Temporal Matching under Uncertainty

Ahmed Y. Tawfik
School of Computer Science
University of Windsor
Windsor, ON N9B 3P4, Canada
atawfik@uwindsor.ca

Greg Scott
University of Prince Edward Island
Charlottetown, PE C1A 9J8, Canada
gscott@upei.ca

Abstract

Temporal matching is the problem of matching observations to predefined temporal patterns or templates. This problem arises in many applications including medical and model-based diagnosis, plan-recognition, and temporal databases. This work examines the sources of uncertainty in temporal matching and presents a probabilistic technique to perform temporal matching under uncertainty. This technique is then applied to the problem of finding the onset of infection with *Toxoplasma Gondii.*

1 Introduction

Temporal matching is the process of matching a limited set of observations to known temporal evolution patterns in order to identify the stage of evolution or determine the onset of the temporally evolving pattern. Given a sequence of observations, and some temporal evolution patterns, a temporal match consistent with the sequence of observations maps observation times to particular points in a pattern.

Formally, S is a set of states, T_R is the set of time points measured relative to a pattern onset time t_0, E is the set of known temporal evolutions between states defined as

$$E \subseteq S \times S \times T_R$$

and the set of observation O is in turn defined as

$$O \subset S \times T_A$$

where T_A is the set of actual time points measured with respect to an arbitrary time origin. The temporal matching problem is that of finding a matching mapping $M : T_R \leftarrow T_A$ to obtain a pattern instance consistent with O.

Temporal matching is needed for many applications. For example, model-based diagnosis of physical systems uses temporal matching to check temporal constraints and to identify failure patterns [9]. Fuzzy temporal matching algorithms are used to determine the onset of infections in medical diagnosis [11]. These techniques have been used to identify the maturity of a given technology by analyzing relevant patent entries and their rate [7]. In computer vision, probabilistic and stochastic techniques have been applied to track moving objects [5]. Humans use different forms of temporal matching in their daily lives. For example, we can determine the time-of-day by observing the angle and direction of the sun. In this example, the E consists of the motion pattern of the sun, by observing the sun, it is possible to match the observation to the known pattern and approximately deduce the time.

Note that temporal matching in this work differs from matching in temporal databases in that the latter aims at finding instances in the database that match a temporal query [3]. If the database contains all possible matches, the results would coincide. Work on bidirectional persistence [4, 12] can be viewed as a special case of temporal matching discussed here. Bidirectional persistence considers simpler evolution patterns corresponding to the persistence of a fluent.

This work is organized as follows: Section 2 examines the sources of uncertainty in temporal matching. Section 3 presents a probabilistic approach to the problem. Section 4 shows how to apply the proposed solution to the problem of finding the onset time of infection with toxoplasmosis. In the remainder of this section we introduce an example of temporal matching.

Example The evolution pattern of a state variable over time is shown in Figure 1-A. If a time t_{obs} this variable is as shown in Figure 1-B, then there are two possibilities: either the variable has reached this level

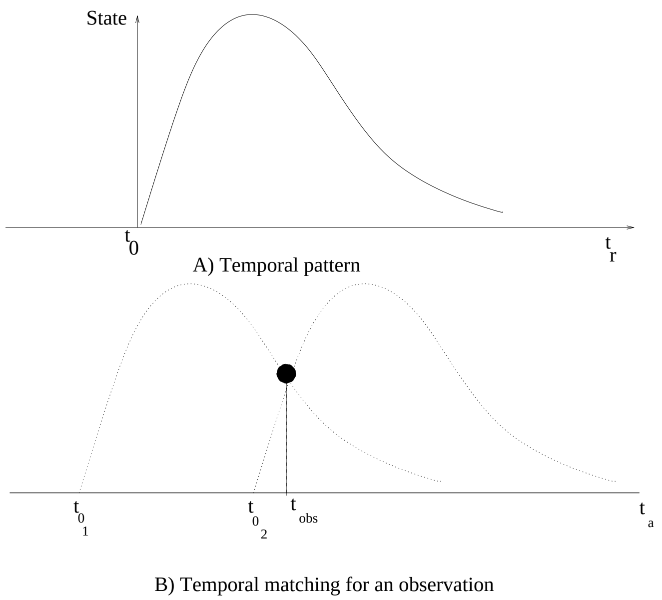

Figure 1: Temporal pattern and matches to an observation

on its way down after starting its evolution at time t_{0_1} or on its way up after starting at t_{0_2}. Note that the observed level occurs twice during the pattern. In this example, it is therefore possible to match the observed level to two instances of the same pattern with different start times.

