((E_k = y, E_{k-1} = n \mid A) OR (E_k = y, E_{k-1} = y \mid A)) \end{aligned}$$

3) since the set of states {E_{k-1}=n , E_{k-1}=y} is exhaustive, we have:

$$\begin{aligned} &P((E_k = y, E_{k-1} = n \mid A) OR (E_k = y, E_{k-1} = y \mid A)) = \\ &P(E_k = y \mid A) \end{aligned}$$

On the basis of these considerations let us rewrite the (1) as follows (for short the sequence $HC1_i$, …, HCN_i, …, $HC1_k$, …, HCN_k is represented by $HC1_i$, …, HCN_k):

$$P(E_k = y \mid E_{i-1} = n, HC1_i, ..., HCN_k) =$$

$$\begin{aligned} &P(E_k = y \mid E_{k-1} = n, \\ &E_{i-1} = n, HC1_i, ..., HCN_k) \cdot \end{aligned} \quad (2)$$

$$P(E_{k-1} = n \mid E_{i-1} = n, HC1_i, ..., HCN_k) + \quad (3)$$

$$\begin{aligned} &P(E_k = y \mid E_{k-1} = y, \\ &E_{i-1} = n, HC1_i, ..., HCN_k) \cdot \end{aligned} \quad (4)$$

$$P(E_{k-1} = y \mid E_{i-1} = n, HC1_i, ..., HCN_k) \quad (5)$$

Let us consider the (2). Every causal path connecting the nodes E_{i-1} , $HC1_i$,…, HCN_i ,…, $HC1_{k-1}$,…, HCN_{k-1} to the node E_k is a serial structure in which E_{k-1} is the last but one node. Since E_{k-1} is instantiated to a state (i.e. the state n), each of its antecedents (that is the nodes E_{i-1} , $HC1_i$,…, HCN_{k-1}) does not affect E_k so they can be neglected (it is the so-called “evidence-propagation rule for serial structures”) and therefore the (2) is equivalent to

$$P(E_k = y \mid E_{k-1} = n, HC1_k, ...HCN_k)$$

which is L_k

Let us consider the (3). The value of the (3) is complementary to the value of the (5).

Let us consider the (4). The value of the (4) is 1. In fact if E_{k-1}=y, then E_k=y independently of the combination of context states in session k: if when we are in time-slice k-1 we know that E has occurred, then the knowledge of that fact does not change for all the subsequent time-slices.

Finally let us consider the (5). The nodes E_{k-1} , $HC1_k$,…, HCN_k are all direct causes of the node E_k (there is a causal structure converging to E_k). Since E_k is not instantiated to any of its states, its causes are all independent (it is the so-called “evidence-propagation rule for converging structure”). Therefore the nodes $HC1_k$,…, HCN_k does not affect E_{k-1} , and as a consequence they can be neglected. The ultimate consequence is that the (5) is equivalent to

$$P(E_{k-1} = y \mid E_{i-1} = n, HC1_i, ..., HCN_{k-1}) \quad (6)$$

But the value of the (6) is the prediction value calculated for session k-1.

Putting all together, it can be stated that:

$$P(E_k = y \mid E_{i-1} = n, HC1_i,..., HCN_k) =$$
$$L_k \cdot (1 - X_{k-1}) + X_{k-1}$$

where X_{k-1} stands for the prediction value calculated for session k-1.
In conclusion:

$$P(E_k = y \mid E_{i-1} = n,$$
$$HC1_i,...HCN_i,..., HC1_k,..., HCN_k)$$

$$=$$

$$L_k \qquad \text{if } k=i$$

$$L_k \cdot (1 - X_{k-1}) + X_{k-1} \qquad \text{if } i < k \leq m$$

End of proof [Theorem 1]

B. Conditioned probability based reasoning: global joint probability-table based approach

We know that from the global joint probability table of a network we can calculate the probability values of all the nodes of the network. Such a joint probability table can be built by applying the so-called Chain Rule. The Chain Rule, which is an application of the Product Rule, is defined as:

$$P(A_1,..., A_n) = \prod_{i=1}^{n} P(A_i \mid pa(A_i))$$

where $pa(A_i)$ stands for "direct parents of A_i".
Let us look at Figure 3 again and let us build the following theorem (Theorem 2) for calculating the value of

$$P(E_k = y \mid E_{i-1} = n,$$
$$HC1_i,...HCN_i,..., HC1_k,..., HCN_k)$$

Theorem 2, based on Chain Rule application, defines a general algorithm, that is equivalent to the one defined in Theorem 1.

Theorem 2 [reasoning based on global joint probability-table]

$$P(E_k = y \mid E_{i-1} = n,$$
$$HC1_i,...HCN_i,..., HC1_k,..., HCN_k) =$$
$$L_k \cdot (1 - L_{k-1}) \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i) +$$
$$L_{k-1} \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i) +$$
$$...$$
$$L_{i+1} \cdot (1 - L_i) +$$
$$L_i$$

where L_j (where j is such that: $i \leq j \leq k$), is given by

$$L_j = \frac{E_y\, prof_j}{E_y\, prof_j + E_n\, prof_j}$$

where $prof_j$ represents the profile instantiating the set of history variables $HC1_j$,…,HCN_j related to time-slice j (i.e. age j)

Proof

For short, the sequence $HC1_i$, …, HCN_i, …, $HC1_k$, …, HCN_k is represented by $HC1_i$, …, HCN_k . Before entering the reasoning of the proof let us define the following five general rules:

RULE 1) Given that $P(E_{i-1}=n)=1$, and the symbols $HC1_i$,…, HCN_k denote context instantiations (that is $HC1_i$ means $HC1_i$ = st, etc.) we have that: $P(HC1_i)=1$,…, $P(HCN_k)=1$. As a consequence:

$$P(E_{i-1} = n) \cdot P(HC1_i) \cdot ... \cdot P(HCN_k) = 1$$

RULE 2) Given a time-slice j ($j \geq 2$) it can be stated that:

$$P(E_j = y \mid E_{j-1} = n, HC1_j,..., HCN_j) = L_j$$

RULE 3) On the basis of Rule 2 it can be stated that:

$$P(E_j = n \mid E_{j-1} = n, HC1_j,..., HCN_j) = (1 - L_j)$$

RULE 4) As already noticed, if E_{j-1}=y, then E_j=y independently of the combination of context states. As a consequence:

$$P(E_j = y \mid E_{j-1} = y, HC1_j, ..., HCN_j) = 1$$

RULE 5) On the basis of Rule 4 it can be stated that:

$$P(E_j = n \mid E_{j-1} = y, HC1_j, ..., HCN_j) = 0$$

Let us now enter the proof.
The proof is defined by an algorithm structured in three basic sequential steps.

