Bennett CC and TW Doub (2014) Temporal modeling in clinical artificial intelligence, decision-making, and cognitive computing: Empirical exploration of practical challenges. ***Proceedings of the 3rd SIAM Workshop on Data Mining for Medicine and Healthcare (DMMH).*** **Philadelphia, PA, USA. In Press.**

Temporal Modeling in Clinical Artificial Intelligence, Decision-Making, and Cognitive Computing: Empirical Exploration of Practical Challenges

Casey C. Bennett[1,2] and Thomas W. Doub[2]

Abstract

Temporal modeling holds great promise for healthcare, where treatment decisions must be made over time, and where continually re-evaluating ongoing treatment is critical to optimizing clinical care for individual patients. Tremendous advances have been made in data mining and temporal modeling of healthcare data, but practical challenges exist in moving these advances from the laboratory/theoretical setting to applied settings with real patients. In this paper, we address a number of these challenges. First, we provide empirical evidence for calculating the optimal trade-off between costs and outcomes in temporal modeling, suggesting that it may be a dynamical system of *relative* values of costs and effects *between* treatment actions (rather than absolute values). Such an approach may allow optimal reward functions to be derived from clinical data. Second, we evaluate the effects of finite horizon levels on both cost effectiveness and outcome change. Finally, we provide a proof-of-concept application for integrating machine-learning-classifier-based (ML) transition models into temporal models (e.g. Markov Decision Processes). The results showed that even a relatively poor classifier can produce small gains in performance and highlights the potential of such an approach for further exploration. Individualized transition models via such ML integration provide a potential practical avenue for implementation of personalized medicine approaches in EHRs and real-world clinical practice. We also discuss a number of future directions for research, such as inclusion of patient safety and treatment non-adherence, and temporal modeling of the clinical process as a basis for cognitive computing.

Keywords- Data Mining; Clinical Artificial Intelligence; Markov Decision Process; Medical Decision Making; Reinforcement Learning

1. Background and Motivation

1.1 Background

Previous research has shown the potential for using temporal modeling, such as reinforcement learning approaches, as a tool for understanding patterns of clinical change in patients over time in the healthcare domain [1-3]. Such modeling can facilitate treatment planning and enhance clinical decision making, e.g. functioning as a key component of dynamic treatment regimes [4]. This has potential implications for providing more sophisticated clinical decision support tools, both to patients and providers [5]. From an artificial intelligence (AI) perspective, such temporal modeling holds promise to better support human decision-making processes as a sort of cognitive scaffolding (see Discussion).

One common approach to conceptualizing such temporal, dynamic models is Markov Decision Processes (MDPs), as well as their partially observable cousins POMDPs [6,7]. These models allow one to reason efficiently about actions/decisions over time, taking into account the probabilistic nature of action effects, outcomes, and other unforeseen events. A number of methods also exist to find optimal solutions (i.e. decisions) in these models: Q-learning, temporal-differencing (TD), SARSA, Dynamic Decision Networks/Dynamic Bayesian Networks (DDNs/DBNs) [8]. These solution methods vary in different ways, but all essentially boil down to estimating/learning the cumulative effects of particular action sequences over time. In other words, if an agent makes a series of decisions (and/or performs the associated actions), what will be the resulting outcome, probabilistically speaking. Such decision-making can also be performed online, so that the agent is constantly re-evaluating its predictions/choices as new information is received [9].

Critically, such temporal modeling can build on existing single decision time-point models, such as a neural network or support-vector machine trained to classify, say, a group of patients into “likely to respond to treatment X” and “likely to respond to treatment Y”. In that sense, temporal modeling using reinforcement learning or MDPs can be seen as a natural extension to the tremendous advances in data mining over the last twenty years, as well as machine learning methods in their own right.

In previous work, we have shown that temporal modeling combining POMDPs and DDNs (using real patient data) can out-perform current treatment-as-usual case-rate/fee-for-service models of healthcare [10]. However, a number of practical challenges remain for

[1] School of Informatics and Computing, Indiana University, Bloomington, IN, USA, cabennet@indiana.edu

[2] Department of Informatics, Centerstone Research Institute, Nashville, TN, USA

SIAM 3rd Workshop on Data Mining for Medicine and Healthcare 2014, April, 2014, Philadelphia, PA.

moving these advances into real-world clinical applications.

1.2 Current Work

In this paper, we focus on addressing some of the practical issues that arise when moving these artificial intelligence methods from the research setting to real-world clinical environments. For instance, all MDP/POMDP and DDN methods have by definition some sort of reward function as well as a cost/utility function (see Section 2.1). In a research setting, these are typically set to contrived values for outcome states. For example, if moving a robot on a discrete grid around a laboratory, some "goal" state can be set to +100, and every non-goal state can be set to -1. Thus the robot receives a reward of +100 for reaching the goal, and incurs "costs" of -1 for every action that it takes till reaching the goal. This allows the robot to learn efficient behaviors that optimize rewards versus costs, although there can be challenges in designing a good reward function [11]. However, in healthcare, it is an even more delicate balance between treatment *costs* and treatment *outcomes* (i.e. rewards). Is a treatment that produces twice the outcome improvement (on some standardized outcome scale) worth ten times the cost of an alternative treatment? What about a treatment that produces better outcomes but takes twice as long to do so (ergo, potentially exposing the patient to greater risks or side effects)? As a practical matter, we need methods that allow us to consistently calculate the trade-offs between the costs and rewards in clinical settings and to determine optimal solutions for treatment planning. Moreover, given the number of disorders in existence, such methods need to be adaptable across clinical domains (i.e. non-disease-specific). Elsewhere, Lizotte et al. provide a theoretical approach to this problem [12]; here we come at it from an empirical direction.

