Reasoning about unpredicted change and explicit time

Florence Dupin de Saint-Cyr and Jérôme Lang

IRIT - Université Paul Sabatier -
31062 Toulouse Cedex(France)
e-mail: {dupin, lang} @irit.fr

Abstract. Reasoning about unpredicted change consists in explaining observations by events; we propose here an approach for explaining time-stamped observations by surprises, which are simple events consisting in the change of the truth value of a fluent. A framework for dealing with surprises is defined. Minimal sets of surprises are provided together with time intervals where each surprise has occurred, and they are characterized from a model-based diagnosis point of view. Then, a probabilistic approach of surprise minimisation is proposed.

1 Introduction

Reasoning about time, action and change in full generality is a very complex task, and much research has been done in order to find solutions suited to specific subclasses of problems, obtained by making some simplificaying assumptions. Sandewall [15] provides a taxonomy for positioning the various subclasses, based on a lot of ontological and epistemological assumptions that may be present or absent in the specification of the subclass. In this paper we consider an integer representation of time and as in [15] we define *fluents* as propositions whose truth value evolves over time, *observations* as pieces of knowledge about the value of some fluents at some time points, *changes* as pairs made of a fluent f and a time point t such that the value of the fluent f at t differs from that at $t-1$. Changes may be either caused by *actions*, which are initiated by the agent, or by *events*, which are initiated by the world. This latter form of change will be called *unpredicted changes*; as they are unrelated to the performance of any action by the agent, they are totally out of the control of the agent. While many approaches in the literature focus on reasoning about change caused by actions[1], there has been significantly less attention on reasoning about unpredicted change.

In this article we propose a framework for detecting unpredicted changes from observations at different time points: the task consists thus in *explaining* observations by a set of elementary unpredicted changes. Note that it is analogous in many aspects to temporal diagnosis (an elementary change corresponding to a faulty component). We do not explicitly consider actions, since this is not our purpose; however, as we will show, our framework can handle actions as well,

[1] This is for instance the case for all approaches to solve the frame problem.

provided that they have a deterministic, unconditional result – thus the class of problems we consider is $\mathcal{K} - IS$ of Sandewall's taxonomy (see Section 2). Then, we propose some qualitative and quantitative criteria for ranking explanations. We propose a formal basis for handling problems in the class K-IS; namely, we argue that what has to be computed is the set of *minimal compact explanations*, where minimality means that these explanations do not contain any unnecessary change, and compactness means that they are given as concisely, as user-friendly as possible. Then we focus on computational issues and for this purpose we relate our framework to (temporal) model-based diagnosis: a problem of K-IS is translated into a system to diagnose and we show that minimal explanations correspond to minimal fault configurations. Thus, computational tools from model-based diagnosis can be reused for reasoning about unpredicted change. In Section 3, we compute a probability distribution on these explanations, which takes account of (i) the intrinsic tendency of each individual fluent to persist, and (ii) the time interval during which the change may have taken place – following the principle that the longer the time period since we last knew the value of a fluent, the fewer reasons we have to believe that this value has persisted. We illustrate this on a small example: *two cats live in a house, namely, Albert and Bart. At 6 p.m. , somebody told me that cat A. was sleeping inside the house. At 6h15 p.m., I knew that cat B. was sleeping inside the house. Now, it is 6h20 p.m., I can see a cat walking in the garden, but I can't recognize from where I am, if it is Albert or Bart.* The two intuitive minimal explanations we wish to get are the following: either *cat A. left the house between 6p.m. and 6h20 p.m.*, or *cat B. left the house between 6h15p.m. and 6h20*. Now, imagine we know the (subjective) prior probabilities for cat A. and cat B. to leave the house in a time unit (one minute); from these, and from the observations, we may compute the probabilities of all explanations. Intuitively, if the probabilities relative to cat A. and cat B. are identical, then the explanation *cat A. got outside* is preferred to *cat B. got outside*, since more time has passed between the last information we got and now – thus, a change is more likely to have occurred. However, the result may not be the same if the prior probability of cat A. getting outside is much lower than cat B's (cat A is lazier than B). What we propose in Section 3 is a simple method to compute most probable explanations in the case where probabilities of change are small within the time period considered, with a methodology similar to computing preferred diagnoses. We end up by positioning our approach w.r.t. related work and mention possible further extension.

2 Minimizing unpredicted change in K-IS

2.1 Background

From now on, the class of problems we consider coincides with the class $\mathcal{K} - IS$ in Sandewall's taxonomy. The symbol $\mathcal{K}$ means that all observations reported to the agent are correct, *i.e.*, true in the real world. I is for *inertia with integer time* and means that fluent values generally tend to persist. S is for *surprises*;

a surprise, or unpredicted change, is a change of the value of a fluent during a given interval, without any reason known by the agent; S means that surprises may occur (and are thus considered when computing completions of scenarios). Inertia implies that surprises occur with a low frequency, and should thus be *minimized* in order to be in accordance with observations. At this point we should say something about actions. In this paper we are not concerned with actions, and the general assumption we make throughout the paper is that the agent is totally passive (no action is performed)[2].

