Hidden Markov Models for Medical Diagnosis

Laurent JEANPIERRE, François Charpillet

Abstract—**This paper shows how the experiment of the Diatelic Project, a Continuous Ambulatory Peritoneal Dialysis monitoring system, has taught us rules concerning the creation of intelligent agents based on a similar architecture. In particular, we explain how a fuzzy discretization of continuous sensors allows for a great model simplification while keeping some good precision in the diagnosis. The adaptation of this architecture to an anaesthesia-monitoring problem is developed, highlighting the qualities and the drawbacks of this kind of models.**

Index Terms—**Hidden Markov Models, Gradient Descent Optimization, Medicine, Model-Based Diagnosis**

I. Introduction

THE Diatelic project is born from the cooperation between the ALTIR (Association Lorraine pour le Traitement de l'Insuffisance Rénale) and the LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications). It is aimed at improving the life quality of renal insufficient patients. More precisely, we focus on the patients who chose continuous ambulatory peritoneal dialysis (CAPD). A good introduction to the different dialysis problems can be seen on [5].

These patient are treated at home, thanks to a surgical modification of their peritoneum. This natural bag rests in the abdomen, and it is very well irrigated by small blood vessels. The modification consists in the addition of a catheter at the bottom of the bag. The patient is then able to fill his peritoneum with some physiological serum. Next, the catheter is closed, and the peritoneum stays filled for several hours. Some osmosis between this poach content and its blood vessels will drain more or less water, depending on the concentration of the injected fluid. From this point, the process of draining water from the whole corpse is done as usual: Blood drains excess water from the various organs it flows through. Next, this water is drained by the peritoneum instead of the kidney.

From a medical point of view, this method is clearly better than the standard haemodialysis, since water is drained in a more progressive way. Haemodialysis, on the other hand, requires the patient to be at hospital a whole day, and sometimes more. During this time, as much as 10 litres may be drained. Moreover, the patient generally have 2 dialyses a week to ensure a minimal stability. This implies that the patients better accept peritoneal dialysis than haemodialysis.

For the sake of safety, a nurse comes and see each patient regularly, and the patient comes back to the hospital once a month to see his doctor. Between two consultations, each patient has to fill a sheet of paper with his medical parameters like his weight, his blood pressures and the kind of physiological serum he has injected into his peritoneum.

The improvement brought by Diatelic on this situation is based on the use of computer sheets instead of paper sheets. The patient enters his medical parameters directly into the computer every day. These data are then sent to a server through the patient's phone line. [6] shows a good overview of the network architecture needed by such a platform. There, an *intelligent* system looks for anomalies to report and stores them in a database. On the other side, the medical team may connect to the same server and look at any data from any patient, for any past day. Since this represents a very large amount of data, the latest alerts generated by the system are displayed on the first page, efficiently driving the doctor's focus onto the patients whose health seems to worsen.

In the end of this paper, we will focus more precisely on this particular system, explaining the reasons for its design. Next, we will show this knowledge can be reused for a different problem. Finally, a short discussion on the possible improvement of this architecture.

II. The Diatelic system

A. *Human cooperation constraints*

In a medical treatment context, the problem of responsibility appears quickly. Let's imagine a patient, one day, enters false data due to a measuring device fault. This can generate some conditions that lead the system into mis-evaluating the patient hydration level. To compensate, the system would suggest to increase the serum concentration to drain this excess water until the weight falls back to its normal value. Since the patient was not really hyperhydrated, he will lose quickly large amounts of water, and he will probably die. In such a case, who is responsible for the patient death? Is this the instrument maker, the doctor, the patient, or the system inventor? To avoid such a case, we decided not to interfere directly with the treatment. The system just helps the doctor in making his own diagnosis. This implies that the system recommendations must be understood by the doctor. Moreover, it should not address the only result, but also the whole deciding process. This eliminates lots of the classically used algorithms.

