MODELING FATIGUE

Walton Sumner, II, MD (1), Jin Zhong Xu, Ph.D. (2)
1. Washington University in St. Louis, Missouri
2. Assessment Technologies, Inc, Lexington, Kentucky

ABSTRACT

The American Board of Family Practice is developing a patient simulation program to evaluate diagnostic and management skills. The simulator must give temporally and physiologically reasonable answers to symptom questions such as "Have you been tired?" A three-step process generates symptom histories. In the first step, the simulator determines points in time where it should calculate instantaneous symptom status. In the second step, a Bayesian network implementing a roughly physiologic model of the symptom generates a value on a severity scale at each sampling time. Positive, zero, and negative values represent increased, normal, and decreased status, as applicable. The simulator plots these values over time. In the third step, another Bayesian network inspects this plot and reports how the symptom changed over time. This mechanism handles major trends, multiple and concurrent symptom causes, and gradually effective treatments. Other temporal insights, such as observations about short-term symptom relief, require complimentary mechanisms.

INTRODUCTION

The American Board of Family Practice plans to begin using a patient simulation program as part of its recertification process by 2004. The program stochastically simulates patients from a knowledge base in an effort to meet security and reusability goals[1-3]. Bayesian networks define health states, control queries, and maintain consistency during stochastic operations.

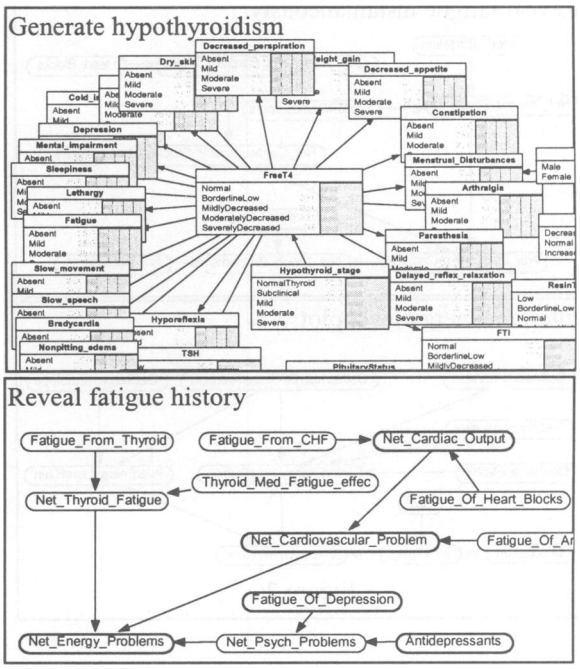
To support testing of fine diagnostic and management skills, the simulator must report temporally and physiologically reasonable symptom histories. For instance, a simulated patient should report on demand the duration of fatigue, the onset of a fever, or the progression of pain. The simulator design intended to support temporal reasoning[4 5] by plotting the presence and absence of diseases, findings, and interventions over time. Attaching these data to a patient would allow the simulator to directly inspect durations of events, and then to produce temporally reasonable reports. For instance, the simulator could check the duration of hypothyroidism or depression, and cause the patient to report feeling fatigued for that amount of time.

Proliferating findings

As the project progressed, we found it expedient to simplify disease definitions and create increasingly complex queries. For instance, clinical findings in hypothyroidism include fatigue and mood disturbance, but depression may cause the same clinical findings. Knowledge acquisition teams defining health states routinely draw Bayesian networks[6] with multiple finding nodes as shown at the top of figure 1, some sharing the same names. However, to model diseases concurrently, the knowledge base and simulator must distinguish the fatigue findings resulting from each disease. In addition, the simulator must distinguish the treatment implications of different fatigue findings. Fatigue from hypothyroidism resolves with thyroid replacement, but not with selective serotonin reuptake inhibitors. Consequently, if we model fatigue as a direct consequence of a health state, the model may require multiple fatigue findings. Queries about fatigue must survey the patient for all possible fatigue findings and corresponding treatments, as shown at the bottom of figure 1.

Disconcerting consequences for the knowledge base follow. Disease generation requires multiple findings named after "fatigue" and, more generally, a plethora of findings represent specific causes of multi-factorial symptoms. Similarly, a fatigue query replicates the list of fatigue findings and corresponding interventions. This approach is awkward, in spite of the initially appealing complete health state definition. Fatigue should not be many findings and a question.