Let $\mathscr{L}$ be a propositional language built on a set of propositional variables $V = \{v_1, \ldots, v_n\}$. A *fluent* is a variable or its negation, *i.e.*, $f = v_i$ or $f = \neg v_i$. The set of all fluents is denoted by F; thus $F = \{v_1, \neg v_1, \ldots, v_n, \neg v_n\}$. Time is assumed to be discrete: the time scale T is a sequence of integers $\{0, 1, \ldots, t_{max}\}$ (it is assumed that between two consecutive time-points the system is completely inert).

If $\varphi \in \mathscr{L}$ and $t \in T$ then $[t]\varphi$ is an *elementary timed formula* (meaning that φ holds at t). (Complex) timed formulas are built from elementary timed formulas and usual connectives.

A *timed model* M is a mapping from $V \times T$ to $\{0, 1\}$: M assigns a truth value to each fluent at each time point. $M \models [t]v_i$ iff $M(v_i, t) = 1$. Satisfaction is extended to elementary timed formulas in the usual way, *i.e.*, $M \models [t]\varphi \wedge \varphi'$ iff $M \models [t]\varphi$ and $M \models [t]\varphi'$, etc., and then to complex timed formulas: $M \models [t]\varphi \wedge [t']\varphi'$ iff $M \models [t]\varphi$ and $M \models [t']\varphi'$, etc. Note that $[t]\varphi \wedge \varphi'$ is equivalent to $[t]\varphi \wedge [t]\varphi'$, $[t]\varphi \vee \varphi'$ is equivalent to $[t]\varphi \vee [t]\varphi'$, etc.

Lastly, we abbreviate $[t]\varphi \wedge [t+1]\varphi \cdots \wedge [t']\varphi$ by $[t, t']\varphi$.

Definition 1. A *scenario* Σ in K-IS consists in a set of timed-formulas

Definition 2. Let $f \in F$, $t \in T$, $t \neq 0$.

1. $\langle f, t \rangle$ is a *change* in M (notation: $C_M(f, t)$), iff $M \models [t-1]f$ and $M \models [t]\neg f$.
2. The *set of all changes in* M is $C(M) = \{\langle f, t \rangle \mid C_M(f, t)\}$.

When the exact time-point where the change occurred is not known, we affect an interval to a change occurrence and call it a *surprise*.

Definition 3. Let $f \in F$, $t, t' \in T$, $t < t'$. $\langle f, t, t' \rangle$ is a *surprise* with respect to M (notation: $S_M(f, t, t')$), iff $M \models [t]f$ and $M \not\models [t, t']f$.

Intuitively, $S_M(f, t, t')$ means that the truth value of f changed at least once between t and t' (note that it does not necessarily imply that $M \models [t']\neg f$, since f may have changed its truth value several times within $[t, t']$).

[2] However, the class $\mathcal{K} - IS$ does not exclude actions, but assumes that actions have deterministic and unconditional results, have only instantaneous effects, and that there is no dependency between fluents (which avoids the ramification problem). Under these assumptions, it is not hard to see that a deterministic, unconditional action can be seen as an observation: for instance, performing the action `Load` at time point t can be considered equivalent to observing its deterministic, unconditional result *loaded* at t.

If $t' = t+1$, $S_M(f,t,t')$ is equivalent to $C_M(t')$; otherwise $(t'-t>1)$ the surprise is said to be *time-ambiguous*.

The following property is straightforward:

Proposition 4. *$S_M(f,t,t')$ iff $C_M(f,t+1)$ or $C_M(f,t+2)$ or ... or $C_M(f,t')$.*

2.2 Minimal explanations

The problem that we address is: given a scenario Σ, find the preferred timed-models satisfying Σ, *i.e.*, find the timed-models in which the changes are minimal: $\{M, M \models \Sigma$ and there is no M' such that $C(M') \subsetneq C(M)\}$.

Definition 5. A *pointwise explanation* PE for Σ is a set of changes $\{\langle f_i, t_i\rangle, i = 1\cdots p\}$ such that $\Sigma \cup \{[t{-}1]f{\rightarrow}[t]f \mid f{\in}F, t{\in}T, t{\neq}0,\ \langle f,t\rangle \notin PE\}$ is consistent.

Intuitively, PE consists in specifying which fluents change their value and when so as to be in accordance with Σ.

Example 6. Let us consider a set of pointwise observations Σ_1 such that:

$$\Sigma_1 : \begin{cases} [0]\ a \wedge b \\ [5]\ \neg a \end{cases}$$

$\{\langle a,2\rangle\}$ is a pointwise explanation for Σ_1: in order to explain the non-inert behaviour of the system, we must assume that the value of the fluent a changed between 0 and 5, (for instance at time point 2).