2 Uncertainty in Temporal Matching

In the absence of uncertainty, temporal matching can be formulated as a constraint satisfaction problem. Temporal constraint networks [1, 10] represent temporal patterns using graphs where the nodes correspond to states (variables) and edges correspond to temporal precedence constraints. The temporal constraints have a set of intervals associated with them indicating the duration separating the states. These networks can be used to test if a set of temporal assignments is consistent. Finding a match corresponds to finding such temporal assignments to observed states. However, in many cases both the pattern and the observations are uncertain. In such cases constraint satisfaction techniques do not yield conclusive results. The sources of uncertainty may include:

- Pattern uncertainty: A pattern may exhibit temporal uncertainty and state uncertainty. If the pattern exhibits temporal uncertainty, the duration separating consecutive states and the persistence of each state is uncertain. In such cases, a probability distribution represents this uncertainty. A particular state in the pattern may be reached with a certain probability.

- Observation uncertainty: Observations may be uncertain in their timing or may not reflect the state with certainty.

- Low sampling rate: A low sampling rate makes it impossible to fully reconstruct the observed pattern from the observations alone. If the sampling rate is low, the number of possible matches increases, threatening the effectiveness of the temporal matching.

- Repeated states: If a pattern includes repeated states, observing such states is less informative. It is preferable in these situations to add to the state definition additional attributes to discriminate between similar states if possible.

- Pattern estimation: The amount of historic data available may limit our ability to accurately infer the pattern and the probability distribution.

- Pattern combination: Sometimes, it is necessary to compose and cross validate a group of matching patterns to properly match given observations.

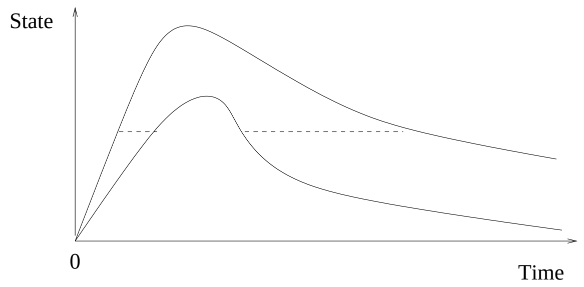

Figure 2: A pattern exhibiting uncertainty

Figure 2 illustrates the fact that in the presence of uncertainty, observing a particular state is less conclusive in determining the possible pattern onset time.

It is desirable in such cases to assign probabilistic beliefs to possible matches rather than just stating that they are all possible. We use here state and transition probabilities. State probability is the probability of a given at a given pattern time. Transition probabilities represent the probability of moving from one state to another at different pattern times.

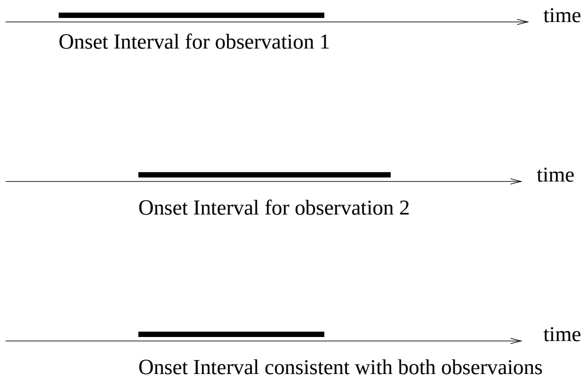

Figure 3: An onset interval consistent with two observations obtained by intersecting the onset intervals

3 Dealing with Limited Uncertain Persistence

Temporal matching under limited uncertain persistence can be defined as follows: given some observation(s) at times $t_i \in T_A$ and a temporal pattern which allows limited uncertain persistence of X, it is required to find the probability that t_i matches a particular pattern relative times $t_R \in T_R$ consistent with the observation(s) for all possible t_r.

Within the constraint satisfaction formulation discussed earlier, each observation limits the possible matches to a more restricted set of intervals on the timeline. A solution that satisfies all the observations lies at the intersection of these intervals. Figure 3 shows how the intersection of intervals is consistent with different observations. Pattern and observations uncertainties detailed later calls for a pattern representation that allows for variations. For example, instead of using a line to describe a pattern as in Figure 1, a band is used. The envelope of this band defines the pattern. Consequently, the onset time corresponding to an observation is an interval (or a set of intervals). Intersecting these intervals, we get a set of possible onset times consistent will all observations. Frequently, the set of possible onset times thus obtain contains candidates that are very improbable. State probabilities identify improbable states and transition probabilities identify unlikely temporal evolutions.