Figure 1



Semi-physiologic models

A more physiologic approach might be simpler. We could consider the defining finding in hypothyroidism to be a low thyroxine level, without pituitary or hypothalamic defects. The thyroxine level, adjusted for any supplementation, is one determinant of fatigue. Although less precisely defined, depression is another determinant of fatigue, requiring adjustments for antidepressant treatment. Queries about fatigue still must survey all possible causes and corresponding treatments, but explicit fatigue findings are irrelevant, and a readable Bayesian network realistically describes what might cause fatigue. In addition, very simple Bayesian networks can define many health states. Fatigue is not a finding at all, only a question.

However, physiologic queries bring a temporal description problem into sharp relief. Falsely reassuring temporal histories of findings in the previous approach obscured the difficulties in combining related findings and treatments to create histories. The complexity of most symptom queries guarantees that they can only describe the instantaneous status of a symptom. The variable duration of concurrent causal diseases, relieving treatments, and exacerbating treatments often defy combination in a single readable Bayesian network. Thus, obtaining temporal information about fatigue from the Bayesian network in figure 1 is extraordinarily difficult, even though temporal information about each contributing patient feature is readily available. Figure 2 illustrates the revised architecture, in which disease definitions are very sparse, and the first level of query calculates instantaneous symptom status. Additional steps must produce the temporal description.

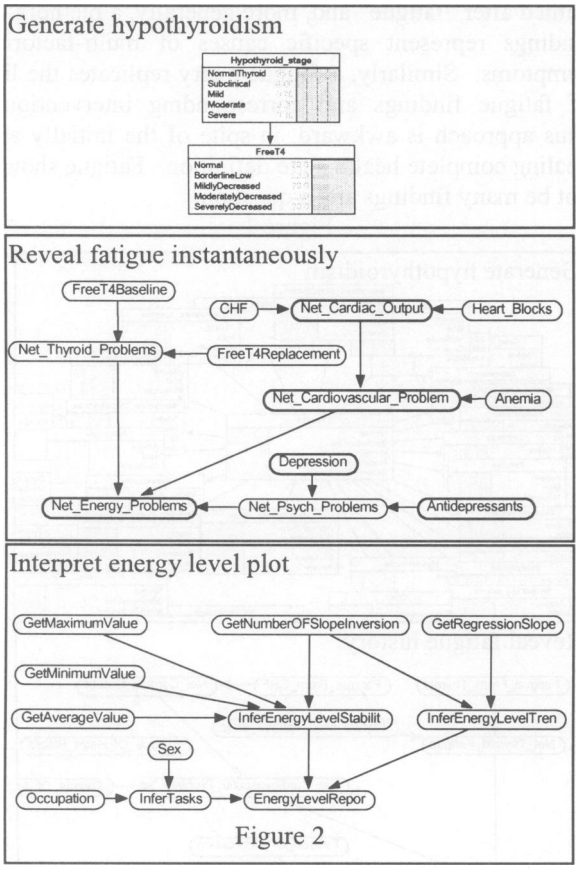
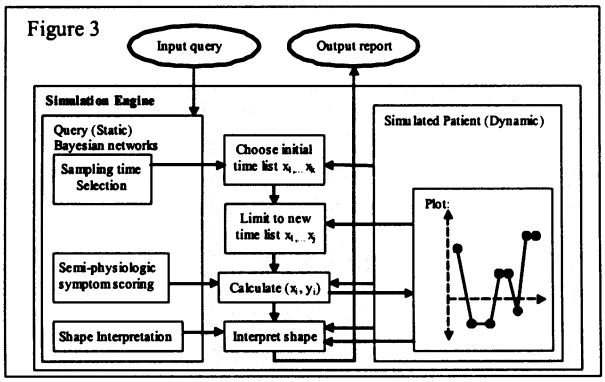


Figure 2

METHODS

Figure 3 illustrates a symptom history algorithm with three major steps. The first step establishes time points to "remember," adjusted for patient characteristics as necessary. The second step calculates an instantaneous symptom status score at each of these time points, adjusted for the treatments and diseases that pertain at that moment. The third step evaluates a plot of symptom scores over time to produce a general description of the symptom history.