STEP 1) Let us consider the global joint probability table of the network in Figure 3. The value of

$$P(E_k = y \mid E_{i-1} = n, HC1_i, ..., HCN_k)$$

is calculated by adding the values of all the table rows containing $E_k = y$ and $E_{i-1} = n, HC1_i, ..., HCN_k$.
In more formal terms:

$$\begin{aligned}
&P(E_k = y \mid E_{i-1} = n, HC1_i, ..., HCN_k) = \\
&P(E_k = y, \\
&E_{k-1} = n, E_{k-2} = n, E_{k-3} = n, ..., E_i = n, \\
&E_{i-1} = n, HC1_i, ..., HCN_k) + \\
&P(E_k = y, \\
&E_{k-1} = y, E_{k-2} = n, E_{k-3} = n, ..., E_i = n, \\
&E_{i-1} = n, HC1_i, ..., HCN_k) + \\
&....
\end{aligned}$$

Let us notice that the number of addenda is 2^{k-i} . In fact between E_k and E_{i-1} there are the k-i E nodes: E_{k-1} , E_{k-2} ,…, E_{i+1} , E_i . Such nodes are not instantiated and since each of them has two states {n, y}, the number of possible combinations for the set of states of these nodes is 2^{k-i} . As a consequence we have 2^{k-i} addenda.

STEP 2) Let us apply the Chain Rule to each addendum. Let us consider the first addendum. We get:

$$\begin{aligned}
&P(E_k = y, E_{k-1} = n, ..., E_i = n, E_{i-1} = n, \\
&HC1_i, ..., HCN_k) = \\
&P(E_k = y \mid E_{k-1} = n, HC1_k, ..., HCN_k) \cdot \\
&P(E_{k-1} = n \mid E_{k-2} = n, HC1_{k-1}, ..., HCN_{k-1}) \cdot \\
&P(E_{k-2} = n \mid E_{k-3} = n, HC1_{k-2}, ..., HCN_{k-2}) \cdot \\
&... \\
&P(E_{i+1} = n \mid E_i = n, HC1_{i+1}, ..., HCN_{i+1}) \cdot \\
&P(E_i = n \mid E_{i-1} = n, HC1_i, ..., HCN_i) \cdot \\
&P(E_{i-1} = n) \cdot P(HC1_i) \cdot ... \cdot P(HCN_k)
\end{aligned}$$

By applying rules 1, 2, 3 the first addendum becomes:

$$\begin{aligned}
&P(E_k = y, E_{k-1} = n, E_{k-2} = n, ..., E_i = n \\
&E_{i-1} = n, HC1_i, ..., HCN_k) = \\
&L_k \cdot (1 - L_{k-1}) \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i)
\end{aligned}$$

Let us now consider the second addendum. By applying the Chain Rule and then rules 1, 2, 3, 4, we get:

$$\begin{aligned}
&P(E_k = y, E_{k-1} = y, E_{k-2} = n, ..., \\
&E_{i-1} = n, HC1_i, ..., HCN_k) = \\
&L_{k-1} \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i)
\end{aligned}$$

And so forth for the remaining addenda.

STEP 3) Let us sum the results obtained by applying the chain Rule and then the above five rules. At the end we get:

$$\begin{aligned}
&P(E_k = y \mid E_{i-1} = n, \\
&HC1_i, ...HCN_i, ..., HC1_k, ..., HCN_k) = \\
&L_k \cdot (1 - L_{k-1}) \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i) + \\
&L_{k-1} \cdot (1 - L_{k-2}) \cdot ... \cdot (1 - L_{i+1}) \cdot (1 - L_i) + \\
&... \\
&L_{i+1} \cdot (1 - L_i) + \\
&L_i
\end{aligned}$$

End of proof [Theorem 2]

Let us notice the equivalence between the two algorithms defined by the two theorems respectively.

For example, let us suppose that k = i+2. By applying the algorithm of Theorem 2 we have:

$$
\begin{aligned}
&P(E_{i+2} = y \mid E_{i-1} = n, PC1_i, \ldots, PCN_{i+2}) = \\
&L_{i+2}(1-L_{i+1})(1-L_i) + \\
&L_{i+1}(1-L_i) + \\
&L_i
\end{aligned}
$$

By applying the algorithm of Theorem 1 with k=i+2 we have:

$$
\begin{aligned}
&P(E_{i+2} = y \mid E_{i-1} = n, PC1_i, \ldots, PCN_{i+2}) = \\
&L_{i+2} \cdot (1 - X_{i+1}) + X_{i+1}
\end{aligned}
$$

where

$$X_{i+1} = L_{i+1} \cdot (1 - X_i) + X_i$$

where

$$X_i = L_i$$

We can notice that even if the two results appear to be formally different, they express the same conclusion.

V. General tool for robabilistic predictive monitoring

So far we have defined a hierarchy of three conceptual levels: the first level (at the bottom) is a collection of the basic concepts underlying the whole model, the second level (in the middle) defines the concept of time-slices based causal network using the concepts of the first level, the third level (at the top) consists in the probabilistic network definition, definition obtained by associating a conditioned probability value to each causal relation of the causal network defined in the second level. The conditioned probability concept has been so embedded in a proper structure and probabilistic reasoning has taken place producing two equivalent prediction algorithms. Having now at our disposal a general probabilistic prediction algorithm, the following question arises: how such an algorithm could, in practice, be used inside a predictive tool? We are therefore prompt to define a general monitoring tool in which the general probabilistic prediction algorithm is embedded. Let us now pass to define the structure and the features of such a tool.

A. The basic algorithm of the general predictive monitoring tool

The tool works according to the following basic algorithm. At the beginning of a monitoring session of a subject X the model asks the expert to enter historic facts concerning the subject: has E occurred to X? In which context-states time has elapsed for X? If E has occurred to X, then the model memory is updated on the basis of the entered facts. Vice versa, if E has not occurred to X, the expert can ask the model for predictions about the probability that E occurs to X in the future. Then before the session ends the model memory is updated on the basis of the entered facts. More formally, the basic algorithm is the following.

Beginning of the current session
  ACQUIRE SUBJECT DATA
  If E has not occurred to X **then**
    While the expert has not yet clicked on "Session end"
      SIMULATE THE FUTURE AND
      PRODUCE PREDICTIONS
    End of While
    UPDATE MODEL MEMORY
  End of If
  If E has occurred to X **then**
    UPDATE MODEL MEMORY
  End of If
End of the current session

The following sub-sections illustrate the algorithm in a deeper way.

B. Acquire subject data

This sub-algorithm concerns the acquisition of both the state of E and the history segments. Let us examine the two acquisitions separately.