For a temporal matching problem to have a solution within this framework, conditions similar to those specified by Nőkel [8] have to be met. These conditions are:

1. **Probabilistic pattern:** The evolution pattern E is specified in terms of states and transition probabilities.
2. **Availability of observations:** The set O is not empty.
3. **Validity of observations:** Observations in O are valid states.
4. **Compatibility of Observations:** The intersection of all possible pattern time onset intervals is not empty.
5. **Pattern composition:** All observations belong to a single pattern occurrence or an appropriate combination of patterns is used.

In addition to the above conditions, we assume the process is Markovian so that the transition probability depends only on the previous state, and observations are conditionally independent so that the probability of a sequence of observations O_t up to time t, given a state sequence X_t up to t is:

$$p(O_t|X_t) = \Pi_{i=1}^{t} p(o_i|x_i).$$

Temporal pattern matching therefore requires to be performed in two phases: a learning phase and a matching phase. During the learning phase, the pattern and the probability distributions are acquired from data. During the matching phase, observations are assigned pattern times.

3.1 Temporal Pattern Acquisition

The training set here is assumed to consist of a sequence of time stamped records. Each record describes the state of a member of the training set at a particular time. This time is measured with respect to an onset time. To acquire a temporal pattern, both the prior probability of being in state s at time t and the probability that a transition from state s_i to state s_j have to be evaluated. The transition probabilities capture the probabilistic persistence as well as the probabilistic change possible.

3.2 Temporal Matching Algorithm

Given a set of observations and a temporal pattern as described above, the algorithm aims at finding how the observations can match the pattern.

1. For each observation, identify all possible pattern times at which the observed state is possible. A set of possible pattern onset times is formed from the identified pattern times. Each onset time is assigned a prior probability equal to the probability of the state at the corresponding pattern time.

2. By intersecting the sets of possible onset times for individual observations and evaluating the probability distribution over time points (or intervals) in the intersection, we obtain a set of onset times consistent with all observations.

3. For each possible onset time in the intersection, the transition probabilities are used to estimate the likelihood of reaching all observed states starting from this particular onset time. The probability that the pattern started at a particular onset time and progressed through all the observed states is thus obtained.

Hence, the algorithm evaluates the posterior probabilities of each onset time given all available observations.

This temporal matching algorithm has been applied to the problem of finding the onset time of an infection with toxoplasma gondii.

4 Application: Finding the Onset of Infection with Toxoplasmosis

Toxoplasmosis is a medical diagnosis indicating that a person has been infected with toxoplasma gondii. Toxoplasmosis is asymptomatic in immune-competent individuals but it can be very dangerous for an immune-compromised individual or a fetus particularly during the first trimester of pregnancy. Serological tests measuring the immunoglobulins G, M and A are used to diagnose toxoplasmosis. A recent infection is usually characterized by high concentrations of IgG and IgM antibodies. Tests measuring the IgG antigen-binding avidity are also used. Routine screening for toxoplasmosis during pregnancy is common in many countries. Diagnostic guidelines have been developed by organizations such as the Canadian Pediatric Society, specifying how to determine the recency of infection using a combination of standardized tests such as the Sabin-Felman dye test, IgG avidity test, IgM, IgA and IgE tests. Interpreting the test results is usually complicated if the sera are collected late in pregnancy [2].

Statistical course of infection data during pregnancy are difficult to obtain for obvious ethical and health reasons. A study involving 27 pregnant women, for whom it has been possible to establish the time of primary infection, shows a large variance in immune responses which could be attributed to factors such as the strain of toxoplasma, the severity and quantity of infecting organisms, and other individual variations [6]. The 27 women in the study have been followed up for durations that ranged from few weeks to a year. Serological test results have been obtained.

4.1 Data Preparation

Due to the small number of individuals (27), and the limited number of sera (126) we have grouped the test results into four levels and we used linear interpolation and extrapolation to obtain additional data points before calculating state and transition probabilities for these patterns.Due to a lack of international standards, the training data does not use the same units as the test data and a unit conversion had to be conducted during the data preparation phase.