Sampling patterns

The first step obtains a series of time points appropriate to the query. During simulation, the simulator receives a query from the user. For queries about symptom history, the knowledge base stores a list of sampling time points representing offsets from the current time. The list may be context sensitive. For instance, a Bayesian network could select a list after inspecting the patient for health problems that impair memory. The simulator subtracts each offset from the current age to find the age for the next symptom status query. For instance, if the sampling times are 0, 0.1, 1, 2, and the current age is 40, the ages to sample are 40, 39.9, 38.9, and 36.9.

Next, the simulator locates any previously calculated data about the same concept. For instance, all queries about fatigue and energy could generate data about a concept called "energy level", regardless of the time horizons inspected by the individual queries. For each concept, a patient has at most one plot of symptom status over time. The simulator obtains this plot, if available. It compares the point list with times already

plotted, and discards points close to previously established points. The simulator calculates closeness relative to the intervals in the sampling pattern. For instance, an existing point at 39.95 need not change the sampling pattern, but existing data at age 39 would cause deletion of the new sampling point at 38.9. For each remaining point, the simulator calculates and plots the instantaneous symptom status at that age.

Instantaneous symptom status

Calculations of instantaneous symptom status typically require information about interventions and findings. Users select interventions to apply to their simulated patients. Interventions include prescriptions, which in turn hold dose and frequency information. The simulator can convert this information to administration events on demand, e.g. to specify that at age 39.8987 the patient consumed 50 micrograms of levothyroxine. Interventions do not hold information about their consequences, and interventions generally have no effect on the simulation until time elapses and a query occurs.

Findings normally specify a baseline value and describe what other events can perturb the value and the temporal course of the perturbation. For instance, a finding of normal thyroid function may set a baseline value between 3 and 5, specify a circadian or lunar fluctuation, and illustrate the rises and falls expected after giving 20 and 200 microgram doses of levothyroxine to a 70 kg person. That is, the finding object enumerates medications and a few representative doses that trigger defined responses in the finding's value. For instance, the dose-response curve for 200-micrograms of levothyroxine might show a rise in thyroxine levels for 2 days, with a peak level on the first day. An intervention can trigger any number of dose-response curves to represent serum levels, tissue levels, or clinical efficacy.

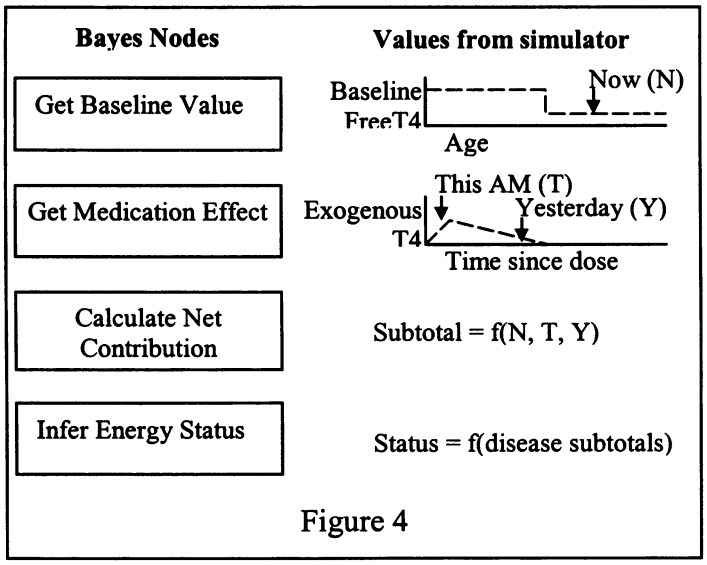


Figure 4

The knowledge base stores a Bayesian network that specifies how to generate the instantaneous symptom status, as figure 4 illustrates. Many nodes in this network are continuous rather than discrete. These either obtain raw data from a simulated patient at run time, or calculate values from raw data and previously calculated values. The raw data for some nodes will be the baseline value of a finding, for instance the baseline thyroxine level. Other nodes will obtain the current contributions of perturbing rhythms and interventions, for instance, the rise in thyroxine level caused by recent levothyroxine doses.