Laurent JEANPIERRE is with the Laboratoire Lorrain de Recherche en Informatique et ses Applications, 54506 Vandoeuvre-lès-Nancy, FRANCE (telephone: 0383592095, e-mail: laurent.jeanpierre@loria.fr).

François CHARPILLET is with the Laboratoire Lorrain de Recherche en Informatique et ses Applications, 54506 Vandoeuvre-lès-Nancy, FRANCE (telephone: 03835920xx, e-mail: francois.charpillet@loria.fr).

B. The first version

The first version of the system has been created from the rules given by the various doctors. Rules were relatively easy to understand by doctors since they were based on those used by the doctors. The trouble is that the detail of the rules was a bit difficult to read. Moreover, the interactions among several rules were difficult to predict and necessitated the use of artificial priorities to ensure proper functionalities.

The experiment has shown that this version had some limitations: Small variations of the parameters gave rather large impact on the diagnosis; The rules were so much linked one to another that it was nearly impossible to have a small modification without revalidating the whole model from scratch. These two points are closely related to the third one: the system was not adaptable from one patient to another. This implied that the only utilizable model was a generic one. So the system is not really bad for any patient, but it does not behave very well neither.

This is why a second version has been realized.

C. The working version

The rule-based version has shown that a few rules could handle the problem of monitoring a patient hydration level. These employed fuzzy logic that gave some flexibility to the system, limiting the amount of false detection by smoothing a little the diagnosis evolution.

The new system uses the same ingredients. However, we have separated the different steps of the diagnosis for the sake of simplicity. To handle the time evolution, we tried some Markov models. These are known for a very correct behavior when they are applied to an evolving system that shows some uncertainty. In our case, this uncertainty may overcome several parameters we miss.

Considering that the state of the patient is unknown, and that the medical signals we receive are only clues that should allow us to uncover some of this state, we have naturally chosen some variant of Hidden Markov Model (HMM). Reference [1] gives a very good tutorial on the various Markov models and the associated algorithms.

The main problem consists in dealing with continuous signals. Actually, the greatest majority of the medical signals is numerical. To keep the model understandable by the doctors, we wanted to keep the symbolic part of the rule-based system. Moreover, since fuzzy logic has proven its usefulness upon smoothing state transitions, we applied these to obtain a fuzzy partition of each signal into symbolic data. We chose three values to conform with doctor’s way of thinking: A given signal may be low, normal, or high.

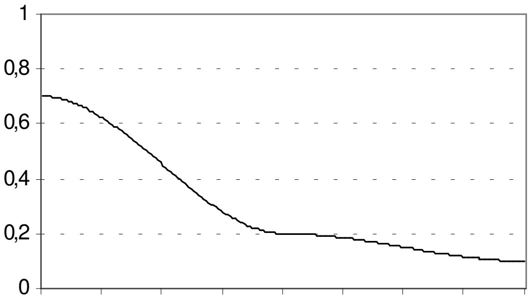
There, our assumption is that these sensors are not intrinsically linked one to another, but that the hidden state of the patient influences several of them simultaneously. This is why we can consider that sensors are statistically independents. Thus, the aggregation formula necessary to compute the probability of the observation O in the state S from the probability of each sensor c_i giving the symbolic value sv reduces to:

$$\mathrm{P}(O \mid S) = \prod_{\mathrm{i}} \sum_{sv \in c_{\mathrm{i}}} \mathrm{P}\,(c_i = \mathrm{s}v \mid S)\,.\,\mathrm{P}\,(c_i = \mathrm{s}v \mid O) \tag{1}$$

This method has been shown in [2] for use in Xavier' s control. The main advantage of such a simple formula is that it can be easily split sensor per sensor, and state per state, knowing the relationship between the numerical value *val* and each symbol *sv*:

$$\mathrm{P}(c_i \mid S)(\mathrm{val}) = \sum_{\mathrm{s}v \in c_{\mathrm{i}}} \mathrm{P}\,(c_i = \mathrm{s}v \mid S)\,.\,\mathrm{P}\,(c_i = \mathrm{s}v \mid c_i = val) \tag{2}$$

These model parameters can then be displayed with a graph showing the probability of the continuous observation in a given state. Here is an example of the resulting plot.