When temporal modeling is applied to healthcare, contrived values are often used either in whole or part for rewards/costs [13]. While such approaches can provide useful information about the application of reinforcement learning or machine learning techniques to healthcare in a general sense, they do make it challenging to create practical applications in real clinical domains from such techniques.

Other practical issues concern how to incorporate long-term change vs. short-term change (e.g. a patient having a sudden upswing), treatment non-adherence, patient safety or risk, and multivariate models of patient outcomes while retaining the tractability of the model.

Here, we address several of these many practical issues - deriving optimal reward functions from clinical data, evaluating horizon effects, and integration of personalized transition models via machine learning classifiers – via building simulations based on actual patient data from a real-world electronic health record (EHR). The goal is to elucidate potential approaches for overcoming these issues in an applied setting, as well as how they may support cognitive tasks in clinical decision-making.

2. Methods

In this section, we first describe the model and/or algorithms used. We then describe an application, including the data and simulations used, designed to provide empirical evidence towards the three practical issues of concern

2.1 Model/Algorithms

We have previously described the POMDP and DDN algorithms in detail we use in this work [10], and for brevity provide only a brief description here. Plentiful descriptions of MDPs and POMDPs can also be found in the literature [2,6,7-9]. In short, any MDP is generally a tuple containing States (s), Actions (a), Observations (o), Rewards (r), and Costs (c). The model occurs over some discrete time-steps (t), which in healthcare typically are patient visits. A transition model (TR) encodes the probabilistic effects of various treatment actions, $P(s_{t+1} \mid s_t, a_t)$, i.e. given an action and current state the probability of ending up in some future state (also see Equation 1 in [10]). The transition model serves as the basis for decision-making, allowing an agent to estimate future rewards and costs for sequences of decisions. We additionally have some probabilistic relationship between observations and the actual underlying state (which is unobservable). This is referred to as a *belief state*. In many domains, since we cannot directly observe the underlying state, we must reason in the realm of belief states. This is often true in healthcare, since we typically do not directly observe a patient's disease (e.g. diabetes) but rather infer it from their symptoms (e.g. blood glucose readings). Furthermore, in a POMDP model, observations may sometimes be missing or noisy, meaning that the current belief state must be inferred from previous belief states. Reasoning over beliefs also provides similarities to human cognition, which must address the same challenges [14]. A basic example of a POMDP/DDN is shown in Figure 1, where CPUC=Rewards (r).

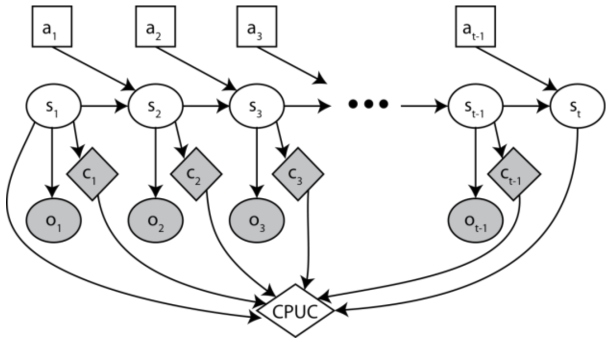

Figure 1: Example of a POMDP/DDN for clinical decision-making (from [10])

Rewards are calculated as Cost-per-Unit-Change (CPUC), which has been previously described [10, 15]. It is basically a measure of cost-effectiveness of a given treatment/action, calculated as a ratio of outcomes and costs. It provides a unit-scaled measure of rewards, and could be calculated with any outcome measure or any disease. Here, we use a functioning outcome (CDOI-ORS, see Section 2.1 in [10] for description) applicable to patients with co-occurring chronic mental and physical illness, as our dataset contains. This outcome tracks a patient's improvement/deterioration in functioning in daily life, a critical aspect in treatment of chronic illness where a "cure" is often not available.

It should be noted that the transition model has a direct correlation to the probabilistic output of many single decision time-point models in typical data mining applications. For instance, a neural network can output the probabilities of a number of possible outcomes for a number of possible input actions, which is in effect a transition model for a given point in time. In fact, by synthesizing alternate input for individual patients into a learned classifier, nearly any data mining model can produce such transition models for reinforcement learning. This could be used in tandem with something like Q-learning during its early stages of training, or as an alternative approach. The advantage is that the state-space can be reduced since classification over relevant variables (e.g. patient age, gender, diagnosis, outcome delta, etc.) can be dealt with outside the temporal model. Only variables that change rapidly over patent-visit time-scales (e.g. weeks, months) need be considered in the temporal model (e.g. gender typically would not change). We show an example application of this approach (Section 3.4). In another sense, this segregation can be thought of the separation of invariant and variant features in perception (e.g. vision in a Gibson-ian sense). We return to this notion in our discussion of cognitive computing in Section 4.1.

A final note is that the framework is configured to run as a multi-agent system (MAS), so that each physician and patient can be thought of as interacting agents within the model. An example of this can be seen in Figure 2 (see also Sections 1.5 and 2.2 in [10]).

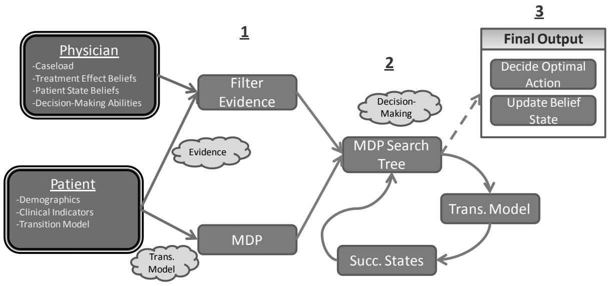

Figure 2: Types of agents are shown in double-line borders. Other boxes represent various aspects of the model. The general flow is: 1) create patient-specific MDPs/physician agent filters evidence into existing beliefs, 2) recurse through MDP search tree to determine optimal action, and 3) perform treatment action and update belief states. From [10].