4.2 Results

A test dataset consisting serological test results (dye test, IgM and IgG avidity) for 394 patients has been used to evaluate the performance of the temporal matching algorithm. A clinician classified the data as preconceptional infection (latent), postconceptional infection (acute), uncertain (suspected acute), or inconsistent. The temporal matching algorithm estimated the probability for all possible onset times, and compared the onset times to the week of conception to determine if the infection is preconceptional or postconceptional. Probabilities in the .4 to .6 range are considered uncertain. In some cases different tests gave conflicting results and these were labeled inconsistent. The temporal matching algorithm agreed with the clinician on 87% of the cases and has been able to make conclusive decisions on about 10 % of the cases that the clinician found uncertain. The program may therefore be correct on up to 97% of the cases.

We then used C5.0 decision tree induction algorithm to fine tune the range of the uncertain class and determine if some tests are more conclusive. We found that the IgG avidity test is the most conclusive but it is only available for about 10% of the patients. After revising the program to use this information, the performance of the program reached the 89% to 99% range.

TempMatch $\Downarrow$ Clinician $\Rightarrow$	Acute	Uncertain	Latent	Total
Acute	5	0	0	5
Uncertain	17	10	21	48
Latent	3	2	336	341
Total	25	12	357	

Table 1: Temporal matching results

Table 1 indicates that the temporal matching algorithm did not miss any of the cases diagnosed by the clinician as acute. It also agrees with the clinician on 336 latent cases. However, most uncertain cases have been classified as acute or latent. This is not necessarily a misclassification as each infection is really acute or latent (with respect to the date of conception). Three cases classified as acute by the algorithm are considered latent by the clinician. The algorithm is uncertain about two cases considered latent by the clinician.

5 Conclusions

This work has presented a methodical approach to temporal matching under uncertainty. This approach relies on Markov models and the use of probabilities to perform quantitative matching. The approach has been applied to the problem of determining the onset of infection with toxoplasmosis and the results obtained are comparable to other systems custom built to perform this task like Onset [11].

The temporal matching algorithm presented here can be viewed as a probabilistic extension to the constraint satisfaction solution. The transition probabilities impose weak constraint on the solution. Moreover, probability distributions over the set of values for each variable allow the temporal matching to select a solution based on its likelihood.

Acknowledgments

Research of the first author is supported by Natural Sciences and Engineering Research Council. The authors would like to thank Dr. Friedrich Steimann and Dr. Michael Hyde for providing the test data.

References

[1] R. Dechter, I. Meiri, and J.Perl. Temporal constraint networks. *Artificial Intelligence*, 49(1/3):61–95, 1991.

[2] Canadian Paediatric Society Infectious Diseases and Immunization Committee. Common questions about the diagnosis and management of congenital toxoplasmosis. *Paediatrics and Child Health*, 4(2), 1999.

[3] M. Dumas, M. Fauvet, and P. Scholl. Handling temporal grouping and pattern-matching queries in a temporal object model. In *Proceedings of the Seventh Conference on Infromation and Knowledge Management (CIKM-98)*, 1998.

[4] S. Goodwin, E. Neufeld, and A. Trudel. Probabilistic regions of persistence. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches for Uncertainty*, Marseille, France, 1991.

[5] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International journal of Computer Vision*, 29(1), 1998.

[6] P. Jenum and B. Stray-Pedersen. Development of specific immunoglobins G, M and A following Toxplasma Gondii infection in pregnant women. *Journal of Clinical Microbiology*, 36(10):2907–2913, October 1998.

[7] M. Mogee. Patent analysis for strategic advantage: Using intenational patent records. *Competitive Intelligence Review*, 5(1):27–35, 1994.

[8] K. Nökel. *Temporally Distributed Symptoms in Technical Diagnosis*, volume 517 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1991.

[9] K. Nökel and H. Lamberti. Temporally distributed symptoms in a diagnostic application. *Artificial Intelligence in Engineering*, 6(4):196, 1991.

[10] E. Schwalb and R. Dechter. Processing disjunctions in temporal constriant networks. *Artificial Intelligence*, 93(1/2):29–, 1997.

[11] F. Steimann and M. Hayde. A method to derive the time of onset of infection from serological findings. *Methods of Information in Medicine*, 36(1):51–58, 1997.

[12] A. Tawfik and E. Neufeld. Changing times: A causal theory of probabilistic temporal reasoning. *Journal of Experimental and Theoretical Artificial Intelligence*, 11:3–21, 1999.