When the simulator encounters such a node, it makes a series of calculations to ensure consistency between queries. First, it determines the duration of the perturbation caused by the intervention, which is 2 days in this levothyroxine example. The window of opportunity for an intervention to affect the instantaneous value of this node is the 2 days prior to the sampling time. For instance, doses of levothyroxine taken between ages 39.8945 and 39.9 influence the fatigue symptom status at age 39.9. Second, it inspects the patient to determine whether the relevant intervention occurred during this window of opportunity. If so, its third step uses the signature information to calculate the patient's age at any missing dosing events during that window of opportunity. For instance, if a dose at age 39.8987 already exists, the simulator adds a dose of levothyroxine at 39.8960. These instances of doses become a permanent part of the simulator's record of the patient's interventions. If subsequent queries require information about levothyroxine doses, the simulator will re-use previously recorded dosing instances as necessary.

Next, the simulator must add the contributions of all doses in the window of opportunity to the finding value at the sampling time. First, it generates a dose-response curve for the dose actually given. For instance, it interpolates between the dose-response curves following 20 and 200-microgram doses to estimate the effect of a 50-microgram dose. Second, the simulator uses the new dose-response estimate to calculate the contribution of each dose taken during the window of opportunity on the dose-response at the sampling time. For instance, at the sampling age 39.9, the levothyroxine dose at 39.8987 is within the past 24 hours, and contributes more to the exogenous thyroxine rise than the dose at 39.8960. This provides a crude approximation of pharmacokinetic behavior of drugs or drug effects. For instance, a knowledge editor can use a shallow, two-week long dose-response curve to model a gradual onset of antidepressant action. A shorter, taller dose-response curve can model the same drug's anticholinergic actions.

A second tier of nodes in the Bayesian networks combines raw data to calculate intermediate results, such as net thyroxine level. Eventually, cascading calculations or Bayesian revisions and stochastic selections establish the state and value of a node that specifies the symptom score at the specified moment in time. In a query about fatigue, the value represents an energy level on a -10 to 10 scale, where 0 is a normal energy level, and positive and negative numbers are elevated and depressed energy levels, respectively. If the 50-microgram dose of levothyroxine was too low for the hypothyroid patient at age 39.9, the numeric energy status may be -3. Finally, the simulator adds the points for each pair of sampling time and symptom status to the plot of symptom status over time. For instance, it would add the point (39.9, -3) to the plot of energy status.

Temporal History

The final step inspects the plot of symptom status over the time interval specified in the original query to produce a report about symptom status during the interval. A report-generating Bayesian network collects information about the shape and the patient as primary evidence, then revises state probabilities and stochastically selects states for unspecified nodes. The node states in the solved network specify a report to return to the user.

The report-generating Bayesian network can use a number of simple attributes of the plot, such as duration, average value, and number of slope sign changes to create a report. Forty reusable Bayes nodes provide access to these attributes. For instance, the plot's duration, average value, time spent below zero, and slope of a linear regression suffice to create a report that the symptom has had a general level of severity, frequency of fluctuation, and trend over the specified interval. For instance, the report on fatigue could be "I have been tired most of the time for the past few years. I am improving a little." The average value drives selection of the phrase "tired," duration selects "the past few years," time spent below zero drives selection of "most of the time," and regression slope selects "improving a little."

Others' experience with temporal abstractions suggests that we will find the current list of plot attributes incomplete.[4 5 7 8] We expect to add other abstractions about temporal trends as new reusable Bayesian nodes, and anticipate that these networks could become quite elaborate. These Bayes nodes require no adjustments for use in shape interpreting Bayesian networks.

These Bayes nodes only operate on the portion of the plot that the simulator obtained from the sampling point list specification. The simulator may truncate distant points in the shape if the sampling pattern specifies too many points. The nodes anticipate a shape with -10 and +10 Y-axis boundaries and 0 to 100 X-axis boundaries.

RESULTS

Initial experiments with a hypothyroidism model demonstrate that this process permits the simulator to produce a variety of general reports about the temporal course of a patient's symptoms. The hypothyroidism model lacks an explicit fatigue finding, and generates only the free thyroxine and drug responses, as described in methods. We generated a series of mildly hypothyroid patients and left them untreated for one month. Each patient then received 200 micrograms of levothyroxine daily for one week, had medication withheld for one week, and then received 50 micrograms of levothyroxine daily for one week. A query about fatigue then produced a plot of energy levels over the preceding three months, demonstrating rapid, dose-responsive resolution of fatigue following the administration of medication. Bayes nodes correctly detected slopes, durations, and other features of the shape, and in turn set the states of other nodes, producing reports that describe several temporal features of the patients' fatigue. The complete sequence executed in less than 50 milliseconds for our current models running on single processor Pentium III class computers.