Note that $\{\langle a,2\rangle, \langle b,3\rangle\}$ is also a pointwise explanation, but, since it is not necessary to assume that b changed in order to explain Σ_1, this pointwise explanation is not minimal:

Definition 7. A pointwise explanation PE is *minimal* iff there is no pointwise explanation PE' strictly included in PE

Definition 8. An *explanation* for Σ is a set of surprises $\{< f_i, t_i, t'_i >, i = 1\cdots p\}$ such that $\forall(t''_1,\ldots,t''_p) \in [t_1+1, t'_1] \times \cdots \times [t_p+1, t'_p]$, $PE = \{\langle f_i, t''_i\rangle, i = 1\cdots p\}$ is a pointwise explanation for Σ. We say that each of these PE is *covered* by E.

For any two explanations E and E', we say that E *covers* E' iff any pointwise explanation covered by E' is also covered by E.

Note that an explanation has to be understood as a disjunction of pointwise explanations. In the example above, $\{\langle a,2,5\rangle\}$ is an explanation for Σ_1; corresponding to the pointwise explanations $\{\langle a,2\rangle, \langle a,3\rangle, \ldots, \langle a,5\rangle\}$.

Definition 9. An explanation E is *minimal* if any pointwise explanation which is covered by E is minimal.

The previous explanation $\{\langle a, 2, 5\rangle\}$ is minimal, but for instance $\{\langle a, 2, 5\rangle, \langle b, 1, 3\rangle\}$ or $\{\langle a, 2, 5\rangle, \langle a, 0, 5\rangle\}$ are explanations but they are not minimal.

Proposition 10. *$E = \{\langle f_i, t_i, t'_i\rangle, i = 1 \ldots p\}$ is an explanation (resp. a minimal explanation) for Σ iff*
$\Sigma \cup \{[t]f \rightarrow [t+1]f \mid f \in F, t \in T, t \neq t_{max},$ and $\nexists \langle f_i, t_i, t'_i\rangle \in E$ s.t. $t \in [t_i, t'_i - 1]\}$ is consistent.

It is a minimal explanation iff the above set of formulas is maximally consistent in $\Sigma \cup \{[t]f \rightarrow [t+1]f \mid f \in F, t \in T, t \neq t_{max}\}$

Definition 11. A minimal explanation E is compact iff there is no explanation E' which strictly covers E.

In the previous example, the only compact minimal explanation for Σ_1 is $\{\langle a, 0, 5\rangle\}$.

Proposition 12. *If there is no disjunction in Σ (*i.e.*, each pointwise observation $[t]\varphi$ is a conjunction of fluents) there is only one compact minimal explanation for Σ.*

Example 13.

$$\Sigma_2 : \begin{cases} [0]\ a \\ [5]\ a \vee c \\ [10]\ b \\ [15]\ \neg a \vee \neg b \\ [20]\ \neg c \end{cases}$$

There are 3 minimal compact explanations:
$E_1 = \{\langle a, 5, 15\rangle\}$, $E_2 = \{\langle b, 10, 15\rangle\}$ and $E_3 = \{\langle a, 0, 5\rangle, \langle c, 5, 20\rangle\}$ covering $10 + 5 + (5 \times 15)$ pointwise explanations.

- $\{\langle a, 10, 15\rangle\}$ is an explanation for Σ_2 but it is not compact since $\{\langle a, 5, 15\rangle\}$ is an explanation for Σ_2.
- $\{\langle a, 0, 15\rangle\}$ is not an explanation for Σ_2 since, for instance, $\{\langle a, 1\rangle\}$ is not a pointwise explanation for Σ_2.
- $\{\langle a, 0, 15\rangle, \langle c, 5, 20\rangle\}$ is an explanation for Σ_2 but it is not minimal since $\{\langle a, 6\rangle, \langle c, 7\rangle\}$ is covered by it and is not a minimal pointwise explanation (since $\{\langle a, 6\rangle\}$ alone is a pointwise explanation).

Definition 14. Let $Cme(\Sigma)$ be the set of all compact minimal explanations for Σ.

Proposition 15. *For each minimal pointwise explanation PE for Σ, there is a* unique *compact minimal explanation E which covers it.*

Thus, $Cme(\Sigma)$ is the most concise expression covering all minimal pointwise explanations.

Proposition 16. $Cme(\Sigma) = \{\varnothing\} \Leftrightarrow \Sigma \cup \{[t-1]f \rightarrow [t]f \mid f \in F, t \in T, t \neq 0\}$ *is consistent.*[3]

Intuitively, $Cme(\Sigma) = \{\varnothing\}$ means that assuming that all fluents kept their value throughout T does not lead to any inconsistency (with Σ).

2.3 Computing minimal explanations

The method consists in completing the set of observations Σ by a set of persistence assumptions in order to express that, by default, fluents persist during the interval between two consecutive observations. In order to add only "relevant" persistence assumptions, we identify, for each variable, the "relevant time points" where we have some observation regarding it.

Definition 17. The *set of relevant variables* $V(\Sigma)$ with respect to the set of observations Σ is defined by: $V(\Sigma) = \{v \in V \mid v \text{ appears in } \Sigma\}$.