2.2 Application

Averse to our previous work [10] where we used real patient data completely (real outcomes, actual treatments, etc.), here we take a simulation-based approach. The reason for this is that we wish to be able to compare across a range of scenarios unconstrained by the data [3], e.g. treatments having stronger/weaker effects, patients receiving different combinations of treatments, alteration of patient characteristics such as treatment adherence likelihood, etc. This allows us to freely manipulate parameters in the model and perform something akin to sensitivity analysis. However, it should be noted that, unless otherwise stated, data input into the model was actual data derived from actual EHR patient data.

Data was derived from clinical data from the EHR at Centerstone, a large, clinic-based behavioral and primary-care healthcare provider. Variables extracted were the same as described in [10]; however, we used an updated set of patient data containing 9,735 patients seen between 2010 and early 2013. This resulted in some slight changes to average costs or outcomes, though the relative patterns are the same. Similar to the previous study, the utilized patients' primary diagnosis was clinical depression with a majority exhibiting co-occurring physical illness, including hypertension, diabetes, chronic pain, and cardiovascular disease. The sample is typical of Medicaid/Medicare populations in the United States, largely comprised of patients with multiple, co-occurring chronic physical and mental illnesses [16,17].

Treatments were classified into three categories at a gross-level as Psychotherapy, Medication, and combined Therapy/Meds. Actions were similarly grouped into these three actions, henceforth labeled as Therapy, Meds, and Therapy+Meds, along with the option to *Not Treat* (i.e. end treatment or move to maintenance treatment, for the patient at that timepoint). There were thus four possible treatment actions (including not treating) to choose from at each timepoint. The ability exists to replicate this approach for, say, specific medications or for the inclusion of augmenting services such as case-management/care-coordination, though we leave that for future consideration. We focus here on the feasibility using the base services, at a relatively high level of abstraction.

Similar to the previous study, five states were calculated from outcome deltas (change in outcome from baseline, t=0, to current timepoint, t). These states were based on clinically validated thresholds [10, 18]. Critically, the use of outcome deltas, averse to clinical outcomes themselves, provides a convenient history meta-variable for maintaining the central Markov assumption: that the state at time t depends only on the information at time $t-1$ [19]. As noted in Section 2.1, reasoning is done over *belief states*, rather than deterministic states.

Treatment decisions were considered over the course of eight sessions, T = 8 treatments (based on the typical average number of sessions amongst Centerstone's outpatient population). The AI physician agent must make a treatment decision for each patient at each timepoint over the course of seven sessions (plus baseline/intake, max total sessions = 8). The transition model for such decision-making was treated as stationary, finite-horizon, and undiscounted.

Patient treatments and outcomes were simulated by taking a *test set* of 500 patients (sampled randomly from the total dataset), and starting with their actual patient data/characteristics at baseline (e.g. their baseline outcome score). Then, depending on the treatment action chosen by the AI model at each timepoint, the patient's next outcome was sampled from a Gaussian distribution using the ~9,000 other patients *not included in the test dataset*. The Gaussian distribution used for a given timepoint varied based on both the treatment chosen as well as the patient's outcome delta state at the given timepoint (see above). Thus, the process is similar to slicing from a histogram in particle filtering, given some history of change. The outcomes were probabilistically "hidden" from the visibility of the AI, to create a partially observable environment where clinical observations are sometimes missing or unavailable. In all results presented here, that was set to 30% of the time (the same observed in the real dataset).

All other probabilities and parameters for modeling were estimated directly from historical EHR data, as we have done previously [10]. That included the average cost per service, expected values of outcome improvement and deterioration, and transition model probabilities *for each type of treatment*. It would also be possible to add a further learning element – for instance by updating the initial transition model probabilities calculated from historical EHR data using something like temporal differencing, (see Section 4.2), though we have not done so here

A couple things are important to make clear here. First, the information made available to the "AI doctor" at a given timepoint was only the same information that would be available to a human doctor at that time (e.g. patient characteristics/diagnosis, patient history, available current observations). Second, the simulation of each next outcome for a given patient occurred only after the action/treatment was decided for the current timepoint. In other words, the AI doctor was not allowed to peek at future information, nor did such information even exist at the time of decision-making.

Obviously, a critical first step is showing that this simulation approach produces similar patient outcome and treatment cost values as the original model based entirely on real patient data. We do this in Section 3.1.

2.3 Simulation Experiments

We performed several experiments using the application described in Section 2.2. All experiments were performed over 10 replications, and all values reported here are averages over trials. The AI framework (including the MDP) is written in Python 2.7 (www.python.org). The code is parallelized using the multiprocessing package, so that multiple patients are run at once. In theory, the speed is thus only constrained by the number of processors available. An evaluation of ground truth for the simulation-based approach is provided in Section 3.1

First, we evaluated the trade-off between costs and outcomes (Section 3.2). Costs were derived from CPUC, which is in effect a measure of cost-effectiveness rather than raw costs. This is essential to creating models that are non-disease specific, because scaling costs relative to their cost-effectiveness allows us to compare across clinical domains. We also note that we are only considering short-term, immediate treatment costs here.