Probability of observing blood pressure values for a dehydrated patient

Since doctors are very used to using such graphs, they are really at ease with this representation.

The transition matrix dictates how the model behaves when there are no available data. Another constraint is that it must converge to a state of uncertainty. The system (that is the patient and his environment) is so noisy that no clear prevision can be made with enough certainty on long periods. This implies a very specific structure. Here is the equation ruling the evolution of the state probability S, starting from time t, when no data are received:

$$\mathrm{P}\big(S(t+1) = q \mid S\,(t)\big) = \sum_{r \in S} \mathrm{P}(S_{t+1} = q \mid S_t = r)\,.\,\mathrm{P}(S_t = r) \tag{3}$$

We impose that the Markov chain converges on a unique stationary point. We impose also that this belief-state must be uniform, i.e. that each one of its _n_ states has the same probability as the others:

$$\begin{bmatrix} \mathrm{P}(1\mid 1) & \dots & \mathrm{P}(1\mid \mathrm{n}) \\ \vdots & & \vdots \\ \mathrm{P}(\mathrm{n}\mid 1) & \dots & \mathrm{P}(\mathrm{n}\mid \mathrm{n}) \end{bmatrix} . \begin{bmatrix} 1/\mathrm{n} \\ \vdots \\ 1/\mathrm{n} \end{bmatrix} = \begin{bmatrix} 1/\mathrm{n} \\ \vdots \\ 1/\mathrm{n} \end{bmatrix} \tag{4}$$

$$\Leftrightarrow \begin{bmatrix} \mathrm{P}(1\mid 1)\text{-}1 & \dots & \mathrm{P}(1\mid \mathrm{n}) \\ \vdots & & \vdots \\ \mathrm{P}(\mathrm{n}\mid 1) & \dots & \mathrm{P}(\mathrm{n}\mid \mathrm{n})\text{-}1 \end{bmatrix} . \begin{bmatrix} 1/\mathrm{n} \\ \vdots \\ 1/\mathrm{n} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

$$\Leftrightarrow \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{P}(S_{t+1} = i \mid S_t = j) = 1 \ \forall \mathrm{i} \tag{6}$$

In definitive, converging on an uniform stationary state implies that each line and each column of the transition matrix must sum to 1.

D. *Advantages and criticisms of this model*

One of the interests in this model is that it is easily understandable by doctors. We showed that each parameter was considered separately from the others, and that allowed us to produce a visual representation of this, signal per signal, and state per state.

From a computer scientist point of view, this model has relatively few parameters. This allows for some model learning in a reasonable time. Additionally, these parameters may be clustered in two groups: those which are specific to a patient, and those which are specific to the monitoring problem. This allows for a specialization of the model from every patient, and it increases the global efficiency of the system. Finally, its algorithms are well known and easy to implement.

The trouble is that standard learning algorithms do not work anymore with this kind of model. The first point is that observations are treated in a specific way …So they should be learned in a specific way also. The second problem is that doctors must be able to understand the model. This implies that the states of the model must have a precise semantic. Simply aggregating similar data in similar states is not sufficient. Aggregated data must be related to the same pathology.

Another drawback of this model concerns the transition matrix: Since each line and each column must sum to 1, it is nearly impossible adapting it to a given patient. An alternative would be to choose an even more specific matrix, such as a symmetric one. This would allow for its learning. Nevertheless, it is not realistic to impose that, in a given situation, the probability to go from state *A* to state *B* is the same that going from state *B* to state *A*. Considering the fact that so many factors may influence the state of a given patient, we finally chose our transition matrix to be uniform, except for diagonal factors:

$$\begin{bmatrix} x & y & \dots & \dots & y \\ y & x & y & \dots & y \\ \vert & & & & \vert \\ y & \dots & y & x & y \\ y & \dots & \dots & y & x \end{bmatrix} \text{with } y = \frac{(1\text{-}x)}{(n\text{-}1)} \qquad (7)$$

E. *Learning algorithm*

To ensure a proper semantic for every state, we used a semi-supervised algorithm. It is based on a sample diagnosis given by a doctor concerning a given patient. However, to prevent over learning of the doctor's diagnosis, we had to optimize the parameters according to a compromise function. This one is minimal when the obtained diagnosis is similar to the doctor's one, and that the reliability of the model (the probability that the observation sequence could be generated by the model) is maximal.