The *set of relevant time points* $RT_\Sigma(v)$ to a variable $v \in V(\Sigma)$ w.r.t. Σ is defined by: $RT_\Sigma(v) = \{t \in T \mid [t]\varphi \in \Sigma \text{ and } v \text{ appears in } \varphi\}$

Let $RT^*_\Sigma(v)$ be the set of relevant time points to a variable v except the last one: $RT^*_\Sigma(v) = RT_\Sigma(v) \setminus \{max\{t \in T, t \in RT_\Sigma(v)\}\}$

In example 13: $RT_{\Sigma_2}(a) = \{0, 5, 15\}$, $RT_{\Sigma_2}(b) = \{10, 15\}$ and $RT_{\Sigma_2}(c) = \{5, 20\}$.

Definition 18. The set of persistence axioms of Σ is defined by:

$$PERS(\Sigma) = \bigcup_{v \in V(\Sigma), t \in RT^*_\Sigma(v)} \{[t]v \rightarrow [next(v,t)]v, \quad [t]\neg v \rightarrow [next(v,t)]\neg v\}$$

where $next(v,t) = \min\{t' \in RT_\Sigma(v), t' > t\}$ is the next relevant time point for v after t.

Definition 19.

- Let $MaxCons(\Sigma)$ be the set of all maximal subsets of $PERS(\Sigma)$ consistent with Σ
- let $Cmc(\Sigma) = \bigcup_{X \in MaxCons(\Sigma)} (PERS(\Sigma) \setminus X)$
- let $Candidates(\Sigma) = \big\{\{\langle f, t, t'\rangle \mid [t]f \rightarrow [t']f \in X\}, X \in Cmc(\Sigma)\big\}$.

Proposition 20.

- *Any set of surprises in $Candidates(\Sigma)$ is a minimal explanation for Σ.*
- *Every pointwise explanation for Σ is covered by a unique set of surprise in $Candidates(\Sigma)$*

[3] Similarly, in model-based diagnosis, when the system description and the observations are consistent with the non-failure assumptions, the preferred diagnosis is $\{\varnothing\}$ (no faulty component).

This principle of finding minimal explanations corresponds exactly to finding minimal sets of faulty components in model-based diagnosis (or candidates in ATMS [4] terminology). Hence, the computation of minimal explanation can be done by well-known algorithms from these fields.

Example 21. (Example 13 continued) The relevant time-points for Σ_2 are : $RT_{\Sigma_2}(a) = \{0, 5, 15\}$, $RT_{\Sigma_2}(b) = \{10, 15\}$ and $RT_{\Sigma_2}(c) = \{5, 20\}$.

$$PERS(\Sigma_2) : \begin{cases} [0]a \rightarrow [5]a \\ [5]a \rightarrow [15]a \\ [0]\neg a \rightarrow [5]\neg a \\ [5]\neg a \rightarrow [15]\neg a \\ [10]b \rightarrow [15]b \\ [10]\neg b \rightarrow [15]\neg b \\ [5]c \rightarrow [20]c \\ [5]\neg c \rightarrow [20]\neg c \end{cases}$$

When computing the minimal set of persistence assumptions which has to be removed from $PERS$ in order to make it consistent with Σ_2, we obtain the 3 minimal candidates $\{\langle a, 5, 15\rangle\}$, $\{\langle b, 10, 15\rangle\}$ and $\{\langle a, 0, 5\rangle, \langle c, 5, 20\rangle\}$.

Note that it is not always the case that the computed explanations are compact: consider $\Sigma_3 = \{[0]a, [5]a \vee b, [10]\neg a\}$. We find the two minimal explanation $\{\langle a, 0, 5\rangle\}$ and $\{\langle a, 5, 10\rangle\}$ which are not compact (since they can be compacted into $\{\langle a, 0, 10\rangle\}$).

If we wish to obtain the *compact* minimal explanations from $Candidates(\Sigma)$, we may define a compactification procedure, where precise description is outside the scope of the paper. Intuitively, this procedure consists in compacting explanations pairwise and iterate until no pair of explanations is compactable. This operation is confluent and its result is exactly $Cme(\Sigma)$.

3 Probabilities and unpredicted change

We assume now that probabilistic information about fluents truth values and persistence is available. This will enable us to compute the probability of each explanation, and thus will help us ranking them. We will take account of 2 important factors which can influence quantitatively the persistence: time duration and the intrinsic tendency of each fluent to persist (some fluents persist longer than other, e.g., *sleeping* persists usually longer than *eating an apple*).

3.1 Markovian fluents

For the sake of simplicity, throughout this Section we assume fluents are *mutually independent* and *Markovian*[4]. We will consider stationary fluents in some cases but we will not impose this restriction to the whole section.

[4] These assumptions mean that timed-variables can be structured in a very simple temporal Bayesian network, which has the following form (where $\{v_1, \ldots, v_n\} \in V$

Definition 22.

- fluents are *mutually independent* iff $\forall t, t', \forall f \in F, \forall f' \in F \setminus \{f, \neg f\}$,
$$Pr([t]f \wedge [t']f') = Pr([t]f).Pr([t']f').$$
- f is *stationary* iff $\forall t, t', \forall f \in F, \quad Pr([t]f) = Pr([t']f) = p_f$.
- f is *Markovian* iff $\forall t, \quad Pr([t+1]f \mid [t]f \wedge H_{0\to t-1}) = Pr([t+1]f \mid [t]f) = 1 - \varepsilon_f$.
  Where $H_{0\to t-1}$ is the history of the system from 0 to $t-1$.