The trade-off between costs and outcomes was adjusted using the *outcome scaling factor* (OSF), which was previously described (see Section 2.6 and equation 7 in [10]). In short, the significance of the outcomes was varied via the OSF that adds in scaled outcome values {0-1} as an additional component of the utility metric. The outcomes (current delta) are scaled and flipped based on the maximum possible delta for a patient at a given timepoint ($\text{delta}_{\text{max}}$), so that higher values (near 1) are worse than lower values (given that we are attempting to minimize CPUC):

$$CPUC_{Final} = (1 - OSF) * CPUC + OSF * \left(\frac{delta_{max} - delta}{delta_{max}}\right) \qquad (1)$$

Where *delta* refers to outcome delta. Averse to previous work, we have altered equation 7 from [10] here so that OSF always ranges between 0-1. The equations are mathematically equivalent, but having OSF scaled 0-1 provides easier control. When OSF is set to 0, outcomes are considered equally important as costs, because outcomes are already accounted for in the basic CPUC utility calculation, even when this factor is set to 0. When OSF is set to 1, however, only outcomes are considered. The case where only costs are considered is not evaluated, since it is a trivial case (if we are simply trying to minimize costs, at least in the short-term, we would not provide any treatment services at all).

We also investigated the effects of different finite horizon levels (Section 3.3). Previous research has indicated potential issues with "look-ahead pathologies" related to horizon levels and other aspects of MDP models that can contribute to accumulating prediction errors into the future [20]. However, too small of a horizon (at its most extreme a greedy one-step look-ahead function) does not always take full advantage of temporal knowledge. We have a limited horizon in this case since patients are only seen over a set number of clinical sessions (which limits potential gains), but an evaluation of horizon effects is still worthwhile.

Finally, we evaluated a proof-of-concept for integrating machine learning (ML) classifiers into the temporal modeling application (Section 3.4). This entailed making *individualized* transition models for use by the MDP (see Section 2.1) based on the output probabilities of the ML classifier for each individual patient at each timepoint. For simplicity, we reduced the number of patient states down to three from five (Deterioration, Flatline, Improvement) and used a single classification scheme with no tuning. This was an ensemble classifier based on max-probability “voting by committee” [21] similar as to done with the CDOI previously in [22], using five underlying algorithms: Naïve Bayes, Multi-layer Perceptron neural network, Random Forests, K-nearest neighbors, and logistic regression. The ML classifer was constructed using Knime 2.8 (www.knime.org). No tuning of algorithm parameters was performed. The same features were used as in [22], although that model was meant for only a binary prediction (deterioration vs. improvement). Discretization of features was done using CAIM, a form of entropy-based discretization [23].

The classifier was pre-trained on patients from the broader dataset not included in the test set (Section 2.2) in Knime. The AI framework would then communicate with Knime via a backend data warehouse (Postgres 9, www.postgres.org) containing EHR data. In short, individual patients at a given timepoint (and a given outcome delta) could be sent to the DW, have various clinical indicators tagged on, and then queried to the classifier in Knime. Knime would apply the pre-trained classifier, then write the results back into the data warehouse, which could then be retrieved by the AI framework (i.e. physician agent).

We emphasize here that the goal was not to build a highly polished classifier, but simply to provide a proof-of-concept for the approach in a real-world application, using all the components that such applications typically contain.

3. Empirical Evaluation

3.1 Simulation Ground Truth

A first step in any simulation-based approach is to compare the simulated model with real-world data in order to provide some ground truth. Tables 1 and 2 provide a comparison of patient outcome and treatment cost values between the original model based entirely on real patient data (taken from [10]) and the current simulated model (which are based on real patients but include simulated outcomes). These tables show results for the Hard Stop (where the AI always decides to end treatment after the third visit), Raw Effect (where the AI always optimistically assumes patients will improve regardless of history), and the MDP models. We thus have a lower threshold where cost containment is primary (Hard Stop), an upper threshold where patient outcomes are primary (Raw Effect), and a more balanced model considering both (MDP). Details of these various models are provided in [10].

Table 1: Real Patient-Data Results (from [10])

					Std Dev		
	Transition	Missing		Avg Final	Final	Avg # of	% Patients
Decision Model	Model	Obs	CPUC	Delta	Delta	Services	Max Dosage
Hard Stop	N/A	Yes	305.53	2.56	8.07	3.00	0%
Raw Effect	0th Order	Yes	497.00	4.73	8.45	8.00	100%
MDP	Global	Yes	189.93	5.59	6.44	4.11	9%

Table 2: Current Simulation Results

					Std Dev		
	Transition	Missing		Avg Final	Final	Avg # of	% Patients
Decision Model	Model	Obs	CPUC	Delta	Delta	Services	Max Dosage
Hard Stop	N/A	Yes	352.14	3.22	7.80	3.00	0%
Raw Effect	0th Order	Yes	455.14	5.72	8.12	8.00	100%
MDP	Global	Yes	182.10	6.58	6.02	4.30	12%

We first note that the *relative* patterns across the models are consistent for both costs (as measured by CPUC) and outcomes (Average Final Delta), as well as the standard deviations in outcomes and max dosage (where max dosage equals the percentage of patients receiving treatment for all eight sessions). The *absolute* values are also comparable, although outcomes (Average Final Delta) are higher due to the use of the updated patient dataset with higher outcomes (see Section 2.2). In short, the simulation data closely approximates the patterns seen in models based entirely on real patient data.

3.2 Cost/Outcome Trade-offs

Next we evaluated the trade-offs between treatment costs and patient outcomes using the OSF parameter (see Section 2.3). This entailed varying the OSF between 0 (where cost effectiveness is emphasized) and 1 (where outcomes are emphasized). The main results using real cost values that varied by treatment type ($80-115 USD) can be seen in Figures 3 and 4.