The core of the algorithm is a derivative free gradient descent algorithm with a relaxation inspired from Powell's work. Since most of the parameters are strictly bounded, we used a dichotomical search for the minimum for each parameter to be optimized. These points have been very well explained in [3], along with other alternatives.

Powell's relaxation consists in the replacement of one of the parameters with a new vector. This one is composed of the various updates made during the previous optimization cycle. This allows for a good speed boost because parameters with very little influence are replaced by a "meta-parameter" modelling the best expected descent direction. The trouble with this method is that the search space quickly degenerates. Some parameters tend to become collinear to other ones, while some simply disappear. Brent suggests resetting the parameter space after a while in [4], along with other optimizations. We chose to simply add this meta-parameter to the search space instead of replacing another one. This allows for some interesting acceleration, but it limits its influence on the model.

Additionally, one must pay attention to the domain of this meta-parameter. Its optimization must ensure that none of the model parameters will leave its validity domain. This task is difficult, because some of the parameters are linked together. The modification of a parameter can also modify the domain of other ones. For example, setting the *medium* value of a signal also modifies the lower bound of the *high* value and the upper bound of the *low* value of this signal.

III. Application to other problems

A. *Useful teachings*

The Diatelic system is being experimented for 3 years. Its interactions with patients and doctors have underlined some points which we consider worthy for similar systems.

There is no need of gigantic systems for handling monitoring problems: the Diatelic model consists in only 5 states for modelling hydration troubles and weight anomalies.

Fuzzy filters are a good way to cope with continuous signals that have a known standard behavior. It is a good alternative between discrete observations and probability density functions. The latter generally are finite sums of normal probability functions. These are infinite support functions that require few parameters. The trouble is that they always tend to zero for both upper and lower limits. On the other hand, discrete observations allow for flexible control of the probability of each generated symbols. The trouble is that their transition is steep. Let's imagine a weight signal. With 50 kg, it is *normal*. If the weight increases by 2 kg, it is a *high* weight. Now, 51.9 kg is then categorized as *normal*. The only solution is then to discretize more precisely the interval between *normal* and *high* values. The model then has lots of observation symbols. The use of fuzzy filters would give us more progressive transitions: 51.9 kg is certainly *high* (98%), but there is a small probability (2%) it would be *normal*. This contributes to a model with few parameters but great expressivity.

Impossible transitions and observations should be avoided: Nothing is really impossible in a real problem. The occurrence of an event categorized as impossible instantly nullifies the whole model efficiency.
It is a drawback of the belief-state updating formula:

$$P(S_{t+1}=q|S_t,O_t)=P(O_t|S=q)\sum_{r\in S}P(S_{t+1}=q|S_t=r).P(S_t=r) \qquad (8)$$

Let’s imagine a situation where the observation can only be generated into an unreachable state. This implies that all the reachable states have a null probability. The other ones are unreachable, whichever can be their probability of having this observation. In definitive, no state is reliable anymore, and the situation cannot be repaired by the following observations.

B. Transposition to Anaesthesia monitoring

The objective of this system is to ensure a good anaesthesia quality with a minimal drug injection. Actually, injected drugs quantity is directly responsible for post-operation problems. As with Diatelic, responsibility conditions prevent us from directly driving the anaesthesia. Instead, we can suggest the anaesthetist some modifications in the drug injection program. Before this, it is necessary to obtain a reliable diagnosis upon the sleeping state of the patient: is he sleeping? Does he risk coma? Will he wake up if the surgeon operates?