The independence and Markovian assumption imply that the only necessary data are, for each propositional variable v and each time-point t, a *prior probability* of v, $p_{t,v}$ and the *elementary change probabilities* ε_v and $\varepsilon_{\neg v}$. We will show later that in many cases prior probabilities are not necessary. Generally $\varepsilon_v \neq \varepsilon_{\neg v}$, *i.e.*, the persistence of a fluent may be different from the persistence of its negation (think of f =alive or f =bell-ringing).

If fluents are stationary then for each variable it is enough to consider a prior probability p_v which is independent of the time-point. Moreover, in this case, there is a relationship between the persistence of a fluent and its negation:

Proposition 23. *If f is stationary then $\varepsilon_f .p_f = \varepsilon_{\neg f} .p_{\neg f}$*

Proof. $Pr([t]f \wedge [t+1]\neg f) = Pr([t+1]\neg f \mid [t]f).Pr([t]f) = \varepsilon_f .p_f$. And $Pr([t]\neg f \vee [t+1]f) = Pr([t]\neg f) + Pr([t+1]f) - Pr([t]\neg f \wedge [t+1]f) = p_{\neg f} + p_f - \varepsilon_{\neg f} .p_{\neg f} = 1 - \varepsilon_{\neg f} .p_{\neg f}$.

Obviously, some fluents are *not* Markovian:

- either because they do not tend to persist independently of t: consider a clock which always rings around 7 a.m., then the fluent *ringing* is not Markovian (nor stationary). However, the fact that the clock is on the bedside table is usually a Markovian fluent (unless you throw it away each time it rings).
- or because their tendency to persist depend on their history: consider a bus which period is 15 minutes, the probability that the bus will come depends on the length of the interval since it last came. If you know that no bus has came for 10 minutes, you have more chance than the bus will arrive immediately than if you know that no bus has come for 5 minutes. Clearly, $Pr([t]\neg bus\_coming \mid [t-1]\neg bus\_coming) = \frac{14}{15}$, while $Pr([t]bus\_coming \mid [t-k, t-1]\neg bus\_coming) = \frac{k}{15}$. And thus *bus_coming* is not Markovian as all periodic fluents (however it may reasonably be assumed stationary without considering the rush hour) .

---

are the propositional variables):
$[0]v_1 \to [1]v_1 \to \cdots \to [t_{max}]v_1$
$[0]v_2 \to [1]v_2 \to \cdots \to [t_{max}]v_2$
$\vdots$
$[0]v_n \to [1]v_n \to \cdots \to [t_{max}]v_n$ (Dean and Kanazawa 1989 [6]).

Among all Markovian fluents, some can be distinguished:

- persistent fluents, such that $\varepsilon_f = 0$ (e.g. f =dead). A stationary and persistent fluent is a degenerate case (either it never change, or the probability of its opposite is null).
- chaotic fluents, such that: $Pr([t+1]f \mid [t]f) = Pr([t+1]f)$ (these fluents have no tendency to persist, since knowing this fluent true at t does not make it true at $t+1$ more probable). It means that $\varepsilon_f = 1 - p_{t+1,f}$, hence, chaotic fluents must be stationary.
- switching fluents (provided that the time unit is well chosen) such that $\varepsilon_f = 1$

Now, given the prior probability distribution p_f and the switch probability $Pr([t+1]\neg f \mid [t]f) = \varepsilon_f$ (which are constants, independent of t), it is possible to compute the prior probability of surprise for each stationary Markovian fluent between any time points.

Definition 24. Notation: $Pr(\langle f, t, t' \rangle) = Pr([t]f \wedge \neg[t, t']f)$ is the probability of the surprise $\langle f, t, t' \rangle$.

Proposition 25. *If f is stationary and Markovian, then*

$$Pr(\langle f, t, t+n \rangle) = (1 - (1 - \varepsilon_f)^n).p_f$$

Proof.

$$\begin{aligned} Pr(\langle f, t, t+n \rangle) &= Pr(\langle f, t, t+n \rangle) \mid [t]f).Pr([t]f) \\ &\quad + \quad Pr(\langle f, t, t+n \rangle) \mid [t]\neg f).Pr([t]\neg f) \\ &= (1 - Pr([t]\neg f \vee [t+1, t+n]f \mid [t]f)).p_f + 0 \\ &= (1 - Pr([t+2, t+n]f \mid [t+1]f).Pr([t+1]f \mid [t]f)).p_f \\ &= (1 - (1 - \varepsilon_f)^n).p_f \end{aligned}$$

Indeed, it looks natural that, knowing the truth value of a fluent at a given instant, the more time has passed, the more likely the fluent had its truth value changed.