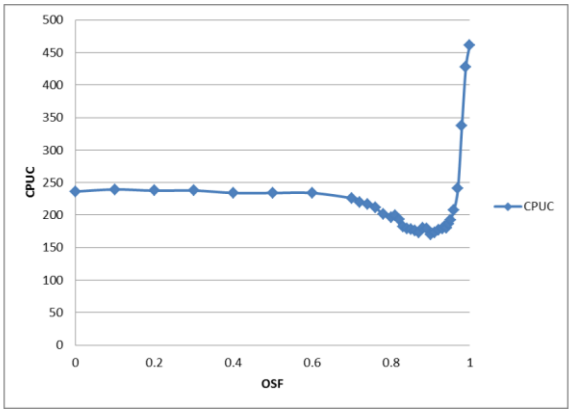

Figure 3: Average CPUC, actual costs

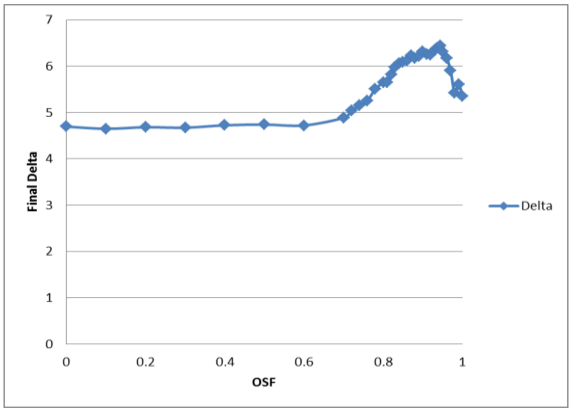

Figure 4: Average Outcome Delta, actual costs

As can be seen in those figures, there appears to be an optimal point (minimal CPUC and maximal outcome delta) around 0.94 for the OSF value, using actual cost for each treatment action as calculated from the EHR. Lower values of OSF result in lower outcomes and higher CPUC. Higher values of OSF also produce suboptimal results, in particular the over-emphasis on outcomes results in skyrocketing costs and CPUC. The reduced optimality of higher OSF even for outcomes falls in line with the notion of "look-ahead pathologies" [20] and can reflect, for example, the problems of over-treatment. Of note, setting OSF to 1 results in values of OSF similar to the Raw Effect model (see Table 2), where the physician agent always optimistically assuming more treatment produces better outcomes. This is not entirely surprising, but does provide some internal validity for the results.

A primary question is whether the patterns seen in Figure 3 and 4 are simply artifacts related to the absolute cost and outcome values, or perhaps the ratio between them. In order to test this, we ran further simulations where we varied only cost values, only outcome values, and both cost and outcome values simultaneously. For example, we halved the cost value for each treatment - keeping the relative values between treatments the same – while holding outcome values constant. The results can be seen in Figures 5 and 6.

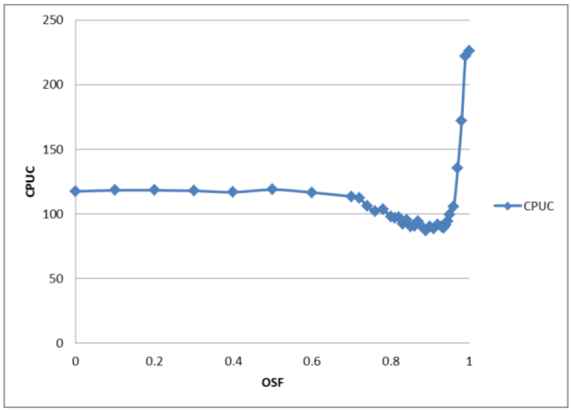

Figure 5: Average CPUC, costs halved

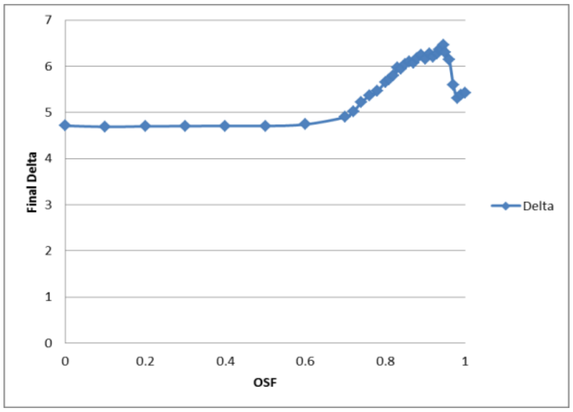

Figure 6: Average Outcome Delta, costs halved

Halving the treatment cost (which also simultaneously changed the ratio between costs and outcome), had no effect on the optimal OSF value – it remained at approximately 0.94. We performed other simulations where we doubled the costs, increased/decreased average outcome deltas, simultaneously altered both costs and outcomes, and altered the horizon level. All of those gave the same result (data not shown for brevity). In short, the optimal OSF value – and thus the optimal trade-off between costs and outcomes – is not a function of the *absolute* value of costs or outcomes, nor the ratio between the two.

A secondary question then is what other aspects might affect the optimal trade-off point. We know from prior research that different datasets can produce different optimal OSF values (e.g. our previously published results had an optimal OSF of 3-4, which equates to around 0.7-0.75 on a 0-1 scale, see [10]). So the question is why? Previous research did not consider different treatment costs, i.e. the costs were the same for all actions. We investigated this by running simulations on our current data where we artificially set all treatment actions to have the same costs (the overall average cost was maintained). Results can be seen in Figures 7 and 8.