At the beginning of the current monitoring session of a subject X, the algorithm, in order to be able to asks the expert to enter, for each context C, the current history segment $X_{C,A1,A2}$, needs to establish the values of the two subject ages: A1 and A2. As for A1, if the current session is the first one, A1 = 0, else A1 = "age of the subject at the preceding session". As for A2, let us distinguish between the case in which E has non occurred to X, from the case in which E has occurred to X. If the expert enters E= not-occurred, A2 = age of X at the time of the current session. If the expert enters E= occurred, then the expert is asked to specify the "age of X at the time of E occurrence". Such value is acquired and stored in a variable, say AEO (Age of the subject at the time of E Occurrence). According to Feature 6 in section III.A , let us conclude that A2= AEO + 1. These considerations explain why the first step of the algorithm is: E state acquisition.

After the E state acquisition phase the algorithm has the values of A1 and A2 and can therefore pass to the second phase, the "history segments acquisition" phase. The expert is asked to enter, for each context C, the proper history segment $X_{C,A1,A2}$. Precisely, the time interval defined by the couple (A1, A2) is the temporal length between: the birth (in the case A1 equals 0) or the birthday related to age A1 (in case A1 is different from 0), and the birthday related to age A2. For example, referring to Figure 2, “age of X at the time of the current session”= 2, AEO= 1, A2= 2, A1=0. The expert is asked to enter the history segment $X_{C1,0,2}$ that covers the time interval between the birth and the birthday 2 (according to Feature 6 in section III.A, we approximate by supposing that the condition C1=s2 covers the whole sub-interval between birthday 1 and birthday 2).

By formalizing all these considerations let us define the following sub-algorithm

ACQUIRE SUBJECT DATA
CA = current age of the subject X at the current session
If the current session is the first one **then**
A1= 0
End of If
If the current session is not the first one **then**
A1= “age of the subject at the preceding session”
End of If
Ask and Acquire the state of E
If E has not occurred to X **then**
A2= CA
End of If
If E has occurred to X **then**
Ask and Acquire the age of the subject when E occurred
Store the acquired age in the variable AEO
A2= AEO + 1
End of If
For each context C
Ask and Acquire the history segment $X_{C,A1,A2}$
Calculate the whole history $X_{C,A2}$
End of For

C. *Simulate the future and produce predictions*

If in a monitoring session of subject X the entered state of E is “not-occurred”, the tool provides the expert with the possibility of getting probabilistic predictions about the occurrence of E to X in the future, supposing that the future time of X elapses in conditions of C1=stc1, C2=stc2, …, where stc1, stc2, … are states, of C1, C2, ... respectively, selected by the expert. More precisely, let i-1 (where $i \geq 2$) be the current age of X. Let m be a future age of X. The expert is provided by the tool with the possibility of:

1) selecting for each context C the state that is supposed to be constant in the time interval between i-1 and m ;

2) asking the tool to calculate the probability values of occurrence of E to X for all the future ages k (where $i \leq k \leq m$) supposing that the time interval between i-1 and m elapses in conditions of C1=stc1, C2=stc2, … .

The tool, in order to provide these possibilities, performs a sub-algorithm that is structured in three sequential steps:
1) Ask and acquire future context states
2) Simulate the future
3) Produce predictions

More formally, the basic scheme of such sub-algorithm is the following

SIMULATE THE FUTURE AND PRODUCE PREDICTIONS
ASK AND ACQUIRE FUTURE CONTEXT STATES
SIMULATE THE FUTURE
PRODUCE PREDICTIONS

C.1 Ask and acquire future context states

This step simply consists in asking the expert to select context states and then in acquiring them. More formally:
ASK AND ACQUIRE FUTURE CONTEXT STATES
- Ask the expert to select, for each context C, the state that is supposed to be constant in the future
- Acquire the selected states {st_{C1} , st_{C2} , …} of the contexts C1, C2, … respectively

C.2 Simulate the future

This step is in turn compound of two sequential sub-steps: simulation and check. Let C1=stc1, C2=stc2, … be the states selected by the expert (in step 1). The tool, for each future age k of X,

a) creates, on the basis of the selected states, the following simulated history-segments: $X_{C1,A2,k}$ = (stc1.D), $X_{C2,i-1,,k}$ = (stc2.D), … where D = k – A2
b) builds, on the basis of these simulated history-segments and $profX_{A2}$ (the profile of X at the age A2), the new profile $profXsim_k$ (simulated profile of X that relates to the future age k).
c) checks if the simulated profile $profXsim_k$ is present in the profiles set (stored in model memory) in a statistically significant number. To be more explicit: if the simulated profile is already present in model memory and the number of cases is ≥ Thr (where Thr is the threshold value required to make significant probabilistic inferences), then the simulated profile $profXsim_k$ can be used to calculate probabilistic prediction for the age k.

The process stops at the first future age for which the check results not-OK. Let us notice that simulated profiles are created just for prediction purposes. They are not profiles got by real cases, as a consequence they are not added to the set of profiles stored in model memory (that is, they do not update real statistical data). Let us now pass to formalize this step.

SIMULATE THE FUTURE
Let Thr be the threshold of minimum number of cases required for producing statistically significant inferences
Let LFA (where LFA > A2) be the Last Future Age (of X) considered for predictions
k = A2
STOP = “no”
While (k < LFA) and (STOP = “no”) do
k = k + 1

D = k – A2
Create the following simulated history segments of X:
$X_{C1,A2,k} = (st_{C1}.D)$, $X_{C2,A2,k} = (st_{C2}.D)$, …
Calculate the following whole histories of X:
$X_{C1,k} = (X_{C1,A2}, X_{C1,A2,k})$,
$X_{C2,k} = (X_{C2,A2}, X_{C2,A2,k})$, …
Create the following simulated profile of X:
$profXsim_k = \{HC1= X_{C1,k}, HC2= X_{C2,k}, \dots\}$
If $profXsim_k$ is not present in the model **then**
STOP = "yes"
End of If
If $profXsim_k$ is present in the model **then**
If $[E_y profXsim_k + E_n profXsim_k] < Thr$ **then**
STOP = "yes"
End of If
If $[E_y profXsim_k + E_n profXsim_k] \geq Thr$ **then**
Put the couple (k, $profXsim_k$) in the list LSIM
End of If
End of If
End of While

C.3 Produce predictions

At the end of the execution of the step 2 we have the list LSIM (possibly empty) of the couples: (age, simulated profile) related to the profiles that have passed the check. Let such a list be: (i, $profXsim_i$), (i+1, $profXsim_{i+1}$), … (k, $profXsim_k$), … (m, $profXsim_m$), where i and m are the first and the last future age respectively. The profiles stored in LSIM are used for instantiating the probabilistic network of Figure 3 and, as a consequence, for calculating probabilistic predictions. More explicitly, let us consider the network of Figure 3 and the list LSIM. For each future age k (i≤k≤m) the history variables (of the time-slices tsi,… tsk): $HC1_i$,… HCN_i, … $HC1_k$,… HCN_k, are instantiated by the profiles $profXsim_i$, … $profXsim_k$ respectively and the prediction algorithm defined in Theorem 1 (or Theorem 2, since the algorithms are equivalent) is applied, producing this way the probabilistic prediction for the time-slice k. Let us notice though that the dynamic creation of the network of Figure 3 and the related nodes instantiation represent a conceptual view of the algorithm. In practice this third step is formalized in the following way.