3.2 Highly persistent fluents and probabilities of explanations

From proposition 25 and the independence assumption it is also possible to compute the prior probability of an explanation (minimal or not) :

$$Pr(\{\langle f_i, t_i, t'_i \rangle, i = 1 \ldots n\}) = \prod_{i=1}^{n} (1 - (1 - \varepsilon_{f_i})^{t'_i - t_i}).p_{f_i}$$

Now, these prior probabilities have to be conditioned on the observations (Σ) in order to get posterior probabilities of all explanations : if E is an explanation, then $Pr(E \mid \Sigma) = \frac{Pr(E \wedge \Sigma)}{Pr(\Sigma)}$.

In the general case these probabilities are not straightforward to compute, mainly because $Pr([t]f \wedge [t']\neg f)$ may be much lower than $Pr(\langle f,t,t'\rangle)$[5]. Moreover, non-minimal explanations may have a significant probability, even in the case where $\{\varnothing\}$ is an explanation, which contradicts the principle of minimal change. This is because elementary probabilities of change (the ε_f's) may be large. Computing probabilities of change over an interval ($Pr([t]f \wedge [t']\neg f)$) requires computing the probability that an odd number of pointwise surprises occurred between t and t'. This is left outside of the scope of the paper which focuses on infrequent change (thus to be minimized). As expected intuitively, change is minimized if and only if the ε_f's are all infinitely small; more precisely, if for any considered interval $[t,t']$, $(t-t')\varepsilon_f \ll 1$.

Informally, f is *highly persistent* w.r.t. interval $[t,t']$ iff $\varepsilon_f \ll \frac{1}{t'-t}$.

Proposition 26. *Assume that all fluents are highly persistent w.r.t.* $[0,t_{max}]$, *then:*

1. $Pr(\langle f,t,t'\rangle) \approx Pr([t]f \wedge [t']\neg f) = (t'-t)\varepsilon_f.p_f + o(\varepsilon_f)$.
2. $\sum_{E \in Cme(\Sigma)} Pr(E \mid \Sigma) \approx_{\varepsilon_f \to 0, \forall f} 1$.

Item 26.2 means that non-minimal explanations for Σ have a very low posterior probability and thus they can be neglected. In particular, if $\{\varnothing\}$ is an explanation for Σ then $Pr(\{\varnothing\} \mid \Sigma) \approx 1$. Proposition 26.1 intuitively means that if f changed at least once its truth value within $[t,t']$ ($\langle f,t,t'\rangle$) then it changed exactly once (and thus $\neg f$ holds at t'). Now probabilities of minimal explanations are easy to compute (provided that all fluents are highly persistent). We start by giving two detailed examples.

Example 27. (cat A and B, pure prediction)

$$\Sigma_4 : \begin{cases} [0]\ a \\ [15]\ b \\ [20]\ \neg a \vee \neg b \end{cases}$$

This means that at 6h, cat A is sleeping inside the house, at 6h15, cat B is also sleeping in the house, and at 6h20, one cat is not sleeping inside the house.
$Pr(\{\langle a,0,20\rangle\}) \approx 20\varepsilon_a p_a$, $Pr(\{\langle b,15,20\rangle\}) \approx 5\varepsilon_b p_b$.

$$\begin{aligned} Pr(\{\langle a,0,20\rangle \mid \Sigma_4\}) &= \frac{Pr(\langle a,0,20\rangle \wedge \Sigma_4)}{Pr(\Sigma_4)} \\ Pr(\langle a,0,20\rangle \wedge \Sigma_4) &= Pr([0]a \wedge [20]\neg a \wedge [15]b) \\ &= Pr([0]a \wedge [20]\neg a).Pr([15]b) \\ &\approx 20\varepsilon_a.p_a.p_b. \end{aligned}$$

[5] Indeed $\langle f,t,t'\rangle$ is true if f holds at t and changed its value *at least once* between t and t', while $[t]f$ and $[t']\neg f$ is true if f changed its value *an odd number of times* between t and t'.

Similarly, $Pr(\{\langle b, 15, 20\rangle \mid \Sigma_4\}) \approx 5\varepsilon_b.p_b.p_a$

$$\begin{aligned}
Pr(\Sigma_4) &= Pr([0]a \wedge [15]b \wedge [20]\neg a \vee \neg b) \\
&= Pr([0]a \wedge [15]b \wedge [20]\neg a) + Pr([0]a \wedge [15]b \wedge [20]\neg b) \\
&\quad - Pr([0]a \wedge [15]b \wedge [20]\neg a \wedge [20]\neg b) \\
&\approx 20\varepsilon_a.p_a.p_b + 5\varepsilon_b.p_b.p_a - (20\varepsilon_a.p_a.p_b)(5\varepsilon_b.p_b.p_a) \\
&\quad \text{(the last term being negligible)} \\
&\approx Pr(\langle a, 0, 20\rangle \wedge \Sigma_4) + Pr(\langle b, 15, 20\rangle \wedge \Sigma_4)
\end{aligned}$$

$Pr(\langle a, 0, 20\rangle \mid \Sigma_4) \approx \frac{20\varepsilon_a.p_a.p_b}{\varepsilon_a.p_a.p_b + 5\varepsilon_b.p_b.p_a} = \frac{1}{1+\frac{\varepsilon_b}{4.\varepsilon_a}}$
and $Pr(\langle b, 15, 20\rangle \mid \Sigma_4) \approx \frac{1}{1+\frac{4.\varepsilon_a}{\varepsilon_b}}$.