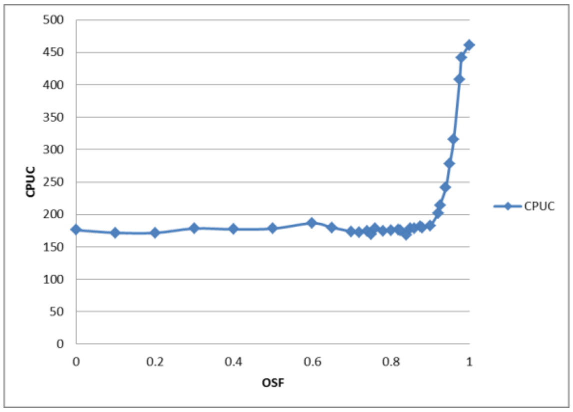

Figure 7: Average CPUC, equal costs for each treatment

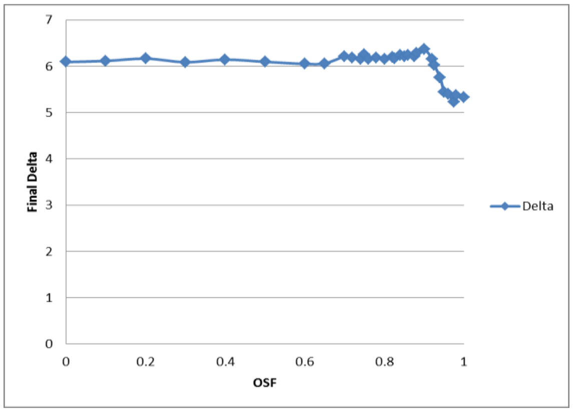

Figure 8: Average Outcome Delta, equal costs for each treatment

As can be seen, the optimal value for OSF is now distributed across a wide range. There is no clear optimum at any value, let alone the 0.94 value we saw above. Again, nothing was changed except the treatment costs were artificially set to be the same. Thus altering the *relative* cost between treatments did affect the optimal trade-off. Preliminary results (data not shown for brevity) suggest the same is also true of *relative* values of treatment effects on outcomes, both for transition probabilities (probability of improvement/deterioration) and average effect amounts (average amount of improvement/deterioration). In other words, it is the *relative* distribution/variation of these values *between* treatments that is key. Moreover, we suspect there may be interactions between these components of the system.

In short, the optimal trade-off value between costs and outcomes in a reinforcement learning model in healthcare may be a function of a dynamical system of *relative* values of costs and effects *between* treatment actions (rather than absolute values). It may be possible to calculate this function in a closed-form solution if enough is understood about the behavior of such systems.

3.3 Horizon Effects

We also investigated the effects of different finite horizon levels (with an optimized OSF=0.94, see Section 3.2). The results are shown in Table 3.

Table 3: Horizon Level Effects

Decision Model	Transition Model	Missing Obs	Horizon	CPUC	Avg Final Delta	Std Dev Final Delta	Avg # of Services	% Patients Max Dosage
MDP	Global	Yes	1	189.08	6.01	6.10	3.98	8%
MDP	Global	Yes	2	188.91	6.28	6.11	4.21	11%
MDP	Global	Yes	3	186.42	6.45	6.04	4.29	12%
MDP	Global	Yes	4	186.67	6.50	6.06	4.30	12%
MDP	Global	Yes	5	182.10	6.58	6.02	4.30	12%

There is gradual increase in outcome deltas (~10%, 6.01 vs. 6.58), though CPUC stays relatively flat. This is not surprising, given that lower outcomes can be compensated for by lower costs of less treatment or choosing less expensive treatment actions, so a more myopic algorithm can still be relatively cost effective. It should be noted that the simulation-based outcomes here follow a Gaussian distribution (by definition, see Section 2.2), while real-world outcomes are typically more unpredictable, which can result in higher costs for more myopic models (which we observed previously [10]).

It is unclear whether outcomes would increase further over larger horizon times or using different horizon models, such as receding horizons or forecast horizons [24,25]. Potential gains are limited here because we only make recommendations over 8 sessions, and no decision is made on the first or last session. It should be noted that longer horizon times do increase the run-time, although using parallelized programming code largely mitigates that issue (see Section 2.3). A typical patient, even at a horizon of 5, can be processed in about 2.2 seconds on a robust desktop personal computer.

One of the great advantages of temporal modeling is that it could potentially take into account the added risks of over-treatment. The inclusion of patient safety/risks into such modeling may reveal further increased horizon-level benefits (See Section 4.2).

3.4 Machine-Learning-Based Transition Models

We also evaluated the use of machine learning (ML) classifiers to dynamically create personalized transition models for each patient at each timepoint. As described in Section 2.3, the goal wasn't to build a highly polished classifier, but simply to provide a proof-of-concept for the approach. As such, we constructed an ensemble classifier based on max-probability "voting by committee" [21] similar as to done with the CDOI previously in [22], using five underlying algorithms: Naïve Bayes, Multi-layer Perceptron neural network, Random Forests, K-nearest neighbors, and logistic regression (see Section 2.3 for details). Importantly, the application of the ML classifier was performed on-the-fly from the EHR's backend data warehouse. Results can be seen in Table 4.

Table 4: ML-Classifier-Based Transition Models

Decision Model	Transition Model	Missing Obs	CPUC	Avg Final Delta	Std Dev Final Delta	Avg # of Services	% Patients Max Dosage
MDP	Global	Yes	182.10	6.58	6.02	4.30	12%
MDP + ML	Global	Yes	177.12	6.71	6.03	4.25	10%

The results show that the incorporation of ML-classifier-based transition models does slightly improve decision-making, with lower CPUC and higher outcomes. The gains were small, however. What is important here though – and what we emphasize – is that the approach showed modest success even with an admittedly less-than-optimal classifier. Overall classification performance – based on the ability to accurately predict three classes of the patients' outcome: deterioration, flatline, or improvement – was roughly 50% (over 33% random chance, area-under-curve: AUC=0.59). Incorporation of a more polished classification scheme – and/or other data points (e.g. genetics) – would likely hold promise to enhance the results seen in the proof-of-concept here.