PRODUCE PREDICTIONS
Let i = A2 + 1 (and therefore i ≥ 2)
If LSIM is empty **then**
Print: "It is not possible to produce predictions"
End of If
If LSIM is not empty **then**
Let m be the maximum age in LSIM
For k = i to m
Get the couple (k, $profXsim_k$) from the list LSIM
Select the row in model memory (e.g. Fig. 1) with
Age= k and {HC1,…, HCN} = $profXsim_k$
Get the related values of $E_y prof$ and $E_n prof$, and
calculate $L_k = E_y prof / (E_y prof + E_n prof)$
End of For
For k = i to m
Calculate (using Theorem 1 or Theorem 2), the value of
$P(E_k = y \mid E_{i-1} = n, HC1_i, .., HCN_i, \dots, HC1_k, .., HCN_k)$
Print the couple: (age= k , E probability = $P(E_k = y \mid \dots)$)
End of For
End of If

At the end of the "Produce Predictions" step the expert can repeat the "Simulate the future and produce predictions" sub-algorithm again. That is, the expert can go back and select other context states that are supposed to be constant in the future. Then he/she can activate a new simulation and prediction production process. Finally the expert, after having gathered a sufficient amount of data from such cycle of "simulation and prediction", clicks on "Session end". However, before the session actually ends, the model memory is updated on the basis of the entered facts.

D. Update model memory

A history segment $X_{C,A1,A2} = (st_C, A2\text{-}A1)$ implies: $X_{C,A1,A1+1} = (st_C, 1)$, $X_{C,A1,A1+2} = (st_C, 2)$, … , $X_{C,A1,A2} = (st_C, A2\text{-}A1)$. In other words, acquiring, in the present session, the history segment $X_{C,A1,A2}$ is equivalent to acquire the above A2-A1 sub-segments in A2-A1 virtual sessions respectively (i.e. at the age A1+1 there is a (virtual) monitoring session in which the expert enters (st_C, 1), at the age A1+2 there is a (virtual) monitoring session in which the expert enters (st_C, 2), and so forth). Let us therefore update the model memory as if we had to do with A2-A1 sub-sessions. Let us notice that for the virtual sub-sessions related to the ages A1+1, A1+2, …, A2-2, A2-1, it is certain that E_{A1+1}= not-occurred, E_{A1+2}= not-occurred, … , E_{A2-2}= not-occurred, E_{A2-1}= not-occurred. It is only for the age A2 that the state of E_{A2} may be "occurred" or "not-occurred", depending on what the expert has entered. By formalizing these considerations let us define the following sub-algorithm.

UPDATE MODEL MEMORY
If A2 > (A1+1) **then**
For I = A1+1 to A2-1 do
For each context C
Calculate $X_{C,I}$
End of For
Consider the current profile
$prof_{X,I} = \{HC1_I = X_{C1,I}, HC2_I = X_{C2,I}, \dots\}$
If in the set of profiles stored in model memory there is not a profile equal to $prof_{X,I}$ **then**
Add, in model memory, the new row:
"age=I , $prof_{X,I}$, $E_n prof$ =1 , $E_y prof$ =0"
End of If
If in the set of profiles stored in model memory there is a a profile equal to $prof_{X,I}$ **then**
Increment (of 1) the related counter variable $E_n prof$
End of If
End of For

End of If
Consider the profile
$prof_{X,A2}$ = {$HC1_{A2} = X_{C1,A2}$, $HC2_{IA2} = X_{C2,A2}$, … }
If in the set of profiles stored in model memory there is not a profile $prof_{A2}$ equal to $prof_{X,A2}$ **then**
Add, in model memory, one of the following two rows:
"age= A2 , $prof_{X,A2}$, $E_n$prof =1, $E_y$prof=0", if E=not-occurred
"age= A2 , $prof_{X,A2}$, $E_n$prof =0, $E_y$prof=1", if E=occurred
End of If
If in the set of profiles stored in model memory there is a profile $prof_{A2}$ equal to $prof_{X,A2}$ **then**
Update, in model memory, the row with age= A2 and $prof_{A2}$ = $prof_{X,A2}$ by incrementing (of 1) the counter variable $E_n$prof if E=not-occurred , $E_y$prof if E=occurred
End of If

VI. A SIMULATED-CASE STUDY

In order to better understand how the general tool for probabilistic predictive monitoring works, let us apply it to a simulated example. Let us consider the above example where E = "First cardiac infarct", and the considered contexts are: Smoke (with states "yes less than 10 cigarettes a day" (s1), "yes 10 or more cigarettes a day" (s2), "no" (s3)), Genetic predisposition (with states "yes" (s1), "no" (s2)), Obesity (with states "yes" (s1), "no" (s2)). Let *tu* (time unit) = year.

A. Acquiring subject data in the case in which the current session is the first one

Let us suppose that for the subject X the current session is the first one and the current age is 60 years. As for the state of E let us suppose that at the time of the current session E has not occurred to X, and therefore the expert enters E_{60}=not-occurred. The expert is then asked to enter the history segments $X_{C1,0,60}$, $X_{C2,0,60}$, $X_{C3,0,60}$, history segments concerning the time interval from the birth to the birthday 60 (i.e. A1= 0, A2= 60). As for the context "Smoke" (C1) let us suppose that X has begun to smoke at the age of 20 years. Then for five years X has been smoking less than 10 cigarettes a day. Then for 25 years X has been smoking more than 10 cigarettes a day. Finally, at the age of 50 years X stopped smoking. As for the context "Genetic predisposition" (C2) let us suppose that X does not have any genetic predisposition regarding cardiac infarct. Finally, as for the context "Obesity" (C3) let us suppose that X has been obese since the age of 40 years. The expert enters such case description by filling in suitable user-friendly forms. The tool acquires this information and converts it into formal history segments: $X_{C1,0,60}$ =(s3.20, s1.5, s2.25, s3.10), $X_{C2,0,60}$ =(s2.60), $X_{C3,0,60}$ =(s2.40, s1.20). Since E= not-occurred, the expert is provided with simulation and prediction facilities.

B. Asking future context states

Let us now suppose that the expert asks the model for predictions. The model shows the following list and asks the expert to select, for each context, the state that has to be kept constant in the simulated future time of X.
"Cigarette smoke" (C1)
- "yes under 10 cigarettes a day" (s1)
- "yes 10 or more cigarettes a day" (s2)
- "no" (s3)

"Genetic predisposition" (C2)
- "yes" (s1)
- "no" (s2)

"Obesity" (C3)
- "yes" (s1)
- "no" (s2)

Let us suppose that the expert (interested in simulating that X begins again to smoke less than 10 cigarettes a day and continues to be obese) selects: C1=s1, C2=s2 (obviously this state cannot change), C3=s1.