Remarkably, these probabilities do not depend on the prior probabilities p_a and p_b, which is natural (since a and b are known to hold at 0 and 15, respectively); this will always be the case for pure prediction problems where disjunctive observations appear only at the last time point. Note that, if $\varepsilon_a = \varepsilon_b$, we get $Pr(\langle a, 0, 20\rangle = \frac{4}{5}$ and $Pr(\langle b, 0, 20\rangle) = \frac{1}{5}$.[6]

Example 28. (pure postdiction)

$$\Sigma_5 : \begin{cases} [0]\ a \vee b \\ [5]\ \neg a \\ [20]\ \neg b \end{cases}$$

$$Pr(\langle a, 0, 5\rangle \wedge \Sigma_5) = Pr([0]a \wedge [5]\neg a \wedge [20]\neg b) \approx 5\varepsilon_a.p_a.(1-p_b).$$

Similarly, $Pr(\langle b, 0, 20\rangle \wedge \Sigma_5) \approx 20\varepsilon_b.p_b.(1-p_a)$.

$$Pr(\Sigma_5) \approx Pr(\langle a, 0, 5\rangle \wedge \Sigma_5) + Pr(\langle b, 0, 20\rangle \wedge \Sigma_5).$$

$$Pr(\langle a, 0, 5\rangle \mid \Sigma_5) = \frac{5\varepsilon_a.p_a.(1-p_b)}{5\varepsilon_a.p_a.(1-p_b) + 20\varepsilon_b.p_b.(1-p_a)}.$$

If we suppose the fluents stationary (it is natural to do so considering the usual behaviour of cats) then using proposition 23 we obtain the same kind of results as in the previous case : $Pr(\langle a, 0, 5\rangle \mid \Sigma_5) = 1/5$ and $Pr(\langle b, 0, 20\rangle \mid \Sigma_5) = 4/5$.

Example 29. Take Σ_2 of example 13, we have:
$Pr(\langle a, 5, 15\rangle) \mid \Sigma_2) = \frac{1}{1+\frac{\varepsilon_b}{2\varepsilon_a}+\frac{15}{2}\varepsilon_c.\frac{p_c}{1-p_c}}$
$Pr(\langle b, 10, 15\rangle \mid \Sigma_2) = \frac{1}{1+2\frac{\varepsilon_a}{\varepsilon_b}+15\frac{\varepsilon_a.\varepsilon_c}{\varepsilon_b}.\frac{p_c}{1-p_c}}$
and $Pr(\langle a, 0, 5\rangle \wedge \langle c, 5, 20\rangle \mid \Sigma_2) = \frac{1}{1+\frac{2(1-p_C)}{15(\varepsilon_C)}+\frac{\varepsilon_b}{15\varepsilon_a.\varepsilon_c}.\frac{1-p_c}{p_c}}$

Note that if $\varepsilon_a = \varepsilon_b = \varepsilon_c$ and are very small, $Pr(\langle a, 5, 15\rangle \mid \Sigma_2) \approx \frac{2}{3}$, $Pr(\langle b, 10, 15\rangle \mid \Sigma_2) \approx \frac{1}{3}$ and $Pr(\langle a, 0, 5\rangle \wedge \langle c, 5, 20\rangle \mid \Sigma_2) \approx 0$. These results are independent of the prior probabilities p_a, p_b, p_c.

[6] More generally, for $\Sigma = \{[t_a]a, [t_b]b, [t']\neg a \vee \neg b\}$ and $\varepsilon_a = \varepsilon_b$, we get $Pr(\langle a, t_a, t'\rangle = \frac{t'-t_a}{(t'-t_a)+(t'-t_b)}$ and $Pr(\langle b, t_b, t'\rangle = \frac{t'-t_b}{(t'-t_a)+(t'-t_b)}$.

4 Discussion

The original aspects of our work are that (i) minimal changes (explanations) are provided together with the interval when change may have occurred, which makes the representation concise, and (ii) knowing probabilities of change of fluents from one point to the subsequent one, and fluent prior probabilities, we compute the probability of each explanation and thus we rank them accordingly. We showed that minimizing change is coherent with a probabilistic handling of change only if probabilities of change are very small (with this assumption, only minimal explanations may have significant probabilities). This probabilistic handling of change needs explicit and metric time, and is coherent with the intuitive idea that the longer the time interval during which a change may have occurred, the more likely it has actually occurred within this interval. Many related works share some features with ours. The first approach dealing with unpredicted change ("fluents that may change by themselves") is Lifschitz and Rabinov's [11]; using the situation calculus and circumscription, they minimize the set of unpredicted changes from one state to another, giving, thus, something analogous to our pointwise explanations (without the explicit temporal dimension). Our notion of "surprise" is borrowed from Sandewall's ontology [15] where surprises are defined as unpredicted changes with low frequency; in an earlier version of his book [14], he proposed to select among competing surprise sets by associating to each fluent a penalty (the higher the penalty, the more unlikely the fluent may change unpredictedly), preferred surprises sets minimizing the sum of the penalties of changing fluents. Dealing with penalties is very analogous to dealing with probabilities, especially in the case of infinitely small probabilities [7], thus our approach extends his by taking interval durations into account in defining the penalties.