As we have argued previously, incorporation of such *individualized* transition models through use of ML classifiers provides a potential practical avenue for implementation of personalized medicine approaches in EHRs and real-world clinical practice [10].

4. Significance and Impact

4.1 General Conclusion

We addressed several practical issues related to temporal modeling in an applied setting (healthcare), building simulations based on actual patient data from a real-world clinical electronic health record (EHR) while using a non-disease specific approach. First, we provided ground truth for the simulation-based approach against actual clinical data. Next, we evaluated the trade-off between costs and outcomes and found that the optimal trade-off may be a function of a dynamical system of *relative* values of costs and effects *between* treatment actions (rather than absolute values). Such an approach may allow optimal reward functions to be derived from clinical data. We also evaluated the effects of increasing finite horizon values, which showed a gradual increase in outcomes while cost effectiveness remained relatively flat. Importantly, parallelizing the code (so that multiple patients can be run at once) is essential to maintaining reasonable run-times at higher horizon levels. Finally, we evaluated a proof-of-concept for integrating ML-based transition models into temporal models like MDPs. The results showed that even a relatively poor classifier can produce small gains in performance and highlights the potential of such an approach for further exploration. Individualized transition models via such ML integration provide a potential practical avenue for implementation of personalized medicine approaches in EHRs and real-world clinical practice [10].

Temporal modeling approaches provide the potential to capture certain aspects of human cognition – the dynamic interplay of perception (observation) and action (treatment) over time. To best assist us, our clinical computing tools should approximate the same process. The more immediate goal is to offload certain cognitive tasks into the tools and artifacts around us, rather than providing data in the form of, say, a line graph, which still requires the bulk of computation and interpretation to occur in the human brain. For example, a line graph of a patient's outcome history or predicted trajectory doesn't necessarily indicate "what to do" in terms of treatment. Should treatment be stopped? Changed? Increased? If the patient has shown improvement (for chronic illness), should they be shifted into maintenance treatment [14]?

This, at its most basic level, is the same functionality provided by a notepad and pen, or by a calculator. It is simply an extension of such principles deeper into the realm of cognitive tasks. This approach also fits into the vein of cognitive computing, though more from an algorithmic front than a hardware one [26]. At a broader level, even the human visual system is thought to rely on environmental scaffolding of invariant features [27]. Re-conceptualizing artificial intelligence tools as a form of such cognitive scaffolding thus may provide better synergy with the way our brains already interact with the environment.

However, practical issues, like the ones addressed in this paper, are challenges that must be tackled in order to integrate machine learning or artificial intelligence models into domains such as healthcare. For instance, understanding the optimal trade-off between costs and outcomes is key [2,3,12]. If a system/model cannot effectively evaluate the utility of its decisions, it cannot make/recommend good decisions. Furthermore, if the utility of those decisions is a product of interacting components of some complex system, then characterizing those dynamics is essential (see Section 3.2). This is similar to arguments about trade-offs between exploration/exploitation in the reinforcement learning literature [28].

4.2 Future Work

There are a number of other aspects to temporal modeling that remain to be explored. Notably, incorporation of patient safety/risks into such modeling holds great promise to fully leverage the advantages. One of the greatest problems with over-treatment is that is exposes patients to unnecessary risks, side effects, and complications. Being able to stop treatment - or reduce treatment levels (e.g. maintenance treatment) – at the appropriate time is key. That is something that a myopic, greedy algorithm may or may not be able to address appropriately [1]. Quantification of such safety/risks is, of course, a critical aspect. Further exploration of this topic is warranted, as well as linking empirical results from real-world EHR data to theoretical models [12].

Another issue worth exploring is the effects of variation of the missing observation rates. In this work, we used the same value in all simulations as was calculated from our real dataset (approximately 30%). However, higher levels of missing observations force the AI physician agent to rely more on belief state calculations, which could affect performance.

Treatment adherence variability (e.g. medication adherence) is also a challenge in providing healthcare. For instance, in the EHR data we used here, nearly 75% of patients did not adhere to medication treatment at least once (e.g. missed a dose one day), and that approximately 25% of doses were missed overall. Thus, treatment adherence operates at two levels: the patient and the treatment. However, it is not equally distributed across patients – a small subset of patients who don't adhere make up a bulk of the missed treatments, and many patients just occasionally miss one. The question is what effect this distribution of non-adherence may have on temporal modeling of clinical decision-making. Can an AI agent based on temporal modeling mediate its treatment recommendations based on non-adherence probabilities for individual patients? Such questions could be empirically explored through careful simulation of real-world patient data.

A final issue we would like to point out here is that there are further possibilities of integrating learning, such as temporal-differencing (TD), into the historical-calculated transition models. This in effect is a simple way to take advantage of both prior knowledge (data from historical patients in the EHR) and reinforcement learning principles to produce the AI framework. At the most basic level, we are updating the transition probabilities using current information from patients being treated by the AI framework. At a higher level, there may be some effects from allowing an AI framework to influence clinical decisions – integration of something like TD learning allows the system to learn from its successes and mistakes, similar to "growing batch" reinforcement learning [29]. This approach is extensible to transition models based on machine learning classifiers (see Section 3.4). It may also be extensible to patterns gleaned from temporal data mining [30-32]. Optimally, the TD learning adjustments could be held out as *weights* that would applied to the transition probabilities at run-time (similar to predictive state representations [33]), which would allow the machine learning/data mining classifiers to be updated over time.