The signals available to the system are similar to those of Diatelic. Most are numeric ones. At the moment, we can exploit data from 2 devices. The first one, Anemon, monitors heart problems through some fractal analysis. Its data are the heart rate, and a computed index reflecting the patient pain level.

The second instrument is Aspect 2000. It receives electro-encephalogram signals (EEG) and it uses a bi-spectral analysis to compute an *anaesthesia index*. Generally, this index is reliable, but it is somewhat experimental and there is no proof it is really related to the patient consciousness level. Data provided by this device are the BIS index, the electro-myogram (EMG), and the suppression ratio (The percentage of flat EEG epochs in the computing window).

The objective is to compute two indices: The consciousness level and the pain level of the patient. These two levels are correlated since pain tends to wake up the patient and generate muscular reflexes. Muscular activity prevent EEG reading, and generates artefacts that disturb the BIS index computation. Another problem worth mentioning is that Anemon works correctly only if the patient is deeply asleep. Otherwise, sympathic and para-sympathic systems interfere with each other. When this situation occurs, Anemon index is no longer reliable.

C. Differences between the two problematics

The very first thing that worsens the situation is the time scale. In Diatelic, a given patient inserts one set of data every day. On the other hand, in the anaesthesia monitoring, each device sends data once every 5 seconds. Normally, all devices should be synchronized; In fact, there is always a small offset that induces some delay between packets of data from different devices. Since the period between two packets' arrival is much shorter, the system should be less driven by observations. The state at the previous time step has much more influence on the current belief-state.

Another main difference between Diatelic and anaesthesia is the patient knowledge. In the former, the same patient is followed by the system during whole years. This allows for a fine adaptation of the model for each patient. In the monitoring of anaesthesia, the patient typically has no known profile. Perhaps this is its first surgical operation, or perhaps the previous one was handled the standard way, without the computer being connected to medical devices. All of the adaptation must be made in real time, during the anaesthesia itself. This limits seriously the computing power available for learning the model. The primary goal is to achieve safe anaesthesia; the problem of having a good model is secondary.

Moreover, the anaesthetist has less time for teaching the system. Correcting the diagnosis given by the computer in less than 5 seconds, while monitoring other devices and listening to the surgeon is not realistic. Additionally, each modification of drug injection has to be written down for administrative purpose.

Nevertheless, an anaesthetist normally monitors only one operation at a time. This implies that he is always present and conscious of the anaesthesia’s condition. In Diatelic, one single doctor may monitor several patients at the same time. This prevents him from being aware of slight modification of a given patient medical parameters.

D. Model adaptations

The sensor part of the model is rather correct. The only modifications consist in the number and type of sensors. The acquisition mode of these data is a bit different because data are not entered by the patient anymore … They are received through serial ports that must be acquired with the highest priority. For safety purposes, it is necessary that this subsystem is separated from the computing part of the model, so that an over lengthy computation might not prevent data acquisition.

The system dynamics is handled mainly by the transition matrix of the HMM. Since the interval between two packets of data is not clearly fixed, it is necessary that this matrix depends on the elapsed time since the previous observation. Moreover, since some data may not be reliable, the transitions may be merged one another. Actually, the reception of an unreliable packet of data should not have impact on the diagnosis. For example, during the use of an electric scalpel, the electromagnetic radiations are such that no measure is possible because of artifacts presence. This lowers the signal quality index (SQI) given by the Aspect device. The system takes this into account to prevent the interpretation of incorrect data. In the worse case, the data are totally spoiled and unusable. However, this packet is sent anyway by the device. This packet should not infer with the diagnosis. More precisely, this implies that two subsequent transitions must be equivalent to a single one, provided the duration of this one equals the sum of two it replaces:

$$S(t+\delta 1+\delta 2) = T(\delta 1)\ .\ T(\delta 2)\ .\ S(t) = T(\delta 1+\delta 2)\ .\ S(t) \qquad (9)$$

By extension, a transition that occurs in no time must be conservative. Thus, the resulting transition matrix should be the identity matrix. The easiest way of handling such constraints is based on powers. Actually, these have plenty of interesting properties. However, the computation of such a transition matrix may encounter some troubles. The easiest way is to discretize the time scale precisely enough, so that integer powers may be used without real loss of precision. Another way of handling such a situation is based on fractional powers. This allows the sample transition matrix to be expressed for somewhat large periods so that it may be understood by the medical team. This is much in the way of our approach: The doctor must be able to understand each part of the diagnosis process.