A probabilistic handling of unpredicted change using explicit, metric time has been proposed for temporal projection by Dean and Kanazawa [6] and later on by Hanks and McDermott [10] in a more general setting. In [6], unpredicted change is modelled with exponential survivor functions, asserting that $Pr([t + \Delta t]f \mid [t]f) = \exp(\frac{-\Delta t}{\lambda_f})$. They do not assume that probabilities of change are small, and thus do not minimize change. We recall that the reason why we want change to be minimized is that it leads to a concise list of explanations, since only minimal explanations may have significant probabilities. This has to be related to the fact that, in [6], probabilities of change are only used for temporal projection, namely, computing the probability of given fluents at the latest time point, but they are not used to deal with explaining scenarios and the reasoning is only forward (it has for instance no postdiction capabilities). Lastly, [6] framework assumes that observations are atomic (no disjunctive observations are allowed while our approach allows the observation of any propositional formula). The same remarks apply to [10].

Other related approaches include work in the domains of probabilistic abduction and model-based diagnosis, temporal diagnosis, and also belief update. Probabilistic (and cost-based) abduction ([12], [8], [5]) attach probabilities to explanations from prior probabilities of hypotheses (faults in model-based diagnosis) and generally an independence assumption amongst them. In the domain of temporal model-based diagnosis, Friedrich and Lackinger [9] attach time intervals to fault configurations, meaning that a given component is in a failure mode during the whole interval – contrarily to our surprises which are instantaneous (recall that the interval in a surprise is understood disjunctively, not conjunctively); probabilities are attached to temporal diagnoses, the evolution of the system is modelled by Markov chains (with transition probabilities for the different modes of a component). Console et al. [2] attach time points to fault configurations; their approach is extended in a probabilistic setting by Portinale [13] who also models the evolution of the system by a Markov process. Cordier and Thiébaux's event-based diagnosis [3] consider sequences of events and compute probabilities over diagnoses from priori probabilities of events (without metric time). Boutilier's generalised update operators [1] handles unpredicted change, explaining observations by events, ranked according to their plausibility, but without explicit time (he only considers two time points). If we specialize our framework to a time scale with only two time points t_{before} and t_{now}, then it is possible to show that we obtain something very close to a generalized update operator, where the allowed events are the elementary surprises.

Further work will include the handling of dependent fluents (speciality D in Sandewall's ontology) – in the probabilistic case they may for instance be structured in a Bayesian network – and the integration of surprise minimisation together with a handling of actions with alternative effects (speciality A in Sandewall's ontology).

Acknowledgement

We would like to thank Didier Dubois and Henri Prade for their comments which helped improve this paper.

References

1. C. Boutilier. Generalised update : belief change in dynamic settings. In *Proc. of the 14^{th} IJCAI*, volume 2, pages 1550–1556, 1995.
2. L. Console, L. Portinale, D. Theseider Dupré, and P. Torasso. Diagnostic reasoning across different time points. In *Proc. of ECAI'92*, pages 369–373, 1992.
3. M.O. Cordier and S. Thiébaux. Event-based diagnosis for evolutive systems. In *Proc. of DX'94*, 1994.
4. J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.
5. J. de Kleer and B. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97–130, 1987.
6. T. Dean and K. Kanazawa. Persistence and probabilistic projection. In *Proc. of IEEE Trans. on Systems, Man and Cybernetics*, volume 19(3), pages 574–585, 1989.
7. F. Dupin de Saint-Cyr, J. Lang, and T. Schiex. Penalty logic and its link with Dempster-Shafer theory. In *Proc. of the 10^{th} Conf. on Uncertainty in Artificial Intelligence*, pages 204–211. Morgan Kaufmann, July 1994.

8. T. Eiter and G. Gottlob. The complexity of logic-based abduction. In *Proc. STACS'93*, 1993.
9. G. Friedrich and F. Lackinger. Diagnosing temporal misbehaviour. In *Proc. IJCAI'91*, pages 1116–1122, 1991.
10. S. Hanks and D. McDermott. Modelling and uncertain world i : symbolic and probabilistic reasoning about change. *Artificial Intelligence*, 66:1–55, 1994.
11. V. Lifschitz and A. Rabinov. Things that change by themselves. In *Proc. of the 11^{th} IJCAI*, pages 864–867, Detroit, Michigan, 1989.
12. D. Poole. Representing diagnostic knowledge for probabilistic Horn abduction. In *Proc. of the 12^{th} IJCAI*, pages 1129–1135, 91.
13. L. Portinale. Modeling uncertain temporal evolutions in model-based diagnosis. In *Proc. Uncertainty in AI'92*, pages 244–251, 1992.
14. E. Sandewall. Features and fluents: a systematic approach to the representation of knowledge about dynamical systems. Technical Report LITHIDA-R-92-30, Linköping University, 1992.
15. E. Sandewall. *Features and Fluents.* Oxford University Press, 1994.