ACKNOWLEDGMENT

This research was funded by the Ayers Foundation and the Joe C. Davis Foundation. The opinions expressed herein do not necessarily reflect the views of Indiana University, Centerstone, or their affiliates.

REFERENCES

[1] Hauskrecht M, Fraser H (2000) Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artificial Intelligence in Medicine*. 18(3): 221–244.

[2] Alagoz O, Hsu H, Schaefer AJ, Roberts MS (2010) Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*. 30(4): 474–483.

[3] Shortreed SM, Laber E, Lizotte DJ, Stroup TS, Pineau J, Murphy SA (2011) Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Machine Learning.* 84(1-2): 109–136.

[4] Chakraborty B, Murphy SA (2014) Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application*. 1(1).

[5] Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. (2009) The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*. 46(1): 5–17.

[6] Schaefer AJ, Bailey MD, Shechter SM, Roberts MS (2005) Modeling medical treatmentusing Markov decision processes. In: Brandeau ML, Sainfort F, PierskallaWP, eds. *Operations research and health care*. Kluwer Academic Publishers: Boston, MA. pp. 593–612.

[7] Littman ML (2009) A tutorial on partially observable Markov decision processes. *Journal of Mathematical Psychology*. 53(3): 119–25.

[8] Wiering M, van Otterlo M, eds (2012). *Reinforcement Learning: State-of-the-Art*. Springer: Berlin.

[9] Ross S, Pineau J, Paquet S, Chaib-draa B (2008) Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*. 32: 663–704.

[10] Bennett CC, Hauser K (2013) Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*. 57(1): 9-19.

[11] Kober J, Peters J (2012) Reinforcement Learning in Robotics: A Survey. In: Wiering M, Otterlo M van, eds. *Reinforcement Learning: State-of-the-Art*. Springer: Berlin. pp. 579–610.

[12] Lizotte DJ, Bowling MH, Murphy SA (2010) Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. *Proceedings of the 27th International Conference on Machine Learning (ICML)*. pp. 695–702.

[13] Zhao Y, Kosorok MR, Zeng D (2009) Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*. 28(26): 3294–3315.

[14] Elstein AS, Schwarz A (2002) Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 324(7339): 729–732.

[15] Bennett CC (2011) Clinical productivity system: a decision support model. *International Journal of Productivity and Performance Management*. 60(3): 311–319.

[16] Wolff JL, Starfield B, Anderson G (2002) Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Archives of Internal Medicine*. 162(20): 2269–2276.

[17] Jones DR, Macias C, Barreira PJ, Fisher WH, Hargreaves WA, Harding CM (2004) Prevalence, Severity, and Co-occurrence of Chronic Physical Health Problems of Persons With Serious Mental Illness. *Psychiatric Services*. 55(11): 1250–1257.

[18] Miller SD, Duncan BL, Brown J, Sorrell R, Chalk MB (2006) Using formal client feedback to improve retention and outcome: making ongoing, real-time assessment feasible. *Journal of Brief Therapy*. 5(1): 5–22.

[19] Stahl JE (2008) Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics*. 26(2): 131–148.

[20] Peret L, Garcia F (2004). On-line search for solving Markov decision processes via heuristic sampling. *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI)*. 530-534.

[21] Kuncheva L (2004) *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience.

[22] Bennett CC, Doub TW, Bragg AD, Luellen J, Van Regenmorter C, Lockman J, et al. (2011) Data mining session-based patient reported outcomes (PROs) in a mental health setting: toward data-driven clinical decision support and personalized treatment. *Proceedings of the IEEE Conference on Health Informatics, Imaging, and Systems Biology (HISB)*. pp. 229–236

[23] Kurgan LA, Cios KJ (2004) CAIM discretization algorithm. *IEEE Transactions on Knowledge and Data Engineering*. 6(2): 145–153.

[24] Hopp WJ (1988) Identifying forecast horizons in nonhomogeneous Markov decision processes. *Operations research*. 37(2): 339–343.

[25] Chang HS, Marcus SI (2003) Approximate receding horizon approach for Markov decision processes: average reward case. *Journal of Mathematical Analysis and Applications*. 286(2):636–651.

[26] Modha DS, Ananthanarayanan R, Esser SK, Ndirango A, Sherbondy AJ, Singh R (2011) Cognitive computing. *Communications of the ACM*. 54(8): 62–71.

[27] Gibson JJ (1979) *The Ecological Approach to Visual Perception*. Houghlin Mifflin: Boston, MA.

[28] Wang T, Lizotte D, Bowling M, Schuurmans D (2005) Bayesian sparse sampling for on-line reward optimization. *Proceedings of the 22nd ACM International Conference on Machine learning (ICML)*. pp. 956–963.

[29] Lange S, Gabel T, Riedmiller M (2012) Batch Reinforcement Learning. In: Wiering M, Otterlo M van, eds. *Reinforcement Learning: State-of-the-Art*. Springer: Berlin. pp. 45-73.

[30] Keogh E, Kasetty S (2003) On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. *Data Mining and Knowledge Discovery*. 7(4): 349–371.

[31] Bellazzi R, Larizza C, Magni P, Bellazzi R (2005) Temporal data mining for the quality assessment of

hemodialysis services. *Artificial Intelligence in Medicine*. 34(1): 25–39.

[32] Laxman S, Sastry PS (2006) A survey of temporal data mining. *Sadhana*. 31(2): 173–198.

[33] Wingate, D (2012) Predictively Defined Representations of State. In: Wiering M, Otterlo M van, eds. *Reinforcement Learning: State-of-the-Art*. Springer: Berlin. pp. 415-439.