The trouble is to compute such a power. We chose to decompose this matrix into its eigenvectors. The resulting space changing operation is then

$$T = Q \, . \, D \, . \, Q^{-1} \tag{10}$$

with D the diagonal matrix made of the eigenvalues of T. This way, expression of the matrix power is reduced to:

$$T^{n} = Q \, . \, D^{n} \, . \, Q^{-1} \tag{11}$$

This transformation requires that the matrix is not singular, and also that none of its eigenvalues is negative. Such a case would actually result in complex matrices with no more medical semantics.

The learning part of the model requires to be able to determine the relative quality of a given diagnosis without the anaesthetist help. We saw previously this one had not enough time to teach the system. This evaluation must be totally objective. This work is in progress.

IV. Conclusion

We showed in this paper the architecture of the Diatelic intelligent system. This structure is easily adaptable to other problems related to uncertain system monitoring. Particularly, the search for pathological cases, or deviations from a *normal* situation is well addressed by such a system.

The handling of continuous sensors through fuzzy discretization is well adapted to the transposition of some expert system rules. These rules keep on being understandable by experts, since parameters are relatively straightforward.

The evolution rules of the system may be simply expressed through a Markovian model; particularly if the system is subject to many unpredictable influences. Noise influences concerning the sensors are inherently included into the model, since all is expressed as probabilities of observation. As the system evolution is better known, the model structure can be upgraded so that it contains this information and so that the monitoring quality will be even better.

Depending on the time scale of the system to look after, expert advices may be used to train the model. Alternatively, some quality function has to be known to allow such a training in real time.

One problem of interest that can enhance a lot the system reliability is taking into account some *action* that may influence the system. Currently, no action is taken into account for the computing. In Diatelic, this action could take the form of precisions concerning the medical treatment, like anti-hypertensor pills' quantity and the concentration of each bag of physiological serum injected into the peritoneum of the patient. Regarding the anaesthesia monitoring, the knowledge of the concentration of drugs injected to the patient would allow for a better transition matrix, conditioned by the anticipated effects such drugs should have. The trouble is that most of these actions are continuous; so much care may be taken to incorporate them into the model properly.

Acknowledgment

We would like to thank doctors from the ALTIR for the help they provided, preparing learning corpora, validating the various models we proposed and suggesting new parameters.

Additionally, we thank anaesthetists from the Brabois Hospital for helping us with the anaesthesia-monitoring project. Their knowledge of the dynamics of the anaesthesia has been an invaluable addition to the project. Even more, they test our models directly in their daily work, providing us with interesting feedback, and allowing us to enhance our models.

References

[1] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, 77 , no. 2, pp 257-285, February 1989.

[2] S. Koenig and R.G. Simmons. Unsupervised learning of probabilistic models for robot navigation. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp 2301-2308.

[3] Press, W.H., et al, Numerical Recipes in C: The Art of Scientific Computing, Chapter 10, Cambridge University Press, 1988-1992.

[4] R.P. Brent, Algorithms for minimization without Derivatives, Prentice-Hall, 1973.

[5] The Kidney Foundation of Canada, Peritoneal Dialysis Publication. http://www.kidney.ca/peritoneal_en.html.

[6] J.P. Thomesse, D. Bellot, A. Boyer, E. Campo, M. Chan, F. Charpillet, J. Fayn, C.Leschi, N. Noury, V. Rialle, L. Romary, P. Rubel, N. Selmaoui, F. Steenkeste, G. Virone, Integrated Information Technologies for patients remote follow-up and homecare, HealthCom 2001