DISCUSSION

A major advantage of the ABFP simulation system is accurate time management of the simulated patients. Existing structures support queries of any patient on any finding at any time. Although not directly supporting temporal symptom queries, this feature made this three-step algorithm for symptom recall practical.

A continuing shift of detailed medical domain content from health state models to query models continues to characterize this simulation project. The discovery that we might generally avoid modeling symptoms as findings, and instead model symptoms entirely as responses to queries maintains the trend. Although we continue to have compelling reasons to avoid strictly physiologic models, queries driven coarsely by physiologic patient features are becoming a very desirable modeling goal as the knowledge base expands.

Limitations

In spite of the progress obtained thus far, a number of important limitations in our temporal modeling remain. First, our experience with this technique is limited, and we may need to expand the range of plot attribute queries that we support.

Second, a human respondent might take very different steps to answer a symptom history question: (a) based on the question, decide how long to recall; (b) recall characteristic event times such as starting and stopping medications, surgery, onset of diseases, exposure to hazard environments, etc.; (c) describe the symptom status associated with the characteristic events. The report would then focus on specific events set in time. We do not yet know whether we need to recreate this logic to achieve the realism required for ABFP recertification testing. The simulator might use process control methods to compare plots of symptom scores against external events. It could then describe symptom changes in relation to any external event.

Third, this algorithm does not directly address a group of temporal queries regarding recurring exacerbating and alleviating factors. For instance, this process can reliably report increasing fatigue with advancing hypothyroid states, and improvement following levothyroxine treatment, but it cannot directly generate an association between levothyroxine treatment and improving energy levels. This could be very useful information for recertification candidates, and will require additional development efforts. For instance, the simulator might calculate correlation coefficients between a drug dose and a symptom score then generate text that summarizes the strength of the association.

Both process control charts and arbitrarily selected correlation coefficients could produce interesting misleading responses. For instance, the coincidental prescription of an ineffective drug at about the same time as the spontaneous resolution of a self-limited illness could lead to an incorrect conclusion that the events are associated.

CONCLUSIONS

These results suggest that patient simulations in general could avoid explicitly defining most, and perhaps all, symptoms as explicit findings. Temporal information about the presence of explicitly defined symptoms and treatments is difficult to combine to create a history of the symptom. Using a physiologic model of instantaneous symptom status to sample the patient's history is a versatile option, allowing arbitrarily detailed review of symptom history. Although we use a computationally intense approach, performance is acceptable. Hardware and software improvements should keep pace with increasingly complex models.

References

1. Sumner W, 2nd, Truszczynski M, Marek VW. A formal model of family medicine. *J Am Board Fam Pract* 1996;91):41-52.
2. Sumner W, 2nd, Truszczynski M, Marek VW. Simulating patients with Parallel Health State Networks. *Proc AMIA Symp* 1998:438-42.
3. Sumner W, Hagen MD, Rovinelli R. The item generation methodology of an empiric simulation project. *Advances in Health Sciences Education* 1999;4:49-66.
4. Shahar Y, Tu SW, Musen MA. Temporal-abstraction mechanisms in management of clinical protocols. *Proc Annu Symp Comput Appl Med Care* 1991:629-33.
5. Shahar Y, Musen MA. A temporal-abstraction system for patient monitoring. *Proc Annu Symp Comput Appl Med Care* 1992:121-7.
6. Charniak E. Bayesian networks without tears. *AI Magazine* 1991;12(4):50-63.
7. Shahar Y, Chen H, Stites DP, Basso LV, Kaizer H, Wilson DM, et al. Semi-automated entry of clinical temporal-abstraction knowledge. *J Am Med Inform Assoc* 1999;6(6):494-511.
8. Chakravarty S, Shahar Y. Acquisition and analysis of repeating patterns in time-oriented clinical data. *Methods Inf Med* 2001;40(5):410-20.