C. Simulating the future

Let us consider the sub-algorithm "Simulate the future". In the present case we have A2 = 60, $profX_{60}$ = {$HC1_{60}$ = (s3.20, s1.5, s2.25, s3.10), $HC2_{60}$ = (s2.60), $HC3_{60}$ = (s2.40, s1.20)}, set of selected states = {C1=s1, C2=s2, C3=s1}. Let us suppose we are interested in the 10 future years, so LFA= 70. Let us now perform the While statement (see sect. V.C.2). Let us consider the first loop (k = 61 years, D= 1 year). Let us create the following simulated history segments: $X_{C1,60,61}$ = (s1.1), $X_{C2,60,61}$ = (s2.1), $X_{C3,60,61}$ = (s1.1). Let us calculate the following simulated histories: $X_{C1,61}$ = ($X_{C1,0,60}$, $X_{C1,60,61}$) = (s3.20, s1.5, s2.25, s3.10, s1.1); $X_{C2,61}$ = ($X_{C2,0,60}$, $X_{C2,60,61}$) = (s2.61) ; $X_{C3,61}$ = ($X_{C3,0,60}$, $X_{C3,60,61}$) = (s2.40, s1.21). Let us create the simulated profile for the age 61, $profXsim_{61}$ ={HC1= (s3.20, s1.5, s2.25, s3.10, s1.1), HPC2= (s2.61), HC2= (s2.40, s1.21)}.

Let us check if $profXsim_{61}$ is already present, the model and, if it is, let us check if [$E_y profXsim_{61}$ + $E_n profXsim_{61}$] ≥ Thr. Let us suppose that the model memory contains the rows illustrated in Figure 1. If, for example, Thr = 100, we can conclude that the check outcome is OK, so let us put the couple (61, $profXsim_{61}$) into the list LSIM. And so forth for all the remaining future ages (i.e. for k= 62, D= 2; …. k= 70, D= 10). For example, in the second loop the simulated profile is created starting from the history segments: $X_{C1,60,62}$ = (s1.2), $X_{C2,60,62}$ = (s2.2), $X_{C3,60,62}$ = (s1.2). That is: $profXsim_{62}$ ={HC1= (s3.20, s1.5, s2.25, s3.10, s1.2), HPC2= (s2.62), HC2= (s2.40, s1.22)}. Considering Figure 1 again, the check is OK and so we can put (62, $profXsim_{62}$) into LSIM. And so forth. Let us notice that even if LFA = 70, in LSIM we might have a number of elements less than 10. If, for example, the check outcome related to age 63 is not OK, then LSIM only contains the two couples: (61, $profXsim_{61}$), (62, $profXsim_{62}$).

D. Producing predictions

After the future simulation phase has been performed, the prediction procedure is applied to each couple in LSIM, producing this way the sequence of as many probabilistic values as the elements of LSIM. Let us consider the sub-algorithm “Produce predictions” and Figure 1. Let us suppose that LSIM only contains two elements. As a consequence: i= 61, m= 62. Let us get, from LSIM, the couple (61, $profXsim_{61}$). Since 61 is the first future age, we have that $P(E_{61} = y \mid E_{60} = n, profXsim_{61}) = L_{61} = E_y profXsim_{61} / (E_y profXsim_{61} + E_n profXsim_{61}) = 3 / (3+124) = 0.024$.
Let us now get, from LSIM, the couple (62, $profXsim_{62}$). Let us calculate $L_{62} = E_y profXsim_{62} / (E_y profXsim_{62} + E_n profXsim_{62}) = 4 / (4+117) = 0.033$. We are now ready to calculate prediction for age = 62, i.e. $P(E_{62} = y \mid E_{60} = n, profXsim_{61}, profXsim_{62})$. Let us use, for example, the algorithm of Theorem 1. The prediction for age 62 is given by $L_{62} (1 - X_{61}) + X_{61} = 0.033 (1 - 0.024) + 0.024 = 0.056$.
In conclusion, the model provides the expert with the following list concerning the subject X.

Future age	*Occurrence Probability of “First cardiac infarct”*
61	0.024
62	0.056

Let us suppose we are not interested in selecting different context states and then activate a further simulation-prediction cycle. Let us suppose we click on “Session end”. The tool, before ending the current session, updates model memory on the basis of the entered subject-data.

E. Updating model memory

The entered history segments are: $X_{C1,0,60}$ =(s3.20, s1.5, s2.25, s3.10), $X_{C2,0,60}$ =(s2.60), $X_{C3,0,60}$ =(s2.40, s1.20). Let us consider the sub-algorithm “Update model memory”. Let us perform the external statement “For I = … “ (knowing that, in the current case, A1= 0, A2= 60). Let us consider the first loop (I = 1). For age = 1 year, we have $X_{C1,1}$ =(s3.1), $X_{C2,1}$ =(s2.1), $X_{C3,1}$ =(s2.1) and as a consequence: $prof_{X,1}$ = {$HC1_1$=(s3.1), $HC2_1$=(s2.1), $HC3_1$=(s2.1)}. If such a profile is not already present in the sub-set of profiles related to age 1, then let us store the row “age=1, HC1=(s3.1), HC2=(s2.1), HC3=(s2.1), $E_n$prof=1, $E_y$prof=0” in model memory. If the profile is already present and has $E_n$prof= n $E_y$prof= m, then no new row is created but the counter variable $E_n$prof is updated: $E_n$prof=n+1. And so forth for all the remaining ages (i.e. for I= 2, …, I= 59). For example, at the loop identified by I = 41 we have $prof_{X,41}$ = {$HC1_{41}$ = (s3.20, s1.5, s2.16), $HC2_{41}$ = (s2.41), $HC3_{41}$ = (s2.40, s1.1)}, and we check if it is already present in model memory, etc. As for age = 60, since E has not occurred to X, if the profile is not present in memory already, we add the new row “age=60, HC1=(s3.20, s1.5, s2.25, s3.10), HC2=(s2.60), HC3=(s2.40, s1.20), $E_n$prof=1, $E_y$prof=0”. Vice versa, if the profile is present in memory already, let us increment (by 1) the related variable $E_n$prof.

F. Acquiring subject data in the case in which the current session is not the first one

Let us suppose that after a certain time, say 4 years, X has the second monitoring session. The current age of X is therefore 64. Let us consider the case in which E has not occurred to X. The expert is asked to enter the history segments $X_{C1,60,64}$, $X_{C2,60,64}$, $X_{C3,60,64}$, history segments concerning the time interval from the birthday 60 to the birthday 64 (i.e. A1= 60, A2= 64). Alternatively, let us consider the case in which E occurred to X, say 2 years before (age of X = 62). The expert is then asked to enter the history segments $X_{C1,60,63}$, $X_{C2,60,63}$, $X_{C3,60,63}$, history segments concerning the time interval from the birthday 60 to the birthday 63 (i.e. A1= 60, A2= 63). (As for the year between the birthday 62 to the birthday 63 let us remember Feature 6 in sect. III.A).

VII. General environment for probabilistic predictive monitoring

The general tool for probabilistic predictive monitoring (that has been defined in the preceding session) needs to be equipped with specific domain knowledge in order to be usefully applied to real world problems. More precisely, in a specific application, contexts C1, C2, … and related states s1, s2, … are instantiated by precise names that depend on the specific application field. Even the generic term *tu* (time-unit) is instantiated by a precise time-unit (year, month, week, day)

Environment for BUILDING predictive monitoring TOOLS

Environment for USING predictive monitoring TOOLS

Environment for ADMINISTERING predictive monitoring TOOLS

Environment for ADMINISTERING predictive monitoring SUBJECTS

Environment for ADMINISTERING predictive monitoring ENVIRONMENTS

Fig. 4 The home-page of the general environment for probabilistic predictive monitoring depending on the specific application. In conclusion, for each specific application the general tool for probabilistic predictive monitoring has to be instantiated by creating proper contexts, states, etc. We are therefore prompted to define a general environment for probabilistic predictive monitoring, i.e. a general environment in which users can easily build, use and administer specific probabilistic predictive monitoring tools. Such a general environment is in turn structured in five target environments (Fig. 4). The first two environments concern the construction and use of specific tools. The remaining three environments concern administration activities.

A. Environment for Building predictive monitoring Tools

The environment for building tools for specific applications, for short, the *Tools Building environment*, provides a set of functions for building a predictive monitoring tool in a friendly

<u>Cancel the tool</u>

<u>Show the tool</u>

<u>EDIT the tool name</u>

<u>EDIT the tool category</u>

<u>EDIT contexts</u>

<u>EDIT context states</u>

<u>EDIT probability levels</u>

<u>TOOL READY TO DEFINITELY ENTER THE USING ENVIRONMENT</u>

<u>TOOL READY TO BE TESTED IN THE USING ENVIRONMENT</u>

Fig. 5 The home-page of the Building environment

and effective way. The home-page of the Portal Building environment consists in a set of functions (Fig. 5) that allow the builder to create and edit the various application-oriented components of the tool (contexts, context-states, etc.). Let us examine the last two functions in Figure 5.

A.1 The function: Tool ready to definitely enter the Using environment

At the end of the building work, the new tool is like a new ship ready to leave the shipyard, i.e. the Building environment, enter the sea, i.e. the Using environment, where it will be used by the crew, i.e. the domain expert, to serve passengers, i.e. the subjects of the population considered by the current application, and the launching of the ship is carried out by executing the function "Tool ready to definitely enter the Using environment". The effects of such a function execution are:

1) a set of new database tables, that will be used by the new tool, are created in the Using environment
2) the information (contexts, states, etc.) collected in the Building environment are copied into the new database tables of the Using environment and all the database tables, created and used in the Building environment during the tool building process, are eliminated.

A.2 The function: Tool ready to be tested in the Using environment

This function provides the possibility of testing the tool under construction before declaring it finished and ready to definitely enter the Using environment. Such a possibility is very useful since the testing phase might reveal some imperfections concerning the domain knowledge entered during the building phase of the tool (for example: the name of a context is not completely appropriate, a new state should be added to the set of states of a context, etc.). Such imperfections can then be removed by resuming the building phase. The function "Tool ready to be tested in the Using environment" inserts the tool into the Using environment in TEST mode and as a consequence the database tables of the specific tool in the Building environment are not eliminated. Moreover, the Using environment, in order to make the specific tool usable in all its functions, provides a set of dummy subjects and simulates that in any session the specific tool has already collected, for any combination of context states, a number of cases greater than the threshold value required to make probabilistic inferences. As a consequence the user of the specific tool can test how the new tool works and looks in the Using environment, knowing, of course, that probabilistic values displayed during the testing phase are dummy.

B. Environment for Using predictive monitoring Tools

This environment performs the basic algorithm of the general probabilistic predictive monitoring tool (illustrated in section V.A) by applying it to the specific contexts, context states, etc. (defined in the Building environment) related to the current specific application. The Using environment is equipped with several facilities for presenting, in both quantitative and qualitative terms, probabilistic predictions. For each future age of the subject the related E occurrence probability is shown along with the related qualitative judgment (Fig. 6). Moreover, qualitative presentations are also carried out through histograms pointing out the trends of future probabilities. Such predictions may be useful in order to take suitable measures in advance, measures personalized to the subject under consideration. Moreover, the possibility to compare different predictions resulting from different simulated cases may help decision making in trade-off problems (Fig. 7). The user is also provided with the possibility of getting explanations about the procedure used to produce the predictions that have been displayed (which simulation plan has been used? Where do the displayed probability-numbers come from?).

Age	**Prob**	**Level**
71	0.015	very-low
72	0.286	low
73	0.4074	low-middle
74	0.5141	middle
75	0.6064	middle-high
76	0.6851	middle-high
77	0.7512	high
78	0.8059	high
79	0.8505	high
80	0.8864	high

Fig. 6 An example of table showing predictions in both quantitative and qualitative terms

```
     ^  Prob                                 ^ Prob
     |                                       |
 100 |                                   100 |
  90 |                                    90 |
  80 |                                    80 |
  70 |                                    70 |                 * *
  60 |                                    60 |               * * *
  50 |                                    50 |               * * *
  40 |                 * *                40 |           * * * * *
  30 |             * * * *                30 |         * * * * * *
  20 |           * * * * *                20 |     * * * * * * * *
  10 |     * * * * * * * * *              10 |   * * * * * * * * * *
     ..*..*..*..*..*..*..*..*..*..*…..>      ..*..*..*..*..*..*..*..*..*..*…..>
     71       75        80                   71       75        80
        Future ages (years)                     Future ages (years)

          Case  A                                  Case B
```

Fig. 7 An example of histograms representing predictions for a subject X in two different cases. If predictions concern an undesired event (e.g. First cardiac infarct) the case A shows probabilistic predictions in the hypothesis that the ten future years of X elapses in rather good conditions (e.g. no cigarette smoke, no hypertension, etc.), whereas the case B concerns the the hypothesis that the future of X elapses in worse conditions (e.g. yes cigarette smoke, yes hypertension, etc.). Of course if predictions concern a desired event, the better situation is represented by the case B.

C. Administering Environments

The general environment for probabilistic predictive monitoring contains three administering environments: the environment for administering predictive monitoring tools (for short: Tools Administering environment), the environment for administering monitored subjects (for short: Subjects Administering environment), the environment for administering environments (for short: Environments Administering environment).

The Tools Administering environment concerns the administration of tools that are in the Using environment. The environment provides the tool administrator with several utility functions. Among such functions there is the one that allows the administrator to eliminate a subject by the list of the subjects monitored by a tool. This function also re-establishes the statistical situation in the database of the tool as if the eliminated subject had never entered the tool.

The Subjects Administering environment contains functions to manage the subjects database and to operate on subjects independently from the specific tools to which they are assigned by tool administrators.

The Environments Administering environment is used by the Super-administrator only. The Super-administrator plays the role of general supervisor of the general probabilistic predictive monitoring environment. Among the many functions that the Environments Administering environment makes available to the Super-administrator let us notice the authorization functions, i.e. authorization to use a certain environment, or in other words, authorization to play the role of tool builder, or tool user, or tool administrator, or subjects administrator. The basic rules are:

- a single subject can be monitored by n different tools (for example, in medicine a subject can be monitored with respect to various undesired pathologic events, and there is a tool for each undesired event respectively)
- a user can use n tools (for example, an organization can study the occurrence probability of n undesired events)
- a single tool can be used by m users (for example, different organizations, possibly international, can co-operate in monitoring a large subjects population relatively to an undesired/desired event)

These features contribute to make the proposed general environment for probabilistic predictive monitoring a product suitable to be successfully used even in large co-operation contexts, facilitating and structuring co-operation among working groups.

VIII. Related work and discussion

The great number of works concerning predictive monitoring, published in scientific journals and conferences both in past and in recent years, gives evidence of both the modernity of the topic and the remarkable effort so far accomplished by the researchers community.

Industry is a typical world in which predictive monitoring, mostly intended as preventive monitoring, has find numerous applications with a variety of approaches. Twenty years ago already, preventive monitoring was a crucial theme for manufacturing processes (typically, for example, in the world of the large car manufacturing companies [1]). In manufacturing industries there is a considerable attention to reduce costly and unexpected breakdowns. As a consequence preventive maintenance is becoming more and more important. Maintenance should abandon the traditional "fail and fix" approach to pass to the more modern "predict and prevent" one [2]. As a consequence the fundamental need is monitoring degradation instead of detecting faults. A predictive performance and degradation monitoring is what is needed for an effective proactive maintenance to prevent machines from breakdown. The theme of degradation monitoring for failure prevention applied to vehicle electronics and sensor systems is faced in [3] where the authors propose a unified monitoring and prognostics approach that prevents failures by analyzing degradation features, driven by physics-of-failure. The need, for manufacturers of complex systems, to optimize equipment performance and reduce costs and unscheduled downtime, gives rise to system health monitoring. System states monitoring is augmented with prediction of future system health states and predictive diagnosis of possible future failure states [4]. Predictive monitoring has been also applied to flexible manufacturing systems. In [5], the main objective is to manage progressive failures in order to avoid breakdown state for the flexible manufacturing system. The approach to predictive monitoring proposed in [6] uses predictions from a dynamic model to predict whether process variables will violate an emergency limit in the future (predictions are based on a Kalman filter and disturbance estimation). Predictive monitoring has also been applied in many specific industry worlds like, for example, cold extrusion and forging processes [7] and chemical plants [8], [9]. In many industrial applications predictive monitoring assumes the meaning of preventive monitoring and aims to enhance the effectiveness of preventive maintenance by making it proactive. In some cases though, predictive monitoring is finalized to early intervening to maintain a system at a high level of performance. It is the case of a predicting monitoring application for wireless sensor networks: "...by monitoring and subsequently predicting trends on network load or sensor nodes energy levels, the wireless sensor network can proactively initiate self-reconfiguration…" [10]. In most industry applications the acquisition of monitoring data is carried out through sensors [11].

Predictive monitoring has found many applications in medicine too. Applications concern both clinical trials [12] and several specific fields. For example, interesting applications have been carried out in the field of diabetes therapy. In [13] and [14], continuous glucose monitoring devices provide data that are processed by mathematical forecasting models to predict future glucose levels in order to prevent hypo-/hyperglycemic events. Many other specific applications of preventive monitoring may be found in medicine. For example, in [15] the authors present the experience of predictive monitoring applied to some patients exposed to gentamicin (a commonly used antibiotic medication) ototoxicity: the most common single known cause of bilateral vestibulopathy. Patients undergoing exercise rehabilitation therapy were tested repeatedly during follow-up visits to monitor changes in their vestibulo-ocular reflex. Predictive monitoring turned out to be useful for continuing or modifying the course of vestibular rehabilitation therapy.

Very recently predictive monitoring has found many applications in the field of environment pollution [16], [17],

[18].

Literature shows that, in general, prediction has been intended in the sense of prevention, that is as a means for preventing undesired events. Actually the possibility of getting early warnings before an undesired event may occur has always been very appealing. Let us think, for example, of prevention of high risk events for health, or serious faults or anomalies of costly and strategic industrial equipments or plants. The proposal presented in the paper has the ultimate purpose (monitoring and prediction) that is in common with all the cited applications, but, at the same time, it has many aspects that distinguish it from them. The proposal, is neither a predictive monitoring application nor a general prognostics tool for preventing undesired events in some fields like, for example, manufacturing industries, medicine, etc. In fact the proposal concerns a general environment for building, using and administering specific application oriented predictive monitoring tools and concerns predictive monitoring applied to both preventing undesired events and favoring desired events. The proposal considers human operators using the various environments (Building, Using, Administering) by playing various roles (Tool builder, Tool user, Tool administrator). In particular, monitoring subjects is an activity carried out by a domain expert. In other words, it is the domain expert (i.e. a human agent) that carries out monitoring sessions and enters data about the current subject situation (context states, etc.). Again, it is the domain expert that defines the starting conditions for simulating the future (which context states are supposed to be constant in the future), and it is the domain expert that reads the simulation results and takes suitable measures. Let us notice though that the conceptual structure of the proposal concerning the theoretical model and the algorithm of the general Tool for probabilistic predictive monitoring (presented in section V), might be embedded in a software program. In such a case it is an automatic agent that plays the role of a human user of a specific predictive monitoring tool. For example, a software agent might periodically gather data about a subject (e.g. a machine) by means of sensors or other software programs interfacing a database. It is the software agent that activates simulations and then examines the results and as a consequence takes suitable measures (for example, if the probabilistic value related to a certain future age is greater than a given threshold, then the software agent activates suitable sub-programs). Finally, let us consider that, with respect to other approaches to predictive monitoring, the general predictive tool considered in the proposal is probabilistic. This means that the tool becomes predictive only after having collected a number of cases that is statistically significant, i.e. sufficient to be able to produce probabilistic inferences. Before reaching that condition the tool works like a mere monitoring tool (subject data acquisition and data entry into the database).

IX. Conclusion

In this last decade predictive monitoring is an emerging theme of great social importance. In many real world fields the possibility for domain experts to have at their disposal tools that may support decisions and help take measures in advance is a crucial need. The paper has faced this problem by presenting a proposal that can be considered complete in that it includes both the theoretical model and its practical utilization inside a general tool for probabilistic predictive monitoring, tool that is in turn integrated in a general environment providing numerous and effective facilities for probabilistic predictive monitoring: facilities for creating new application oriented tools, monitoring subjects and simulating possible future probabilistic scenarios, administering tools and subjects and regulating co-operation among working groups. In this sense the proposal may represent a contribution to promote the use of predictive monitoring in heterogeneous domains.

A. Future work

The proposal implementation is in progress (the author is building a software prototype of the proposal). The prototype will be usable at the web address www.cheerup.it and should be ended by may 2013. Once the prototype will be available, it will be interesting to experiment the proposal by applying it to real world problems. To this end it will be crucial to select suitable application fields and co-operate with interested domain experts.

References

[1] S. Spiewak and M. Szafarczyk, “A Predictive Monitoring and Diagnosis System for Manufacturing”, *CIRP Annals - Manufacturing Technology*, vol. 40, no. 1, pp. 400-403, 1991.

[2] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu and H. Liao, “Intelligent prognostics tools and e-maintenance”, *Computers in Industry*, vol.57, no. 6, pp. 476-489, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.compind.2006.02.014

[3] H. Liao and J. Lee, “Predictive Monitoring and Failure Prevention of Vehicle Electronic Components and Sensor Systems”, *SAE 2006 World Congress & Exhibition, April 2006*, Detroit, MI, USA, Session: Automobile Electronics and Systems Reliability (Part 1 of 2).

[4] R. Kothamasu, S. H. Huang and William H. VerDuin, “System health monitoring and prognostics—a review of current paradigms and practices”, *The International Journal of Advanced Manufacturing Technology*, vol. 28, no. 9-10, pp. 1012-1024, 2006. [Online]. Available: http://dx.doi.org/10.1007/s00170-004-2131-6

[5] F. Ly, A. K. A. Toguyeni and E. Craye, “Indirect predictive monitoring in flexible manufacturing systems”, *Robotics and Computer-Integrated Manufacturing*, vol. 16, no. 5, pp. 321-338, 2000. [Online]. Available: http://dx.doi.org/10.1016/S0736-5845(00)00015-6

[6] B. C. Juricek, D. E. Seborg and W. E. Larimore, “Predictive monitoring for abnormal situation management”, *Journal of Process Control*, vol. 11, no. 2, pp. 111-128, 2001. [Online]. Available: http://dx.doi.org/10.1016/S0959-1524(00)00043-3

[7] S. A. Spiewak, R. Duggirala and K. Barnett, “Predictive Monitoring and Control of the Cold Extrusion Process”, *CIRP Annals - Manufacturing Technology*, vol. 49, no. 1, pp. 383-386, 2000.

[8] J Jeng, C Li, H Huang, “Dynamic Processes Monitoring Using Predictive PCA”, *Journal of the Chinese Institute of Engineers*, vol. 29, no. 2, pp. 311-318, 2006.

[9] G. Chen, T. J. McAvov, “Predictive on-line monitoring of continuous processess”, *Journal of Process Control*, vol. 8, no. 5-6, pp. 409-420, 1998.

[10] A. Ali, A. Khelil, F. K. Shaikh and N. Suri, "MPM: Map based Predictive Monitoring for Wireless Sensor Networks", presented at Autonomic Computing and Communications Systems, Third Int. ICST Conf. Autonomics 2009, Limassol, Cyprus, September 9-11, 2009.

[11] J. Chen, T.-Y. Hsu, C.-C. Chen, and Y.-C. Cheng, "Online Predictive Monitoring Using Dynamic Imaging of Furnaces with the Combinational Method of Multiway Principal Component Analysis and Hidden Markov Model", *Industrial & Engineering Chemistry Research*, vol. 50, no 5, pp. 2946-2958, 2011. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ie100671j

[12] S. C. Choi and P. A. Pepple, "Monitoring Clinical Trials Based on Predictive Probability of Significance", *Biometrics*, vol. 45, no. 1, pp. 317-323, 1989. [Online]. Available: http://www.jstor.org/stable/2532056

[13] J. Reifman, S. Rajaraman, A. Gribok and W. K. Ward, "Predictive Monitoring for Improved Management of Glucose Levels", *Diabetes Science Technology*, vol. 1, no. 4, pp. 478-486, 2007.

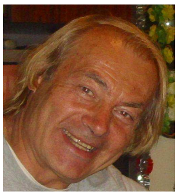
[14] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E.J. Gómez, M. Rigla, A. de Leiva and M.E. Hernando, "Artificial Neural Network Algorithm for Online Glucose Prediction from Continuous Glucose Monitoring", *Diabetes Technology & Therapeutics*, vol. 12, no. 1, pp. 81-88, 2010. [Online]. Available: http://dx.doi.org/10.1089/dia.2009.0076

[15] D. P. O'leary, L. L. Davis and S. Li, "Predictive Monitoring of High-frequency Vestibulo-ocular Reflex Rehabilitation Following Gentamicin Ototoxicity", *Acta Oto-Laryngologica*, vol. 115, no. S520, pp. 202-204, 1995.

[16] Y. Kim, M. Kim, J. Lim, J. T. Kim, C. Yoo, "Predictive monitoring and diagnosis of periodic air pollution in a subway station", *Journal of Hazardous Materials*, vol. 183, no. 1-3, pp. 448-459, 2010.

[17] M. M. Zamzam, V. B. Reddy, K Al Bisher, "Reducing Energy and Emissions through Predictive Performance Monitoring System ",*Abu Dhabi International Petroleum Conference and Exhibition*, 11-14 November 2012, Abu Dhabi, UAE.

[18] E. Pathmanathan, R. Ibrahim, V. S. Arsivadm, "CO2 emission model development employing particle swarm optimized — Least squared SVR (PSO-LSSVR) hybrid algorithm", *International Conference on Intelligent and Advanced Systems (ICIAS*), 12-14 June 2012.



Silvano Mussi received the degree in Physics in 1975 from the University of Milan, Italy. He has been with a telecommunication company (ITALTEL) for over 10 years, working in the Research and Development department, in the fields of Software Engineering and Functional Discrete Simulation of Real-time Systems. Then, for over 20 years he has been with an interuniversity consortium for information technology (CILEA) where for over 10 years he has played the role of marketing manager. For over 10 years he has done research in the fields of Knowledge Engineering and Expert Systems. For some years he has been contract-professor of Artificial Intelligence at the University of Brescia. His current research interests are in the fields of probabilistic expert systems, predictive monitoring and decision support systems. His scientific publications can be found at the Web-page: www.cheerup.it